University of Southampton

Faculty of Social Sciences

Southampton Education School

Cognitive Processes in Writing-to-Learn

by

Amy Peters

ORCID ID 0000-0001-9833-7709

Thesis for the degree of Doctor of Philosophy

December 2025

University of Southampton

Abstract

Faculty of Social Sciences

Southampton Education School

Thesis for the degree of Doctor of Philosophy

Cognitive Processes in Writing-to-Learn

Amy Peters

Writing-to-learn is the idea that writing is not only an effective tool for communication, but it also has the capacity to shape learning and facilitate the development of knowledge (Emig, 1977; Klein, 1999). Specifically, it is *writing* that facilitates learning, and not another medium such as thinking of speaking (Galbraith et al., 2025). This thesis explored writing-to-learn in the context of closed-book source-based writing, under the lens of the dual-process model of writing (Galbraith & Baaijen, 2018). It also aimed to compare different measures of learning from writing.

In Paper 1, 78 higher-education students read a source text passage about solar activity (Arnold et al., 2017), created either an outline or synthetic plan, and wrote an essay or wrote down everything they could remember from the text (free recall). Two days later, participants returned to answer text comprehension questions to assess their knowledge of the source material. They also rated their subjective knowledge of the source text before writing, after writing, and two-days later. In support of a previous study by Arnold et al. (2017), there were no differences in text comprehension scores across the planning or writing conditions, because all conditions involved retrieval practice. However, participants subjective knowledge decreased from pre- to post-writing, regardless of condition. Writing about the source helped participants to identify knowledge gaps, consistent with the experience of the illusion of explanatory depth (Rozenblit & Keil, 2002). The knowledge-telling process was re-interpreted to suggest it can have a role in learning through metacognitive monitoring. Importantly, participants subjective knowledge ratings positively correlated with their text comprehension scores, supporting the validity of the measures.

Paper 2 investigated the effects of written and non-written (mental) source-based activities on text comprehension and participants' subjective knowledge. Experiment 1 was a methodological check of a new source text passage. In the main experiment, Experiment 2, 116 participants read one of two source texts, and then wrote or thought about writing either an essay or free recall. Two-days later, participants completed text comprehension questions, and rated their subjective knowledge before writing, after writing, and two-days later. Results showed that writing about the source text led to better performance on short-answer questions than thinking. Participants in the writing condition also experienced a greater decrease in their subjective knowledge than participants in the thinking conditions. These results highlighted that writing, and not another medium, has a unique value for enhancing text comprehension and metacognitive monitoring. Subjective knowledge ratings were positively correlated with short-answer question scores and factual multiple-choice question scores, replicating the results from Paper 1.

Paper 3 aimed to test the reliability and validity of a new methodological tool for assessing learning from writing in a replication and extension of Silva and Limongi (2019). Experiment 1 was a direct replication of Silva and Limongi's study, which compared the effects of writing versus speaking summaries of short texts on participants' reaction time on an episodic memory word-recognition task. Experiment 2 adapted the procedure to control for a potential confounding factor (modality effect). Across both experiments, results replicated Silva and Limongi's findings. Drift-diffusion modelling of reaction time distributions revealed that after

writing summaries, participants had faster reaction times than after speaking summaries, represented by the non-decision parameter (encoding and/or motor processes).

Finally, Paper 4 aimed to investigate the cognitive processes underlying writing-to-learn in a closed-book, multiple source-based writing task. A second aim was to assess how the three measures of learning established in Papers 1 to 3 were related to one another. 40 university students read two source texts before writing either a synthesis essay or writing down everything they could remember from the two texts. After writing, they completed a word-recognition reaction time task, and a text comprehension test. Participants also rated their subjective knowledge before reading, before writing, and after writing. The most important finding was that whilst a subjective knowledge decrease was observed for participants in the free recall condition, participants in the synthesis condition did not experience such a decrease. This suggests the synthesis task enabled participants to carry out knowledge-transforming, whereas the free recall participants were essentially told do carry out knowledge-telling. Crucially, subjective knowledge ratings were significantly positively correlated with short-answer question scores, a finding that has replicated across these papers.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: AMY PETERS

Title of thesis: Cognitive Processes in Writing-to-Learn

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: ..... Date: .....

Acknowledgements

I would like to express my sincere gratitude to my supervisor, David Galbraith. I would not have been able to complete this thesis without your expertise, guidance and support. Thank you also to Veerle Baaijen for all of your help and advice, particularly when it came to analysing keystroke data!

I am also grateful to the ESRC South Coast Doctoral Training Partnership (SCDTP) for providing me with the funding to undertake a PhD. As a result, I have been lucky enough to attend overseas conferences and visits which have been invaluable in shaping this research. Thank you to the EARLI SIG Writing community for providing a welcoming space to present my work and meet other writing researchers.

I would also like to say a huge thank you to Roberto Limongi and Angelica Silva for hosting me at Western University in Canada. I am truly grateful for their contributions that helped me carry out this research and hope to collaborate again in the future.

My appreciation also goes to Southampton Education School for providing a brilliant place to work, not just for my PhD but also during my undergraduate study. I have been able to get to know so many wonderful academics, too many to name here!

To the friends I have met during this PhD, thank you for always being there during the highs and the lows. To Josette, you have been the best PhD friend I could have asked for, and I am really going to miss our unplanned study breaks for a coffee and a sweet treat! Here's to many more years of friendship. To Sophie, I not only thank you for creating reproducible keystroke logging analysis methods but more importantly for being such a fantastic 'PhD sibling'! To Olivia, thank you for always being only an Instagram message away, whether that be to have a moan about PhD life or to clown about yet another Taylor Swift theory. To Hazel, I will always be grateful for your advice and encouragement. I look forward to many more Pret lunches! To Furkan and Sotiria, thank you for being not just the most fun office mates but even better friends. To Alice, thank you for being such a brilliant housemate for a big chunk of our PhDs, I will always value the laughs during our evening chats at the most stressful times!

Last, but by no means least, thank you to Mum, Dad, Lauren, and Josh for your encouragement and belief in me when I needed it most.

Acronyms and Abbreviations

ANOVA ..... Analysis of Variance

DV ..... Dependent Variable

HDDM..... Hierarchical Drift-Diffusion Model

ICC ..... Intraclass Correlation Coefficient

IV..... Independent Variable

MCQ ..... Multiple-Choice Question

ms ..... Milliseconds

MSE ..... Mean Squared Error

RT ..... Reaction/Response Time

SAQ ..... Short-Answer Question

SD ..... Standard Deviation

SEM ..... Standard Error of the Mean

UoS ..... University of Southampton

Chapter 1 Introduction and Thesis Overview

1.1 Research Background, Aims, and Contribution

Writing-to-learn is the idea that writing is not just an effective tool for communication and language production, but that it also has the capacity to shape learning and facilitate development of knowledge (Emig, 1977; Klein, 1999). Traditionally, research on writing has focused on the quality of the final written product rather than the learning processes that occur during text production. This means the cognitive processes underlying writing are not well understood.

Classic models of writing (e.g., Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) suggest knowledge is constructed during writing through a goal-directed, problem-solving process. Hayes and Flower's (1980) model presented three key processes underlying writing: planning, translating, and reviewing. These processes are situated within the task environment and the writer's long-term memory. However, the model has been critiqued for its descriptive nature which later led Bereiter and Scardamalia (1987) to propose the writing process as comprising of a knowledge-telling model and a knowledge-transforming model. Support for both Hayes and Flower's (1980) and Bereiter and Scardamalia's writing models largely comes from comparisons of novice and expert writers. Younger, novice writers tend to engage mostly in knowledge-telling, simply retrieving ideas from memory and translating them into text, whereas older, expert writers produce more elaborate texts guided by rhetorical goals (e.g., Flower & Hayes, 1980; Bereiter & Scardamalia, 1987; Hayes & Flower, 1986).

Nevertheless, Bereiter and Scardamalia have also been criticised for making distinct assumptions about knowledge transformation during writing and provide limited explanation as to how knowledge is represented in long-term memory (Applebee, 1984; Baaijen & Galbraith, 2018). Recently, Galbraith and Baaijen (2018) have conceptualised writing as a dual process. The goal-directed knowledge-transforming process is responsible for the retrieval of the writer's thoughts and ideas from episodic memory; and the knowledge-constituting process guides the writer to synthesise content using a semantic network of knowledge in memory. When writers effectively manage these two processes, a development in knowledge is said to occur. Empirical support for the dual-process model comes from studies in which participants write spontaneously about their own personal existing knowledge (Baaijen & Galbraith, 2018; Hall, 2023). In these studies, participants rate their subjective knowledge pre- and post-writing, and in support of the dual-process model, findings suggest that discovery of knowledge is related to synthetically planned, spontaneous text production.

However, in educational settings, writing is often preceded by reading. The claims of the dual-process model have not been explored in the context of source-based writing. Meta-analyses have shown that writing about a source can enhance reading comprehension and learning of content (e.g., Graham & Hebert, 2011; Graham, Kiuhard & MacKay, 2020). Research on source-based writing-to-learn has explored summarising (e.g., Friend 2002; Brown & Day, 1983; Spirgel & Delaney, 2016) and synthesis writing (e.g., Wiley & Voss, 1999; Vandermeulen et al., 2020). Other research using source materials has shown that self-explaining can improve learner's understanding (e.g., Chi et al., 1994; Roelle et al., 2023), although some suggest this is because it helps learners identify gaps in their knowledge (e.g., the illusion of explanatory depth; Rozenblit & Keil, 2002). Typically, research on source-based writing takes place in an open-book environment, where writers can access source materials during composition.

In this thesis, source-based writing is explored in a closed-book context, where participants cannot access the source material(s) whilst they are writing. Therefore, this generative learning activity is combined with retrieval practice. A large amount of literature has produced a robust finding on 'the testing effect', where simply retrieving information from memory (retrieval practice) can enhance the retention of information (Glover, 1989; Roediger & Karpicke, 2006a, 2006b; Karpicke, 2012). The idea that retrieval practice may be an important underlying process in the effects of writing on learning was suggested by Arnold et al. (2017), who found greater text comprehension resulting from source-based writing activities involving retrieval (essay and free recall) than activities not involving retrieval (notetaking and highlighting).

The advantage of source-based writing activities for researching writing-to-learn is that participants' knowledge can be assessed through objective text comprehension tests about the source materials. Such recall measures are dominant in the writing-to-learn literature (Ackerman, 1993). However, given the focus of the dual-process model on knowledge discovery, it is important to incorporate a measure of participants' subjective development of knowledge. This may also help to distinguish between a view of learning as factual memory of information and learning as gaining understanding or knowledge. Therefore, Papers 1, 2, and 4 of this thesis incorporate both a measure of text comprehension and a measure of subjective knowledge to assess learning from writing. In addition, Paper 3 explores a new measure of learning from writing: long-term memory consolidation measured by reaction time (RT) in a word-recognition task (Silva & Limongi, 2019). This represents the first shift in writing-to-learn research from using response accuracy (text comprehension) and subjective knowledge measures to RT measures.

Overall, this thesis aims to explore the underlying cognitive processes in writing-to-learn from sources in a closed-book environment. Learning from writing is measured with text comprehension tests, participants' subjective ratings of their understanding, and RT on a word-

recognition test. Therefore, a second goal for this thesis is to compare different measures of writing-to-learn. In Paper 4, the writing process is explored using keystroke logging. Therefore, this thesis has novel theoretical and methodological contributions. Theoretically, it builds on the limited research exploring closed-book source-based writing-to-learn and considers whether the claims of the dual-process model are supported in such a context. This is achieved by comparing a variety of writing activities (such as essay writing, free recall, and synthesis writing) against each other, and against non-writing activities (such as thinking and speaking). Methodologically, this thesis explores how different measures can be used to assess the learning effects of writing, and how these measures relate to one another. It also aims to validate a novel measure of writing-to-learn (RT). These aims are achieved across four empirical papers.

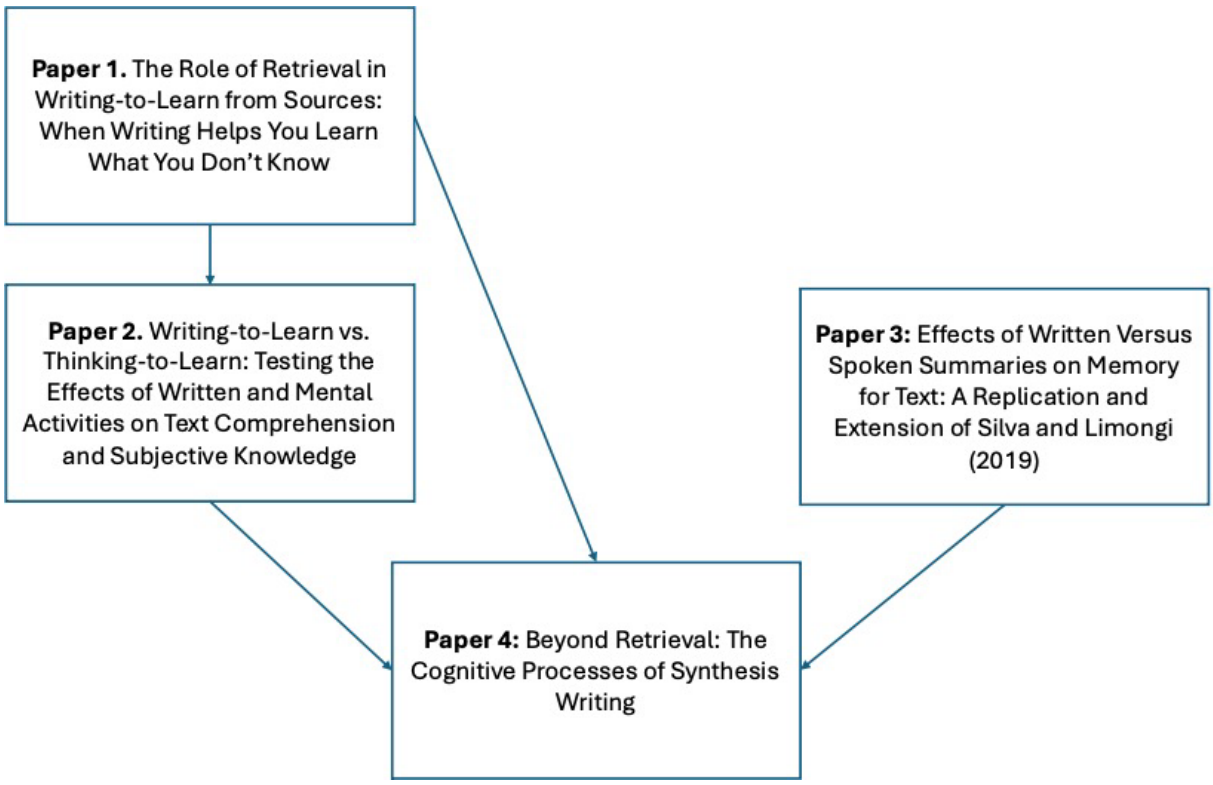
1.2 Thesis Outline and Overview of Papers

This thesis adopts a research paper format, with four empirical, distinct (though closely linked) papers. Taken together, the findings contribute to enhancing theoretical and methodological knowledge on the cognitive processes in writing-to-learn from sources.

Chapter 2 provides introductory material. It explains the theoretical background, including classic models of writing to learn, and the more recent dual-process model in which this project is situated. It reviews existing literature on writing-to-learn, source-based writing, and retrieval practice and generative learning. In Chapter 3, the overarching methodology of the thesis is explained, including the experimental approach, measures of learning and the writing process, as well as sampling and ethical considerations.

Chapters 4, 5, 6, and 7 present the four empirical papers central to this thesis. Although these are written as stand-alone papers, they have clear links between them, shown in Figure 1.

Figure 1. Thesis structure: Links between Papers 1-4.



Chapter 4 presents Paper 1, which aimed to explore the claims of the dual-process model of writing (Galbraith & Baaijen, 2018) in the context of closed-book source-based writing. Participants were asked to read one source text passage on solar activity (the same as one of the texts used by Arnold et al., 2017), create a plan in either an outline or synthetic condition, and then write an essay or write down everything they could remember from the text (free recall). These are the two retrieval conditions from Arnold et al.'s (2017) study. Two-days later, participants completed a text comprehension test based on the source text comprising of multiple-choice questions (MCQs) and short-answer questions (SAQs). Participants also completed a subjective knowledge rating scale (Galbraith et al., 2023) before writing, after writing, and two-days later (before completing the text comprehension questions). Results showed no differences between planning and writing conditions on either participants' text comprehension scores or subjective knowledge ratings, supporting Arnold et al.'s (2017) claim that retrieval practice has an important role in the learning effects of closed-book writing from sources. This was also reinforced by a novel finding that participants' subjective knowledge significantly decreased from pre- to post-writing. It appeared that engaging in retrieval practice helped writers identify gaps in their knowledge. Participants' subjective knowledge ratings were moderately correlated with their text comprehension scores. It was concluded that in this type of writing task, participants primarily engaged in knowledge-telling as opposed to the knowledge-constituting and knowledge-transforming processes of the dual-process model.

Chapter 5 presents Paper 2, which is comprised of two experiments aiming to address limitations and additional research questions that arose from the findings of Paper 1. Paper 1 was conducted online due to the Covid-19 pandemic, using the University of Southampton tool 'iSurvey'. This survey platform was unable to accurately control the time participants spent on the writing task during the experiment. Therefore, data collection for Paper 2 was conducted using an alternative online platform, Qualtrics. Two additional limitations were also addressed. First, in Paper 1, all participants read the same source text on solar activity. In Paper 2, a new source text was added (deserts and savannas) to assess whether findings could be generalised to a new topic. Experiment 1 of Paper 2 tests a new set of text comprehension questions for the new deserts and savannas text. Second, Paper 1 did not include a non-writing control condition, so it could not be concluded whether it was the act of writing itself that lead to the decrease in participants' subjective knowledge ratings.

In Experiment 2 of Paper 2, two non-writing control conditions were included. Participants were randomly allocated to a source text topic (solar activity or deserts and savannas), and then randomly allocated to one of four conditions: written essay, written free recall, mental essay, or mental free recall. The non-writing, mental, conditions, required participants to think about how they would carry out the writing task without writing anything down. Two days later, participants completed a text comprehension test on the source text (MCQs and SAQs). Participants also completed the subjective knowledge rating scale (Galbraith et al., 2023) before writing, after writing, and two-days later. In support of the notion that *writing*-to-learn, and not another medium (Galbraith et al., 2025) can enhance text comprehension, participants in the writing conditions performed better on SAQs than participants in the non-writing conditions. Furthermore, participants in the writing conditions experienced larger decreases in their subjective knowledge from pre- to post-writing than participants in the non-writing conditions.

Chapter 6 presents Paper 3, which explores a new methodological tool proposed by Silva and Limongi (2019) for measuring learning from source-based writing. In previous writing-to-learn research, 'learning' is typically measured through response accuracy measures like text comprehension tests (e.g., Arnold et al., 2017; Atasoy, 2013; Gingerich et al., 2014; Drabick et al., 2007) or measures of subjective knowledge (e.g., Baaijen & Galbraith, 2018; Galbraith et al., 2023). These are the measures of learning adopted in the first two papers of this thesis. In cognitive psychology, learning is often assessed through measures of long-term memory consolidation, for example RT in word-recognition tasks. Silva and Limongi (2019) suggested that writing-to-learn research could benefit from such a measure, to identify the underlying neural processes of how writing facilitates learning. In their study, participants wrote or spoke summaries after reading short text passages, and then completed a word-recognition memory task in which their RT to indicating whether a word appeared in the source text or not was

measured. A writing superiority effect was found whereby RTs were quicker following written summaries than spoken summaries. A potential confound in the design of Silva and Limongi's study means an alternative explanation of this writing superiority effect could be a priming or modality effect. To assess the reliability and validity of the measure, Experiment 1 is a replication of Silva and Limongi's study. To address the possible confound in the design, Experiment 2 modifies the procedure to control for the modality effect. The findings replicated Silva and Limongi's original findings and ruled out the potential confound, supporting the use of this measure in Paper 4 and other future experiments.

Chapter 7 presents Paper 4, the final paper of this thesis. Paper 4 brings together the methodological tools developed in Papers 1-3 for measuring learning from writing (text comprehension, subjective knowledge ratings, and RT) in a multiple source-based writing task. In Papers 1 and 2, participants only read and wrote about one source text passage, and the effects of retrieval practice appeared to lead to a decrease in participants' subjective knowledge. It is possible that reading and writing about more than one (related) text in a synthesis writing task involves more than just retrieval practice (knowledge-telling), and may reflect the processes invoked by the dual-process model.

Therefore in Paper 4, participants read two source texts on different modes of therapy for the treatment of neuroses, and were then randomly assigned to either a synthesis writing task (comparing and contrasting the two texts), or a free recall task (writing down everything they could remember from the sources). Participants' keystrokes were logged during the writing task to provide insight into their writing processes. After the writing task, participants completed a word-recognition memory test measuring their RT, as well as a text comprehension test comprising of MCQs and SAQs about the sources. They also completed subjective knowledge ratings before reading, after reading, and after writing. The main finding was that participants in the free recall condition experienced a decrease in their subjective knowledge from pre- to post-writing, but participants in the synthesis writing condition did not experience this decrease, supporting the notion that synthesis writing goes beyond the simple process of retrieval practice. Participants' subjective knowledge ratings were significantly positively correlated with their scores on SAQs. Overall, there were no differences between conditions on participants' text comprehension scores or RT. The writing process measures identified by Baaijen and Galbraith (2018) to represent components of the dual process model were not found to predict change in participants subjective knowledge in either condition.

Finally, Chapter 8 concludes the thesis with a general discussion and final remarks. The novel findings of each of the four empirical papers are summarised and synthesised, to consider how they relate to existing theory and literature and how they contribute to current understanding of

the cognitive processes in writing-to-learn. The chapter also considers the limitations of the research and sets out recommendations and implications for theory and practice.

Chapter 2 Theoretical Background and Empirical Context

Writing is a fundamental skill at all levels of education and in professional contexts, and it is a complex task involving several cognitive processes across memory, language, and thinking (Alamargot & Chanquoy, 2001). The earliest research on writing focused on the finished product, exploring the characteristics of good quality texts free of grammatical and mechanical mistakes (Klein et al., 2014; Leggette et al., 2015). Since the 1970s, writing research has shifted its focus from the product to the process (i.e., how writing is produced). This cognitive approach to research on writing gives greater attention to what the writer does during text production as opposed to the writing they produce (Applebee, 1984), moving from a text-based perspective to a writer-based perspective (Spelman Miller & Sullivan, 2006).

The cognitive approach views the writing process as a dynamic one, where learners have to continuously engage in a variety of processes such as planning, translating, and revising to clarify knowledge and identify connections between new information and existing ideas in memory (Alamargot & Chanquoy, 2001; Bereiter & Scardamalia, 1987; Hayes & Flower, 1980). *Writing-to-learn* captures the idea that writing is not only a tool for communication, but it can also be a tool for facilitating learning (Emig, 1977; Klein, 1999; Kellogg, 1994). The most recent conceptualisation of writing-to-learn describes its unique feature as developing the writer's own understanding (discovery of knowledge), rather than to achieve a communicative purpose for a reader (Galbraith et al., 2025).

In addition to developing a writer's knowledge and understanding, another fundamental feature of theories of writing-to-learn is the claim that it specifically involves *writing* (Galbraith et al., 2025). Although writing processes are similar to thinking processes (searching for ideas, organising, and reviewing; Klein & Boscolo, 2016), writing can provide an advantage for learning as it creates an external representation of the writer's knowledge. In comparison to speaking, writers have a record of what has already been transcribed and are not restricted to these ideas fading in short-term memory (Baaijen, Galbraith & de Glopper, 2012). This allows the writer to monitor the ideas represented in their texts (Olive, 2010) and can therefore improve learning outcomes. For example, Drabick et al. (2007) found undergraduate students improved their performance on factual and conceptual multiple-choice questions (MCQs) by engaging in in-class writing compared to in-class thinking.

There is a variety of evidence showing that writing-to-learn activities are effective for learning outcomes in educational settings. Bangert-Drowns, Hurley and Wilkinson's (2004) meta-

analysis of 48 school-based writing-to-learn interventions found consistent positive effects of writing-to-learn. This was especially due to the support such activities provide for metacognition and self-regulation during learning. Furthermore, Gunel, Hand and McDermott (2009) explored the use of writing-to-learn in secondary school classrooms in the USA. Participants scores on a writing task were the strongest predictor of performance on a content test formed of MCQs, true/false questions, and essay questions. They also acknowledged the importance of having an audience in mind when engaging in writing-to-learn activities, for example writing to a younger audience was helpful for students to transform their knowledge into a product appropriate for the target reader.

The benefits of writing-to-learn have also been demonstrated in university settings. Gingerich et al. (2014) compared writing-to-learn assignments in a large introductory psychology course (924 students) to copy assignments. The writing-to-learn assignments involved active learning processes such as generation and application, and this improved outcomes on a multiple-choice retention test, with benefits sustaining eight-weeks after the course. Qualitative studies have also shown that university students valued writing-to-learn activities to guide them in understanding the course content, although they found such activities time-intensive (Fry & Villagomez, 2012). Individual differences can also moderate the relationship between writing and learning, for example learners' beliefs about writing can moderate the effectiveness of different writing-to-learn strategies (Baaijen, Galbraith & de Glopper, 2014; Galbraith & Baaijen, 2018; White & Bruning, 2005; Sanders-Reio et al., 2014).

Since the 1980s, multiple theories and models have emerged to explain the writing process and how it can enhance learning. In an early overview of writing-to-learn research, Klein (1999) distinguished between four theories. This included early research conducted by Britton and colleagues (1975) who focused on spontaneity in writing; forward-search hypotheses focusing on the permanence of written text for reflection (e.g., Young & Sullivan, 1984); and genre theories which demonstrate how text structure can facilitate connections amongst ideas (Newell, 1984). However, the most widely cited and accepted theories of the writing process are the classic problem-solving models, proposed by Hayes and Flower (1980) and Bereiter and Scardamalia (1987). These are explained next, followed by the dual-process model of writing (Galbraith & Baaijen, 2018), which this thesis is focused on.

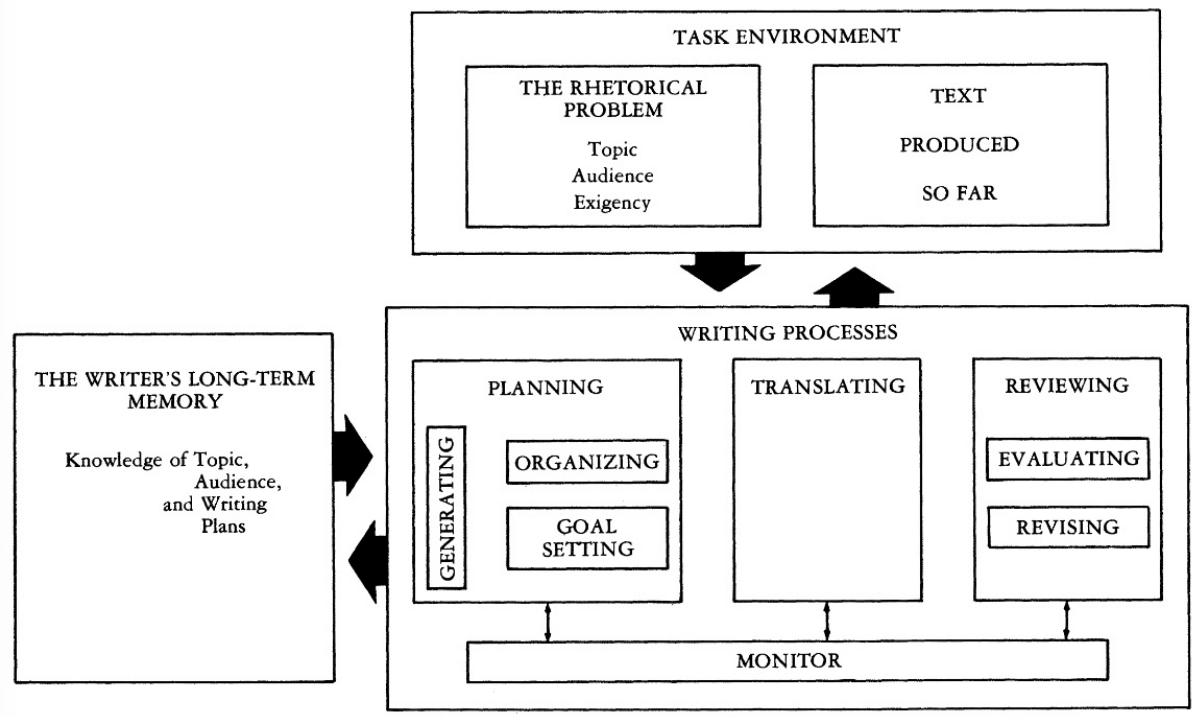
2.1 Classic Models of Writing

Classic models of writing suggest knowledge is generated during writing through a goal-directed, problem-solving process (Hayes & Flower, 1980; Bereiter & Scardamalia, 1987). Flower and Hayes (1980) observed how, during composition, writers harness a variety of

processes including planning, retrieving ideas from long-term memory, forming inferences and connections between concepts, and communicating to an audience. They suggest that during writing, individuals are trying to solve a rhetorical problem and that this guides production of text. Writers differ in how they define and elaborate on the rhetorical problem.

Hayes and Flower's (1980; Flower & Hayes, 1981) model of the writing process proposes that writers adapt to rhetorical goals through three processes: planning, translating, and reviewing. These processes are situated within the task environment (anything outside of the writer themselves that can influence performance) and the writer's long-term memory, as shown in Figure 2. The writer's long-term memory provides a store for knowledge about the topic at hand and knowledge about writing and the audience (Flower & Hayes, 1981). It is important to note that the model is not a stage model; writers engage in these processes recursively such that any process can occur at any time during writing.

Figure 2. Hayes and Flower's (1980) model of writing, as shown in Flower and Hayes (1981). Reproduced with permission of the rights holder.



Planning. In the planning process, writers use information from the task environment and their long-term memory to create goals and a plan to guide composition (Hayes & Flower, 1980). Planning is formed of three sub-processes. Generating is responsible for retrieving relevant thoughts and ideas from long-term memory. The organising process then identifies the most important information and organises it into a writing plan to help the writer give structure to their ideas. Goal setting allows the writer to identify the rhetorical goals for the text. The planning

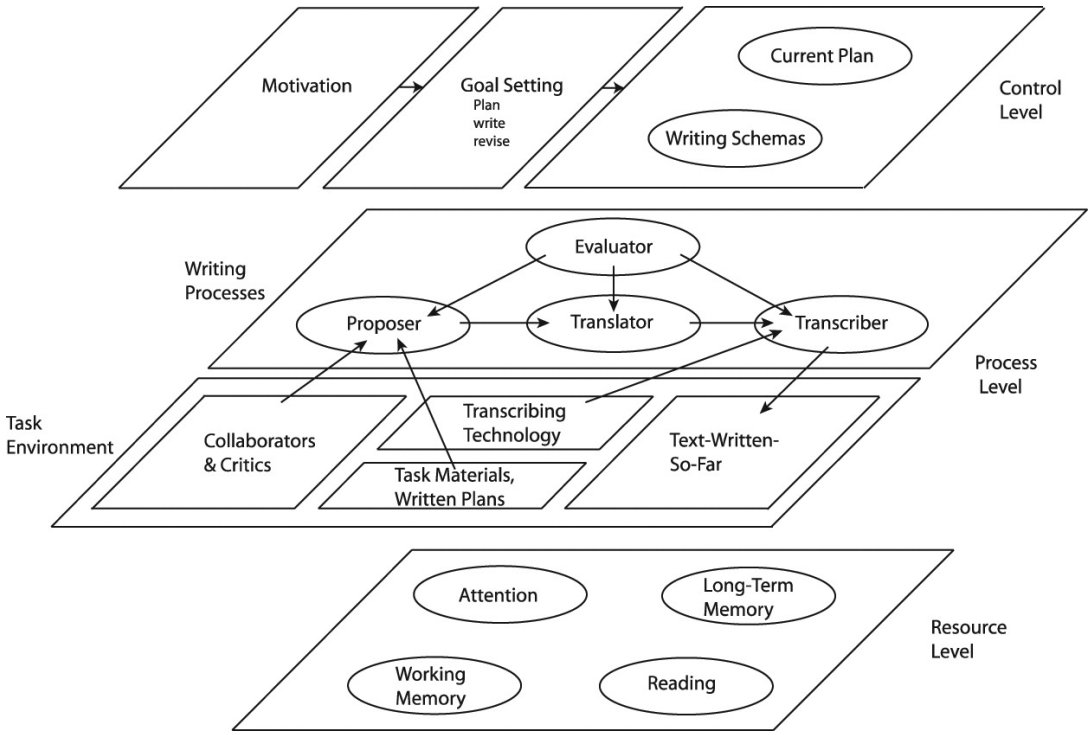
process can promote learning as it involves generating new ideas and making connections between new knowledge and prior knowledge in long-term memory (Applebee, 1984).

Translating. In the translating process, writers produce visible language corresponding to the ideas they have retrieved from long-term memory under the guide of the writing plan (Hayes & Flower, 1980; Flower & Hayes, 1981). This process is underspecified in the original model, as demonstrated by the empty box for translating in Figure 2.

Reviewing. The reviewing process can either be a conscious process or an unplanned action through evaluation of the emerging text (Flower & Hayes, 1981). Through the sub-processes of evaluating and revising, the reviewing process allows writers to improve the quality of their texts by identifying and correcting errors as well as review the extent that the text they are producing achieves the rhetorical goals.

In an updated model (Hayes, 1996), writing processes are thought of as taking place within working memory (Baddeley & Hitch, 1974). The roles of motivation and affect are also incorporated and there was some reorganisation of the cognitive processes. Galbraith, van Waes and Torrance (2007) identified the main differences as reflection replacing the planning process; text production replacing the translation process; and revision involving reading, reflection, and text production as opposed to being a distinct process (Hayes, 1996). The most recent update to the model (Hayes, 2012; Figure 3) maintains the overall writing process as a goal-directed, problem-solving activity, but planning, and reviewing and revision are no longer included as basic writing processes, and working memory, transcription skills, and motivation have been emphasised.

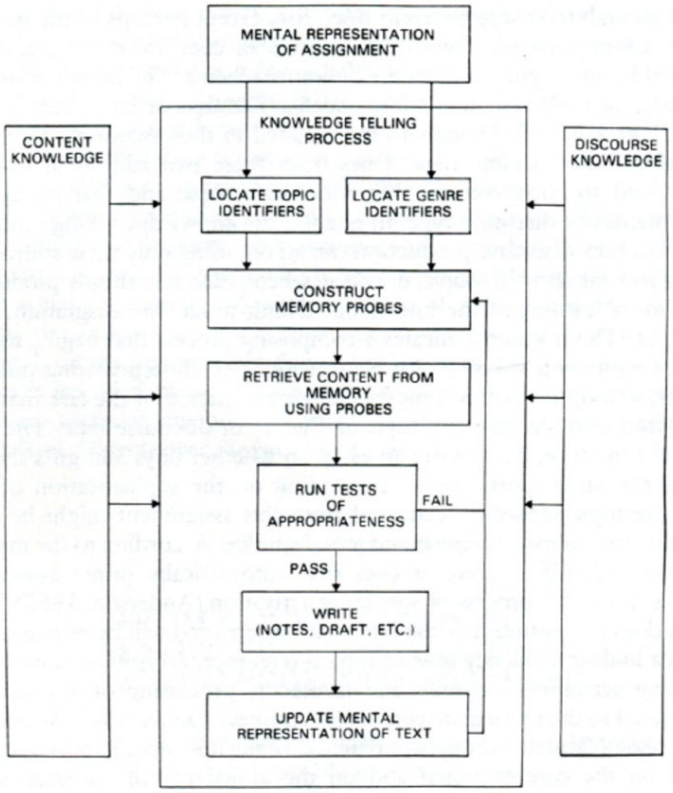
Figure 3. Hayes' (2012) model of writing. Reproduced with permission of the rights holder.



Hayes and Flower's (1980) model has been useful, and widely used, for identifying the key features of the writing process (Applebee, 1984), however it has received criticism for being too descriptive. Long-term memory, and the translating process, are not very well specified in the original model. In comparison, Bereiter and Scardamalia (1987) have proposed a developmental model of writing formed of two processes: knowledge-telling and knowledge-transforming. These processes can be seen as showing how younger, developing writers and older, more experienced writers implement Hayes and Flower's model.

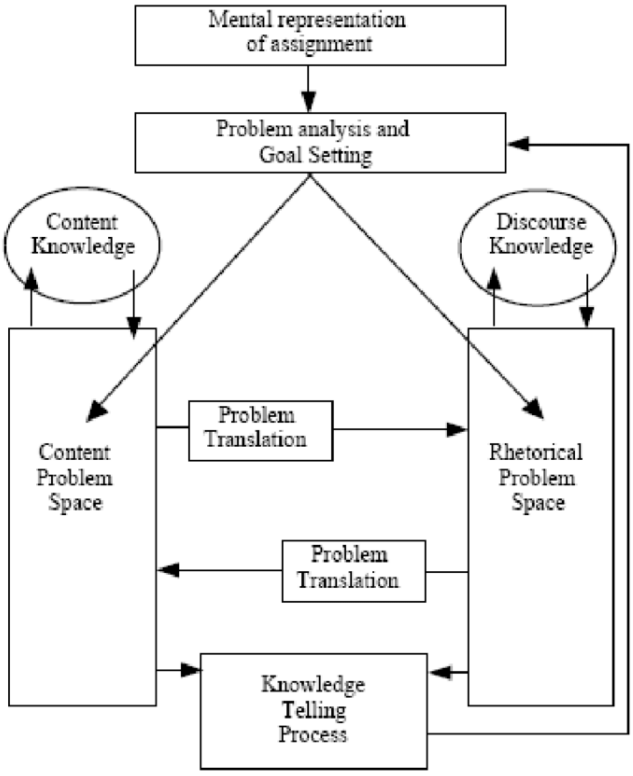
Knowledge-telling (Figure 4) is essentially the translating process in Hayes and Flower's (1980) original model. This process involves translating information retrieved from long-term memory into text, without elaborate organisation of the content or guidance by rhetorical goals (Bereiter & Scardamalia, 1987). Retrieval is assumed to be an automatic process through the spread of activation in memory. Galbraith (1999) described the knowledge-telling process as a 'think-say' method of text production, and because its role is simply to externalise ideas from memory, it is limited in its effect on the development of knowledge (Bereiter & Scardamalia, 1987; Atasoy, 2013). This is not to say that knowledge-telling does not produce good quality texts; if a writer has a clear representation of ideas already organised and easily accessible in memory, they can still produce a high-quality text through knowledge-telling.

Figure 4. Bereiter and Scardamalia's (1987) knowledge-telling process. Reproduced with permission of the rights holder.



The knowledge-transforming process (Figure 5) is distinguished by its focus on goal-directed problem-solving. It achieves this through the interaction and management of two 'problem spaces' (Bereiter & Scardamalia, 1987). The content problem space ("what do I mean?") comprises of the writers' beliefs, thoughts and ideas, and knowledge, and the rhetorical space ("what do I say?") comprises of the goals for the emerging text. When writers draw connections between these two spaces, they engage in reflective thought, and this is what enables development of knowledge.

Figure 5. Bereiter and Scardamalia's (1987) knowledge-transforming process. Reproduced with permission of the rights holder.



Knowledge-telling, as a translation process, is assumed not to involve learning. Discovery of knowledge is assumed to occur through knowledge-transforming (Galbraith, 1999). This is a specific theory of how learning occurs as a consequence of goal-directed problem-solving.

Bereiter and Scardamalia (1987) conducted empirical research on the knowledge-telling and knowledge-transforming processes by comparing the think-aloud protocols of younger, novice writers and older, expert writers. Knowledge-telling is characteristic of novice writers who simply translate ideas retrieved from memory into written text. Knowledge-transforming is characteristic of expert writers who produce more elaborate texts, directed by rhetorical goals. Experts use the planning process to formulate goals and organise content for the emerging text (Flower & Hayes, 1980).

The classic models of writing (Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) have been critiqued for overlooking the writer's development of understanding. They take for granted that writing will inevitably lead to knowledge discovery (Applebee, 1984; Baaijen & Galbraith, 2018), without empirical studies measuring the writer's understanding or knowledge change. Furthermore, the models describe writing processes and the interactions amongst them, but do not adequately explain how knowledge is represented in long-term memory (Galbraith & Baaijen, 2018). They also have little to say about how thoughts and ideas are constituted as the

text is produced (Galbraith, van Waes & Torrance, 2007). In light of the criticisms of these classic models, the dual-process model of writing (Galbraith & Baaijen, 2018) was presented. This model centres around the development of knowledge through writing and provides a more adequate explanation of how knowledge is represented in memory.

2.2 The Dual-Process Model of Writing

Galbraith and Baaijen's (2018) dual-process model of writing explains the writing process as a process of knowledge discovery. It accepts the claims of the classic problem-solving models (Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) but proposes that writing is not just a process of translating thoughts and ideas from memory to achieve rhetorical goals, but is an active knowledge-constituting process in its own right (Galbraith et al., 2025). The model is based on parallel-distributed processing theory and connectionist principles to provide a more detailed explanation of how knowledge is represented in long-term memory (a clear strength of the model over the classic models of writing).

Parallel-distributed processing theory (Rumelhart et al., 1986) proposes that knowledge is stored implicitly in memory through the strength of connections between sub-symbolic neural units. As activation moves through these units, information is synthesised (Galbraith & Baaijen, 2018). Learning strengthens and modifies the connections over time, and therefore these connections represent the writer's disposition towards the topic (Galbraith & Baaijen, 2018).

The parallel-distributed processing approach effectively captures the representation of a writer's general knowledge, but it tends to overlook the specific experiences that originally informed those representations. Consequently, the dual-process model also adopts Complementary Learning Systems theory (McClelland, McNaughton & O'Reilly, 1995; Kumaran et al., 2016; O'Reilly et al., 2011) which distinguishes between the episodic and the semantic memory systems. The semantic system captures the knowledge representation of parallel-distributed processing theory, consisting of an individual's general knowledge extracted from their experiences. This system therefore supports recognition and reasoning, allowing learners to make inferences and solve problems. The episodic system stores memory of specific events and experiences, and therefore supports retrieval processes.

According to the dual-process model (Galbraith & Baaijen, 2018), the writing process is an interaction between the semantic and episodic memory systems. The semantic system is *"a system designed for action"* (Galbraith et al., 2025, p. 291), where the writer's thoughts and ideas are synthesised (or constituted) during composition. The episodic system is *"a system designed for reflection"* (Galbraith et al., 2025, p. 291), where the writer's thoughts and ideas are

retrieved from memory and evaluated as to how they meet the rhetorical goals for the text. Galbraith and Baaijen (2018) argue that the knowledge-transforming process operates on knowledge representations in episodic memory, and knowledge-constituting takes place within the semantic memory system.

The knowledge-constituting process (Galbraith, 1999; Galbraith & Baaijen, 2018), implies that the writer's disposition is not explicitly retrievable but guides the synthesis of information during writing. This initiates further activation of semantic units in memory. As ideas are written down and synthesised, the writer's knowledge implicit within these units becomes more explicit. For this reason, their understanding increases as a result of knowledge discovery. Galbraith (1992) found that increases in participants' subjective knowledge were experienced by writers who direct their texts towards dispositional goals when they planned their text synthetically (summing up the main idea of the text). This suggests knowledge development is associated with a spontaneous, dispositionally guided writing process, in line with the knowledge-constituting process.

The knowledge-transforming process (Galbraith & Baaijen, 2018) reflects the retrieval of information stored in episodic memory which is then evaluated according to the writer's rhetorical goals in working memory. This is different to the knowledge-telling process in Bereiter & Scardamalia's (1987) model as it involves a more thoughtful search than merely recalling content from memory without evaluation of rhetorical goals. Knowledge-transforming is responsible for producing good quality texts as it allows the writer to review rhetorical constraints. This is supported by studies showing organisation and rhetorical goals are associated with improved text quality (Galbraith et al., 2005, 2009). Developments in understanding are not associated with knowledge-transforming, as the process is responsible for evaluating retrieved content to rhetorical goals rather than generating new knowledge.

Both the knowledge-constituting and knowledge-transforming processes are required for effective writing which aids learning and produces good quality text (Galbraith & Baaijen, 2018). However, because of the difference in the way knowledge is organised in the two processes, this leads to potential conflict during writing. The retrieval and evaluation of ideas aligned to rhetorical goals in the knowledge-transforming process can prevent a writer from developing their understanding; and dispositionally guided text production in the knowledge-constituting process can prevent the production of an effectively organised and rhetorically guided text. When a writer is able to effectively manage this conflict, they are able to both write a text which satisfies rhetorical goals and develops their implicit understanding of the topic (Galbraith & Baaijen, 2018). For this reason, the dual-process model would promote a writing process whereby first, the writer produces a rough, spontaneous first draft which allows them to

constitute their understanding, followed by identifying key ideas and a plan to revise the text so it achieves rhetorical goals (Galbraith & Baaijen, in press).

The main empirical support for the dual-process model comes from a study by Baaijen and Galbraith (2018). 78 undergraduates, who were either high or low self-monitors, wrote either an outline or synthetic plan for 5 minutes, before spending 30 minutes writing a well-structured newspaper article. Participants rated their subjective knowledge before and after writing, and the quality of their texts was assessed. The researchers also analysed participants' writing process using keystroke logging to identify two component measures: 1) global linearity, representing the extent to which participants produced linear or non-linear texts; and 2) sentence production, representing the extent to which participants' text production was controlled or spontaneous.

The findings of Baaijen and Galbraith's (2018) study indicated that increases in knowledge were associated with synthetically planned spontaneous text production, and non-linear revision of the global structure. This was further supported by a negative relationship between global linearity and change in subjective knowledge. Furthermore, outline planning prevented spontaneous writers from developing their understanding. Presumably this was because the pre-determined structure from the outline plan supported a writing process involving simply translating ideas from episodic memory. The best quality texts were produced by participants who had more controlled sentence production and non-linear global revision. In summary, the results support the dual-process model as discovery of knowledge was related to synthetically planned, spontaneous text production (knowledge-constituting) and reorganisation of the global structure (knowledge-transforming), with the latter associated positively with text quality. This effect has been replicated in the doctoral thesis of Hall (2023).

The main strength of the dual-process model is its explanation of how knowledge is represented in memory, and there is some empirical support for its claims. The knowledge-transforming process represents the retrieval of pre-existing information from episodic memory and translates them into written text. It is not a process of knowledge discovery since rather than formulating new information, retrieved ideas are reorganised towards the goals for the emerging text. Knowledge discovery occurs through knowledge-constituting which is reflected by spontaneous sentence production and non-linear global revision. The knowledge-constitution process is the active synthesis of information guided by units in semantic memory. However, research so far investigating the dual-process model (Baaijen & Galbraith, 2018) has focussed on writing from prior knowledge (e.g., writing about self-chosen topics). Writing is rarely completed as an isolated activity, it often involves the pre-reading of source material(s),

especially in higher education. Therefore, the dual-process model needs to be explored in the context of source-based writing.

In summary, classic models (e.g., Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) describe writing as a problem-solving process, guided towards rhetorical goals for the text. They implicitly assume that this will result in learning, but do not provide an adequate explanation of how this takes place, nor about how knowledge is represented in long-term memory. In response, the dual-process model of writing (Galbraith & Baaijen, 2018) proposes knowledge-transforming and knowledge-constituting processes, which are represented in episodic and semantic memory respectively. Evidence for the dual-process model comes from Baaijen and Galbraith's (2018) study in which participants wrote from their personal knowledge. This thesis aims to explore the claims of the dual-process model when writers construct text based on source material. It is possible that source-based writing offers less opportunity for learning since the writing process may be limited to reporting information from sources which already have a pre-defined structure. The next section explains the current empirical context on writing-to-learn from sources.

2.3 Writing-to-Learn from Sources

Source-based writing is a common activity in educational settings and involves writing about pre-read material. A source-based approach has been growing steadily in the writing-to-learn literature and many studies now use reading to write tasks due to their ecological validity (Wiley et al., 2014; Klein, 1999). Higher education in particular places emphasis on students' reading comprehension and their capability to integrate sources to produce written assignments (Rouet et al., 1996; Cumming, 2013). Despite this, students often experience difficulties with writing from sources (Barzilai & Strømsø, 2018), although they are able to develop strategies especially in response to pedagogical instruction (Cumming, Lai & Cho, 2016).

A fundamental feature of source-based writing is the reading-writing relationship (Hayes, 1996), and as such activities can be described as 'hybrid' tasks (Spivey, 1990; Klein et al., 2014). Reading provides writers with topic knowledge, as well as an understanding of the message of the original author and a representation of text structure (Hayes, 1996). Furthermore, Fitzgerald and Shanahan (2000) argue reading and writing have similar cognitive processes. The learner, as a reader and a writer, has metacognitive knowledge (about the general aims of reading and writing, and about how to monitor their own knowledge); topic and subject knowledge; knowledge about texts; and knowledge about reading and writing skills.

Source-based writing has the capability to affect learning as writers must select content from the source, use their pre-existing knowledge to organise this content, and create links between the source and their own knowledge as a new text emerges (Spivey & King, 1989). Two meta-analyses provide evidence for the learning effects of writing from sources. First, Graham and Hebert (2011) found that writing about source material enhanced understanding of source material for students in Grades 2 to 12, including students identified as struggling writers. This finding was true for a range of text types, including expository and narrative texts, as well as for a range of subject areas. Several types of source-based writing activities were shown to be effective, including summary writing, extended writing, note-taking, and answering or generating questions.

Graham and Hebert (2011) suggested writing can aid reading comprehension as it provides an external and permanent product to record ideas. Writing from reading also encourages learners to produce a text for an audience (e.g., in response to rhetorical goals). Klein, Haug and Arcon (2017) found that when students in Grades 5 to 7 were prompted with a rhetorical goal when writing source-based texts, they produced better quality text and improved scores on a multiple-choice post-test, compared to students prompted to construct valid reasons and a control group not receiving any prompt.

A more recent meta-analysis by Graham, Kiuhard and MacKay (2020) found source-based writing in Grades 1 to 12 improves learning in science, social studies, and mathematics, with a statistically significant average weighted effect size of 0.3. The authors noted that writing-to-learn literature often provides little information about what students' actually do during writing, but identified the main activities as writing answers to open-ended questions about sources, creating descriptions or explanations, writing summaries, and metacognitive reflection. Hebert et al. (2014) and Hebert, Gillespie and Graham (2013) also found source-based writing can enhance reading comprehension, but argue this typically depends on the activity. Research on source-based writing typically focuses on summary writing (of one or more texts), and synthesis writing (of multiple texts).

2.3.1 Summarising

A summary can be defined as "a brief statement that represents the condensation of information accessible to a subject and reflects the gist (central ideas or essence) of the discourse" (Hidi & Anderson, 1986, p.473). More simply put, a summary is a "condensation of the topic's main points" (Spirgel & Delaney, 2016, p.171). Summarising can enhance learning by requiring the learner to select the most important content from a source and condense it into a clear, focused overview of the topic (Friend, 2002; Wade-Stein & Kintsch, 2004). This

reconstruction and connection of ideas can help individuals gain a better understanding (Hidi & Anderson, 1986; Gelati, Galvan & Boscolo, 2014) as they integrate new information with their existing prior knowledge (Spirgel & Delaney, 2016).

Similar to other types of source-based writing, summary writing has been found to be a difficult activity (Hidi & Anderson, 1986). Older high-school and university students have been found to be better at writing summaries than younger children because of their ability to select the most important information from a source text, plan, and reduce the information into a summary (Brown, Day & Jones, 1983). When younger children are asked to produce summaries, they tend to use a copy-delete process comparable to the knowledge-telling process in Bereiter & Scardamalia's (1987) model (Brown & Day, 1983). However, summarising a text has been found to be a useful learning activity for enhancing comprehension and recall of the text in children. Doctorow, Wittrock and Marks (1978) found asking pupils in Grade 6 to produce sentences about short story paragraphs improved both recall of information and comprehension. A meta-analysis on the teaching of writing carried out by Graham and Perin (2007) recommended summary writing should be taught to children and adolescents because it can enhance their skills in accurately representing information they have read in their writing. Teaching summary writing (e.g., Gelati, Gelvan & Boscolo, 2014), as well as the development of tools to support the production of summaries (e.g., 'Summary Street'; Wade-Stein & Kintsch, 2004) have been shown to be effective in educational settings.

Another study providing support to the idea that summary writing can enhance learning was conducted by Thiede and Anderson (2003), although their study shows this can depend on the nature of the summary. Undergraduate students were randomly allocated to either a no-summary group (read six texts and completed a comprehension test), a delayed-summary group (read six texts, wrote a summary of each, and completed a comprehension test), or an immediate-summary group (read one text and wrote a summary of it, followed by each of the six texts, and then completed a comprehension test). The participants also rated their own comprehension of the text. The results showed that writing summaries increased participants' meta-comprehension accuracy, but only for those in the delayed-summary group; a finding that also replicates to at-risk readers (Thiede et al., 2010).

Summarising has also been found to be advantageous over other types of writing such as argument writing. Gil et al. (2010) found participants who wrote summaries received higher scores on a test of inferential comprehension than participants who wrote argument texts. Summary writers also used more of the information from the source text, and integrated and transformed this information more than argument writers. Despite these findings, Le Bigot and Rouet (2007) found contradicting evidence in that argument writing resulted in more

transformations and connections of information than summary writing. This is not to say that summarising is not an effective learning tool; Delaney (2008) suggests such writing activities could reflect different dimensions of the reading-writing relationship. In a study with 139 undergraduates, Delaney found writing an essay in response to a prompt invoked more critical thinking whereas writing a summary involved selecting important content and organising and condensing it.

Most of the literature on summary writing compares it to re-reading, or other types of writing, as opposed to non-writing generative activities such as speaking. Although their paper is mostly focused on the development of a new methodological tool for measuring learning from source-based writing, Silva and Limongi's (2019) study provides interesting findings on the comparison between written and spoken summaries. In eight experimental blocks, participants read a short text passage, wrote or spoke a summary of it, and completed an episodic memory word-recognition task in which they had to identify whether target words had appeared in the original text passage. Silva and Limongi used drift-diffusion modelling to compare participants' reaction time (RT) on the word-recognition task after writing and speaking summaries. They found that participants' RTs were faster after writing summaries than after speaking them, and this was explained by a parameter of the drift diffusion model representing encoding and motor processes (non-decision time). It was therefore concluded that the superiority effect of writing summaries was associated with fast retrieval of information from episodic memory.

Despite the favourable evidence for the effects of writing summaries on learning, a review of the literature by Spirgel and Delaney (2016) claimed "summary writing is usually a passive and nonreflective process" (p.172). Furthermore, Dunlosky et al. (2013) considered summarisation as a low-utility tool for learning. Spirgel and Delaney argued that although summarising helped individuals remember information, it was not a particularly good learning tool relative to re-studying information. They concluded from this that writing a summary is not a useful learning technique when learners do not know what to write about and favoured writing strategies such as essays which promote knowledge transformation. This notion is consistent with models of writing-to-learn claiming that for learning to occur, writers must transform their knowledge instead of retrieving information in a knowledge-telling process (Bereiter & Scardamalia, 1987; Galbraith & Baaijen, 2018).

2.3.2 Synthesising

In contrast to the potentially passive nature of summarising, synthesis writing may offer a greater potential for writers to constitute and transform their knowledge. Synthesis writing can be defined as "a type of source-based writing that requires writers to synthesise the information

from different sources to a new and meaningful text" (Vandermeulen, van Steendam & Rijlaarsdam, 2022, p.747). In producing a synthesis, a writer has to plan and establish an organisational structure which combines information from multiple sources, which can encourage knowledge-transforming (Mateos et al., 2014). In comparison, when producing a summary, writers can often adopt the same structure as the source texts and therefore reduce the process to retrieving and translating information from memory (knowledge-telling) (Mateos et al., 2014).

Synthesis writing is a cognitively demanding task (Spivey & King, 1989; Segev-Miller, 2004; McGinley, 1992; Primor & Katzir, 2018). The process of writing a synthesis text requires organising, selecting, and connecting to achieve goals for the emerging text which can function independently of the source materials (Nelson & King, 2022). In the organising process, writers must assess the structure of the source materials and develop their own organisation; during selecting, writers must identify information from the source materials to include in their synthesis; and in connecting, writers must create connections between the source materials and their own prior knowledge.

Research supports the beneficial effects of synthesis writing on learning through knowledge transformation processes (organising, selecting, and connecting). Wiley and Voss (1999) found that when undergraduates wrote argumentative syntheses after reading multiple history texts, their performance was better on comprehension and analogy tests than undergraduates who wrote summaries, narratives, or explanations. Writing an argument synthesis encouraged students to elaborate on, organise, and integrate information. Castells, Minguela and Nadal (2022) also explored synthesis writing in undergraduate students and found improved text comprehension for students who wrote a better synthesis with respect to organisation and relevance and accuracy of ideas.

Furthermore, Vandermeulen et al. (2020) conducted a large study with 658 Dutch upper-secondary school students who wrote four synthesis texts (two argumentative texts and two informative texts) based on three to five source materials. The best quality texts were produced by older students who revised and switched between their own text and the sources more often, and wrote more fluently. In another study of 15- and 16-year-olds (Solé et al., 2013), engaging in elaborative processes through writing a synthesis of three history texts produced both better quality texts and enhanced comprehension of material. Participants who wrote poorer syntheses texts typically engaged in a more linear writing process. Together, these studies reinforce the process of integration as fundamental to effective learning from synthesis writing.

2.3.3 Learning by Explaining

Although much of the literature on writing from sources focuses on summarising and synthesising, research on learning by explaining also provides valuable insights on the benefits of writing from sources. Early research suggested that asking learners to self-explain concepts they read about can improve their understanding (Chi et al., 1994). Similarly to the process of synthesis writing, it is thought that producing an explanation requires learners to integrate, organise, and elaborate on source material to connect it with already existing ideas in their knowledge (Coleman, Brown & Rivkin, 1997; Roelle et al., 2023; Ainsworth & Burcham, 2007; Pan et al., 2024). Coleman, Brown and Rivkin (1997) found self-explaining produced better outcomes on a transfer test than summarising. Hoogerheide et al. (2016) suggest this is because explaining involves reflective knowledge-building where the learner has to monitor their own understanding, whereas summarising involves knowledge-telling. In addition, writing an explanation can be spontaneous, shaped 'at the point of utterance' (Britton et al., 1975), or directed towards rhetorical goals (Rivard, 2004).

One current approach building on the learning by explaining literature to encourage learning from writing is the self-regulated journal writing approach (Nückles et al., 2020). In their research, Nückles and colleagues ask students to produce learning journals, which are reflections on source materials from lectures or courses (Hübner, Nückles & Renkl, 2010). Laboratory and field studies have found that writing learning journals is beneficial for both comprehension and conceptual knowledge (Nückles et al., 2020). The production of learning journals encourages writers to engage with cognitive and metacognitive strategies, and therefore supports a self-regulated view of writing-to-learn. Writers can apply the effective cognitive processes of organisation and elaboration (Nückles et al., 2009), as well as metacognitive processes of monitoring and regulation. This helps them to identify gaps in their understanding (Chi et al., 1989). Writing also allows the learner to externalise and preserve their thoughts which further supports metacognitive processes (Nückles et al., 2020).

Writing activities that encourage metacognitive processes are important for enabling learners to monitor their own understanding. Individuals are typically poor at judging their performance in cognitive activities (Mills & Keil, 2004) and are often overconfident when evaluating their own comprehension of material (Glenberg, Wilkinson & Epstein, 1982). For example, Glenberg and Epstein (1985) found participants' predictions of their performance in on a recall task was poorly related to their actual performance.

Following a series of studies on self-explaining, Rozenblit and Keil (2002) introduced the 'illusion of explanatory depth' as the idea that *"most people feel they understand the world with far greater detail, coherence, and depth than they really do"* (p.522). The basic procedure of these

studies was as follows: participants (undergraduates) are asked to rate their understanding of a device (e.g., how a sewing machine works), phenomenon (e.g., how the heart pumps blood), or procedure (e.g., how to make pasta). They then write down a detailed explanation of the device, phenomenon, or procedure, before rating their understanding again. The consistent finding was that participants attempts to explain led to a decrease in their understanding for devices and phenomena (facts), but not for procedures (narratives). This result has been replicated in children. Mills and Keil (2004) asked kindergarteners, second graders, and fourth graders to rate their knowledge of devices (e.g., a piano, toaster, tricycle) and procedures (e.g., making a pizza, fishing) and give a spoken explanation of them before re-rating their knowledge. Children also experience the illusion of explanatory depth for devices but not for procedures. This difference is likely due to the fact that explanatory knowledge (e.g., of devices) involves complex causal patterns (Rozenblit & Keil, 2002; Keil, 2003).

Attempting to explain something can help to develop a conceptual framework which helps learners to realise they have an incomplete understanding (Keil & Wilson, 2000). Johnson, Murphy and Messer (2016) suggest the decrease in understanding ratings through the illusion of explanatory depth reflect knowledge calibration as a result of metacognitive accuracy. Rozenblit and Keil (2002) and Mills and Keil (2004) propose a number of factors which may contribute to this illusion. First, self-explaining is a difficult task that people have little experience of undertaking. This means they have little knowledge of what a successful and unsuccessful explanation looks like. Second, explanatory knowledge is complex and not as well-defined as other kinds of knowledge. It is possible that people confuse their broad, high-level understanding with the lower-level detailed understanding. Third, people may experience a confusion between their internal and external representations of their knowledge, such that the support from the written product can reveal gaps in their understanding.

Overall, an important reason as to why explaining can be effective for learning is because it combines generative learning and retrieval practice (Roelle et al., 2023), which support both cognitive and metacognitive processes. This is typically true of closed-book situations where learners are expected to generate their explanations without access to the source material(s). Much of the literature on writing from sources has focused on writing in open-book situations where learners can revisit the source material during writing. This reduces the direct memory load of the writing task (Hidi & Anderson, 1986), allowing working memory resources to be used for higher-level processes such as elaboration and organisation. Closed-book source-based writing is a common task in educational settings, although it is typically used as an assessment tool. However, a wide body of literature on the beneficial effects of retrieval practice has shown that closed-book environments also provide the opportunity for learning.

2.4 Retrieval Practice and Writing

Retrieval practice simply involves recalling information from memory. Traditionally, activities involving retrieval are used in educational settings as assessment methods (such as comprehension tests) (Blunt & Karpicke, 2014). However, a wide body of literature on 'the testing effect' has found that engaging in retrieval practice can also enhance the retention of information and create meaningful learning (e.g., Glover, 1989; Roediger & Karpicke, 2006a, 2006b, Karpicke, 2009, 2012; Butler, 2010; Carpenter, 2012).

Studies consistently find that completing a recall test improves long-term retention and transfer of information compared to restudying or re-reading (Karpicke, Butler & Roediger, 2009). In Roediger and Karpicke's (2006a) first experiment, 120 undergraduates studied short text passages, and then either took a recall test about the text passages or re-studied it, before taking a final test of retention after five minutes, two days, and one week. Initially, after five minutes, participants in the re-study condition performed better on the final test than participants in the recall condition. However, after two days and one-week, this result was the opposite: participants in the recall condition (56% recalled after one week) achieved better results on the final test than participants in the re-study condition (42% recalled after one week). The findings were supported by a second experiment in which 180 undergraduates were randomly allocated to one of three conditions: 1) studied a text once and then took three tests; 2) studied a text three times and then took one test; or 3) studied a text four times. Participants indicated how well they thought they would remember the text one week later, and then took a final retention test after five minutes and after one-week. Although after five minutes those who studied the text four times performed best; after one week, participants who studied the text once and took three tests performed better than participants in the other two conditions.

Further evidence for these results was established by Karpicke and Roediger (2007) who found that participants who were tested recalled more items on word lists one week later than participants who repeatedly studied. Butler (2010) found not only did testing result in better retention, but it also supported transfer of learning to related concepts in different subjects. A recent review of the literature on retrieval practice and the testing effect (Pan et al., 2024) concluded that it is one of the most effective learning activities across age groups, for a wide range of materials, and in diverse contexts. Additionally, there is strong evidence from meta-analytic reviews that testing in the classroom is beneficial for learning (Yang et al., 2021). More recent studies have suggested that spaced retrieval practice (distributing testing over multiple sessions) (e.g., Higham et al., 2023) and successful relearning (combining retrieval practice and spaced practice with corrective feedback) are particularly effective (e.g., Higham et al., 2021).

It is widely believed that retrieval practice supports learning because recalling knowledge from memory strengthens activations between units, making it easier for the learner to recall the same knowledge in future (Karpicke, 2012). Retrieval practice is seen as a desirable difficulty (Bjork, 1975, 1998), meaning that although retrieval can initially be a challenging task, it results in long-lasting and transferable learning (Murphy, Little & Bjork, 2023). Testing can support the development of retrieval plans (or 'schemas') on the basis of a learner's knowledge and their attempts at retrieval (Zaromb & Roediger, 2010). In sum, these are known as the direct benefits of retrieval practice.

Retrieval practice can also have indirect benefits due to the role it can play in metacognitive processing. Retrieval fluency is the speed or probability that information is accessed and recalled from memory (Benjamin, Bjork & Schwartz, 1998). Research (e.g., Koriat, 1995; Bjork, 1998; Karpicke, 2009) suggests that students often base their judgements of learning or comprehension on retrieval fluency and the ease with which they feel they can process the information. Little and McDaniel (2015) found university students' predictions about their future performance were more accurate when they engaged in retrieval after reading text passages than re-reading. These students used such predictions to allocate the amount of time they would spend re-studying the material. Glenberg and Epstein (1987) also showed that judgements of learning were used to help students decide if they were ready to be assessed or if they needed to continue to revise, and Barenberg and Dutke (2019) found retrieval could act as a meta-memory cue to encourage learners to reflect on the quality of their knowledge and understanding. This reinforces the educational value of retrieval practice not only for the direct benefits to outcomes on recall and comprehension tests but also for the indirect benefits of gaining an understanding about one's own learning.

Even though retrieval practice has clear and well-established benefits for learning, these are not always recognised by students and teachers. It has been found that students largely believe restudying to be more effective than testing. Karpicke (2009) observed that whilst studying, undergraduates typically chose to restudy material instead of testing themselves on the material. Whilst testing, these students often disregarded information once they could retrieve it once rather than undertake further testing. A large survey of 177 undergraduates also implied that retrieval practice is not viewed as an activity that can support learning (Karpicke, Roediger & Butler, 2009). Only 11% of participants reported using retrieval practice as a study strategy. For those that do test themselves during study sessions, it is typically as an assessment tool to gain feedback about their current knowledge level rather than because they understand the task itself to promote learning.

Combining retrieval practice with generative learning activities may be especially advantageous for learning. Generative learning is "the idea that learning involves actively constructing meaning from to-be-learned information by mentally reorganising it and integrating it with one's existing knowledge" (Fiorella & Mayer, 2015, p. 717). In the previous section, self-explaining was discussed as an effective strategy for learning, and this is an example of a tool which has high retrieval and generative learning processes (Roelle et al., 2023). Hinze, Wiley and Pellegrino's (2013) third experiment found explaining produced better performance on a detail and inference test than free recall or re-reading because of the constructive processes it encourages. Engaging in retrieval practice through a generative learning activity could maintain the cognitive benefits of enhancing performance on future tests, but also promote further the metacognitive benefits such as monitoring one's own learning to identify knowledge gaps (Roelle et al., 2023).

Waldeyer et al.'s (2020) study demonstrates the potential value of combining generative learning with retrieval practice. 59 ninth-grade students in a German high-school read an expository text, and then completed a generative learning activity in either an open-book condition (not requiring retrieval) or a closed-book condition (requiring retrieval). In the activity, students received prompts to elaborate and organise information from the text, such as what information they considered to be the most important, interesting, useful, or convincing. Learning was measured by the number of idea units, degree of organisation, and quantity and quality of elaboration in the generative learning task, and by performance on a delayed post-test one week later. Although participants in the open-book condition were able to include more information from the sources in the generative learning activity, they remembered the information less well than participants in the closed-book condition, who included less material from the sources in the generative learning activity. Waldeyer et al.'s second experiment revealed allowing participants to switch between an open- and closed-book environment gave them the opportunity to benefit from both retrieval practice and generative learning. Similarly, Arnold et al.'s (2021) study showed that when undergraduate students wrote essays in a closed-book context, they benefited from retrieval practice, whereas in an open-book context, they benefited from generative learning through elaboration and organisation processes.

According to Fiorella and Mayer's (2015) definition of generative learning, writing-to-learn can be considered as a generative learning activity. Therefore closed-book source-based writing combines generative learning and retrieval practice. Generating a written text from memory based on the reading of source texts may therefore enhance retention of factual information but also allow a writer to monitor their learning using an external and elaborated representation of their knowledge.

Few writing-to-learn studies have considered retrieval processing to be involved in the effects of writing on learning, and this idea has been absent from recent reviews on writing-to-learn (e.g., Klein & Boscolo, 2016). Nevertheless some studies have acknowledged retrieval practice in writing-to-learn. In Klein's (2004) study, around 50 university students were asked to explain a scientific phenomenon, conduct an experiment on the phenomenon, explain the phenomenon again, and then write a journal entry about the experiment before providing another explanation. Think-aloud protocols from participants revealed that they engaged in problem-solving processes, but also in automatic and controlled retrieval, both of which supported learning. For example, writing down a thought or idea helped to trigger participants to reject incorrect ideas (automatic retrieval), but writers also intentionally looked for thoughts and ideas in long-term memory (controlled retrieval).

Arnold et al. (2017) proposed that retrieval practice may underlie the cognitive mechanisms by which source-based writing facilitates learning. 100 undergraduates were randomly assigned to one of four conditions, two of which required retrieval processes (essay and free recall writing) as they took place in a closed-book environment, and two of which did not require retrieval processes (highlighting and notetaking) as they took place in an open-book environment. Participants completed their assigned activity for two source text passages on astronomy, and then returned two days later to complete a text comprehension test comprised of factual and inferential MCQs and problem-solving short-answer questions (SAQs). Better text comprehension performance was achieved by participants who had completed the essay and free recall writing activities requiring retrieval. Although this study demonstrates the value of closed-book source-based writing as a retrieval activity, learning is assessed with a text comprehension test about the source texts (i.e., a measure of response accuracy). It did not consider how participants' subjective knowledge may develop as a result of the activities.

To summarise, theories on the cognitive processes in writing-to-learn span from the classic problem-solving models (Hayes & Flower, 1980; Bereiter & Scardamalia, 1987), to the more recent dual-process model of writing (Galbraith & Baaijen, 2018). So far, the dual-process model has only been evidenced by studies where participants write about their own knowledge (Baaijen & Galbraith, 2018; Hall, 2023). This thesis assesses the claims of the model in the context of source-based writing. Researchers have considered source-based writing as an effective learning strategy, especially when it involves processes of elaboration, organisation, and integration, for example in writing syntheses from multiple sources in an open-book environment. However, the learning by explaining, retrieval practice, and generative learning literature suggests that closed-book source-based writing-to-learn can have value for learning, even in single source environments. The limited research on the cognitive processes underlying closed-book source-based writing-to-learn is the motivation behind this project.

Chapter 3 Overview of Methodology

The two overarching aims of this project are: 1) to investigate the cognitive processes underlying source-based writing to assess the claims of the dual-process model (Galbraith & Baaijen, 2018) in a new context; and 2) to develop and compare different measures of writing-to-learn. These aims are achieved in four independent empirical research papers which report the findings of six studies.

The six studies all have an experimental design in line with the underlying research paradigm for this project as quantitative and positivist. This is because the aim of this thesis is to observe and measure cognitive processes and learning from writing. Quantitative methods offer several advantages, including generalisability and representativeness of a specific population, as well as standardisation of methods (Cohen, Manion & Morrison, 2018). An experimental design also allows for a degree of objectivity in the research and allows the cognitive learning processes to be explored under controlled conditions. In principle, experimental research enables the testing of causal hypotheses; however, in practice, more complex experimental designs involving production tasks can make it more difficult to track causality.

The next sections focus on the common methodological features across the four papers. The measures used to assess learning from writing and analyse the writing process are briefly outlined. The themes of replication and reproducibility that underlie this project are also described. The sampling and recruitment of participants is explained, as well as ethical considerations. A note on the impact of the Covid-19 is included due to the influence this had on the forced online data collection for Papers 1 and 2.

3.1 Measuring Learning from Writing

In order to assess the effects of learning from closed-book source-based writing, this thesis adopts several different measures of learning and assesses how they relate to one another. This includes text comprehension, as a measure of how well the participants have understood the source texts (Papers 1, 2, and 4); subjective knowledge, as a measure of how much the participants feel they have developed their understanding (Papers 1, 2, and 4); and finally reaction time (RT) on a word-recognition test, as a measure of long-term memory consolidation (Papers 3 and 4). An important aim, fulfilled most strongly in Paper 4, is to estimate how these different measures of learning, which have typically been used independently in past research, are related to one another. This also provides an assessment of the validity of the different measures.

Using multiple measures to assess learning from writing is advantageous as learning is multifaceted. Each measure can capture different dimensions of learning. Considering both response accuracy measures like text comprehension as well as the subjective development of knowledge experienced by participants can help to distinguish between 'reproducing knowledge' and 'seeking meaning' (Säljö, 1979). Reproducing knowledge views learning in terms of acquiring factual information or memorising and recalling knowledge. Seeking meaning views learning as a process in which understanding is developed. It is expected that these are moderately related, as they reflect distinct yet potentially overlapping dimensions of learning.

It is important to note here that this thesis purposefully does not include a measure of text quality, which has been dominant in writing-to-learn research. Assessing text quality aligns with the idea of writing as a *product*. The central focus of this thesis is on writing as a *process*. Furthermore, the writing tasks participants complete do not encourage them to produce good quality texts, thus evaluating the quality of their texts would not be a meaningful measure.

3.1.1 Text Comprehension

In writing-to-learn research, learning has typically been conceptualised as "the ability to retrieve and apply the learned information over time and in different contexts" (Murphy, Little & Bjork, 2023). Ackerman's (1993) review of 35 studies on writing-to-learn found recall and comprehension measures to be the most dominant in the research field. A methodological advantage and attraction of exploring source-based writing is that it provides something explicit that knowledge can be tested against. This is also ecologically valid as university environments often assess whether students have retained the information from materials they have read. The materials from which the text comprehension tests in this thesis are based on are expository source texts. These are the types of sources typically used in higher-education settings and are therefore appropriate in the context of this thesis where participants are university students.

The text comprehension tests used in this thesis include both multiple-choice questions (MCQs) and short-answer questions (SAQs). It was important that the tests were not comprised purely of MCQs; open-ended SAQs eliminate the possibility of guessing and could be a more valid indicator of comprehension (Weigle, Yang & Montee, 2013). In Papers 1 and 2, the source text and associated text comprehension questions were adopted from Arnold et al.'s (2017) study. The MCQs include direct (factual) questions which can be answered using explicit information from the source, as well as inferential questions which required integration or extrapolation of information from the source. The SAQs require problem-solving based on factual information in the source. These may relate to the distinction between surface versus deep learning (or processing) (Craik & Lockhart, 1972). MCQs may capture surface learning

based on the memory of facts from source materials, whereas SAQs may capture deeper learning reflecting elaborate semantic encoding.

Paper 2 included an additional source text on a new topic from an undergraduate textbook. Text comprehension questions were designed for this text following the same design principles as Arnold et al. (2017) and were tested in Experiment 1 of Paper 2. In Paper 4, two source texts adapted from Schnotz (1984) were used and again MCQs and SAQs were generated on the same principles. The text comprehension test questions used in all papers are in Appendix B.

3.1.2 Subjective Knowledge

Response accuracy measures of learning, such as multiple-choice tests, have been critiqued as they equate learning with the remembering and recall of factual information, which goes against the original purpose of writing-to-learn activities (Ackerman, 1993). Several authors advocate for more subjective measures of knowledge change in studies on writing-to-learn (Applebee, 1984; Schumacher & Gradwohl Nash, 1991; Galbraith & Baaijen, 2018).

Participants' subjective knowledge is measured in this thesis using a Subjective Knowledge Rating Scale (Galbraith et al., 2023; Appendix A). This is a 12-item scale in which participants rate each item on a 7-point scale, where 1 represents 'very little' and 7 represents 'a great deal'. The tool was designed to align to the dual-process model's (Galbraith & Baaijen, 2018) claims about writing. As the model's claims are about the writer's experience of knowledge discovery, this is inherently a subjective measure (Baaijen & Galbraith, 2018). Eight of the items load onto an 'organisation' subscale (e.g., '*How structured your thoughts about the topic are'*), which are designed to measure knowledge represented in explicit episodic memory. Six load onto an 'understanding' subscale (e.g., *'How much you feel you know about the topic'*), designed to measure knowledge represented implicitly in semantic memory. These scales have high reliabilities and are strongly correlated (r = 0.66; Galbraith et al., 2023). A key advantage of this measure is that it includes multiple items, in comparison to past research which used a single measure for subjective knowledge (Galbraith & Baaijen, 2018; Keil, 2003). To date, just two studies have used this scale (Baaijen & Galbraith, 2018; Hall, 2023), and it is unclear how participants' subjective ratings relate to their text comprehension. This thesis provides the first evidence of the validity of the subjective knowledge scale by assessing how it relates to other measures of writing-to-learn.

3.1.3 Reaction Time

As previously mentioned, response accuracy measures dominate writing-to-learn research. Silva and Limongi (2019) have argued that the field is missing an important measure of learning

through writing at the neural level. Research in cognitive psychology often uses RT to measure the speed at which individuals retrieve information from memory (Whelan, 2008; Ratcliff, 1978). Silva and Limongi have argued that measuring RT following writing-to-learn tasks would "capture the differential effects of dispositionally-guided and problem-solving processes of [long-term memory] consolidation" (p. 217). They designed a task in which participants read short texts, wrote or spoke a summary of them, and then completed an episodic memory word-recognition test in which they had to identify, as quickly and as accurately as possible, whether target words appeared in the original texts.

RT data is often analysed using statistical techniques such as analysis of variance on the sample mean, typically because of the familiarity of these methods to the researcher (Whelan, 2008). However, these methods can be inappropriate for detecting differences in RT between conditions due to a lack of power (Whelan, 2008). Several authors have highlighted that there is more to RT data than the mean (e.g., Balota et al., 2008; Balota & Yap, 2011; Anders et al., 2015). Silva and Limongi (2019) used Bayesian hierarchical drift-diffusion modelling to analyse the RT data in their study. Drift-diffusion models are a type of evidence accumulation model which allow RT distributions and response accuracy to be modelled simultaneously (Anders et al., 2015; Ratcliff, 1978) in two-choice decision-making tasks. Drift-diffusion models are well-established. They can provide detailed insight into the underlying cognitive processes through interpretation of several parameters in the model (Aschenbrenner et al., 2016; Voss, Rothermund & Voss, 2004). This analysis method is fully described and explained in Paper 4.

3.2 Measuring the Writing Process

In Paper 4, keystroke logging is used to measure the writing process as participants compose their texts. This is one of the most commonly used methods in writing research to explore the cognitive processes behind writing. The fundamental advantage of keystroke logging is that it provides an unobtrusive, moment-by-moment record of composition (Baaijen, Galbraith & de Glopper, 2012).

Before the uptake of keystroke logging, researchers typically used think-aloud protocols (concurrent protocols; Levy, Marek & Lea, 1996) to infer cognitive processes taking place during the writing process (Chenoweth & Hayes, 2003; Hayes & Flower, 1980). The assumption behind the method of think-aloud protocols is that writers are able to report their thoughts, and therefore their cognitive processes, during text production itself (Olive, 2010). Although several authors provided extensive justification for think-aloud protocols (Newell & Simon, 1972; Ericsson & Simon, 1980, 1993, 1998), they have received heavy criticism (Smagorinsky, 1998). Think-aloud protocols are often incomplete as there are many processes are writer will utilise

that they are not able to report on (Hayes & Flower, 1980). Think-aloud also increases the attentional demand on writers (Kellogg, 1994) and research has found reactive effects of such methods (Russo, Johnson & Stephens, 1989). For example, Janssen, van Waes & van den Bergh (1996) found think-aloud protocols can increase pause time in writers and can be particularly disruptive in complex writing involving knowledge-transforming.

Keystroke logging software logs and time stamps each button pressed when a person is typing a text, allowing text production to be reconstructed and analysed in relation to cognitive processes (Leijten & van Waes, 2013). In comparison to think-aloud protocols, keystroke logging is unobtrusive and does not rely on self-reporting by writers themselves (Baaijen, Galbraith & de Glopper, 2012). Keystroke logging can reveal information about writing fluency; the amount of text produced in comparison to the final product; and pausing and revision behaviour. It can also be used as a tool for teachers to monitor their students' writing development and for students to learn about their own writing processes underlying their performance (Spelman Miller, Lindgren & Sullivan, 2008). A further advantage of keystroke logging is that it can be combined with other measures, such as eye-tracking, to provide even deeper insights into cognitive processes (Leijten & van Waes, 2013; Wengelin et al., 2009). In Paper 4, Inputlog (Leijten & van Waes, 2006) is used to log keystrokes as it runs in Microsoft Word, a familiar environment to participants.

Researchers typically focus on three features when analysing keystroke logs: pauses, bursts, and revisions. 'Cognitive' pauses are typically of most interest to writing researchers. These are pauses of a sufficient length to allow writing processes to take place (Olive, 2010). These can reflect a number of processes including planning, translating, and revision (Olive, Alves & Castro, 2009; Alves, Castro & Olive, 2008). For this reason pauses can be difficult to interpret (Olive, 2010; Baaijen, Galbraith & de Glopper, 2012; Conijn, Roeser & Van Zaanen, 2019).

Bursts were originally conceptualised by Hayes and colleagues (Chenoweth & Hayes, 2003; Hayes, 2006, 2012; see also Kaufer, Hayes & Flower, 1986) and represent groups of words bounded by breaks in composition. Bursts are typically categorised as P-Bursts (bursts terminated by a pause) and R-Bursts (bursts terminated by a revision). Baaijen and Galbraith (2018) also considered I-Bursts (bursts produced away from the leading edge). Baaijen, Galbraith and de Glopper (2012) hypothesised that the relative amount of P- and R-Bursts reflects the extent to which writers produce controlled or spontaneous text.

Revisions are modifications made during text production. They have been viewed as reflecting a disagreement between the emerging text and the writer's intentions for the emerging text (Leijten, van Waes & Ransdell, 2010). The focus of revisions is typically on when they occur as opposed to what they involve. Baaijen, Galbraith and de Glopper (2012) suggest revisions at the

leading edge capture sentence fluency, and revisions departing from the leading edge can provide measures relating to linearity at different levels of the text (e.g., word or sentence level).

The analysis of keystrokes in Paper 4 utilises the frameworks developed by Baaijen, Galbraith and de Glopper (2012) and Baaijen and Galbraith (2018). Baaijen, Galbraith and de Glopper (2012) separate linear transitions (uninterrupted forward movements) from event transitions (non-linear transitions where the writer engages in another action before continuing). This is important as it allows pauses to be analysed based purely on those associated with forward production of text; and it allows the researcher to distinguish between revisions at the leading edge and revisions elsewhere, which provides global measures of revision. This approach allowed Baaijen, Galbraith and de Glopper (2012) to interpret keystroke logs through combined analysis of pauses, bursts and revisions, which is arguably superior to analysing these features separately.

Baaijen and Galbraith (2018) combined 11 individual measures constructed from pauses, bursts, and revisions into two overall composite scales: sentence production and global linearity. These scales had high reliability (α = 0.80 for global linearity; α = 0.79 for sentence production). The sentence production measure indicated the extent the generation of sentences was planned or spontaneous. Planned sentence production typically involved longer pauses followed by fluent production, whereas spontaneous sentence production typically involved short pauses with increased revision at the leading edge. The global linearity measure indicated how linearly sentences were produced during writing. A linear writing process is indicated by the order of sentences in the final product matching the order in which they were produced. A non-linear process is indicated by the opposite, and is represented by more time spent on I-Bursts and events. These measures were used to predict how the writing process relates to a writer's development of understanding and the quality of the text they produce to provide evidence for the dual-process model of writing. The measures have been reproduced in the doctoral thesis of Hall (2023; see also Hall, Baaijen & Galbraith, 2024). Paper 4 describes in detail the 11 individual measures that comprise sentence production and global linearity, including the methods used to calculate them from participants' keystroke logs.

3.3 A Note on Replication

Paper 3 of this thesis is a direct replication and extension of previous research conducted by Silva and Limongi (2019). It aims to validate a methodological tool for measuring long-term memory consolidation from writing (RT) for use in future studies. It replicates the original study exactly in a new group of participants (Experiment 1) and modifies the design to control for a potential modality effect (Experiment 2). Experiment 1 is therefore a 'direct replication', defined

as collecting data from a different, yet similar, sample, using the same procedure and data analysis methods as the original study (Simons, 2014). Conducting a direct replication is arguably the best method for providing support for a finding and assessing its reliability, especially in the early stages of establishing new methodological tools for use in future studies. Replication is facilitated by the transparency of authors, for example by clearly reporting methods and procedures in detail and providing open access to study materials and data. Although not all of the materials were provided in Silva and Limongi's (2019) published article, they were able to provide these on request. In addition to the replication study (Paper 3), Paper 4 explores the validity of the RT measure by assessing how it relates to other measures of writing-to-learn.

3.4 Participants and Sampling

The source-based writing activities utilised in this thesis are analogous to tasks often required in higher-education settings. Therefore, the cognitive processes in writing-to-learn are explored in the context of source-based *academic* writing. Accordingly, higher-education students were recruited to participate in the research as the most appropriate sample given their experience of academic writing and writing from sources.

Volunteer sampling was the main sampling strategy used across all four papers. In Papers 1 to 3, most participants were recruited from the Psychology Research Participation Scheme at the University of Southampton (UoS). These participants could receive course credits for their participation. As a result, a large proportion of these participants were undergraduate, female psychology students (the limitations and implications of this are acknowledge in Chapter 6). As the data collection for Papers 1 and 2 was conducted online, the study was also advertised on social media with the potential to recruit participants outside of UoS, however, very limited numbers volunteered to participate through these methods. Due to access issues with the research participation scheme at the time of data collection for Paper 4, participants were recruited through social media and posters around the university campus. Rather than course credits, these participants received a monetary voucher as a token of appreciation for their participation. Although this recruitment method was more time consuming than recruitment from the research participation scheme pool, it meant that participants from a wider variety of subject backgrounds were recruited, an advantage for the generalisability of the findings.

3.5 Ethical Considerations and the Impact of Covid-19

All of the studies that form the four empirical papers received ethical approval from the University of Southampton's ethics and research governing body, ERGO (Appendix F). The main

ethical considerations throughout the research were participant informed consent, anonymity, confidentiality, and the right to withdraw. Before taking part in an experiment, all participants were required to read a Participant Information Sheet and sign (or tick a box online) a consent form. The information sheets made clear to participants that their participation would remain confidential and that they could withdraw from the study during data collection at any point. Participant data were linked using a unique identifier number, allowing their data to be anonymised. Participant email addresses were collected in some studies for the purpose of contacting participants with links to take part in a second experimental session, or for them to be entered into a prize draw as a token of appreciation for their participation. Email addresses were never linked to participants data, and were stored in a password-protected file until data collection was complete and they were subsequently deleted. A final ethical consideration was the amount of time required from participants due to the lengthy nature of the experiments where they required extended writing. This was handled by allowing participants to take breaks between activities if necessary.

The Covid-19 pandemic posed significant challenges in undertaking the early stages of this research (Papers 1 and 2), in particular in terms of participant recruitment and the need to conduct online data collection. The recruitment of participants can be a significant hurdle for research in 'normal' circumstances, and these hurdles were heavily exacerbated by the pandemic. On one hand, online data collection presented an opportunity to reach a larger group of potential participants from across the UK, as well as a more time efficient way of collecting data; on the other hand, maintaining experimental control during remote data collection was (and remains), an arguably impossible task. The strengths and limitations of online experimental data collection are discussed in Chapter 8.

The pandemic also resulted in a re-structuring of the plan for this thesis. Initially, the thesis was proposed to comprise of three empirical papers, the first validating the RT measure (now Paper 3); the second exploring the text comprehension and subjective understanding measures (now Papers 1 and 2); and the third paper combining these measures (now Paper 4). Prior to the onset of Covid-19, in January 2020, ethics approval had been obtained, and data collection was scheduled for the study which later became Paper 3 of this thesis. However, the suspension of in-person data collection by UoS in March 2020 meant these plans could not go ahead. After exploring a range of options, it was not possible to conduct this study remotely; as mentioned by Cohen, Manion and Morrison (2018), RT experiments are very difficult to conduct remotely, considering the precise standardisation required. It was possible to conduct what was anticipated to be Paper 2 (now Papers 1 and 2) remotely. However, the only available online platform provided by the University was iSurvey (the University did not have a Qualtrics license at this stage). Problems with iSurvey were experienced which resulted in notable limitations

(most notably the control of time on task), and as such a follow-up study was conducted when a Qualtrics license became available.

Chapter 4 Paper 1. The Role of Retrieval Practice in Writing-to-Learn from Sources: When Writing Helps You Learn What You Don't Know

4.1 Abstract

Writing is an important tool for both communicating information and enhancing learning. The current study explored writing-to-learn in the context of summarising (planning and writing about) a source text passage and analysed the effects on text comprehension and subjective knowledge. 78 participants (higher-education students) read a source text passage about solar activity, created either an outline or synthetic plan, and wrote an essay or free recall text about the source text. They returned two-days later to answer text comprehension questions about the source text. Participants also rated their subjective knowledge of the text pre-writing, post-writing, and two-days later. Results showed no differences between the planning and writing conditions on text comprehension or subjective knowledge, suggesting effects were the result of retrieval practice. Regardless of planning or writing condition, participants subjective knowledge decreased across the three time points, confirming an illusion of explanatory depth. Participants' subjective knowledge ratings were moderately correlated with text comprehension test scores. It was therefore concluded that, in the context of summarising source text passages, the effects of writing on learning result from retrieval practice, which can both aid text comprehension but also help a writer identify gaps in their knowledge.

4.2 Introduction

Writing is a crucially important skill during all stages of education, both for communicating ideas and improving knowledge. The concept of writing-to-learn captures this general idea that writing is not just about effective communication and language production, but also about its capacity to shape learning and facilitate development of knowledge (Emig, 1977; Klein, 1999). Research exploring this development of knowledge through writing has often been overlooked in literature on writing, which typically focuses on the final product rather than the processes that occur during the writing process. Additionally, writing-to-learn research has assumed that to say learning has occurred, an individual must develop their knowledge (i.e., their knowledge increases pre- to post-writing). Here, it is suggested that this is not always the case, and an important feature of writing-to-learn is that writing helps reveals knowledge to the individual that they do not know.

The current study partially adopted the experimental procedure used in previous work by Arnold et al. (2017; details provided below), and asked whether, in the context of summarising (planning and writing about) a source text passage, learning from writing supports a dual-process model of writing (Galbraith & Baaijen, 2018), or whether the effects of writing on learning result from retrieval practice. The study also aimed to explore the extent to which subjective knowledge is related to text comprehension in measuring whether learning has occurred.

Classic writing-to-learn literature has largely focused on theories which assume writing is a problem-solving process (e.g., Hayes & Flower, 1980; Bereiter & Scardamalia, 1987). These have been labelled ‘backward search’ models by Klein (1999) and tend to assume that in expert writing (in comparison to writing by novices), the writing process is directed towards fulfilling the writer’s rhetorical goals for the emerging text. The extent that the writer’s thoughts and ideas are adapted to these goals leads to the transformation and development of knowledge (Galbraith, 1996). For example, Hayes and Flower’s (1980) model suggests writing is formed of three key processes: planning, translating, and reviewing, and that writers engage in these processes according to the goals they set for the text to be written. Bereiter & Scardamalia (1987) later criticised this model for its descriptive nature about what happens during the process of composing. They propose two alternative models of writing, claiming to explain, rather than describe, the processes taking place during composition. The first model, knowledge-telling, typically evident in the writing of novices, involves translating information retrieved from long term memory into written text. In comparison, Bereiter and Scardamalia (1987) suggest experts engage in a knowledge-transforming model which involves a problem-solving process in line with the writers’ goals for the text. However, these problem-solving models make distinct

assumptions about the transformation of knowledge during the writing process, without any measure of subjective knowledge change during writing (Baaijen & Galbraith, 2018; Applebee, 1984) and provide little explanation about how knowledge is represented in long term memory.

Galbraith and Baaijen (2018) therefore present a dual-process model of writing, claiming writing involves two key processes: a dispositionally-guided knowledge-constituting process, and a goal-directed knowledge-transforming process. The knowledge-transforming process is goal-directed, and responsible for the retrieval of the writer's thoughts and ideas stored in episodic memory (Galbraith & Baaijen, 2018). The knowledge-constituting process is represented in a semantic network of thoughts and ideas which guide the writer to synthesise the content to be produced. A development in knowledge is said to occur when writers effectively manage the two processes by reorganising and integrating knowledge representations in long term memory (Galbraith & Baaijen, 2018).

Empirical evidence for the dual-process model is found in studies comparing the effects of different planning strategies on learning outcomes. There is a robust literature implying the benefits of outline planning, which involves creating a hierarchically organised plan corresponding to the text to be written. For example, Kellogg (1990; 1994) found outline planning to improve text quality above techniques such as just jotting down a list of ideas. Baaijen and Galbraith (2018) compared outline planning to synthetic planning (summarising the overall goal for the text in one sentence) on participants' subjective knowledge. They found developments in subjective knowledge were related to increased revision of the overall text structure, and synthetic planning combined with spontaneous text production. Whilst participants' subjective knowledge increased in both conditions pre- to post-writing, the increase was small in the online planning condition and was associated with controlled sentence production. These findings support the dual-process model; in the absence of a plan in the synthetic condition, writers rely on synthesising implicit knowledge, whereas outline planning is associated with retrieving information from episodic memory to meet rhetorical goals and thus development in knowledge is more limited.

Like most other studies on writing-to-learn, Baaijen and Galbraith's (2018) research asks participants to write based on their own knowledge, for example opinion-based articles, generally showing that participants' subjective knowledge increases after writing. The current study will explore whether these same results occur in support of a dual-process model when participants summarise (plan and write about) a source text they read prior to writing. Thus, the study will adopt Baaijen and Galbraith's (2018) planning condition manipulation. The effects of outline or synthetic planning of a source text passage on learning has not been investigated

before and so the question of whether knowledge-transforming and knowledge-constituting will take place is an open one.

Despite consistent findings that subjective knowledge increases after writing, other literature measuring subjective knowledge has found participants' ratings decrease after writing. Rozenblit and Keil (2002) refer to this as the 'illusion of explanatory depth', which is the idea that "most people feel they understand the world with far greater detail, coherence, and depth than they really do" (p. 522). For explanatory knowledge, people's assessments of how much they feel they know changes after attempting to explain the topic as they re-evaluate their knowledge (Keil, 2003). When asked to rate their knowledge of a topic (for example, how a radio works), provide a written explanation, and re-rate their knowledge, Rozenblit and Keil's (2002) participants' ratings decreased, suggesting they were overconfident in their initial ratings and realised they knew much less when faced with generating an explanation. This finding has been replicated with children (e.g., Mills & Keil, 2004; Markman, 1979) and adults (e.g., Glenberg, Wilkinson, & Epstein, 1982).

Mills and Keil (2004) have suggested the illusion of explanatory depth occurs because (1) it is difficult for an individual to know if they have provided a good explanation; (2) individuals often have little experience in evaluating explanatory knowledge; and (3) they have minimal examples of past performance to compare their success against. Furthermore, individuals may confuse their internal representation of their knowledge with limited environmental support in the form of a written explanation (Rozenblit & Keil, 2002). This leads to the suggestion that writers may rate their knowledge lower when their written text does not correspond to a source text, or they are unable to answer test questions on the topic. In the context of writing-to-learn from source text passages, individuals may judge their understanding on the information they have readily available, or how easily they can retrieve this information from memory. This leads to the prediction in the current study that participants' subjective knowledge will decrease pre- to post-writing as knowledge they do not know is revealed to them through attempting to write about the topic.

This idea that another aspect of learning from writing is that it results from retrieval practice has seldom been considered in the field of writing-to-learn. In comparison, the domain of cognitive psychology has a well-established notion that simply recalling information from memory can foster retention of knowledge and transfer of learning (Arnold et al., 2017). This is known as 'the testing effect', the finding that participants who take practice tests after learning material outperform participants who simply restudy (Glover, 1989), especially enhancing retention after some delay (Roediger & Karpicke, 2006a; 2006b). The testing effect has been explained by the retrieval hypothesis; the idea that active retrieval is a powerful tool for learning new material

(Karpicke, 2012). Retrieval practice can lead to knowledge change as the simple act of recalling information strengthens activation in long term memory, so the information can be more readily retrieved in future (Roediger & Butler, 2011). Engaging in retrieval also provides individuals with feedback about what they do not know (Roediger & Butler, 2011). This corresponds with the illusion of explanatory depth; attempting to recall information during writing reveals to the writer what they do not know, so they rate their subjective knowledge lower after writing.

One study that has considered the concept of retrieval practice within writing-to-learn is that by Arnold et al. (2017). 100 undergraduate students read source text passages about two science topics and assigned to one of four writing activity conditions. Notetaking and highlighting conditions were assumed not to involve retrieval as participants had access to the source text during the writing activity; essay and free recall conditions were assumed to involve retrieval as they took place in a closed-book environment where participants had to write about the source text from memory. After writing, participants returned after a two-day delay to answer multiple-choice questions (MCQs) and short-answer questions (SAQs) about the source text. The researchers found that those who completed the writing activities involving retrieval (essay and free recall) performed significantly better on the MCQs and SAQs than participants who completed the activities that did not involve retrieval (notetaking and highlighting). There were no differences between the essay and free recall conditions on learning effects, leading Arnold et al. (2017) to conclude that the effects of learning from writing summaries of source text passages results from retrieval practice. However, it is possible that no differences were found between the essay and free recall writers, because those in the free recall condition tended to write 'essay-like' responses. Participants also spent notably different amounts of time on the writing activities, which may have confounded the results.

The current study aims to address three research questions that arise from the gaps in the current writing-to-learn literature. First, what effects do planning (outline and synthetic) and writing (essay and free recall) conditions have on text comprehension? To answer this question, the current experiment combines the planning manipulation from Baaijen and Galbraith's (2018) study with the writing manipulation from Arnold et al. (2017). This question seeks to confirm whether Arnold et al.'s (2017) retrieval practice conclusion holds when the essay and free recall writing activities are combined with different planning strategies, as well as discover whether the claims of the dual-process model apply in the context of summary writing based upon a source text.

Second, how is participants' subjective knowledge about the source text affected by the planning and writing conditions? The current study builds on that of Arnold et al. (2017) by including a measure of subjective knowledge. Previous research using such a measure has

found subjective knowledge increases after writing (Baaijen & Galbraith, 2018), however studies on the illusion of explanatory depth indicate subjective knowledge will decrease after writing (Rozenblit & Keil, 2002).

With regards to the first two research questions, if retrieval practice is the key process taking place in this context, no differences will be found between the planning and writing conditions on both text comprehension and subjective knowledge. This would contrast previous writing-to-learn literature which overlooks the possibility that writing based on source texts is different to writing about pre-existing knowledge in the absence of source texts.

Third, is participants' subjective knowledge related to text comprehension? It is expected that a participant who scores well on a test of text comprehension would also indicate higher subjective knowledge, but these correlations will only be moderate with the two measures assessing slightly different aspects of learning.

Briefly, participants in the current experiment read a source text, created either an outline or synthetic plan based on this text, summarised the text by either writing an essay or a free recall piece, and then returned to answer text comprehension questions two-days later. Participants also rated their subjective knowledge of the text pre-writing, post-writing, and two-days later.

4.3 Method

4.3.1 Design, Materials, and Measures

The current study adopted a mixed experimental design, with two between-subjects variables and two within-subjects variables. The between-subjects variables formed two independent variables (IV); planning strategy (outline or synthetic) and writing strategy (essay or free recall). The within-subjects variable, subjective knowledge, was one of the dependent variables (DV), and time of subjective knowledge rating was a within-subjects IV. Test scores on the MCQs and SAQs formed additional DVs (resulting in three IVs and three DVs). In the first part of the study, participants read a source text passage, created either an outline or synthetic plan based on the text, and wrote about the text in the style of an essay or free recall. They then returned for a second part two-days later to answer test questions about the text. Participants rated their subjective knowledge of the source text before and after writing, and on return after the two-day interval.

Source text. The source text passage (Appendix B.1) participants read was a passage on solar activity from Arnold et al. (2017). The text was 809 words long and was randomly selected from the two text passages used in the Arnold et al. study.

MCQs. The MCQs (Appendix B.2), designed to assess text comprehension of the source text, were the same as those used by Arnold et al (2017). Test questions included four factual ('direct') MCQs and four inferential MCQs. The direct MCQs, e.g., 'All of the following are true about sunspots, except?', could be answered using explicit information from the source text. The inferential MCQs, e.g., 'If temporary difficulties with short-wave radio communications were being experienced on Earth, which solar activity might you expect to be occurring?', "required participants to integrate across or extrapolate from" information in the source text (Arnold et al., 2017, p.118). For each MCQ, participants selected one answer out of four options. The number of correct answers were totalled for analysis, where eight was the highest mark available.

SAQs. The SAQs (Appendix B.3), also designed to assess text comprehension of the source text, were the same as those used by Arnold et al. (2017). Participants were instructed to 'pretend you are an amateur astronomer who likes to watch solar activity from your back garden' and answered questions such as 'As part of your hobby you take careful notes of what you observe, on a daily basis. Today you are observing the sun, and you compare your observations with those you included in your research notes two years ago. If the solar cycle has advanced, what two changes should you see today, as compared to what you noted two years ago?'. These questions involved problem-solving based on the factual information in the source text. Two

judges (researcher and main supervisor) rated participants' responses to the SAQs, with a score of two to four (including half marks) depending on the question. For example, in the example question given above, participants could receive a maximum of two marks, one for each correct change in the solar cycle given (see Appendix B.3 for a full breakdown of the scoring criteria). Ratings were screened for any mistakes, which were subsequently corrected. Scores that differed by one or more marks were discussed by the judges who agreed that in these cases, calculating an average was a fair representation of the participants' answer. Judges scores that differed by half a mark were also averaged. Marks on each of the SAQs were then totalled for analysis, where 13 was the highest mark available.

Inter-rater reliability for the SAQ ratings was measured with the Intraclass Correlation Coefficient (ICC). A two-way mixed effects model was used to evaluate the absolute agreement between the two raters. Following the initial ratings there was strong agreement between the two raters (ICC = 0.959), which increased after ratings were corrected for mistakes (ICC = 0.966).

Subjective knowledge. Participants rated their subjective knowledge of the source text on the Subjective Knowledge Rating Scale (Galbraith et al., 2023; Appendix A). This included 12-items in which participants rated each item on a 7-point scale, where one represented 'very little' and seven represented 'a great deal'. Eight of the items load onto an 'organisation' subscale, e.g., 'How structured your thoughts about the topic are', designed to measure subjective knowledge represented in explicit episodic memory. Six of the items load onto an 'understanding' subscale, e.g., 'How much you feel you know about the topic', designed to measure subjective knowledge represented implicitly. These scales have been shown to have high reliability (Galbraith et al., 2023). Participants' ratings were averaged as a whole, and on each subscale, for analysis.

4.3.2 Participants

229 people volunteered to participate in the study. Of these, 117 participants completed the planning and writing activities, and 104 completed the full study. Some drop out of participants was to be expected since the procedure has a two-day interval, however this was likely exaggerated due to the online nature of data collection. Participants who completed the full study were entered into a prize draw to win £50, and participants studying undergraduate psychology degrees at the University of Southampton received 12 course credits for their participation.

Data for the first part of the study were missing for two participants and so they were excluded from analysis. A further 16 participants were excluded for not following the correct procedure

(i.e., returning for the second part more or less than two-days later), and one participant was excluded for indicating they made notes whilst reading the source text. Additional exclusion criteria included spending longer than 480 seconds planning and/or longer than 2000 seconds writing, which led to the removal of five more participants. One participant wrote more than one sentence in the synthetic planning condition and thus they were excluded. Finally, outline plans of participants who spent less than 180 seconds were checked and one participant was subsequently excluded for not making an adequate outline plan. In sum, after these exclusions, 78 participants were included in the analysis. 39 participants were randomly allocated to the outline planning condition, and 39 to the synthetic planning condition. 30 participants were randomly allocated to the essay writing condition, and 48 to the free recall condition (this imbalance of participants in the writing conditions is a clear limitation of the current study). Participants had a mean age of 19.96. 72 of the participants were female and 6 were male. 72 participants spoke English as their first language. All 78 participants were current students, with the majority being undergraduates (n = 75). Most participants were Psychology students, either as a major or minor subject (n = 71); other subjects studied included Education, Sociology, Biomedical Science and English.

4.3.3 Procedure

Ethical approval was obtained from the University of Southampton's ethics committee (Appendix F.1). The study was conducted fully online, using the University of Southampton's platform iSurvey. Participants were recruited using an advert on social media and the University of Southampton's research advertising platform 'eFolio', in which those interested could follow an online link to iSurvey where they could read the participant information sheet. If they decided to participate, the individual provided consent and completed an initial questionnaire of demographic questions (gender, age, whether English was their first language, whether they were a student, their education level, and subject of study). On completion, participants were emailed a link to complete the main study.

Participants had ten minutes to read the source text on solar activity. They then rated their subjective knowledge of the text for the first time (pre-writing). Next, participants were randomly assigned to one of two planning conditions. In the outline planning condition, participants were guided to spend about five minutes creating a structured outline of the text they read about. In the synthetic planning condition, participants were guided to spend about five minutes writing down a single sentence summing up the overall point of the text they read about. After, participants were randomly assigned to one of two writing conditions. In the essay condition, participants received the following prompt, 'Write an essay describing the different types of solar activity, including their properties, their relationships with one another, and their effects

on Earth. Be as clear, detailed, and thorough as possible so that a high school student who has not read the text could understand. Your essay should have an introduction and a clear thesis, and you should make sure to back up your points with supporting details'. In the free recall condition, participants were instructed to 'recall (as close to word for word as possible) everything you can from the passage you read. If you can't remember the exact words, just try to express roughly what the text said'. Participants were guided to spent approximately 15 minutes writing. Both writing and planning were carried out closed book (participants did not have access to the source text), and participants could not look at their plans during writing. Finally, participants rated their subjective knowledge of the text for the second time (post-writing).

Participants returned two-days later where they firstly rated their subjective knowledge of the text for the third and final time. They then completed the eight MCQs about the source text, which were presented in a random order, followed by the four SAQs, which were also presented in a random order. Participants were thanked for their participation and could read a debriefing statement giving further details about the study.

4.4 Results

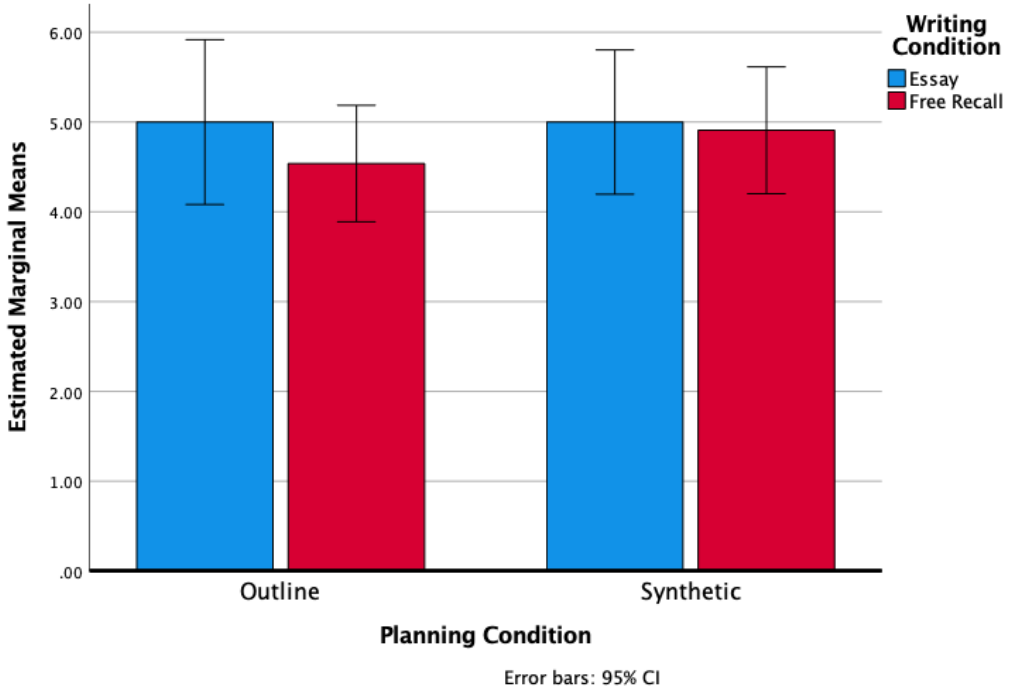
4.4.1 Analytic Approach

Data was analysed using the statistical package IBM SPSS (IBM Corp, 2020. IBM SPSS Statistics for Macintosh, Version 27.0.1.0). Initial exploration of the data showed all variables were normally distributed. First, four two-way between-subjects ANOVAs were carried out to assess the effects of the planning (outline vs. synthetic) and writing (essay vs. free recall) conditions on comprehension test scores. Next, the effects of planning and writing conditions on participants' subjective knowledge were analysed using two mixed ANOVAs. Finally, correlation analysis was conducted to analyse the relationship between participants' subjective knowledge and text comprehension test scores. Descriptive statistics for participants' text comprehension scores and subjective knowledge ratings are presented in Appendix G.

4.4.2 Effects of Planning and Writing Conditions on Text Comprehension

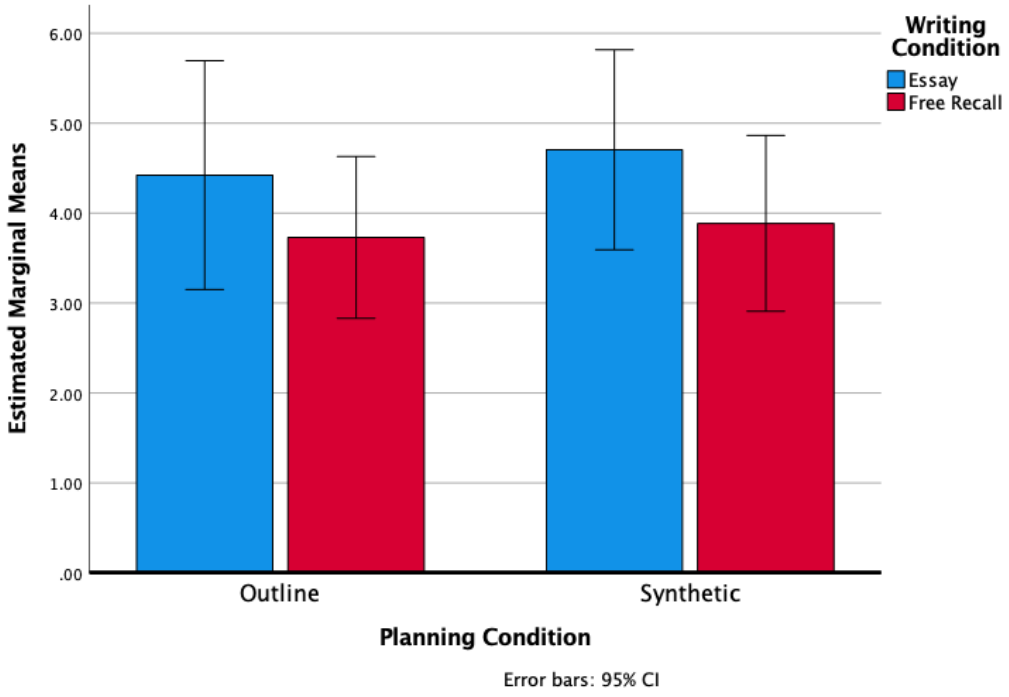
MCQs. The first set of analyses revealed no significant main effects of planning ($F_{(1,74)} = 0.227$, $p = 0.636$, $\eta^2 = 0.003$) or writing ($F_{(1,74)} = 0.503$, $p = 0.480$, $\eta^2 = 0.007$) condition, or an interaction of these conditions ($F_{(1,74)} = 0.227$, $p = 0.636$, $\eta^2 = 0.003$) on participants' total scores on the MCQs. Figure 6 visualises this non-significant difference between conditions on MCQ scores. In addition, no significant main effects or interactions were found when direct and inferential MCQs were considered separately. Participants' direct MCQ scores did not differ by planning ($F_{(1,74)} = 0.891$, $p = 0.348$, $\eta^2 = 0.012$) or writing ($F_{(1,74)} = 0.766$, $p = 0.384$, $\eta^2 = 0.010$) condition, or an interaction of these two conditions ($F_{(1,74)} = 0.114$, $p = 0.737$, $\eta^2 = 0.002$). Participants' inferential MCQ scores also did not differ by planning ($F_{(1,74)} = 0.046$, $p = 0.831$, $\eta^2 = 0.001$) or writing ($F_{(1,74)} = 0.046$, $p = 0.831$, $\eta^2 = 0.001$) conditions, or an interaction of these conditions ($F_{(1,74)} = 0.154$, $p = 0.695$, $\eta^2 = 0.002$).

Figure 6. Mean total score on MCQs as a function of planning and writing conditions.



SAQs. Figure 7 presents participants' mean total score on the SAQs for the planning and writing strategies. The analyses showed no significant main effects of planning ($F_{(1,74)} = 0.165$, $p = 0.686$, $\eta^2 = 0.002$) or writing ($F_{(1,74)} = 1.965$, $p = 0.165$, $\eta^2 = 0.026$) conditions, or an interaction of the planning and writing conditions ($F_{(1,74)} = 0.014$, $p = 0.906$, $\eta^2 < 0.001$) on participants' total scores on the SAQs.

Figure 7. Mean total score on SAQs as a function of planning and writing conditions.



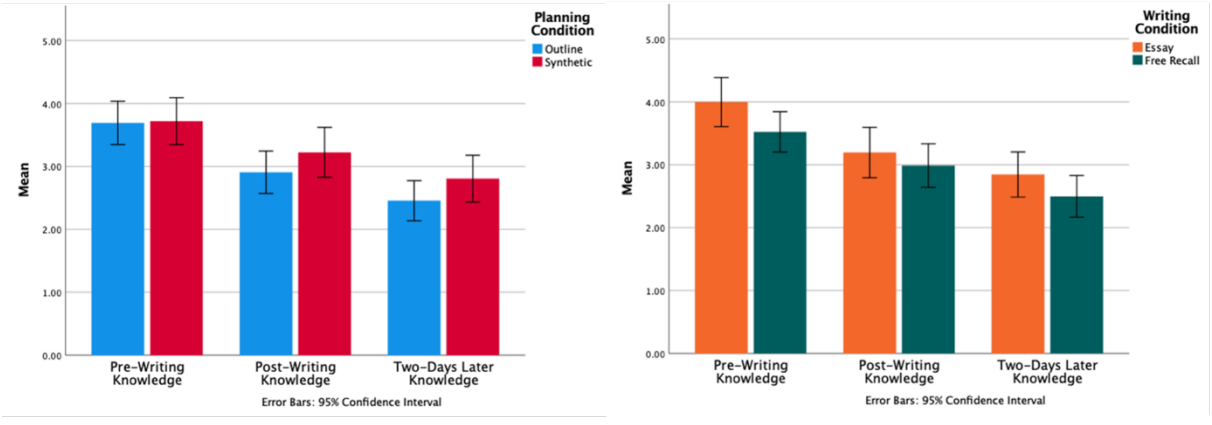
These results confirm Arnold et al.'s (2017) conclusion that the learning effects of writing from source texts are a result of retrieval practice, as both the essay and free recall conditions were assumed to involve retrieval. The finding that there were no differences between the outline and synthetic writing conditions, or an interaction of these with the essay and free recall writing conditions, suggests the effects of planning found by Baaijen and Galbraith (2018) do not operate when summarising a source text.

4.4.3 Effects of Planning and Writing Conditions on Subjective Knowledge

First, to assess how participants' subjective knowledge was affected by the planning and writing conditions, a mixed ANOVA was conducted with one within-subjects factor (subjective knowledge ratings at each time point ('time', three levels: pre-writing, post-writing, and after two-days)) and two between-subjects factors (planning condition (outline or synthetic) and writing condition (essay or free recall).

Figure 8 shows a decrease in ratings at each time point regardless of planning or writing condition. The analysis confirmed there was a significant main effect of time ($F_{(2,148)} = 65.766$, $p < 0.001$, $\eta^2 = 0.471$), and within-subjects contrasts revealed the decrease in participants' subjective knowledge ratings were indeed significant from pre-writing to post-writing ($F_{(1,74)} = 45.541$, $p < 0.001$, $\eta^2 = 0.381$) and post-writing to two-days later ($F_{(1,74)} = 17.591$, $p < 0.001$, $\eta^2 = 0.192$). There were no significant interactions of time with planning ($F_{(2,148)} = 1.835$, $p = 0.163$, $\eta^2 = 0.024$) or writing ($F_{(2,148)} = 1.019$, $p = 0.264$, $\eta^2 = 0.014$) condition.

Figure 8. Mean subjective knowledge ratings at each time point as a function of planning and writing conditions.



As the subjective knowledge rating scale has two subscales, further analysis was conducted to explore any effects in more detail. A mixed ANOVA was conducted with two within-subjects factors (subjective knowledge ratings at each time point, as previous; and the knowledge scale

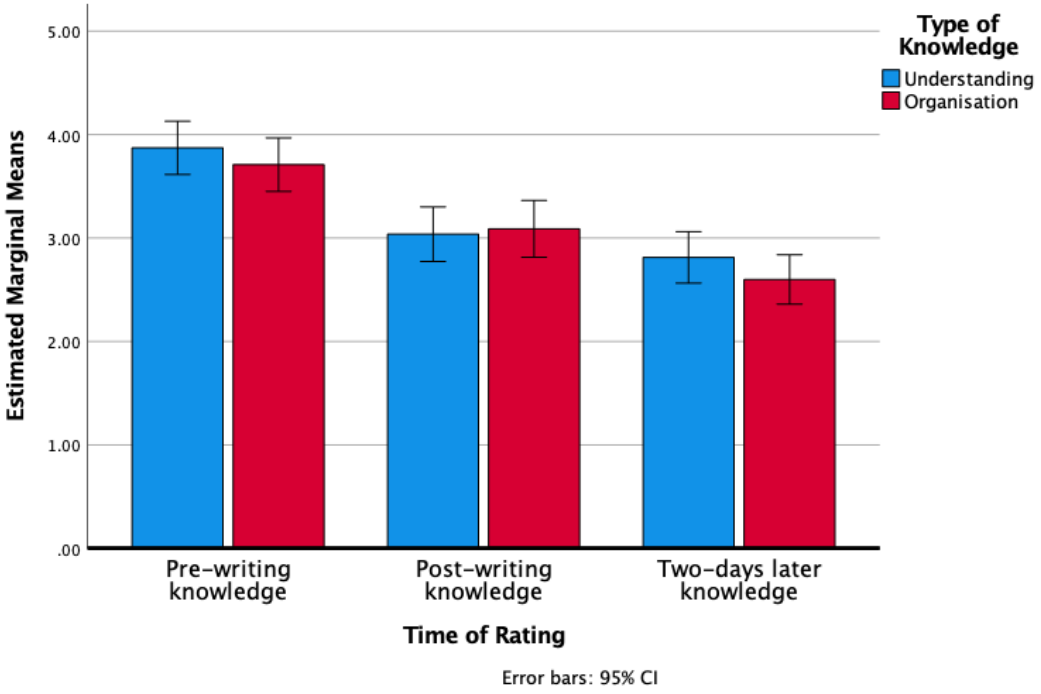
subscales ('type of knowledge', two levels: organisation and understanding)) and two between-subjects factors (planning and writing condition).

This analysis retained the significant main effect of time ($F_{(2,148)} = 68.674$, $p < 0.001$, $\eta^2 = 0.481$), with a significant decrease from pre-writing to post-writing ($F_{(1,74)} = 53.198$, $p < 0.001$, $\eta^2 = 0.418$) and from post-writing to two-days later ($F_{(1,74)} = 14.475$, $p < 0.001$, $\eta^2 = 0.164$). Again, no significant interactions were found of time with planning condition ($F_{(2,148)} = 2.192$, $p = 0.115$, $\eta^2 = 0.029$) or time with writing condition ($F_{(2,148)} = 1.332$, $p = 0.267$, $\eta^2 = 0.018$).

There was also a significant main effect of type of knowledge ($F_{(1,74)} = 7.670$, $p = 0.007$, $\eta^2 = 0.094$), indicating ratings on each subscale were different. This provides validity to the subscales on the subjective knowledge scale, suggesting they measure different aspects of participants' subjective knowledge. There were no significant interactions of type of knowledge with planning condition ($F_{(1,74)} = 1.089$, $p = 0.300$, $\eta^2 = 0.015$) or writing condition ($F_{(1,74)} = 2.446$, $p = 0.122$, $\eta^2 = 0.032$), implying participants' ratings on the organisation and understanding subscales did not vary by either condition.

In addition, a significant interaction was found between time and type of knowledge ($F_{(2,148)} = 9.336$, $p < 0.001$, $\eta^2 = 0.112$), indicating participants' ratings on the organisation and understanding subscales decrease differently over time. Within-subjects contrasts revealed this difference was significant for ratings pre- to post-writing ($F_{(1,74)} = 7.880$, $p = 0.006$, $\eta^2 = 0.096$) and post-writing to two-days later ($F_{(1,74)} = 23.553$, $p < 0.001$, $\eta^2 = 0.241$). These differences are visualised in Figure 9. Follow-up planned contrasts were conducted to explore these differences further.

Figure 9. Mean subjective knowledge ratings at each time point as a function of the organisation and understanding subscales.



Organisation subscale. Participants' subjective knowledge ratings on the organisation subscale decreased significantly from pre- to post-writing ($F_{(1,74)} = 31.921$, $p < 0.001$, $\eta^2 = 0.301$) and from post-writing to two-days later ($F_{(1,74)} = 24.200$, $p < 0.001$, $\eta^2 = 0.246$). The size of this decrease was similar for pre- to post-writing and post-writing to two-days later (a mean decrease of 0.60 and 0.51 respectively).

Understanding subscale. Participants' subjective knowledge ratings on the understanding subscale decreased significantly from pre- to post-writing ($F_{(1,74)} = 64.898$, $p < 0.001$, $\eta^2 = 0.467$) and post-writing to two-days later ($F_{(1,74)} = 5.493$, $p = 0.022$, $\eta^2 = 0.069$). However, differently to the organisation subscale, the size of the decrease was more substantial from pre- to post-writing than from post-writing to two-days later (a mean decrease of 0.78 vs. 0.27 respectively). It is worth noting here that participants had significantly higher pre-writing subjective understanding ratings than subjective organisation ratings ($t_{(77)} = 2.160$, $p = 0.034$, $d = -0.11$).

Taken together, these findings are consistent with the illusion of explanatory depth (Rozenblit & Keil, 2002) and provide further support to Arnold et al.'s (2017) notion that effects are due to retrieval practice as participants' knowledge ratings did not differ by writing condition. The differences between the organisation and understanding subscales support the dual-process model's (Galbraith & Baaijen, 2018) claim that these measure different aspects of knowledge; the organisation component representing explicit knowledge in episodic memory and the understanding component representing implicit semantic memory.

4.4.4 Relationship Between Text Comprehension and Subjective Knowledge

To explore whether, and how well, participants' subjective knowledge and text comprehension were related, correlation analysis was conducted between participants' subjective knowledge ratings and scores on the MCQs and SAQs. Only the subjective knowledge ratings participants completed two-days after writing were included in this analysis because they were completed at the same time point in the experimental procedure as the test questions. The bivariate correlations are displayed in Table 1.

Table 1. Means (M), standard deviations (SD), and Pearson bivariate correlations for comprehension test scores and subjective knowledge ratings two-days later (overall, and for each subscale).

			Correlations		
	M	SD	Knowledge two-days later	Organisation two-days later	Understanding two-days later
Total MCQ score	4.82	1.64	0.251*	0.231*	0.277**
Direct MCQ score	2.78	1.08	0.180	0.157	0.206*
Inferential MCQ score	2.04	1.06	0.204*	0.198*	0.220*
SAQ score	4.10	2.29	0.454**	0.431**	0.488**

Note. $n = 78$, *$p < 0.05$, **$p < 0.01$, (one-tailed).

Correlations between test scores and subjective knowledge ratings were strongest on the understanding subscale, in which all correlations were significant, with the largest occurring between SAQ scores and the understanding subscale. On the organisation subscale, correlations were weak but significant for participants' total MCQ score and inferential MCQ score, and correlations with SAQ score were stronger.

4.5 Discussion

In addressing the first two research questions, the current study found no differences in text comprehension test performance or subjective knowledge as a result of the outline and synthetic planning strategies. This contrasts with Baaijen and Galbraith's (2018) research which found knowledge development was facilitated by spontaneously produced writing in the synthetic planning condition and smaller increases in knowledge in the outline planning condition. Thus, the results of the current study indicate these same effects may not operate in the context of summarising source texts, where writers essentially just have to retrieve content from memory (knowledge-telling).

Results also showed no differences between the essay and free recall writing conditions in agreement with Arnold et al.'s (2017) conclusion that effects result from retrieval practice, as it was assumed that both essay and free recall involve retrieval. However, it is possible that the current experiment lacked enough power to detect potentially very small effects. Additionally, participants only read and wrote about one source text, as opposed to two source texts in Arnold et al.'s (2017) study. Therefore, to confirm the effects are retrieval practice, future research should be conducted using more than one source text passage and ranging across more topics. Here, only one text on solar activity was used, with Arnold et al. (2017) using an additional passage on extra-terrestrial life. Both texts are scientific topics, and so it is important to discover whether our findings can be generalised to texts in other domains.

The most interesting result of the current study was that participants' subjective knowledge significantly decreased from pre- to post-writing and from post-writing to two-days later. This opposes previous writing-to-learn research using the same measure of subjective knowledge, which has typically found subjective knowledge increases pre- to post-writing as writers re-organise their knowledge according to their rhetorical goals (Baaijen & Galbraith, 2018). It is suggested that in the context of writing about source texts, participants experience an illusion of explanatory depth (Rozenblit & Keil, 2002). Writing about the text highlights to the individual what they do not know, as they come to realise they cannot retrieve as much information from the source text as they originally believed, subsequently rating their knowledge lower after writing. The finding that participants rated their subjective understanding higher than their subjective organisation immediately after reading the source text also supports the illusion of explanatory depth as it implies participants may have overestimated their initial understanding. The differences in the nature of the decrease in ratings between the organisation and understanding subscales supports the idea that these measure different aspects of long term memory (Baaijen & Galbraith, 2018). Nevertheless, the current experiment did not feature a non-writing control condition and thus it cannot be concluded that it is writing itself that led to

this decrease in subjective knowledge. A future study could compare writing to thinking or speaking aloud about the text to rule out this possibility.

Furthermore, results supporting an illusion of explanatory depth, and the suggestion that the learning effects found in the current study result from retrieval practice, imply that source-based writing may only engage writers in knowledge-telling, rather than knowledge-transforming or knowledge-constituting. Wiley and Voss (1999) demonstrated that knowledge-transforming is necessary to allow writers to develop a deeper understanding of material. In addition, the dual-process model suggests both knowledge-transforming and knowledge-constituting processes are required for developments in understanding (Galbraith & Baaijen, 2018).

With regards to the third and final research question, the significant but moderate correlations between subjective knowledge ratings and text comprehension test scores suggests these measures are somewhat related. The strongest correlations were between the subjective knowledge ratings (particularly on the understanding subscale) and SAQ scores, indicating performance on these test questions is the best predictor of knowledge ratings after a two-day delay. The open-ended nature of these test questions in comparison to the MCQs suggest they offer participants the greatest opportunity to display their understanding of the source text. However, the finding that the correlations were only moderate implies the subjective knowledge measure and the test questions do assess different learning effects. This has implications for the wider writing-to-learn field where research often uses these objective measures of learning (Klein, 2004) which may only assess mere recall, and a subjective measure of knowledge in addition would allow deeper insights into participants' implicit knowledge representations during the writing process.

4.5.1 Strengths and Limitations

The most notable strength of the current study is its expansion on previous work by Arnold et al. (2017) and Baaijen and Galbraith (2018). The inclusion of a measure of participants' subjective knowledge in addition to the text comprehension test questions has brought to light that discovery through writing does necessarily have to involve knowledge development or knowledge transformation, as much of the writing-to-learn literature suggests. In the source text passage context, it appears that engaging in retrieval practice during writing has a dual effect. If retrieval is successful, writing about the text can aid text comprehension, particularly after some delay. If retrieval is unsuccessful, writing about the text can reveal gaps in the writer's knowledge. In both instances, it could be said that 'learning' has occurred.

The current experiment is not without its limitations, which have largely arisen from the online delivery. Here, the inevitable lack of control meant that whilst participants were guided to spend

five minutes planning and fifteen minutes writing, there was variation in the amount of time spent on these activities between participants. This is especially evident within the planning conditions, where outline participants tended to spend the full five minutes planning, but synthetic participants did not require as much time to construct a plan in the form of one sentence. Thus, it is likely that in this study, the planning manipulation was not a very strong experimental manipulation, which may explain why no differences were found. Furthermore, even within the outline planning condition, for example, participants' plans were highly varied in structure. Future studies should include more precise instructions to participants to reduce this variation. The online nature also means it is unknown whether participants were fully focused throughout the whole task or became distracted. Future research could incorporate keystroke logging which would provide insights into how participants' texts were constructed and provide more detail on the underlying cognitive processes involved.

Finally, participants in the current study were all higher-education students, and most of them were female undergraduate psychology students. Further studies are needed to establish whether the findings generalise to other types of writers, especially to those outside of higher education such as individuals under 18 years old and older adults.

4.6 Conclusion

In the current study, the effects of summarising (planning and writing about) source text passages on text comprehension and subjective knowledge were explored amongst higher-education students. This paper concludes by arguing that the effects of writing on learning in this context result from retrieval practice, which can aid a writer's later text comprehension performance but also help them identify gaps in their knowledge. Participants experienced an illusion of explanatory depth (Rozenblit & Keil, 2002), which contrasts previous research that finds participants' subjective knowledge ratings increase pre- to post-writing (e.g., Baaijen & Galbraith, 2018). These results present important new findings suggesting the processes taking place when writing about a source text are very different to those occurring when writing about pre-existing knowledge in the absence of such material. Future research is required to explore these differences in more detail.

Chapter 5 Paper 2. Writing-to-Learn vs. Thinking-to-Learn: Testing the Effects of Written and Mental Activities on Text Comprehension and Subjective Knowledge

5.1 Abstract

This paper investigated the effects of written and mental source-based activities on text comprehension and participants' subjective knowledge. Experiment 1 aimed to assess the validity of a new source text passage (deserts and savannas) and text comprehension questions, in comparison to the solar activity source text passage and text comprehension questions used in Paper 1 and by Arnold et al. (2017). Participants were randomly allocated to one of the source text topics, followed by one of three test conditions: (1) no text, in which participants answered text comprehension questions without reading the source; (2) text present, in which participants read the source and completed text comprehension questions in an open-book environment; and (3) text absent, in which participants read the source text and completed text comprehension questions in a closed-book environment. Those in the text present condition received the highest comprehension test scores, followed by the text absent condition, with those in the no-text condition performing the worst, and there were no differences between the two text topics. Therefore the new text was suitable for use in Experiment 2, the main experiment. In this experiment, 116 university students were randomly assigned to one of the source text topics, followed by one of four activity conditions: written essay, written free recall, mental essay, or mental free recall. In the mental conditions, participants were required to think about writing an essay or free recall without writing anything down. Two-days later, participants completed text comprehension test questions (multiple-choice and short-answer). They also rated their subjective knowledge of the text pre-writing, post-writing, and two-days later. Findings indicated writing lead to better short-answer question performance, but also a greater decrease in subjective knowledge ratings from pre- to post-writing, than thinking. Subjective knowledge ratings were positively correlated with short-answer question scores and factual multiple-choice question scores. Taken together, these results highlight the unique value of *writing* for enhancing comprehension and the identification of knowledge gaps, beyond the effects of retrieval practice.

5.2 Introduction

Source-based writing tasks involve writing about pre-read material, and are frequently used in educational settings to assess students' understanding of what they have previously read. Beyond the use as an assessment method, however, source-based writing tasks are capable of enhancing learning as writer's generate links between source materials and their own pre-existing knowledge (Spivey & King, 1989; Friend, 2002; Wade-Stein & Kintsch, 2004). This idea that writing can effect learning is known as 'writing-to-learn' (Emig, 1977; Klein, 1999).

Classic models (e.g., Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) describe writing as a problem-solving process in which the writer's thoughts and ideas are directed towards rhetorical goals for the emerging text. Hayes and Flower (1980) explained writing in terms of planning, translating, and reviewing, which are situated within the task environment and the writer's long-term memory. Later, Bereiter and Scardamalia (1987) proposed two processes involved in writing: knowledge-telling and knowledge-transforming. The knowledge-transforming process reflects goal-directed problem-solving, and is responsible for learning through development of the writer's understanding. Writers engage in reflective thought through connecting two problem spaces: the content problem space (thoughts and ideas), and the rhetorical problem space (goals for the text). Knowledge-telling involves translating information retrieved from long-term memory into text, without elaborate organisation of the content (Bereiter & Scardamalia, 1987). For this reason it is implied that knowledge-telling does not have a role in learning and developing the writer's understanding.

A fundamental limitation of the classic models of writing is that they have little to say about how knowledge is represented in long-term memory, and consider translation as the passive reporting of thoughts and ideas (Applebee, 1984; Galbraith & Baaijen, 2018). The dual-process model of writing (Galbraith & Baaijen, 2018) distinguishes between the reflective thought emphasised by the classic models, but also translation as an active knowledge-constituting process. According to Galbraith and Baaijen (2018), learning through writing occurs through the effective management of two processes: knowledge-transforming and knowledge-constituting. The knowledge-transforming process operates on representations within episodic memory, in which the writer's thoughts and ideas are retrieved and evaluated in line with the rhetorical goals for the text. The knowledge-constituting process takes place within the semantic memory system, where knowledge is stored implicitly through the strength of connections between neural units. During text production, this knowledge is made explicit, and the writer's thoughts and ideas are synthesised, resulting in a development in understanding.

Evidence for the dual-process model comes from a study conducted by Baaijen and Galbraith (2018). It was found that when participants wrote a text spontaneously following synthetic planning (summarising the overall goal for the text in one sentence), they developed their understanding by relying on constituting their implicit knowledge during text production. On the other hand, when participants constructed an outline plan before writing their texts, development of understanding was limited as the writing process was comprised of transforming information retrieved from episodic memory to meet rhetorical goals. Despite this empirical support, the dual-process model is based on individual's writing about their own pre-existing knowledge in the absence of source materials.

In Paper 1, the claims of the dual-process model were explored in the context of closed-book source-based writing. Arnold et al. (2017) suggested retrieval practice is an underlying cognitive process in closed-book source-based writing-to-learn. In their study, participants who undertook a writing activity involving retrieval (essay or free recall) performed better on a text-comprehension test two-days later than participants who undertook a writing activity not involving retrieval (notetaking or highlighting). In Paper 1, 78 participants read a source text passage about solar activity (the same text used by Arnold et al.), constructed either an outline or synthetic plan, and then wrote either an essay or free recall text about the source. Two-days later, participants returned to complete a text comprehension test. They also rated their subjective knowledge pre-writing, post-writing, and two-days later. Whilst there were no differences between conditions on participants' text comprehension scores or their subjective knowledge, participants in all conditions experienced a decrease in their subjective knowledge from pre- to post-writing. It appeared that by engaging in retrieval practice, writing helped participants to identify gaps in their knowledge. This is similar to the experience of the 'illusion of explanatory depth' (Rozenblit & Keil, 2002). Before writing, participants felt they understood the text with greater detail and coherence than they actually do.

A limitation of Paper 1 is that it did not include a non-writing control condition, and as such it could not be concluded whether it was *writing* specifically that led to such effects, or if this would also occur for non-writing retrieval tasks. Few studies in writing-to-learn literature have included non-writing control conditions. A handful of studies comparing writing with speaking have found advantages for writing on post-tests of understanding (e.g., Klein, Piacente-Cimini & Williams, 2007), long-term memory consolidation (e.g., Silva & Limongi, 2019), and for transforming thoughts and ideas into structured knowledge (e.g., Rivard & Straw, 2000). However, speaking still requires the learner to generate content. In the current study, thinking is used as a non-writing control condition as it allowed the potential cognitive benefits of writing to be separated from those associated with content generation.

Two previous studies have compared writing to thinking. Kellogg's (1988) second experiment compared the writing process of 20 university students completing a persuasive writing task following either no planning, a written outline plan, or a mental outline plan. No differences were found between written and mental outline planning on participants' writing process. In contrast, Drabick et al. (2007) did find an advantage of writing over thinking. During a 14-week psychology course with 978 undergraduates, students either completed brief (five-minute) in-class writing tasks or were simply instructed to think about the topic for five-minutes per week. Students who completed the in-class writing tasks performed better on a multiple-choice exam at the end of the course.

One reason why writing can have an advantage for learning over other mediums such as speaking or thinking is because of the external representation that comes with written text. This is supported by studies which find removing visual feedback can have an effect on writing processes (Olive & Piolat, 2002). For example, writing with invisible ink can invoke greater processing demands as the writer loses access to an 'external store' of the text they have produced so far (Gould, 1980; Grabowski, 1996; Olive & Piolat, 2002). An alternative explanation is that writing encourages engagement with elaboration and organisation processes to a greater extent than non-writing. Elaboration processes involve *"connecting what one is learning to what one already knows"*; organisation processes involve *"structuring the to-be-learned material often in a way that creates a new structure"* (Arnold et al., 2017, p. 116).

In summary, the current paper aims to extend the findings from Paper 1 by including non-writing control conditions. It also seeks to address two additional limitations from Paper 1. First, is to include an additional source text topic to increase the generalisability of the findings to a new topic. Second, is to more stringently control the time participants spend writing. The relationship between participants' text comprehension scores and subjective knowledge is also analysed to assess the reliability of the results in Paper 1. These research objectives are explained next in an overview of the experiments reported in this paper.

5.2.1 Overview of Experiments

5.2.1.1 Experiment 1

The main focus of Experiment 1 was to validate a new source text passage and its associated text comprehension questions in comparison to the source text used in Paper 1 of this thesis and by Arnold et al. (2017) ('solar activity'), for use in Experiment 2. A new expository source text ('deserts and savannas') of similar reading and knowledge level (undergraduate) and length to the solar activity text was selected. It is predicted that there will be no differences in participants' text comprehension test performance across the two source text topics.

To assess the construct validity of the text comprehension questions, the procedure of Hinze, Wiley and Pellegrino's (2013) first experiment was adopted. Their experiment aimed to evaluate multiple-choice questions (MCQs) linked to five expository texts. Participants were randomly assigned to one of three conditions. In a 'prior knowledge' condition (the 'no text' condition in the current experiment), participants were asked to answer the MCQs without reading the source texts. In a 'text available' condition (the 'text present' condition in the current experiment, participants were given the source texts at the same time as the MCQs and could use the sources to answer the questions. Finally, in an 'immediate test' condition (the 'text absent' condition in the current experiment), participants first read the source texts and then answered the MCQs from memory. It was found that participants only using their prior knowledge to answer the MCQs performed the worst, followed by the participants in the immediate test condition. Participants in the text available condition had the highest performance as they did not need to rely on their memory to answer the questions. These results are what is predicted to be found in Experiment 1 as the text comprehension questions are designed to be answered using information from the source texts.

5.2.1.2 Experiment 2

Following validation of the materials in Experiment 1, Experiment 2 sought to expand on the findings from Paper 1 of this thesis and explore the effects of written and mental source-based activities on text comprehension and subjective knowledge. Briefly, participants read one of two source text passages, and then undertook one of four 'writing' activities (written essay, written free recall, mental essay, or mental free recall). Two-days later, they returned to complete a text comprehension test about the source texts. Participants also rated their subjective knowledge of the source text topics pre-writing, post-writing, and two-days later.

The key finding of Paper 1 was that participants' subjective knowledge decreased from pre- to post-writing, regardless of the writing condition they were allocated to. However, all conditions required the act of writing (i.e., there was not a non-writing control condition). The external representation of knowledge in written form may have contributed to the finding that participants' knowledge decreased from pre- to post-writing, as writers could see their knowledge written down on the page and thus it was clear to them knowledge that was missing. Therefore, the first aim of Paper 2 is to assess whether this effect is unique to writing, or whether the effect will still occur through engaging in retrieval practice without writing. To address this aim, in Experiment 2 of this paper, participants undertook one of four 'writing' activities: written essay, written free recall, mental essay, or mental free recall. In the 'mental' conditions, participants were simply asked to think about the activity without writing anything down. This is a novel manipulation that has rarely been used in previous writing-to-learn research.

The 'writing' activities were conducted in a closed-book format (i.e., participants did not have access to the source text during the activity). This means that all participants are engaged in retrieval processes (Arnold et al., 2021). As a result, it is predicted that subjective knowledge ratings will decrease from pre- to post-writing in all conditions as participants experience an illusion of explanatory depth (Rozenblit & Keil, 2002). However, this extent of this decrease may depend on whether participants completed a written or mental activity. On one hand, participants in the written conditions will have an external representation of their knowledge which reinforces information that they do not know (i.e., provides them with more metacognitive awareness). Therefore, these participants may have a bigger decrease in subjective knowledge ratings than participants in the mental conditions. On the other hand, the lack of external feedback in the mental conditions could lead to a larger decrease amongst these participants. Physically writing as opposed to thinking about writing also requires participants to organise and structure their thoughts.

Experiment 2 addresses two limitations of Paper 1. First, in Paper 1, all participants read the same source text passage about solar activity. The current experiment aims to assess whether the findings of Paper 1 generalise to another source text topic (deserts and savannas). Second, the time participants spent on the writing task in Paper 1 was not well controlled due to the online data collection platform used (iSurvey). This meant participants spent varying amounts of time on their assigned writing activity, which may have confounded the results. In Paper 2, a different online data collection platform was used (Qualtrics), which more reliably controls time on task. However, it should be noted that due to the remote, online data collection in this experiment as a result of the Covid-19 pandemic, it is still possible that participants did not spend all of the allocated time writing or thinking.

To summarise, Experiment 2 aimed to answer two research questions:

1. What are the effects of writing activities (essay vs. free recall) in different modalities (writing vs. thinking) on participants' text comprehension scores and subjective knowledge ratings, and are there any differences depending on the source text topic (solar activity vs. deserts and savannas)?
2. What is the relationship between participants' subjective knowledge ratings and text comprehension scores?

5.3 Experiment 1

5.3.1 Method

5.3.1.1 Design

Experiment 1 aimed to assess the reliability and validity of a new text passage topic (deserts and savannas) and text comprehension questions in comparison to the text passage and text comprehension questions used in Paper 1 and by Arnold et al. (2017) (solar activity).

The design of Experiment 1 was based on Hinze, Wiley and Pellegrino's (2013) first experiment. A 2x3 between-subjects design was used in which participants were first randomly assigned to one of the two text passages (deserts and savannas or solar activity), and then randomly assigned to one of three test conditions. In the 'no text' condition, participants were given the test questions without reading the text. In the 'text present' condition, participants read the text passage for ten minutes, and were then given the MCQs, followed by the SAQs. They were allowed to refer back to the text passage when answering the test questions (i.e., an open-book test). Finally, in the 'text absent' condition, participants read the text passage for ten minutes, and were then given the MCQs, followed by the SAQs, but they had to answer the questions from memory without being able to refer back to the text (i.e., a closed-book test).

It was hypothesised that participants in the text present condition would have the highest scores, as the text comprehension questions are based on content from the source text. Having never seen the source text, participants in the no text condition should score the lowest. Participants in the text absent condition are predicted to score in-between the text present and no text conditions having read the text passage but tested from memory. These results are predicted to be consistent across the two text passages to indicate concurrent validity.

5.3.1.2 Materials and Measures

Source texts. In addition to the solar activity source text (Arnold et al., 2017), a new source text passage on a different topic was sourced. The solar activity text (Appendix B.1) was created by Arnold et al. using an undergraduate level astronomy textbook and was 810 words long. Therefore, another text of roughly the same reading and knowledge level and length was selected for use. The chosen text, 'deserts and savannas' (Appendix B.4), was adapted from a recommended textbook for first-year Biology undergraduates (Hill, Wyse & Anderson, 2018) and was 859 words long.

Text comprehension questions. Like the solar activity passage, eight MCQs (four direct and four inferential) (Appendix B.5) and four SAQs (Appendix B.6) were created for the deserts and

savannas passage, in accordance with Arnold et al.'s (2017) descriptions. Four direct (factual) MCQs could be answered "using information explicitly stated in the text passage" (Arnold et al., 2017, p. 118), and four inferential MCQs "required participants to integrate across or extrapolate from two stated facts" (Arnold et al., 2017, p. 118). Finally, four short-answer problem-solving questions involved "making inferences and drawing connections across several facts from the passage" (Arnold et al., 2017, p. 119).

Correct answers on the MCQs were totalled across all eight questions. SAQ responses were rated by two judges (researcher and main supervisor) (Intraclass Correlation Coefficient = 0.963), out of a score of two to four (including half marks) depending on the question (Appendix B.3 and B.6 show a breakdown of the scoring criteria). Ratings were checked for any mistakes which were corrected. Scores that differed by one or more marks were discussed by the judges, and adjusted accordingly. In most cases it was agreed that an average of the two marks would be taken. Scores that differed by half a mark were also averaged. The SAQ ratings were then totalled, where 13 was the highest mark available.

5.3.1.3 Participants

133 higher-education students at UK universities were recruited to take part. If participants volunteered through the University of Southampton research participation scheme, they received course credits for their participation. All participants could also enter a prize draw to win a £50 voucher. The study was approved by the University of Southampton Ethics Committee (Appendix F.2) and participants checked (ticked) a box before beginning the study to provide consent.

The majority of participants were female (*n* = 116) undergraduate students (*n* = 123) who spoke English as their first language (*n* = 112). The remaining participants were Masters students (*n* = 3) and Doctoral students (*n* = 7). Participant studied a range of subjects, although the majority studied Psychology (*n* = 95). The average age of participants was 21.49 (SD = 5.64).

Out of the 133 participants, 9 did not complete all eight of the MCQs, and so 124 participants were included for analysis. Similarly, 30 participants did not complete all four of the SAQs, and so 103 were included for analysis. A breakdown of the number of participants in each condition is shown in Table 2.

Table 2. Number of participants included in analysis by source text and test condition (Experiment 1).

				Test Condition		
			Overall	No text	Text present	Text absent
Text	Overall	MCQ	124	42	47	35
		SAQ	103	33	32	38
	Solar activity	MCQ	65	22	25	18
		SAQ	53	15	21	17
	Deserts and savannas	MCQ	59	20	22	17
		SAQ	50	18	17	15

5.3.1.4 Procedure

Experiment 1 was conducted online either on UoS iSurvey (*n* = 35) or Qualtrics (*n* = 98). This is because the university-supported online study tool changed during the data collection period. Participants were recruited using an advert on social media or through the UoS research participation platform. Those interested in participating followed an online link to iSurvey or Qualtrics where they could read the Participant Information Sheet and provide consent to participate. Participants then completed some demographic questions (gender, age, whether English is their first language, whether they are a student, their education level, and subject of study), and were randomly assigned to a source text condition (solar activity or deserts and savannas), followed by a test condition (no text, text present, text absent).

Participants assigned to either the text present or text absent condition then received the following instruction: "*You will see a text about [solar activity/deserts and savannas]. Please read the text carefully. You will have 10 minutes to read it*". A timer was visible on screen whilst participants were reading the source text, so they knew the amount of time remaining. After the 10 minutes had elapsed, participants were then instructed: "*You will answer some multiple-choice and short-answer questions about the [solar activity/deserts and savannas] text you have just read. You can take as much time as you want to answer the questions*". Participants in the text present condition also received an additional line of instruction: "*You will find each question below the text, and you may use the text to answer each question".*

Participants in the no text condition did not read a source text; instead, they received the following instruction: "*You will answer some multiple-choice and short-answer questions about*

[solar activity/deserts and savannas]. You can take as much time as you want to answer the questions".

All participants then answered the eight MCQs, followed by the four SAQs, one at a time, for the source text topic condition they had been allocated to. Participants in the text present condition answered in an open-book format and participants in the text absent condition answered in a closed-book format.

After completing all of the text comprehension test questions, all participants were given the option to enter into a prize draw as a token of appreciation for their participation. They were then thanked for their participation and provided with the researchers' details as a contact if they had any further questions.

5.3.2 Results and Discussion

5.3.2.1 Analytic Approach

Data was analysed using the statistical package IBM SPSS (IBM Corp, 2020. IBM SPSS Statistics for Macintosh, Version 27.0.1.0). Continuous variables were assessed using Explore and reasonably adhered to the normal distribution. To compare text comprehension scores between the two source texts, independent samples t-tests were conducted. Next, one-way between-subjects ANOVAs were conducted to explore the differences between conditions (no text, text present, and text absent) on text comprehension scores. Where homogeneity of variance was violated, Welch's F-test was used. Where significant differences were found, post-hoc tests were conducted to determine the source of the effect. Descriptive statistics (means and standard deviations) are available in Appendix Appendix H.

5.3.2.2 Comparison Between Source Texts

Independent samples t-tests revealed there were no significant differences between the two source text conditions (solar activity and deserts and savannas) for either total MCQ scores ($t_{(122)}$ = -1.16, p = 0.125, d = 0.21) or SAQ scores ($t_{(101)}$ = 1.05, p = 0.147, d = 0.21).

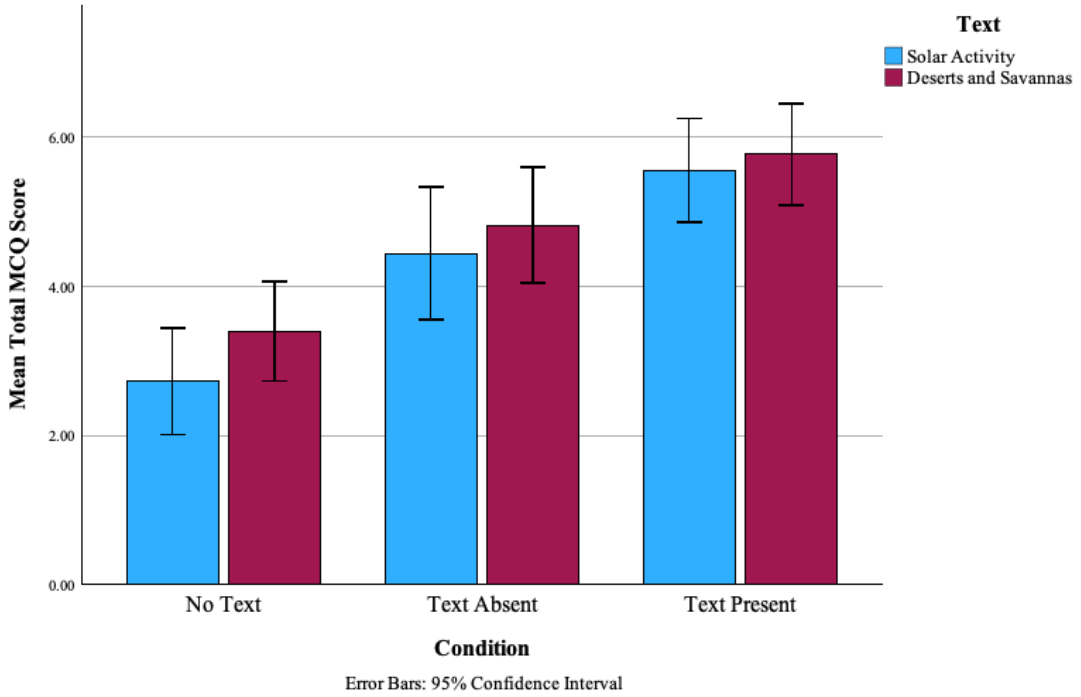
5.3.2.3 Differences Between Conditions

Participants' text comprehension scores for both MCQs and SAQs were in the predicted direction: those in the no text condition have the lowest scores, followed by those in the text absent condition, with those in the text present condition having the highest scores. Figure 10 visualises these differences for participants' total scores on the MCQs.

MCQs. The difference between conditions on participants' MCQ scores was significant ($F_{(2, 121)}$ = 29.86, $p < 0.001$, $\eta^2 = 0.33$). Tukey HSD post-hoc tests showed the no text condition had significantly lower MCQ scores than both the text absent ($p < 0.001$) and text present ($p < 0.001$) conditions; and the text present condition had significantly higher MCQ scores than the text absent condition ($p = 0.013$).

SAQs. The difference between conditions on participants' SAQ scores was also significant ($F_{(2, 61.65)} = 56.67$, $p < 0.001$, $\eta^2 = 0.50$) (Welch's F-test is reported as the assumption of homogeneity of variance was violated). Games-Howell post-hoc tests showed the no text condition had significantly lower SAQ scores than both the text absent ($p < 0.001$) and text present ($p < 0.001$) conditions; and the text present condition had significantly higher SAQ scores than the text absent condition ($p < 0.001$).

Figure 10. Mean total MCQ score by text and condition.



5.3.2.4 Discussion

The findings of Experiment 1 indicate that the source texts and text comprehension test questions are suitable for use in Experiment 2. Analysis comparing text comprehension scores across the two source texts found no significant differences, indicating that participants performed similarly for both the solar activity and deserts and savannas test questions. The results analysing differences between the conditions (no text, text present, and text absent) was as expected: participants in the no text condition had the lowest MCQ and SAQ scores, followed by participants in the text absent condition, with participants in the text present condition scoring the highest for both MCQs and SAQs.

Therefore, the findings provide evidence of the construct validity of the text comprehension questions. The low scores of participants in the no text condition demonstrate that being able to answer the questions does require engagement with the source text rather than general knowledge alone. The result that participants in the text present (open book) condition performed the best shows that the test questions are related to comprehension of the source text. Participants in the text absent (closed book) condition had performance between the other two conditions, which would be expected since these participants had read the source texts but could not refer back to it when answering the text comprehension questions. Overall, this means the two source texts (solar activity and deserts and savannas) and associated text comprehension questions can be used as reliable and valid materials in Experiment 2.

5.4 Experiment 2

5.4.1 Method

5.4.1.1 Design

In short, Experiment 2 was formed of two online sessions. In the first session, participants read a source text passage, and then carried out one of four 'writing' activities (essay writing, free recall writing, mental essay, or mental free recall; described fully in the *Procedure*). Participants also rated their subjective knowledge of the text passage pre- and post-'writing' activity. After a two-day interval, participants completed the second session where they rated their subjective knowledge of the text, and then answered the MCQs and SAQs to measure their text comprehension.

Experiment 2 adopted a mixed experimental design, with three between-subjects variables and one within-subjects variable. Between-subjects independent variables were text (solar activity or deserts and savannas), 'writing' activity (essay or free recall), and 'writing' activity modality (mental or written). The within-subjects variable was the time of subjective knowledge rating (pre-writing, post-writing, and two-days later). Participants' test scores on the MCQs and SAQs, and participants' subjective knowledge ratings were the dependent variables.

5.4.1.2 Materials and Measures

Source texts and text comprehension questions. Two source text passages (Appendix B.1 and B.4) were used in Experiment 2, 'Solar Activity' and 'Deserts and Savannas'. Each text passage has eight MCQs and four SAQs designed to assess participants' comprehension of the source text. Details of the source texts and text comprehension questions are outlined in Experiment 1 Methodology. The ICC of judges' ratings for the SAQs in Experiment 2 was 0.94.

Subjective knowledge ratings. Participants' subjective knowledge of the source text was assessed using the Subjective Knowledge Rating Scale (Galbraith et al., 2023; Appendix A). The scale has 12-items and participants rated each item on a seven-point scale, where one represented 'very little [knowledge]' and seven represented 'a great deal [of knowledge]'. Eight of the items load onto an 'organisation' subscale, designed to measure participants' knowledge represented in episodic memory. Six of the items load onto an 'understanding' subscale, designed to measure participants' knowledge represented in semantic memory.

5.4.1.3 Participants

195 UK higher-education students began the study, however only 148 of these participants completed both sessions. A further 32 participants were excluded as they completed the second session more than two days after the first session. Therefore, data from 116 participants were analysed. The majority of these participants were undergraduate students (*n* = 115), and most were studying Psychology as a main subject (*n* = 113). 96 of the participants were female; 18 were male; and 2 did not disclose their gender. All participants were self-declared fluent English speakers (103 had English as their first language). Participants had a mean age of 19.17. The distribution of participants across conditions is shown in Table 3.

All participants who completed the full study received course credits for their participation. The study was approved by the University of Southampton Ethics Committee (Appendix F.3) and participants checked (ticked) a box before beginning the study to provide consent.

Table 3. Number of participants by source text and writing condition (Experiment 2).

		Writing condition				
		Written essay	Mental essay	Written free recall	Mental free recall	Total
Text	Solar Activity	11	18	14	16	59
	Deserts and Savannas	14	11	17	15	57
	Total	25	29	31	31	116

5.4.1.4 Procedure

Experiment 2 was conducted online using Qualtrics. Participants were recruited through the University of Southampton's research advertising platform 'eFolio'. After reading an information sheet, participants ticked a box to provide consent to participate. The study was then conducted over two online sessions.

First, in session one, participants completed an initial questionnaire of demographic information, including their gender, age, first language, student status, education level, and subject of study. Participants were then randomly assigned to either the solar activity or deserts and savannas source text. They had ten minutes to read the text and could not progress until this time had passed. Participants were able to see how much time was remaining to read the

texts on the screen. Next, participants rated their subjective knowledge of the text on the Subjective Knowledge Rating Scale (pre-writing knowledge).

Participants were then randomly assigned to one of four 'writing' activity conditions: written essay, written free recall, mental essay, or mental free recall. In all conditions, participants had fifteen minutes to spend on the activity, and they could not progress until the time had elapsed. The precise instructions participants received in each condition are stated in Appendix E.1.

Written essay condition. Participants received an essay prompt based on the text they had read and were instructed to write (type) an essay in response to this prompt.

Mental essay condition. Participants received the same essay prompt, but instead of being asked to type their essay, they were asked to think about how they would write the essay. At one-minute intervals, participants were prompted to write down the first word that came to mind, and they had fifteen seconds to do so. This was designed to ensure participants remained on-task.

Written free recall. Participants were instructed to recall and write down (type) everything they could remember from the source text, as close to verbatim as possible.

Mental free recall. Participants were instructed to think about recalling everything they could remember from the source text, as close to verbatim as possible. Like in the mental essay condition, these participants were prompted at one-minute intervals to type the first word that came to mind, and they had fifteen seconds to do so.

After completing their assigned writing activity, participants completed the Subjective Knowledge Rating Scale for a second time (post-writing knowledge).

Two-days later, participants completed session two. First, they completed the Subjective Knowledge Rating Scale for a third time (two-days later knowledge). Next, participants completed the MCQs about the text passage they read in session one. They could take as much time as required to answer the questions. Questions (and answer options) were presented in a random order. Finally, participants completed the SAQs about the text passage they read, also in a random order. Participants were thanked for their participation and provided with the researcher's contact details if they had any questions.

5.4.2 Results

5.4.2.1 Analytic Approach

Data was analysed using the statistical package IBM SPSS (IBM Corp, 2020. IBM SPSS Statistics for Macintosh, Version 27.0.1.0). Initial exploration of the continuous variables showed they adhered reasonably to the normal distribution. To explore the effects of source text topic (solar activity vs. deserts and savannas), 'writing' activity (essay vs. free recall), and 'writing' activity modality (written vs. mental) on participants' text comprehension (MCQ total score, MCQ direct total score, MCQ inferential total score, and SAQ total score), three-way ANOVAs were conducted.

To explore the effects of source text topic, 'writing' activity, and 'writing' activity modality on participants' subjective knowledge, a mixed ANOVA was conducted. The assumption of sphericity was violated, as indicated by Mauchly's test ($\chi^2_{(2)} = 7.106$, $p = 0.029$), so Greenhouse-Geisser corrected tests are reported ($\varepsilon = 0.94$). A follow-up independent-samples t-test was conducted to explore the difference between 'writing' activity modalities on the change in participants' subjective knowledge pre- to post-writing, as this was the only significant interaction in the mixed ANOVA.

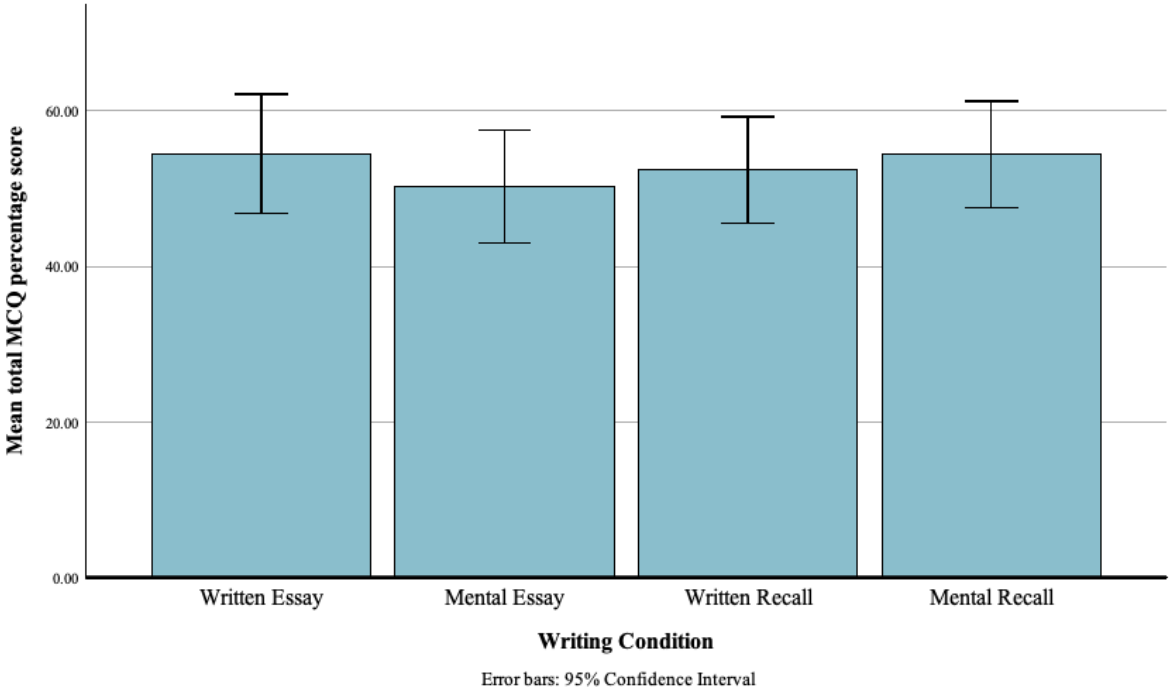
Finally, Pearson bivariate correlations were calculated between participants' text comprehension scores and their subjective knowledge ratings two-days later to assess the relationship between these measures.

Descriptive statistics (means and standard deviations) of text comprehension scores and subjective knowledge ratings across conditions are available **Error! Reference source not found.**.

5.4.2.2 Effects on Text Comprehension

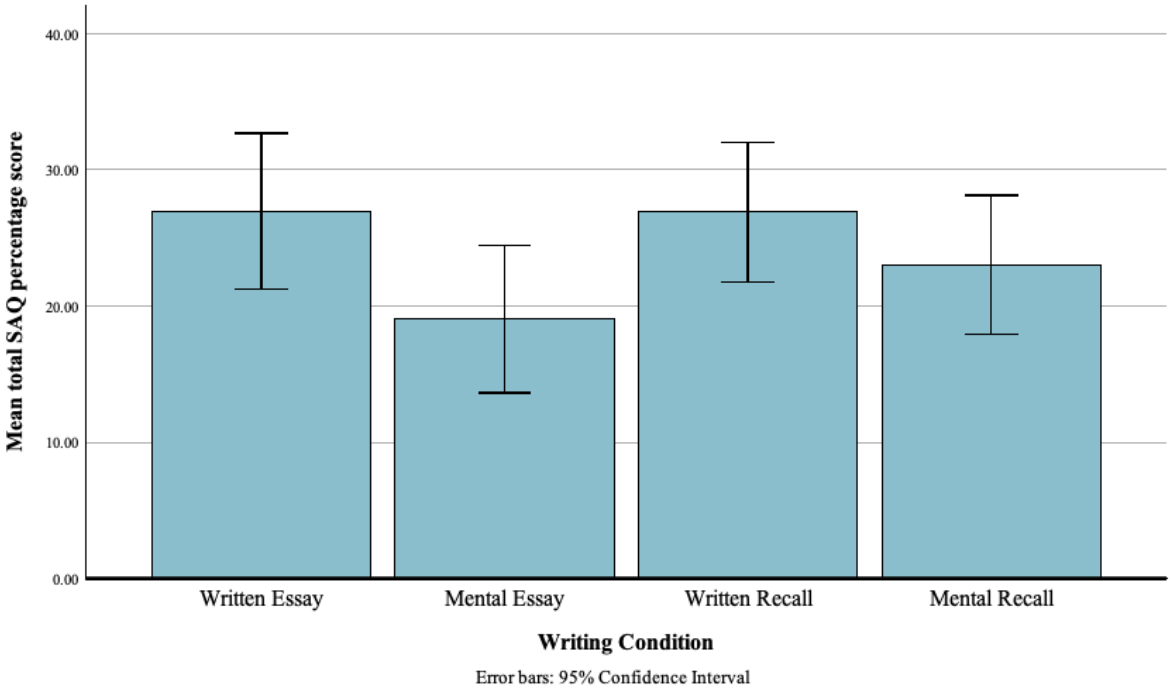
MCQs. There were no significant differences in MCQ total scores by text topic ($F_{(1, 108)} = 0.008$, $p = 0.929$, $\eta^2 < 0.001$), 'writing' activity condition ($F_{(1,108)} = 0.079$, $p = 0.780$, $\eta^2 = 0.001$), or 'writing' activity modality condition ($F_{(1, 108)} = 0.093$, $p = 0.761$, $\eta^2 = 0.001$). There were also no significant differences for either direct or inferential MCQs. This suggests participants performed equally well in the MCQs across conditions, as shown in Figure 11.

Figure 11. Mean total percentage score on MCQs by writing condition.



SAQs. There were also no significant differences in SAQ scores by text topic ($F_{(1, 108)}$ = 3.394, p = 0.068, η^2 = 0.030) or 'writing' activity condition ($F_{(1, 108)}$ = 0.528, p = 0.469, η^2 = 0.005). There was, however, a significant difference in SAQ scores across 'writing' activity modality conditions ($F_{(1, 108)}$ = 4.776, p = 0.031, η^2 = 0.042). As shown in Figure 12, participants in the mental conditions (mental essay and mental free recall) had lower SAQ scores than participants in the written conditions (written essay and written free recall).

Figure 12. Mean total percentage score on SAQs by writing condition.



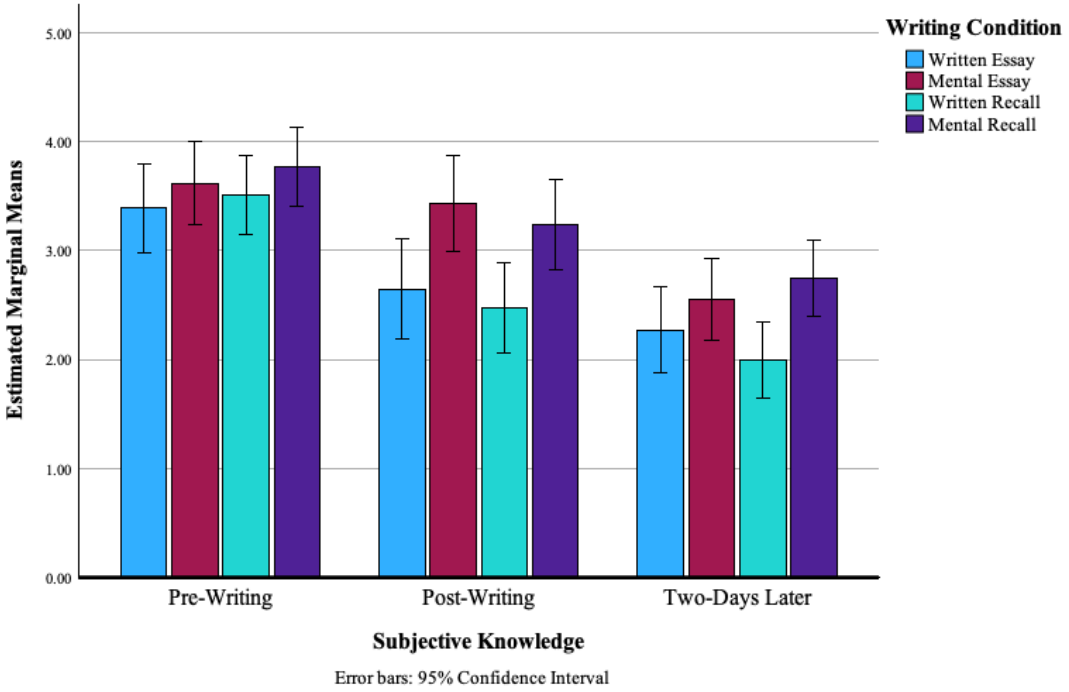
5.4.2.3 Effects on Subjective Knowledge

The mixed ANOVA exploring the effects of the conditions on participants' subjective knowledge revealed a significant interaction between 'writing' activity modality and subjective knowledge ($F_{(1.879, 202.959)} = 4.365$, $p = 0.016$, $\eta^2 = 0.039$). Figure 13 demonstrates that participants in the written conditions (written essay and written free recall) experienced a larger decrease in subjective knowledge ratings pre- to post-writing than participants in the mental conditions (mental essay and mental free recall). To determine the significance of this effect, an independent-samples t-test was conducted which found that participants' who completed the 'writing' activity in the written modality (mean = -0.90, SD = 1.12) had significantly larger decreases in their subjective knowledge ratings than participants who completed the activity in the mental modality (mean = -0.38, SD = 0.97) ($t(114) = 2.68$, $p = 0.004$, $d = 0.496$).

There were no significant interactions between 'writing' activity (essay vs. free recall) ($F_{(1.879, 202.959)} = 1.528$, $p = 0.220$, $\eta^2 = 0.014$) or text topic (solar activity vs. deserts and savannas) ($F_{(1.879, 202.959)} = 0.856$, $p = 0.420$, $\eta^2 = 0.008$) and participants' subjective knowledge.

Effects on participants' subjective knowledge for the understanding and organisation subscales of the Subjective Knowledge Rating Scale (Galbraith et al., 2023) followed similar patterns.

Figure 13. Participants' subjective knowledge at each time point across the four 'writing' conditions.



5.4.2.4 Relationship Between Text Comprehension and Subjective Knowledge

Pearson bivariate correlations between participants' text comprehension scores and their subjective knowledge ratings two-days later are presented in Table 4. There was significant, but weak, positive correlations between participants' scores on the direct MCQs and their knowledge ratings two-days later. There were also significant, and stronger, positive correlations between participants' scores on the SAQs and their knowledge ratings. The strongest correlation was between participants SAQ scores and their ratings on the understanding subscale. Direct MCQs were also more strongly correlated with the understanding subscale than the organisation subscale.

Table 4. Pearson bivariate correlations of text comprehension scores and subjective knowledge two-days later.

	Correlations		
	Knowledge two-days later	Understanding two-days later	Organisation two-days later
MCQ	0.128	0.151	0.113
MCQ direct	0.195*	0.225**	0.167*
MCQ inferential	-0.014	-0.010	-0.007
SAQ	0.249**	0.292**	0.223**

Note. $n = 116$, *$p < 0.05$, **$p < 0.01$ (1 tailed).

5.5 Discussion

This paper builds upon the findings of Paper 1 of this thesis by including a new source text topic and non-writing control conditions, and controlling time on task. It aimed to answer two research questions to assess whether the decrease in subjective knowledge pre- to -post writing observed in Paper 1 is an effect of *writing* rather than another medium (thinking), as well as if this effect replicates to a new source text passage. It also assessed the reliability of the previous finding of whether subjective knowledge ratings are moderately correlated with text comprehension scores.

First, *what are the effects of writing activities (essay vs. free recall) in different modalities (writing vs. thinking) on participants' text comprehension scores and subjective knowledge ratings, and are there any differences depending on the source text topic (solar activity vs. deserts and savannas)?* Regarding text comprehension, the results showed there were no differences between conditions on participants' MCQ scores. This implies that participants who wrote or thought about either an essay or a free recall performed equally well on these questions. However, on the SAQs, participants who wrote either an essay or a free recall received significantly higher scores than participants who thought about writing either an essay or a free recall. Taken together, these results support the finding in Paper 1 that there were no differences in text comprehension scores between the essay and free recall writing conditions (this was also the finding of Arnold et al., 2017), and this extends beyond one source text topic.

The finding that there was no difference between conditions on participants' MCQ scores likely reflects that all conditions involved retrieval practice. Retrieval practice is a valuable strategy for supporting the retention of information for subsequent recall tests (e.g., Roediger & Karpicke, 2006a; Karpicke & Roediger, 2007). Previous research has also shown that the format of retrieval practice does not matter for enhancing long-term retention, for example Blunt and Karpicke (2014) found both concept mapping and writing paragraphs about material produced better performance than re-studying.

It is also possible that the equivalent performance across conditions on MCQs reflects the dimension of learning that MCQs assess. This is in comparison with SAQs, in which participants in the written conditions performed better than participants in the mental conditions. These different types of text comprehension questions may reflect deep versus surface level learning and processing (Craik & Lockhart, 1972), concepts that traditionally are widely referred to in the higher-education literature (Dinsmore & Alexander, 2012; Beattie, Collins & McInnes, 1997). MCQs could measure more superficial surface learning, based on participants' memory of facts

from the source text, whereas SAQs could measure deeper learning reflecting more elaborate semantic encoding.

This explanation is also consistent with the findings related to the second research question, *what is the relationship between participants' subjective knowledge ratings and text comprehension scores?* Participants direct (factual) MCQ scores and their SAQ scores were significantly positively correlated with their subjective knowledge; although still moderate, the correlations between SAQ scores and subjective knowledge were stronger than for the direct MCQs. Since the response boxes were open-ended, SAQs can provide participants with more of an opportunity (than MCQs) to demonstrate their knowledge, and it is not possible for participants to guess an answer like for MCQs (Weigle, Yang & Montee, 2013). In Paper 1, the strongest correlations were also between SAQs and subjective knowledge ratings; this is notable as it reinforces the validity of the subjective knowledge rating scale (Galbraith et al., 2023) against a more objective measure of learning.

Turning to the effects of conditions on subjective knowledge, although there were no differences between the essay and free recall activities, there was a different effect depending on the modality of the activity (written versus thinking). Participants who wrote their essay or free recall experienced a significantly larger decrease in their subjective knowledge pre- to post-writing than participants who thought about writing an essay or free recall. This replicates the finding from Paper 1 that after writing about a source text, participants rate their knowledge much lower than they did before writing, and extends it to show that this decrease is not as prominent when participants do not write. This is an important finding since it demonstrates that the 'illusion of explanatory depth' (Rozenblit & Keil, 2002) participants experience is a unique feature of writing itself.

Taken together, the results support the conclusion from Paper 1 that writing from a single source text in a closed-book setting mostly engages writers in knowledge-telling (Bereiter & Scardamalia, 1987), rather than the knowledge-transforming or knowledge-constituting processes proposed by the dual-process model of writing (Galbraith & Baaijen, 2018). When writing an essay about material from a source text, or writing down everything one can remember from a source text, all participants retrieved information from memory and translated it into written text. This therefore produced equivalent performance on the text comprehension test, but also resulted in a decrease in subjective knowledge ratings from pre- to post-writing. Contrary to previous assumptions, knowledge-telling can play a role in learning as it allows individuals to identify information that they do not know.

Nevertheless, what this paper reveals is that this effect is not solely due to retrieval practice. This is because the thinking (mental) conditions also required retrieval practice, as they also

took place in a closed-book environment. Those who *wrote* an essay or free recall performed better than those who *thought* about writing an essay or free recall on SAQs, but writers experienced a significantly larger decrease in subjective knowledge than thinkers. Therefore writing, rather than another medium such as thinking, serves a unique role in the learning process.

The cognitive and metacognitive benefits of writing may be due to the physical and external written representation of information, thoughts, and ideas. Writing without visual feedback can increase processing demands and as such writers may have poorer recall performance; make more mistakes in their texts; and have difficulties generating inferences and retrieving information (e.g., Grabowski, 1996, 2007; Olive & Piolat, 2002). In the current study, writing an essay or writing down everything recalled from memory provided participants with an external representation of their knowledge, in comparison with those in the thinking conditions who could not physically view the information that had come to mind. Therefore, participants in the writing conditions could more easily reflect on their current knowledge level which supported metacognitive monitoring (identifying gaps in their knowledge). Furthermore, writing requires participants to elaborate and organise retrieved information into written text (Arnold et al., 2017). Such elaboration and organisation processes may have supported performance on the SAQs, resulting in better performance for writers than thinkers.

Overall, these findings have important implications for writing-to-learn theory as they imply that knowledge-telling does have some value for learning and the 'development' of knowledge (i.e., learning what one does not know). Whilst retrieval practice is an important underlying process in these effects, it cannot fully explain the superiority of writing in comparison to other retrieval activities such as thinking. Future research could explore possible explanations of the writing-superiority effect by comparing source-based writing with and without visual feedback as well as in closed-book versus open-book environments.

5.5.1 Limitations

Although the results discussed here provide valuable insights into writing-to-learn from sources, certain limitations must be acknowledged. Data collection for the experiments explained here was conducted online due to the Covid-19 pandemic. This was partially a strength as it allowed a reasonably large sample of participants to be recruited, who could undertake the experiment at a time most convenient to them. However, in both experiments a high number of the recruited participants were not included in the analysis because they did not complete the full procedure. In addition, the online nature meant there is some lack of experimental control as the researcher was not observing participants as they undertook the study. It is possible that

participants did not spend the full allocated time completing the thinking or writing activity, or became distracted whilst completing the experiment. A replication of the study in a laboratory setting would therefore be valuable.

The homogeneity of the sample is another limitation of the current paper. In both experiments, the majority of participants were female, undergraduate psychology students. This may limit the generalisability of the findings. Future research should endeavour to include a more diverse sample in terms of gender and academic background. Furthermore, the study did not consider the range of individual differences that could affect the writing process and learning outcomes, for example participants' beliefs about writing (e.g., White & Bruning, 2005; Baaijen, Galbraith & de Glopper, 2014).

5.6 Conclusion

In conclusion, this study extends the findings of Paper 1 by demonstrating that the act of writing itself plays a distinctive role in the learning process, beyond the effects of retrieval practice alone. Whilst all participants engaged in retrieval practice, those in the writing conditions showed better performance on short-answer comprehension questions and a greater decrease in their subjective knowledge ratings from pre- to post-writing than participants in the thinking (mental) conditions. Participants' subjective knowledge ratings were also significantly positively correlated with their SAQ and factual MCQ scores. These effects were consistent across different source text topics. Importantly, the observed decrease in subjective knowledge ratings following writing about a source text reinforces that writing facilitates the identification of knowledge gaps and therefore knowledge-telling can have a valuable role in learning. Collectively, the insights from this paper reinforce the theoretical and practical importance of *writing*-to-learn and suggest interesting avenues for future research to assess possible explanations of the writing-superiority effect.

Chapter 6 Paper 3. Effects of Written Versus Spoken Summaries on Memory for Text: A Replication and Extension of Silva and Limongi (2019)

6.1 Abstract

This paper explored the effects of writing versus speaking summaries on long-term memory consolidation of source text passages, in a replication and extension of Silva and Limongi (2019). Experiment 1 was a direct replication of Silva and Limongi's study. In a within-subjects (repeated measures) design, across eight experimental blocks, participants read short texts, wrote or spoke a summary of each text, and completed an episodic memory word-recognition task in which they had to identify if target words appeared in the original texts. Experiment 2 aimed to rule out a potential confounding factor in Silva and Limongi's study (a priming, or modality, effect). Participants listened to the short texts (rather than reading them), wrote or spoke a summary of each text, and completed the word-recognition task in an auditory form (participants listened to the target words). Across both experiments, results replicated Silva and Limongi's findings. A writing-superiority effect was found, represented by the non-decision parameter of the drift-diffusion model. After writing summaries, participants had faster reaction times than after speaking summaries.

6.2 Introduction

Writing-to-learn is the idea that the goal of writing is not just to communicate information to a reader, but that it can also facilitate learning and the writer's development of understanding (Emig, 1977; Klein, 1999; Galbraith et al., 2025). Classic models of writing (e.g., Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) describe the writing process, but say little about how knowledge is represented in long-term memory and subsequently *how* writing leads to learning (Baaijen & Galbraith, 2018). It is also assumed that translating, or knowledge-telling (retrieving content from memory), does not lead to developments of understanding. Developments of understanding are said to occur when the writer engages in goal-directed problem-solving processes, or knowledge-transforming.

The dual-process model of writing (Galbraith & Baaijen, 2018) explains the effect of learning through writing as a consequence of two processes: knowledge-constituting and knowledge-transforming. This is built upon complementary learning systems theory (McClelland, McNaughton & O'Reilly, 1995; Kumaran et al., 2016; O'Reilly et al., 2011) which distinguishes between semantic and episodic memory systems. In the knowledge-constituting process, a semantic network of thoughts, ideas, and information guides composition, allowing the writer to synthesise content. The knowledge-transforming processes involves goal-directed retrieval of information, thoughts, and ideas from episodic long-term memory. Studies comparing different drafting strategies provide evidence for the dual-process model and indicate knowledge development occurs when a writer manages these dual processes by reorganising and integrating new information in memory (Baaijen & Galbraith, 2018).

Although writing-to-learn theory suggests that learning only takes place through writing via knowledge-transforming and knowledge-constituting, research outside of the field has demonstrated that simply recalling information from memory (retrieval practice) can enhance learning. 'The testing effect' captures the robust finding that engaging in recall tests improves the long-term retention and transfer of information above restudying and rereading (e.g., Glover, 1989; Karpicke, Butler & Roediger, 2009; Roediger & Karpicke, 2006a, 2006b; Karpicke, 2012). One study that has considered the role of retrieval practice in writing-to-learn was conducted by Arnold et al. (2017). Participants who engaged in retrieval writing activities (closed-book essay writing or free recall writing) performed better on a text comprehension test two-days later than participants who engaged in non-retrieval writing activities (open-book notetaking or highlighting). However, Arnold et al.'s (2017) study relies on a response accuracy measure to assess learning. Silva and Limongi (2019) have argued writing-to-learn research could benefit from a reaction time (RT) measure to capture long-term memory consolidation from writing at the neural level.

Silva and Limongi's (2019) claim is based on a cognitive system proposed by Marr (1982), who suggested human cognition can be studied at the computational, algorithmic, and implementational levels. They suggest that writing-to-learn theory (e.g., Hayes & Flower, 1980; Bereiter & Scardamalia, 1987) is based at the algorithmic level. At this level, emphasis is on identifying cognitive processes and strategies for problem-solving (Griffiths, Lieder & Goodman, 2015). Here, learners consolidate knowledge in long-term memory through rehearsal or elaborative rehearsal (Tulving & Craik, 2000). Recently, scholars have suggested theories of learning should also be based at the neural level (Cooper & Peebles, 2015), linking the algorithmic and implementational levels (Silva & Limongi, 2019). This would more strongly facilitate making inferences about cognitive processes from behavioural data, particularly with the use of neuroimaging research (Cooper & Peebles, 2015). Silva and Limongi (2019) suggest Galbraith and Baaijen's (2018) dual-process model accommodates a link between algorithmic and implementational levels because knowledge representation is based on complementary learning systems theory (McClelland, McNaughton & O'Reilly, 1995; Kumaran et al., 2016; O'Reilly et al., 2011).

In cognitive psychology, long-term memory consolidation is not only measured through response accuracy measures, but also RT measures (also known as mental chronometry; Silva & Limongi, 2019). This is achieved by calculating the speed at which information is retrieved from memory: the time between when a stimulus is presented and when the representation of such stimulus is activated in long-term memory (Silva & Limongi, 2019). One frequently used example is the lexical decision task (Meyer & Schvaneveldt, 1971), in which the speed at which participants recognise a combination of letters as a real word or a non-word is measured. Studies have used the lexical decision task for a range of aims, including to explore effects of word frequency (e.g., Abrams & Balota, 1991; Hudson & Bergman, 1985; Hendrix & Sun, 2021; Mendes, Luna & Albuquerque, 2021; Yap et al., 2006); neighbourhood effects (e.g., Andrews, 1989; Meade, Grainger & Declerck, 2021; Dirix et al., 2017); and to explore cognitive processes involved in reading (e.g., Coltheart et al., 2001; Kuperman et al., 2013).

Silva and Limongi's (2019) experiment comprised of eight experimental blocks, with a repeated measures within-subjects design. In each block, 25 participants read a short text passage and produced a summary of the passage by either writing (typing) or speaking. They then completed an episodic-memory word-recognition (lexical decision) task in which they had to respond, as quickly and as accurately as possible, to whether 60 target words (30 in-text and 30 non-text words) appeared in the original source text. This procedure was repeated for eight text passages. Participants' RT on the word-recognition task was analysed using Bayesian hierarchical drift-diffusion modelling (HDDM), which revealed a writing-superiority effect. Participants RTs were faster are writing summaries than after speaking them, explained by the

non-decision parameter of the drift diffusion model such that writing summaries was associated with faster retrieval of information from episodic memory.

What Silva and Limongi's (2019) study reveals is that writing effects found in previous research (e.g., Arnold et al., 2017) must be the result of more than just retrieval practice. This is because both writing and speaking in Silva and Limongi's study involved retrieval. Silva and Limongi (2019) provide a somewhat vague explanation of these effects in relation to the dual-process model (Galbraith & Baaijen, 2018), hypothesising that "reconstructive processes and text revision could be two writing-specific operations that facilitate LTM consolidation" (p. 230). Essentially, they suggest that because writing takes longer to produce, there is more opportunity for reflection than in speaking.

6.2.1 Drift-Diffusion Modelling

Silva and Limongi (2019) found that the writing-superiority effect over speaking was explained by the non-decision parameter of the drift-diffusion model. Researchers often analyse RT data using statistical methods such as analysis of variance of the sample mean (Whelan, 2008). However, several authors promote methods that analyse the *distribution* of RT data to assess how distributions vary depending on stimuli or condition (Norris, 2013). There is more information that can be extracted from RT data than the mean (Balota et al., 2008; Balota & Yap, 2011; Anders et al., 2015). Since tasks involve task-specific processing, models should incorporate both speed and task performance (Norris, 2013). A main advantage of drift-diffusion models is that they consider response accuracy and RT simultaneously (Ratcliff, 1978). The speed-accuracy trade-off is an important feature of quick decision-making tasks and therefore data should be analysed with methods that account for such interactions between speed and accuracy (Myers, Interian & Moustafa, 2022).

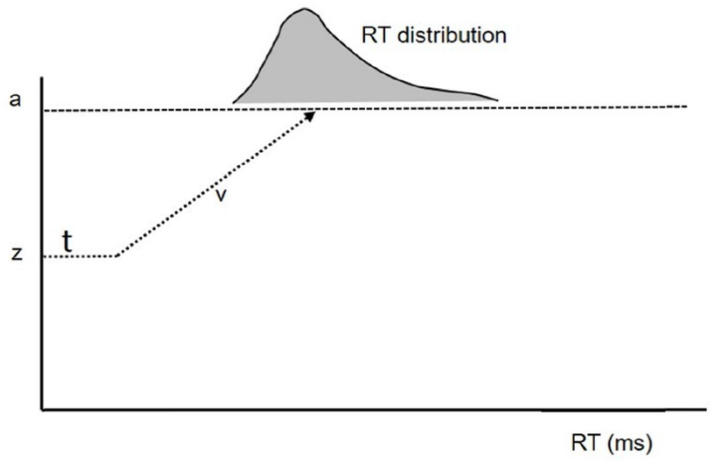
The drift-diffusion model (Ratcliff, 1978; Ratcliff, Gomez & McKoon, 2004) models simple two-alternative forced-choice decision making tasks as an evidence accumulation process. The process underlying a decision, e.g., to recognise a word as an in-text word, is an accumulation of noisy information (evidence) until a threshold is reached and the decision is made (Ratcliff, Gomez & McKoon, 2004; Wiecki, Sofer & Frank, 2013; Norris, 2013). The drift-diffusion model is very well-established and is often the standard analysis method for RT data.

Specifically, there are four parameters in the basic drift-diffusion model (Ratcliff, 1978; Ratcliff, Gomez & McKoon, 2004; Silva & Limongi, 2019), shown in Figure 14: (1) the starting point of the process, *z*; (2) the threshold at which the decision is reached, *a*; (3) the drift rate, *v*, which is the mean rate of approach to the threshold; and (4) non-decision processes, *t*. Also included in the model are parameters of the intertrial variability of the starting point (*sz*), drift rate (*sv*) and non-

decision processes (*st*). Intertrial variability allows the model to explain the speed of correct and incorrect responses (Ratcliff, Gomez & McKoon, 2004). The model also automatically predicts the positive skew typical of RT distributions (Ratcliff, Gomez & McKoon, 2004) in decision-based tasks.

Another advantage of drift-diffusion models is that they can provide insight into the underlying cognitive processes of a task through the interpretation of the parameters, which have been well-established (Aschenbrenner et al., 2016; Voss, Rothermund & Voss, 2004). Voss, Rothermund and Voss (2004) used experimental manipulations to demonstrate the empirical validity of the psychological interpretations of the parameters in the drift-diffusion model. Specifically, the starting point (*z* parameter) can be interpreted as an estimate of bias towards one response over the other. The threshold at which a decision is reached (*a* parameter), measures conservatism, e.g., a higher threshold implies that more information has to be collected to execute a response. The drift rate (*v* parameter) reflects the amount of information that has to be accumulated and is related to task difficulty or perceptual sensitivity. Finally, the *t* parameter can be inferred as the duration of non-decision processes such as encoding of the stimuli or the motor response.

Figure 14. The basic drift-diffusion model (reproduced from Silva & Limongi, 2019, p. 219). This work is licensed under CC BY-NC-ND.



The drift-diffusion model has been applied in a range of previous research measuring RT in two-choice decision tasks. For example, in the lexical decision task requiring participants to identify whether a string is a word or a non-word, Ratcliff, Gomez & McKoon (2004) identified that the drift rate accounted for word-frequency effects. The model has also been used to explain age differences in RT tasks in that older people typically have slower non-decision times and higher thresholds at which decisions are made (Theisen et al., 2021).

Silva and Limongi's (2019) findings indicated the writing superiority effect was explained by the non-decision parameter (*t*). This parameter is reflective of encoding stimuli and executing motor responses (Myers, Iterian & Moustafa, 2022), as such participants in Silva and Limongi's study required less time to carry out these processes after writing summaries than after speaking them. They therefore concluded that "initial stages of learning through writing are associated with fast episodic-memory retrieval" (p. 211).

6.2.2 Writing Versus Speaking

The writing superiority effect found by Silva and Limongi (2019) is aligned with recent conceptions of writing-to-learn that a fundamental feature is that it involves *writing* and not another medium (Galbraith et al., 2025). Few writing-to-learn studies have included non-writing control conditions, such as speaking. However, results from these studies tend to show writing to be more beneficial than speaking. For example, Rivard and Straw (2000) compared speaking and writing in eighth-grade students and found writing to be valuable for transforming thoughts and ideas into structured and coherent knowledge, although speaking was useful for sharing thoughts and ideas with other students. In another study, Rivard (2004) found high achieving eight-grade pupils gained more from writing than talking. Nonetheless, speaking (dictation) can facilitate production of better-quality texts than writing amongst younger and struggling writers who are still learning the motor skills required for transcribing processes (Bourdin & Fayol, 2002).

Klein, Piacente-Cimini and Williams (2007) compared the effects of composing an analogy in writing versus speech. Undergraduate students allocated to the writing condition received higher scores on a post-test of their understanding than students allocated to the speaking condition. The authors suggested this was because writing facilitates more reflection (as it is slower paced than speaking) and encourages semantic elaboration through the explicit externalisation of knowledge. By offloading thoughts and ideas physically in writing, they can be reviewed (Grabowski, 2007) which subsequently re-activates these ideas in memory, supporting inference generation. This eludes to the famous quote by E. M. Forster, *"How can I know what I think until I see what I say?"*.

Several authors highlight the role of the permanency of written products, demonstrating how such an external representation supports writers to review and monitor their texts, which therefore supports learning (Olson, 1996; Bereiter, 1980; Olive & Piolat, 2002; Olive, 2010; Berninger & Winn, 2006). Studies comparing invisible writing (e.g., early studies by Gould, 1980), where visual feedback is removed, to writing or speaking have demonstrated the value of writing in providing an 'external store', so the emerging text does not have to be retained in memory

(Olive & Piolat, 2002). Grabowski (1996) described how invisible writing influenced the frequency that information was repeated, utterance length, reliance on high-level summarisation for condensing content, and the number of mistakes made as a result of incorrect inferences or problems retrieving information. Furthermore, Grabowski (2007) reports a series of experiments with university students comparing writing with invisible writing, dictating, and speaking and found a writing-superiority effect for the recall of information.

It has also been suggested that the external representation of writing supports the cognitive demands of writing and higher-level writing processes such as planning and revision (Olive & Passerault, 2012), as well as goal-directed problem-solving (Berninger & Winn, 2006). Olive and Piolat's (2002) study provides evidence for this explanation. They compared 40 undergraduate students in a handwriting copy task with visual feedback or with an inkless pen. During writing, participants were subject to a secondary RT task where they would have to react as fast as possible to auditory probes. RTs were longer following writing without visual feedback, suggesting this increased the load of execution processes. When writers have visual feedback, execution processes like motor processes can be removed from working memory once complete to facilitate more complex writing processes (Olive & Piolat, 2002).

6.2.3 Aims and Rationale

The current study is a direct replication (Experiment 1) and extension (Experiment 2) of Silva and Limongi's (2019) study. It aims to validate the word-recognition RT task as a novel methodological tool for measuring long-term memory consolidation from writing. This is important as the implications of the measure for the writing-to-learn field are far-reaching. For example, it could facilitate a cognitive neuroscience approach to writing-to-learn through neuroimaging studies using fMRI. Currently, there is very little neuroimaging research on writing, yet this could have high utility for identifying cognitive processes at the neural level (Pugh et al., 2006; Silva & Limongi, 2019).

Conducting a direct replication has been argued to be "the best (and possibly the only) believable evidence for the reliability of an effect" (Simons, 2014, p. 76). This is particularly valuable in the early stages of establishing new methodologies. Experiment 1 reproduces the procedure of Silva and Limongi (2019), in a different sample of undergraduate participants, using the same materials. The only difference to note is that the materials were translated from Spanish to English; Silva and Limongi's study was conducted in a Spanish-speaking country, the current study was conducted in an English-speaking country.

Silva and Limongi's (2019) study, and therefore Experiment 1, has a possible confound in the design that could explain the writing-superiority effect as a priming or modality effect. Both the

source text and the target words in the word-recognition task were presented orthographically. Repeated exposure to orthographic representations may have primed participants to have faster RT after producing written summaries. In Experiment 2, instead of reading the source text and target words, participants listen to them. Previous research suggests learners use visuospatial information to find and retrieve content from a text (Hayes, 1996; Olive & Passerault, 2012), therefore a priming effect is plausible since reading a text provides visuospatial cues which listening to a text does not. Authors have also noted how listening to speech can be more difficult than reading (Eysenck & Keane, 2015; Piolat, 2007). If a writing-superiority effect does exist, RTs will remain faster after writing summaries than after speaking summaries. However, if the findings result from a priming or modality effect, RTs following written summaries will not be faster, and the effect might reverse so that quicker RTs occur after spoken summaries.

6.3 Method

6.3.1 Experiment 1

6.3.1.1 Participants

In Experiment 1, participants were 31 (27 female and 4 male) higher-education students at the University of Southampton. They were mostly undergraduate Psychology students (n = 30) with a mean age of 19.2. All participants spoke English as their first language. Participants volunteered to take part in the experiment, and those studying Psychology at undergraduate level were given course credits for participating. The study was approved by the University of Southampton Ethics Committee (Appendix F.4) and participants signed consent forms prior to participation.

6.3.1.2 Design and Procedure

In a direct replication of Silva and Limongi's (2019) study, Experiment 1 was a within-subjects experimental design with three stages in each procedural block: reading, production, and memory. There was one practice (familiarisation) block followed by eight experimental blocks with two conditions (spoken and written production; four experimental blocks per condition). Only data from experimental blocks were analysed. During reading, participants read a short text passage on screen. Immediately after, they were instructed to either write (type) or speak a summary of the text passage they read. This was followed by an episodic memory word-recognition task. The experiment was run on E-Prime 3.0 software. All materials were provided on request by Limongi (Silva & Limongi, 2019) and translated to English.

Reading. In the reading task, participants were presented with a text passage on screen which had a title at the top of the passage. They had two minutes to read the text with a timer displayed to inform participants of the remaining time. Participants were not allowed to take notes whilst reading. In the practice block, all participants read a text on 'Pets'. In the experimental blocks, texts were chosen randomly without replacement from a pool of eight texts (Appendix C), and each text could appear for either production modality (written or spoken). The text passage disappeared automatically after two minutes had passed.

Production. In the production task, participants received instructions depending on the randomly assigned condition. For written production, participants read that they had to write (type) a summary with the main idea of the text passage. Following this instruction, a text box appeared on the screen for participants to type their response into, with the instruction, *"now write the main idea of the passage"* shown above. Participants had 90 seconds to produce their written summary, and a timer was displayed on screen to inform them of the remaining time. For

spoken production, participants heard that they had to speak a summary with the main idea of the text passage after hearing a tone. Following this instruction, an image of a microphone was displayed on screen alongside the following text, *"now say the main idea of the passage"*. Participants had 90 seconds to pronounce their spoken summary, and a timer was displayed on screen to inform them of the remaining time. In both conditions, for a summary to be considered valid, participants had to produce at least two sentences. Participants were informed of this requirement at the beginning of the study.

Memory. In the episodic memory word-recognition memory task, 30 words (per block) appeared on the screen one at a time, and participants had to indicate as quickly and as accurately as possible if the word had appeared in the original text passage during the reading task. Participants responded on the keyboard by pressing the 'k' key if the word did appear in the text, and the 'l' key if the word did not appear in the text. 15 of the words appeared in the original text, and 15 did not (see Appendix C for the full list of words). Participants RT to each word was recorded for analysis.

6.3.2 Experiment 2

6.3.2.1 Participants

In Experiment 2, participants were 33 (27 female and 5 male) undergraduate students at the University of Southampton, with a mean age of 19.63. All participants were Psychology students, either as a main subject (*n* = 29), or joint honours degree with Criminology (*n* = 2) or Education (*n* = 1). One participants did not disclose their age or subject. Most participants (*n* = 31) spoke English as their first language. Participants volunteered to take part in the experiment and were given course credits for participating as part of the undergraduate Psychology research participation scheme. The study was approved by the University of Southampton Ethics Committee (Appendix F.4) and participants signed consent forms prior to participation. Two participants were excluded from analysis as they did not record any RTs during the memory task, therefore 31 participants were included in the analysis.

6.3.2.2 Design and Procedure

Silva and Limongi's (2019) original study, and the replication, have a key limitation. There is a potential confound in the design: an orthographic priming or modality effect. The source text in the reading task and the words in the memory task were presented orthographically, so when participants produced their summaries in the written condition, they may have been primed by multiple orthographic representations, leading to faster RTs compared to spoken production.

The design and procedure for Experiment 2 is identical to that of Experiment 1, with some minor modifications to the procedure to control for the potential confound.

Reading (listening). Participants listen to the source text from a recording, rather than view it on the screen (a listening task rather than a reading task).

Production. Procedure identical to Experiment 1.

Memory. Participants listen to each target word from a recording, rather than view it on the screen, and then respond on the keyboard if the word did or did not appear in the original text.

6.4 Results

6.4.1 Analytic Approach

Participants' responses in the word-recognition memory task were coded as correct (1) if the response matched the expected response (i.e., the participant correctly identified an in-text word as an in-text word) and incorrect (0) if the response did not match the expected response (i.e., the participant identified an in-text word as a non-text word, or a non-text word as an in-text word). Missing RTs and RTs < 200 ms and > 2000 ms were excluded from analysis (Experiment 1: 3.41%; Experiment 2: 5.87%). Analysis was conducted in Python (Version 2.7.18) using the HDDM library (Wiecki et al., 2022).

Drift-diffusion models were fit to the RT data with the goal of replicating Silva and Limongi's (2019) finding that the non-decision parameter (t) could account for the learning effect of writing. Specifically, Bayesian Hierarchical Drift Diffusion Modelling (HDDM) was used as it allows for modelling at multiple hierarchical levels (both participant and group parameters) and as it considers response accuracy and RT simultaneously. This method of analysis breaks down the decision-making process into distinct components which can be related to cognitive processes; therefore, HDDM models the unobservable cognitive processes behind the RT rather than the RT itself. Adopting a Bayesian approach allows for the incorporation of prior information and a quantification of uncertainty.

In the model, the 't' parameter (non-decision time) was a free parameter; 'a' (decision threshold), 'v' (drift rate), and 'z' (starting point) were fixed parameters; and the intertrial variability of 'z', 'v', and 't' ('sz', 'sv', 'st') were estimated in the model. The chain length of the model was 5000; the number of burn-in iterations was 250; and the chains were generated with no thinning. Model convergence was assessed by calculating the Gelman-Rubin statistic ($\hat{R}$). The parameter estimates of the model are reported at the group and participant level.

6.4.2 Experiment 1

6.4.2.1 Descriptive Statistics

Descriptive statistics of RTs following spoken and written summary production for correct and incorrect responses are shown in Table 5.

Table 5. Means (and standard deviations) of RT (ms) by condition and accuracy (Experiment 1).

		Response Accuracy		
		Overall	Correct	Incorrect
	Overall	897.01 (312.13)	881.78 (299.05)	975.01 (362.23)
Condition	Written	888.53 (316.42)	875.12 (305.48)	958.47 (360.55)
	Spoken	905.55 (307.56)	888.52 (292.28)	991.15 (363.43)

6.4.2.2 Bayesian HDDMs

Overall model convergence was good with all R^ values below 1.1. Table 7 presents the parameter estimates at the group and participant levels. For each participant, the *t* parameter was smaller in the written condition than in the spoken condition. Additionally, as visualised in Figure 15, the posterior distributions of *t* estimates do not overlap, showing they differ. Overall, non-decision time was shorter in the written condition than in the spoken condition. Posterior predictive checks indicated that the model fairly reproduced the observed data (Table 6).

Figure 15. Posterior distributions of *t* (non-decision) estimates (Experiment 1).

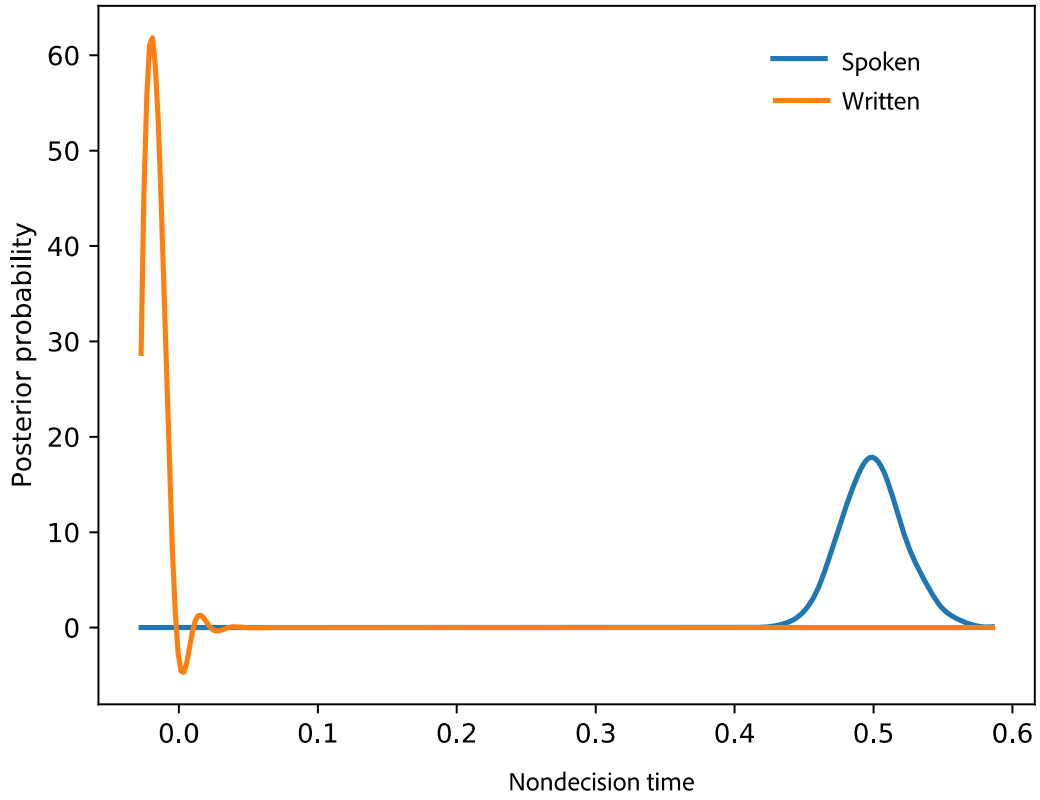


Table 6. Posterior predictive checks (Experiment 1).

	Observed	Model Generated	SD	SEM	MSE	Credible
Accuracy	0.84	0.84	0.05	0.00006	0.0022	True
RT (Upper Bound)	0.88	0.91	0.14	0.00088	0.0196	True
RT (Lower Bound)	-0.98	-0.99	0.20	0.00031	0.0407	True

Note. Upper Bound = Correct Responses; Lower Bound = Incorrect Responses. Credible (True) = in the 95% credible interval. Observed and model-generated values are collapsed across conditions.

Table 7. Parameter estimates of the HDDM at the group and participant level (Experiment 1).

Parameter	*a*		*z*		*v*		*t* (spoken)		*t* (written)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Group	1.583	0.077	0.468	0.014	1.700	0.094	0.495	0.023	0.482	0.027
Participant 1	1.704	0.077	0.501	0.029	1.225	0.158	0.483	0.018	0.469	0.023
Participant 2	1.459	0.056	0.362	0.025	1.607	0.159	0.339	0.008	0.325	0.013
Participant 3	1.707	0.076	0.439	0.027	1.664	0.153	0.367	0.010	0.354	0.014
Participant 4	1.100	0.049	0.502	0.030	1.499	0.180	0.534	0.012	0.520	0.016
Participant 5	1.332	0.063	0.478	0.029	1.525	0.167	0.521	0.017	0.507	0.021
Participant 6	1.890	0.086	0.478	0.028	1.516	0.151	0.407	0.012	0.393	0.016
Participant 7	1.044	0.051	0.560	0.036	1.595	0.191	0.823	0.014	0.810	0.018
Participant 8	1.596	0.079	0.497	0.030	1.800	0.172	0.598	0.014	0.585	0.018
Participant 9	1.670	0.092	0.480	0.034	1.820	0.173	0.601	0.022	0.587	0.026
Participant 10	2.650	0.135	0.317	0.024	2.055	0.166	0.306	0.011	0.292	0.015
Participant 11	1.732	0.092	0.491	0.032	1.582	0.162	0.536	0.022	0.522	0.026
Participant 12	1.301	0.063	0.545	0.033	1.562	0.176	0.608	0.016	0.595	0.021

Parameter	*a*		*z*		*v*		*t* (spoken)		*t* (written)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 13	2.187	0.129	0.429	0.031	1.810	0.162	0.652	0.021	0.639	0.026
Participant 14	2.021	0.095	0.371	0.026	1.937	0.156	0.300	0.009	0.286	0.014
Participant 15	1.285	0.062	0.487	0.031	1.762	0.176	0.553	0.016	0.539	0.020
Participant 16	1.576	0.067	0.363	0.025	2.007	0.177	0.436	0.009	0.422	0.013
Participant 17	1.110	0.062	0.564	0.037	1.918	0.204	0.699	0.017	0.686	0.021
Participant 18	1.131	0.057	0.537	0.038	1.901	0.207	0.515	0.015	0.502	0.019
Participant 19	2.082	0.095	0.355	0.024	1.849	0.154	0.322	0.013	0.309	0.017
Participant 20	1.314	0.067	0.507	0.032	1.720	0.172	0.549	0.018	0.536	0.022
Participant 21	1.406	0.061	0.464	0.0272	1.858	0.176	0.479	0.009	0.465	0.013
Participant 22	1.281	0.065	0.547	0.035	1.712	0.189	0.534	0.017	0.520	0.022
Participant 23	2.167	0.099	0.375	0.024	1.590	0.148	0.348	0.009	0.335	0.014
Participant 24	1.215	0.057	0.532	0.033	1.591	0.176	0.523	0.014	0.510	0.018
Participant 25	1.493	0.079	0.567	0.035	1.714	0.175	0.582	0.019	0.569	0.023
Participant 26	1.885	0.083	0.412	0.026	1.527	0.147	0.391	0.014	0.378	0.018

Parameter	*a*		*z*		*v*		*t* (spoken)		*t* (written)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 27	1.273	0.061	0.495	0.032	1.658	0.173	0.512	0.017	0.499	0.021
Participant 28	1.706	0.087	0.472	0.031	1.720	0.162	0.502	0.020	0.489	0.024
Participant 29	1.442	0.072	0.488	0.031	1.805	0.170	0.452	0.017	0.439	0.021
Participant 30	1.856	0.087	0.458	0.029	1.800	0.159	0.440	0.013	0.427	0.017
Participant 31	1.321	0.058	0.449	0.029	1.375	0.168	0.408	0.015	0.394	0.020

Note. Intertrial variability of *z* (*sz*), *v* (*sv*), and *t* (*st*) were estimated at group level (means (and standard deviations): *sz* = 0.103 (0.049), *sv* = 1.062 (0.090), *st* = 0.269 (0.008).

6.4.3 Experiment 2

6.4.3.1 Descriptive Statistics

Descriptive statistics of RTs following spoken and written summary production for correct and incorrect responses are shown in Table 8.

Table 8. Means (and standard deviations) of RT (ms) by condition and accuracy (Experiment 2).

		Response Accuracy		
		Overall	Correct	Incorrect
	Overall	1235.00 (280.79)	1215.19 (272.26)	1331.97 (301.14)
Condition	Written	1226.71 (284.32)	1202.78 (275.53)	1341.17 (297.78)
	Spoken	1243.25 (277.04)	1227.44 (268.49)	1322.46 (304.53)

6.4.3.2 Bayesian HDDMs

Overall model convergence was good with all $\hat{R}$ values below 1.1. Table 10 presents the parameter estimates at the group and participant levels. For each participant, the *t* parameter was smaller in the written condition than in the spoken condition. Additionally, as visualised in Figure 16, the posterior distributions of *t* estimates do not overlap, showing they differ. Overall, non-decision time was shorter in the written condition than in the spoken condition. Posterior predictive checks indicated that the model fairly reproduced the observed data (Table 9).

Figure 16. Posterior distributions of *t* (non-decision) estimates (Experiment 2).

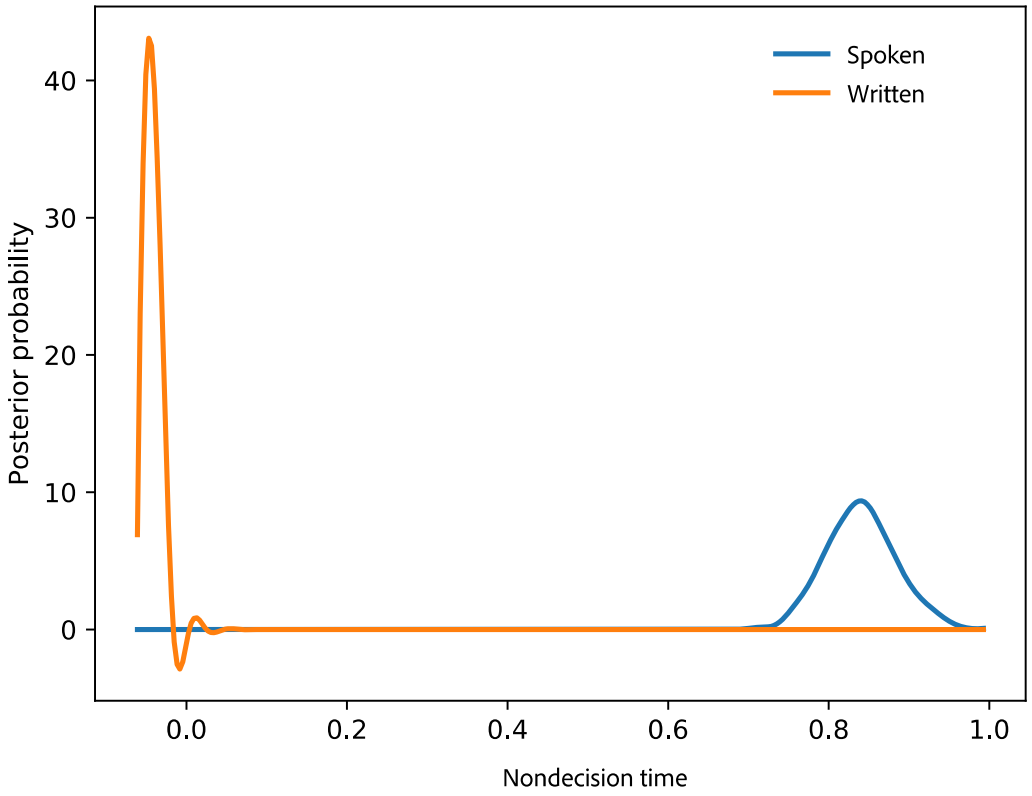


Table 9. Posterior predictive checks (Experiment 2).

	Observed	Model Generated	SD	SEM	MSE	Credible
Accuracy	0.83	0.84	0.04	0.00008	0.0020	True
RT (Upper Bound)	1.22	1.24	0.10	0.00069	0.0109	True
RT (Lower Bound)	-1.33	-1.22	0.15	0.01151	0.0338	True

Note. Upper Bound = Correct Responses; Lower Bound = Incorrect Responses. Credible (True) = in the 95% credible interval. Observed and model-generated values are collapsed across conditions.

Table 10. Parameter estimates of the HDDM at the group and participant level (Experiment 2).

Parameter	*a*		*z*		*v*		*t* (spoken)		*t* (written)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Group	1.426	0.090	0.487	0.017	1.378	0.077	0.833	0.044	0.800	0.051
Participant 1	0.859	0.057	0.580	0.038	1.392	0.170	1.168	0.016	1.135	0.024
Participant 2	1.382	0.068	0.454	0.035	1.457	0.150	0.781	0.014	0.749	0.021
Participant 3	1.735	0.076	0.444	0.031	1.213	0.128	0.595	0.014	0.562	0.021
Participant 4	1.496	0.065	0.409	0.030	1.389	0.132	0.648	0.012	0.616	0.019
Participant 5	0.943	0.071	0.547	0.041	1.321	0.164	1.006	0.034	0.973	0.041
Participant 6	0.973	0.061	0.575	0.043	1.503	0.190	0.947	0.018	0.915	0.025
Participant 7	1.814	0.078	0.370	0.025	1.273	0.120	0.629	0.012	0.596	0.019
Participant 8	1.701	0.084	0.425	0.032	1.516	0.143	0.735	0.015	0.702	0.022
Participant 9	1.084	0.072	0.594	0.049	1.607	0.226	1.162	0.020	1.130	0.027
Participant 10	1.617	0.074	0.427	0.028	1.255	0.124	0.561	0.014	0.529	0.021
Participant 11	0.998	0.067	0.567	0.039	1.392	0.163	1.101	0.025	1.068	0.032
Participant 12	0.931	0.067	0.584	0.040	1.364	0.167	1.227	0.023	1.195	0.030

Parameter	*a*		*z*		*v*		*t* (spoken)		*t* (written)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 13	2.172	0.103	0.396	0.026	1.268	0.116	0.565	0.017	0.532	0.024
Participant 14	0.892	0.064	0.537	0.040	1.312	0.162	1.154	0.027	1.122	0.034
Participant 15	1.553	0.075	0.393	0.031	1.492	0.150	0.623	0.013	0.590	0.020
Participant 16	1.092	0.073	0.564	0.043	1.501	0.179	1.049	0.029	1.017	0.036
Participant 18	0.816	0.061	0.606	0.041	1.477	0.193	1.065	0.020	1.032	0.027
Participant 19	1.343	0.059	0.410	0.031	1.273	0.137	0.791	0.012	0.758	0.019
Participant 21	1.652	0.076	0.466	0.032	1.204	0.133	0.739	0.016	0.707	0.023
Participant 22	0.857	0.071	0.626	0.047	1.566	0.225	1.214	0.023	1.182	0.030
Participant 23	1.776	0.083	0.454	0.031	1.402	0.124	0.676	0.015	0.643	0.023
Participant 24	1.821	0.087	0.427	0.030	1.416	0.125	0.637	0.017	0.605	0.024
Participant 25	1.343	0.076	0.558	0.037	1.279	0.146	0.991	0.027	0.959	0.034
Participant 26	1.691	0.083	0.465	0.034	1.419	0.137	0.775	0.015	0.743	0.022
Participant 27	1.835	0.085	0.469	0.031	1.238	0.126	0.610	0.016	0.577	0.023
Participant 28	1.718	0.077	0.433	0.031	1.300	0.127	0.615	0.015	0.582	0.022

Parameter	*a*		*z*		*v*		*t* (spoken)		*t* (written)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 29	1.720	0.078	0.388	0.030	1.507	0.145	0.691	0.012	0.658	0.019
Participant 30	1.910	0.090	0.424	0.029	1.327	0.119	0.575	0.014	0.542	0.0212
Participant 31	0.927	0.058	0.547	0.040	1.348	0.164	1.057	0.021	1.025	0.028
Participant 32	2.176	0.105	0.392	0.028	1.289	0.116	0.610	0.017	0.578	0.024
Participant 33	0.991	0.068	0.559	0.042	1.410	0.170	0.962	0.030	0.929	0.037

Note. Intertrial variability of *z* (*sz*), *v* (*sv*), and *t* (*st*) were estimated at group level (means (and standard deviations): *sz* = 0.072 (0.045), *sv* = 0.460 (0.120), *st* = 0.819 (0.012).

6.5 Discussion

The goal of this paper was to validate a novel methodological tool to measure long-term memory consolidation from writing, introduced by Silva and Limongi (2019). Experiment 1 was a direct replication of Silva and Limongi's study. Across eight experimental blocks, participants read a short text passage, wrote or spoke a summary of it, and completed an episodic memory word-recognition task. In this task participants had to decide, as quickly and as accurately as possible, whether target words had appeared in the original text passage. In Experiment 2, participants listened to the short text passages and target words to control for a potential priming effect of repeated orthographic representations in Silva and Limongi's original experiment and Experiment 1. RT distributions were analysed with Bayesian hierarchical drift-diffusion modelling (HDDM).

The results replicated those of Silva and Limongi (2019). Following written summaries, participants RT on the word-recognition memory task were shorter than for spoken summaries. The non-decision parameter of the drift-diffusion model (*t*) captured this difference. This means participants required less time to execute non-decision processes, such as stimulus encoding and/or motor responses, after writing a summary than after speaking a summary. Such a finding implies that writing can support learning at the very early stage of episodic memory retrieval by decreasing encoding and/or the motor response to subsequent stimuli. Importantly, these results remained unchanged in the second experiment which rules out orthographic priming as an alternative explanation. Overall, this provides further support to the use of this measure in the context of writing-to-learn.

The writing-superiority effect found supports previous research which finds an advantage for writing over speaking (e.g., Rivard & Straw, 2000; Rivard, 2004; Klein, Piacente-Cimini & Williams, 2007). One reason for this is because writing provides an external representation of participants' thoughts and ideas. This helps the writer to monitor transcription of the emerging text which is not possible in speaking (Olive & Passerault, 2012). In addition, writing, in its physical form, has a visuospatial dimension (Olive & Passerault, 2012) which can support memory for text. For example, Le Bigot, Passerault and Olive (2008) found recall accuracy of word location was worse when participants performed a visuospatial task at the same time as text production. Current research is being undertaken by Galbraith and colleagues to evaluate the visual feedback explanation, by comparing written summaries to summaries written in invisible ink.

Arnold et al. (2017) claimed that retrieval practice is an underlying cognitive process in the effects of writing on learning. Retrieval practice is one important process that can support long-term memory consolidation of information, as evidenced by the vast literature on the testing effect (e.g., Roediger & Karpicke, 2006b). However, the current study, in addition to Silva and Limongi's (2019) original work, highlights that the superiority effect of writing in closed-book source-based tasks is the result of more than just retrieval practice. When speaking a summary, participants were also required to retrieve content from memory. It is unclear precisely why writing summaries would lead to reduced non-decision processing time on the word-recognition task than speaking summaries.

A possible explanation is that writing summaries, rather than speaking summaries, leads to a more consolidated representation of the target word in long-term memory, and therefore subsequent encoding of the stimuli in the word-recognition task would be faster. After writing about a source, the long-term memory representation of its content may be more activated than after talking about a source. Silva and Limongi (2019) provide a loose explanation of the findings in relation to the knowledge-constituting process of the dual-process model (Galbraith & Baaijen, 2018). They suggest reconstructive processes and revision of text can support long-term memory consolidation by helping writers to resolve uncertainty about how well their summaries capture the meaning of the source texts. This makes knowledge explicit in episodic memory. However, according to Galbraith and Baaijen (2018), the knowledge-constituting process takes place within the semantic memory system. Therefore it is unlikely that the word-recognition task reflects knowledge-constituting. It is more likely that knowledge-transforming is responsible for the writing-superiority effect, but exactly how this happens remains an open question. Future research should explore how this RT measure relates to other measures of writing-to-learn (e.g., text comprehension and subjective knowledge) to assess its validity and investigate possible explanations of the writing-superiority effect.

The successful replication of Silva and Limongi's (2019) has important methodological implications for writing-to-learn research. These are the first studies to use a RT measure to assess learning from writing. This is a first step in linking writing-to-learn with cognitive neuroscience to discover the underlying cognitive processes of writing at the neural level. There are no studies which use neuroimaging techniques, such as fMRI, to explore writing-to-learn. This is because of the practical difficulties of having participants write extended texts in an fMRI scanner. However, the simple word-recognition task would be practically easier to implement in a neuroimaging study and therefore could be used to provide insights about brain activity during writing-to-learn.

6.5.1 Limitations

The experiments presented here have highlighted that writing summaries can enhance long-term memory consolidation. This is based on a measurement of the speed of episodic memory retrieval, suggesting that writing helps individuals to remember content from source text passages. Although this is one aspect of learning, it does not consider the individual's subjective interpretation of their understanding or knowledge, or their ability to make inferences and apply their knowledge in new contexts. Future research should assess how participants' RT on word-recognition tasks relates to ratings of their subjective knowledge and performance on inference tests.

The writing task in these experiments may also be somewhat atypical of writing tasks in educational settings. As acknowledged by Silva and Limongi (2019), source-based writing can also take place in an open-book setting where writers are able to access the source materials whilst they are writing. In an open-book setting, the reliance on retrieval of information from memory is removed (Hidi & Anderson, 1986; Arnold et al., 2017). Since this measure captures episodic memory retrieval, it may not be an appropriate method of measuring learning in open-book settings. A future study could explore differences in RT between participants who construct open- versus closed-book summaries and then undertake the word-recognition memory task in a closed-book environment.

Participants were asked to summarise short text passages of around 200 words. From a brief review of participants' summaries it appeared that participants often attempted to recall the texts rather than summarise them. One reason this may have occurred is because the task instructions needed to more fully describe the meaning of a summary (participants were told to "write or speak a summary of the main idea of the text", and that they must produce at least two sentences). Yu (2013) highlighted the importance of instruction in summary tasks. It also shows that summarising can be a difficult task (Hidi & Anderson, 1986), even for advanced writers like undergraduate students. This observation suggests the process of constructing summaries is similar to knowledge-telling (Bereiter & Scardamalia, 1987), i.e., the translation of retrieved ideas into written text. Future research could explore the extent to which participants who produced better quality summaries (by definition, a "condensation of the topic's main points" (Spirgel & Delaney, 2016, p. 171) influences RT in comparison to those who merely recall the original source.

This replication demonstrates the reliability and validity of using a word-recognition task in a relatively simple experiment on writing versus speaking. However, writing-to-learn studies are often more complex, comprising of multiple source texts, multiple writing activities, and/or multiple other outcome measures. Participants are also often asked to produce much longer

texts than summaries. Exploration of the most effective way to incorporate the measure into different types of writing-to-learn studies is important.

Another limitation relates to the sample of participants. In both experiments, participants were higher-education students (predominantly undergraduates) at the same university in the UK (Silva and Limongi's (2019) sample was also comprised of undergraduate students). This is appropriate in the context of exploring the learning processes of academic writing, but the results may not represent learning from writing in non-student writers. Furthermore, the majority of participants were female and studied psychology. Psychology students are typically more familiar with experimental tasks like the word-recognition task due to the frequency with which they participate in psychological research as a result of research participation schemes. The reliability of this methodological tool should be assessed in experiments with writers from different backgrounds, including gender, domain and education level, as well as with children and adolescents.

6.6 Conclusion

In conclusion, the current paper provides an important methodological contribution to the writing-to-learn field by replicating Silva and Limongi's (2019) results and ruling out a potential confounding factor (orthographic priming). The measure was shown to be reliable in assessing the effects of writing summaries on long-term memory consolidation, in comparison to speaking summaries. This affirms the use of the measure in future writing-to-learn studies, although further work is needed to explore how best to implement the measure in more complex writing tasks. It is also important to establish how this novel measure relates to more widely used measures of learning such as text comprehension (Arnold et al., 2017) and subjective understanding (Baaijen & Galbraith, 2018). In comparison to speaking, writing is associated with faster non-decision processes (e.g., encoding and/or motor processes) in the early stages of retrieval. Possible explanations of this effect in relation to the dual-process model of writing (Galbraith & Baaijen, 2018) were outlined, but the precise reasoning of the writing-superiority effect remains an open question.

Chapter 7 Paper 4. Beyond Retrieval: The Cognitive Processes of Synthesis Writing

7.1 Abstract

The aim of this paper was to investigate the cognitive processes underlying writing-to-learn in the context of closed-book, multiple source-based writing. Participants (40 higher-education students) read two source text passages before writing either a synthesis essay or writing down everything they could remember from the two texts (free recall). Keystrokes were logged in Inputlog whilst participants constructed their texts to derive writing process measures (global linearity and sentence production; Baaijen & Galbraith, 2018). After writing, participants completed a word-recognition memory task in which they had to decide whether target words appeared in the original text passages, as well as a text comprehension test comprised of multiple-choice and short-answer questions. Participants also rated their subjective knowledge of the topic before reading, after reading (pre-writing), and after writing (post-writing). Results showed that following free recall (knowledge-telling), participants subjective knowledge decreased from pre- to post-writing, consistent with an illusion of explanatory depth (Rozenblit & Keil, 2002). This decrease was not observed for participants who wrote a synthesis, suggesting these participants engaged in a knowledge-transforming process. Subjective knowledge ratings were significantly positively correlated with short-answer question scores. Although the global linearity and sentence production composite measures were reproduced, these did not differentiate between conditions or predict development of knowledge, suggesting such measures may not generalise to source-based writing. Finally, reaction time data showed no condition effects, nor did it correlate with the other learning measures (text comprehension and subjective knowledge), likely due to design limitations.

7.2 Introduction

Synthesis writing can be defined as "a type of source-based writing that requires writers to synthesise the information from different sources into a new and meaningful text" (Vandermeulen, van Steendam & Rijlaarsdam, 2022, p. 747). As a generative learning task, effective synthesis of two or more source text passages requires both elaboration and organisation processes (Roelle et al., 2023; Nelson & King, 2022). Writers are said to use elaboration processes when they connect the material to be learned with their pre-existing knowledge, and organisation processes to arrange the material into a new coherent structure (Arnold et al., 2017). Synthesis writing is therefore a cognitively demanding activity (Spivey & King, 1989; Vandermeulen et al., 2020) which has the capacity for learning (i.e., it is a writing-to-learn task). For example, in Solé et al.'s (2013) study, adolescents who wrote better quality synthesis texts by engaging in elaboration processes had improved comprehension of source material in history in comparison to poorer quality synthesis texts associated with a more linear writing process.

Previous research has found that writing a synthesis can lead to better performance on comprehension tests of source material than writing summaries, narratives, or explanations (Wiley & Voss, 1999). When producing a summary, the writing process can be comprised of retrieving information from memory and translating it into written text (Mateos et al., 2014), a process known as knowledge-telling (Bereiter & Scardamalia, 1987). In Papers 1 and 2 of this thesis, writing either an essay or writing down everything one could remember from a single source reflected a knowledge-telling process involving retrieval practice, and as such participants subjective knowledge decreased from pre- to post-writing. Prompting writers to produce a synthesis can allow them to select, organise, and connect thoughts, ideas and information (Castells, Minguela & Nadal, 2022; Kiewra, 2019) and move away from knowledge-telling and into knowledge-transforming.

Several authors have highlighted the value of synthesis writing as a learning activity due to the opportunity for knowledge-transforming (Mateos et al., 2014; Vandermeulen, van Steendam & Rijlaarsdam, 2022; Spivey & King, 1989). Knowledge-transforming is a writing process first described by Bereiter and Scardamalia (1987) as a goal-directed, problem-solving process which encourages the writer to make connections and engage in reflective thought. The dual-process model of writing (Galbraith & Baaijen, 2018) conceptualises a knowledge-transforming process as the retrieval of information stored in episodic memory which is then evaluated towards the writers' rhetorical goals for the emerging text. In this model, since it is responsible for retrieving content rather than generating new knowledge, the process is not associated with knowledge discovery. In the dual-process model, knowledge discovery is associated with a

knowledge-constituting process in which the writer's disposition guides the active constitution of information represented in semantic memory.

The dual-process model (Galbraith & Baaijen, 2018) has empirical support from a study in which writers constructed texts based on their prior knowledge in the absence of source material (Baaijen & Galbraith, 2018). The claims of the dual-process model in the context of source-based writing were explored in Papers 1 and 2. However since writers were constructing text based on a single source, the process was reduced to knowledge-telling and did not reflect knowledge-transforming or knowledge-constituting. In the current paper, free recall writing (which essentially prompts writers to knowledge-tell) is compared to synthesis writing which is hypothesised to elicit knowledge-transforming.

Furthermore, studies on synthesis writing have typically been conducted in an open-book environment where participants can access the source material whilst writing their texts (e.g., Vandermeulen et al., 2020; Castells, Minguela & Nadal, 2022). Therefore synthesis writing is often a recursive process of reading and writing (Spivey, 1990; Nelson & King, 2022). In an open-book synthesis task, writers do not need to retain information in memory. In comparison, closed-book source-based writing tasks require the learner to remember information they have read. For this reason, Arnold et al. (2017) highlighted retrieval processes as important in closed-book writing from sources. Few, if any, studies have explored synthesis writing in a closed-book environment, even though such environments are common in educational practice.

Research on source-based synthesis writing has also focused on response accuracy measures of learning such as text comprehension tests based on the source material. Whilst these are useful for testing students' knowledge of an explicit reference point (the source text), they do not measure how well students' themselves feel they know or understand the content. Students' subjective development of knowledge is also an important measure for assessing the claims of the dual-process model. Therefore the current study considers text comprehension performance as well as participants' perceptions of their knowledge of the source materials. Finally, this study also adopts a reaction time (RT) task (Silva & Limongi, 2019) to measure learning to explore how such tasks can be incorporated into writing-to-learn studies to discover the underlying cognitive processes at the neural level.

In summary, this study investigates the cognitive processes underlying closed-book synthesis writing, in comparison to free recall writing, under the framework of the dual-process model. It considers the learning outcomes of these writing tasks on measures of text comprehension and subjective knowledge, as well as RT on a word-recognition task.

7.2.1 Measuring the Writing Process: Global Linearity and Sentence Production

In the current study, participants keystrokes were logged whilst they wrote either a synthesis essay or a free recall piece. Keystroke logging has been used as an alternative method to think-aloud in several recent studies exploring the writing process due to it being an unobtrusive and more detailed measure of text production processes (e.g., Baaijen & Galbraith, 2018). Although it is becoming a commonly used method in writing research, problems remain in linking keystroke data to theoretical components of the writing process (Baaijen, Galbraith & de Glopper, 2012). Reproducibility issues in keystroke logging analysis also exist (Hall, 2023; Hall, Baaijen & Galbraith, 2024).

The three main keystroke features often used in writing research to map keystroke logs to the writing process are pauses, revisions, and bursts. Put simply, pauses are moments of empty time where the writer stops composing text. Pauses can represent many processes, such as re-reading and reflecting on text previously written, or planning the next unit of text or revision (Baaijen, Galbraith & de Glopper, 2012). Revisions are modifications during text production.

Bursts are rapid and uninterrupted moments of text production (Baaijen & Galbraith, 2018), and the type of burst is identified by the way in which it is terminated. P-Bursts are terminated by a pause lasting two seconds or longer; and R-Bursts are terminated by revision (Kaufer, Hayes & Flower, 1986; Baaijen & Galbraith, 2018). I-Bursts are bursts inserted away from the leading edge of the text (Baaijen, Galbraith & de Glopper, 2012; Baaijen & Galbraith, 2018), where the leading edge is “the position immediately after the last written character in the text produced so far” (Lindgren et al., 2019, p. 347).

Baaijen and Galbraith (2018) suggest a high percentage of P-Bursts indicates cleanly produced text; a high percentage of R-Bursts indicates more revision during text production; and the percentage of I-Bursts indicate “the relative amount of text produced away from the leading edge compared to the amount produced during forward text production” (Baaijen & Galbraith, 2018, p. 10).

The value of analysing pauses, revisions, and bursts in combination has been emphasised by Baaijen, Galbraith and de Glopper (2012), as opposed to these keystroke features being viewed in isolation when relating them to cognitive processes. An example of this is Baaijen and Galbraith’s (2018) distinction between linear transitions and event transitions; where linear transitions are “uninterrupted transitions to the next unit of text” and event transitions are “other operations before the production of the next unit” (Baaijen & Galbraith, 2018, p. 9). This approach facilitates the differentiation of linear pauses – often associated with planning what is

to be written next – and pauses associated with revision or insertion of text, as opposed to handling all pauses in the same way.

As highlighted by Hall (2023; Hall, Baaijen & Galbraith, 2024), keystroke analysis in writing research lacks a standardised approach, and methods of analysing keystroke data are often not transparent, raising concerns about the reproducibility of such analysis. Therefore, the definitions and rules applied to the keystroke concepts in this paper are clearly set out in Table 12. These concepts underlie the keystroke analysis framework utilised in this study, which was developed by Baaijen & Galbraith (2018) to map keystroke logs to the cognitive processes in the dual-process model of writing (Galbraith & Baaijen, 2018).

Baaijen and Galbraith (2018) identified eleven keystroke measures through analysis of draft essays (newspaper articles) written by higher-education students. Definitions of these measures and the methods used to calculate them in the current study are detailed in Table 12 in the Methodology. Following this, they identified two independent scales (Global Linearity and Sentence Production) using Principal Components Analysis (PCA) of the keystroke measures (Table 11). High reliability was found for both measures (Global Linearity, $\alpha = 0.80$; Sentence Production, $\alpha = 0.79$; Baaijen & Galbraith, 2018). The scales have also been reproduced in another experimental study comparing types of essay drafting strategies (Hall, 2023), except for one of the eleven measures due to differing methods of calculation of the measures by the researchers.

Six keystroke measures were found to load onto the *Global Linearity* scale. Participants who scored highly on this scale were found to write their texts more linearly, whereas those with lower scores had more non-linear text production involving more revision. The remaining five keystroke measures were found to load onto the *Sentence Production* scale. Participants with high Sentence Production typically paused for longer between each sentence and revised less within sentences, suggesting their text production was more controlled. Participants with low Sentence Production wrote more spontaneously; that is, with shorter pause time within and between sentences and revised more within sentences.

Table 11. Measures for global linearity and sentence production.

Global Linearity	Sentence Production
Sentence Linearity Index	Percentage of bursts terminated by revision at the leading edge
Percentage of I-Bursts	Percentage of words produced in P-Bursts
Percentage of time spent on events	Percentage of >2 second pauses between words
Percentage of linear transitions between sentences	Text Modification Index
Number of production cycles	Mean pause duration between sentences
Percentage of linear transitions between words	

Baaijen and Galbraith's (2018) study also demonstrated how the global linearity and sentence production scales were related to both text quality and participants' subjective knowledge. Subsequently, the scales were useful for interpreting how keystroke measures could relate to the cognitive processes of the dual-process model (knowledge-transforming and knowledge-constituting; Galbraith & Baaijen, 2018). Low and high self-monitors produced either an outline or synthetic plan before writing newspaper article. They also rated their subjective knowledge of the topic before and after writing.

Results of Baaijen and Galbraith's (2018) study showed a significant negative relationship between global linearity and change in knowledge, and the relationship between sentence production and change in knowledge was dependent on the type of planning. When participants synthetically planned their texts, spontaneous sentence production was associated with knowledge development. When participants outline planned their texts, developments of knowledge were more limited. Regarding text quality, global linearity had a negative relationship with better quality text in the outline planning condition; and sentence production was also associated with poorer quality text. These findings support the dual-process model since development of knowledge was related to the construction of new ideas during sentence production; as well as the re-organisation of participants' knowledge according to the global structure of their texts.

The current study uses the global linearity and sentence production scales to explore the writing process of participants constructing either a synthesis essay or a free recall text based on two

source texts. In Papers 1 and 2 of this thesis, writing from single sources was found to predominantly engage writers in knowledge-telling as opposed to knowledge-transforming or knowledge-constituting. Instruction to write a synthesis text is designed to encourage writers to compare and contrast multiple sources, and may therefore go beyond knowledge-telling. The global linearity and sentence production measures are therefore used to assess whether synthesis writing, in comparison to free recall, reflects the processes claimed by the dual-process model (Galbraith & Baaijen, 2018).

7.2.2 Aims and Research Questions

The current study aimed to explore and compare the writing process and learning effects of synthesis essay writing versus free recall, as well as explore how different measures of learning from writing are related to one another. Briefly, participants read two source text passages, wrote either a synthesis essay or a free recall, completed a word-recognition test assessing memory for words in the source texts, and finally completed a text comprehension test (multiple-choice (MCQ) and short-answer questions (SAQ)) about the texts. Participants also rated their subjective knowledge of the topic before the experimental session, after reading the source texts and after writing. The aims of the study are addressed by three research questions.

First, *what effects do the writing conditions (synthesis vs. free recall) have on subjective knowledge, text comprehension, and word-recognition RT?* In Papers 1 and 2 of this thesis, it was found that participants' subjective knowledge decreased from pre- to post-writing, as engaging in knowledge-telling and retrieval practice helped writers to identify gaps in their knowledge (an 'illusion of explanatory depth'; Rozenblit & Keil, 2002). In the current study, writers in the free recall condition are told to write down everything they can remember from two source texts, i.e., they are told to knowledge-tell. Thus it can be predicted that these participants will experience a decrease in their knowledge from pre- to post-writing. In comparison, writers in the synthesis condition are prompted to write an essay comparing and contrasting the two source texts. This may require the writer to create novel content during sentence production, and reorganise their thoughts and ideas in line with the global structure (e.g., engage in knowledge-transforming and/or knowledge-constituting processes). Therefore, these writers may not experience a decrease in their subjective knowledge from pre- to post-writing, in line with the claims of the dual-process model.

In terms of text comprehension, since both writing conditions require retrieval practice, they may perform equally well on the text comprehension test. Previous studies by Arnold et al. (2017) and in Papers 1 and 2 of this thesis generally found no differences overall between different writing conditions, when writing from single sources. Alternatively, it is possible that

those in the free recall condition will have better performance on the text comprehension questions (in particular the MCQs) as the task explicitly requires recalling content from the source texts in comparison to the synthesis task.

This study also incorporates a word-recognition RT measure developed by Silva and Limongi (2019) which was shown to be reliable in Paper 3. This measure is designed to assess long-term memory consolidation of source text material, by measuring participants' RT to deciding whether target words appeared in one of the original source texts. The measure was developed in studies comparing writing and speaking, which found a writing-superiority effect explained by faster encoding or motor processing of the target words following writing summaries than speaking summaries. It is an open question as to whether this measure will capture differences between the synthesis and free recall writing conditions. This is the first study to incorporate the measure into a more naturalistic and complex writing task.

Second, *how are subjective knowledge, text comprehension, and word-recognition RT related to one another as measures of learning from writing?* Consistently across Papers 1 and 2 of this thesis, subjective knowledge has been significantly positively correlated with SAQ scores on the text comprehension test; in Paper 1 subjective knowledge was also significantly positively correlated with MCQ scores and in Paper 2 it was significantly positively correlated with factual MCQ scores. Therefore, here it is predicted that subjective knowledge ratings will at least be significantly correlated with SAQ scores. This study is the first to consider whether participants' RTs on the word-recognition test are related to their subjective knowledge and/or their text comprehension performance. Since this measure is designed to reflect long-term memory consolidation, it could be predicted that faster RTs would be associated with better text comprehension performance and better subjective knowledge.

Third, *how are the two writing process measures, global linearity and sentence production, related to participants' subjective knowledge development (pre- to post-writing) across the two writing conditions (synthesis vs. free recall)?* As previously explained, the global linearity and sentence production keystroke measures were designed by Baaijen and Galbraith (2018) to relate to the dual-process model of writing (Galbraith & Baaijen, 2018). Since it is hypothesised that the synthesis writing condition will elicit the processes claimed by the dual-process model, in this condition it is predicted that developments of knowledge will be associated with more spontaneous sentence production and increased revision of the global structure. By comparison, in the free recall condition, participants' writing processes will be more controlled and linear, associated with a decrease in their subjective knowledge.

7.3 Method

7.3.1 Design, Materials, and Measures

In short, the procedure of this study was as follows. During one experimental session, participants read two source text passages, wrote either a synthesis essay or free recall about the texts, completed a word-recognition test assessing memory for words in the source texts, and finally completed MCQs and SAQs about the texts. Participants rated their subjective knowledge of the topic before the experimental session (pre-experiment), after reading the source texts (pre-writing), and after writing (post-writing).

This study adopted a mixed experimental design, with one between-subjects variable and one within-subjects variable. The between-subjects variable was an independent variable (IV): writing condition (synthesis or free recall). The within-subjects dependent variable (DV) was subjective knowledge (repeated at three time points, pre-experiment, pre-writing, and post-writing). Additional DVs were scores on a text comprehension test (MCQs and SAQs), and RTs on a word-recognition memory test. Keystroke logging data was also collected for participants' writing in both between-subjects conditions in order to measure participants' writing process on two scales: global linearity and sentence production.

Source texts. The source text passages participants read were on the topic of different modes of therapy for the treatment of neuroses from Schnotz (1984). One of the texts was about Psychoanalysis (Appendix B.7), and the other text was about Behaviour Therapy (Appendix B.10). Both passages outlined the theoretical background of the therapy and its treatments, assumptions of the therapy for neurotic disorders and their associated symptoms, and the scientific position of the therapy. The Psychoanalysis text was 615 words long, and the Behaviour Therapy text was 606 words long. These source texts were selected on the grounds that they are independent text passages, but that they can also be compared and contrasted, thus suitable for both synthesis essay and free recall writing conditions.

MCQs. Eight MCQs (Appendix B.8 and B.11) for each source text were created following guidance in Arnold et al. (2017), designed to assess participants comprehension of the passages. Test questions included four factual (or 'direct') MCQs, and four inferential MCQs for each source text, resulting in sixteen questions in total. Direct MCQs (e.g., 'How long does psychoanalytical treatment usually last?') could be answered using explicit information from the source text, and inferential questions (e.g., 'According to psychoanalysis, which statement best characterises the role of the psychological complex in neurosis?') "required participants to integrate across or extrapolate from" information in the source text (Arnold et al., 2017, p. 118).

For each question, participants selected one answer out of four options, and the number of correct answers were totalled for analysis.

SAQs. Four SAQs (Appendix B.9 and B.12) for each source text were created, also designed to assess text comprehension of the source texts. These questions involved problem-solving based on factual information in the source text (guidance from Arnold et al., 2017), for example 'Based on psychoanalysis theory, explain the distinction between causes and symptoms and their role in maintaining neuroses'. Two judges (researcher and main supervisor) rated participants' responses to the SAQs with a score of three to four (including half marks) depending on the question (see Appendix B.9 and B.12) for a breakdown of the scoring criteria). After rating participants' responses independently, ratings were screened for any mistakes, which were subsequently corrected. Ratings that differed by one or more marks were discussed by the judges. On some occasions, this led judges to change their rating, but in the majority of cases, the judges agreed to calculate an average mark as a fair representation of the participants' answer. Judges scores that differed by half a mark were also averaged. Marks on SAQs were then totalled for analysis, where 15 was the highest mark available for each text (30 marks available in total).

Inter-rater reliability for the SAQ ratings was measured with the Intraclass Correlation Coefficient (ICC). A two-way mixed effects model was used to evaluate the absolute agreement between the two raters. Following the initial ratings there was strong agreement between the two raters for both the Behaviour Therapy (ICC = 0.957) and Psychoanalysis (ICC = 0.916) SAQs. Inter-rater reliability increased following correction of mistakes and discussion between the raters for both the Behaviour Therapy (ICC = 0.976) and Psychoanalysis (ICC = 0.956) SAQs.

Subjective knowledge. Participants rated their subjective knowledge of the source text using the Subjective Knowledge Rating Scale developed by Galbraith et al. (2023) (Appendix A). The scale includes twelve items in which participants rate each on a seven-point scale, where one represents 'very little', and seven represents 'a great deal'. Eight of these items load onto an 'organisation' subscale (e.g., 'How structured your thoughts about the topic are'), designed to measure subjective knowledge represented in explicit episodic memory. Six of the items load onto an 'understanding' subscale (e.g., 'How much you feel you know about the topic'), designed to measure subjective knowledge represented implicitly. These scales have been shown to have high reliability (Galbraith et al., 2023). Participants' ratings were averaged as a whole, and on each subscale, for analysis. Change in subjective knowledge pre- to post-writing was also calculated for each participant by subtracting pre-knowledge average ratings from post-knowledge average ratings.

Word recognition memory test (RT). Participants completed a word recognition memory test in which they had to respond as accurately and as fast as possible as to whether a word appeared in one of the original source texts or not. 60 words in total were presented, 30 of which were in-text words from the original sources (15 from each text), and the remaining 30 of which were non-text words which did not appear in the original sources, but were matched for word frequency with the in-text words (Leech et al., 2001). The full list of in-text and non-text words can be viewed in Appendix D. Participants' RTs were recorded as well as whether they responded correctly (e.g., identified an in-text word was indeed an in-text word) or incorrectly (e.g., identified an in-text word as a non-text word). RTs below 200 ms and above 2000 ms were removed for analysis.

Global Linearity and Sentence Production. Inputlog Version 8.0.0.17 (Leijten & Van Waes, 2013) was used to collect keystroke logs of participants' essay or free recall text production. This software was chosen as it can be run within Microsoft Word and is therefore more naturalistic to participants who are familiar with such word processing environments. Using Inputlog, a 'general' output was produced for each participant and converted to a Microsoft Excel file for analysis. In these files, each row represents an individual key press, and the columns provide data on the cursor position, document length, start and end time of each log, and action and pause times.

A set of Visual Basic Application (VBA) Macros, developed by Hall (2023; available at: https://osf.io/q59cj/?view_only=b57d500880a247fcaba84258d91db603\), were run on the Excel files to identify and code pauses and revisions. This allowed for some analysis of the keystroke logs to be automated, which is important for two key reasons. First, it ensures a standardised and reproducible procedure is used in analysis; and second, it reduces the time-consuming and labour-intensive nature of analysing keystroke logs (Hall, 2023). Although the identification and subsequent analysis of pauses and revisions was automated; bursts were coded manually as there is generally large disparities in their creation amongst writers.

The Macros and hand-coding of bursts facilitated the calculation of the eleven keystroke measures which form the two independent scales, Global Linearity and Sentence Production (Baaijen & Galbraith, 2018). The definitions and method of calculation for each of these measures are outlined in Table 12. The written production of titles and planning in participants' keystroke logs were not included in the calculation of the keystroke measures.

Table 12. Writing process measures (global linearity and sentence production).

Measure	Definition	Method of Calculation
Global Linearity		
Sentence Linearity Index	Proportion of sentences in the final product that were produced in the same order as in the keystroke log. Sentence was determined to be produced out of order when the participant moved away from leading edge to insert a new complete sentence or to add text or edit an already existing sentence.	Calculated manually by counting each time the participant moved away from the leading edge to insert a new complete sentence or add text or edit an already existing sentence.
Percentage of I-Bursts	Number of bursts inserted within already written text, expressed as a percentage of the total number of bursts.	The number of I-Bursts divided by the total number of R-Bursts, P-Bursts, and I-Bursts. I-Bursts were identified as periods of uninterrupted text production inserted away from the leading edge. An I-Burst was terminated when either a revision or a pause lasting two seconds or longer was executed.
Percentage of Time Spent on Events	Percentage of the total time spent writing devoted to operations other than producing text or planning the next unit of text.	Calculated with a Macro as the total time on events divided by the total time spent writing.

Measure	Definition	Method of Calculation
Percentage of Linear Transitions Between Sentences	Number of linear transitions at sentence boundaries as a proportion of the total number of sentence transitions.	Calculated with a Macro as the number of identified linear between sentence pauses, divided by the total number of sentence transitions (as identified by a genuine full stop, i.e., not full stops that were typos). When a participant did not use full stops at the end of a sentence, linear between sentence transitions were identified manually.
Number of Production Cycles	A sequence of language bursts produced without interruption, with each break away from the leading edge defined as the start of a new cycle.	Manual count of the number of times participant moves away from the leading edge.
Percentage of Linear Transitions Between Words	Number of linear transitions between words as a proportion of the total number of word transitions.	Calculated with a Macro as the number of identified linear between-word pauses, divided by the total number of 'Before Word' records.
Sentence Production		
Percentage of R-Bursts	Number of bursts revised at the leading edge of the text, expressed as a percentage of the total number of bursts.	The number of R-Bursts, divided by the total number of R-Bursts, P-Bursts, and I-Bursts. R-Bursts were identified as periods of uninterrupted text production terminated by a major revision (higher-level, more substantive edits to text that are likely to change

Measure	Definition	Method of Calculation
		meaning or structure). In instances where revision was followed by a >2 second pause; the preceding burst was coded as an R-Burst as the revision came before the pause.
Percentage of Words Produced in P-Bursts	Total number of words produced in P-Bursts, expressed as a percentage of the total number of words in the keystroke log.	The number of words in a P-Burst divided by the total number of words within the keystroke log (counted by totalling the number of 'Before Word' records which occur during P-bursts, divided by the total number of 'Before Word' records; calculated by a Macro). P-Bursts were identified as periods of uninterrupted text production terminated by a pause lasting two seconds or longer. In instances where a >2 second pause was followed immediately by revision; the preceding burst was coded as a P-Burst as the pause came before the revision.
Percentage of ≥ Two Second Pauses Between Words	Number of extended pauses occurring between linearly produced words, calculated as a proportion of the total number of linear transitions between words.	Calculated with a Macro as the number of identified linear between-word pauses that last for two seconds or longer, divided by the total number of all identified linear between-word pauses in a keystroke log.

Measure	Definition	Method of Calculation
Text Modification Index	Total number of characters in the final text (including spaces) and the total number of non-character keys, divided by the total number of characters produced during the writing process. The lower the number, the bigger the difference between the writing process and the final product.	Calculated by dividing the total number of characters, spaces, and returns within the keystroke log by the total number of characters, spaces, and returns in the final product.
Mean Pause Duration Between Sentences	Mean of all pause durations between sentences.	Calculated with a Macro as the mean of all identified linear between sentence pauses in the keystroke log. Where the macro failed to detect a linear between sentence transition, or when participant did not use full stops at the end of a sentence, linear between sentence pauses were calculated manually by calculating the difference between the start time of the last character in the previous sentence and the start time of the first character in the new sentence.

7.3.2 Participants

40 higher-education students participated in the study, and each participant received a £15 gift voucher as a token of appreciation for their participation. Participants had a mean age of 21.3. 36 participants were female, and 4 participants were male. The majority of participants spoke English as their first language (n = 36), and the 4 participants of whom English was not their first language were self-declared fluent English speakers. Most participants were undergraduate students (n = 34), and remaining participants were doctoral (n = 5) or master's (n = 1) students. Participants studied a range of subjects, mostly from the Social Sciences (n = 28), but participants also studied Humanities (n = 3), Physical Sciences (n = 3), Medicine and Nursing (n = 3), and Engineering (n = 3) (see Appendix I.1 for the full breakdown of subjects studied).

7.3.3 Procedure

Ethical approval was obtained from the University of Southampton's ethics committee (Appendix F.5). Participants were recruited through volunteer sampling using an advert on social media and posters around the university campus. Individuals who expressed interest in participating were provided with the Participant Information Sheet and Consent Form by email. If they decided to participate, the individual provided consent and booked an in-person study participation slot.

Before attending their slot, participants completed a pre-experiment online questionnaire involving demographic questions (gender, age, whether English was their first language, their education level, and subject of study), and the Subjective Knowledge Rating Scale (Galbraith et al., 2023; Appendix A) as a measure of *prior knowledge* on the broad topic of different modes of therapy for the treatment of neuroses. Participants also created a unique identifier number so their data from the pre-experiment online questionnaire could be linked to their data from the in-person experimental session.

The in-person experimental session took place in a private room on the university campus, and the full session took approximately an hour to 1.5 hours. Participants completed the study on a computer. First, they had 10 minutes to read each source text passage (20 minutes in total). The texts were presented in a random order, and participants could not progress until the time had elapsed. Next, participants completed the Subjective Knowledge Rating Scale as a measure of *pre-writing* subjective knowledge. Participants then completed a short summarising task where they were instructed that they had five minutes to 'Summarise the main idea of the texts you have read in one sentence'. The reading task, Subjective Knowledge Rating Scale, and summarising task were completed on Qualtrics.

After the summarising task, participants were randomly allocated to one of two writing conditions: synthesis essay (experimental condition) or free recall (control condition). In the synthesis essay condition, participants were prompted to, 'Write a draft essay comparing and contrasting the two different modes of therapy for neuroses (psychoanalysis and behaviour therapy) and conclude with an overall evaluation'. In the free recall writing condition, participants were prompted to, 'Recall and write down, as close to word for word as possible, everything you can remember from the text passages you read'. In both conditions, participants had 15 minutes to write (type) on a Microsoft Word document. During this task, participants keystrokes were logged in Inputlog.

Following the writing task, participants completed the Subjective Knowledge Rating Scale for the third time (post-writing) on Qualtrics (same prompt as previous).

Participants then completed a word-recognition memory task on PsychoPy. After receiving the task instructions, participants viewed a short fixation cross, after which 60 words appeared one at a time on the screen. Participants had to respond as to whether the word appeared in one of the original source texts or not by pressing either the 'K' or 'L' key on their keyboard. 30 of the words had appeared in one of the original texts (15 from each text), and the other 30 words did not appear in the original text. The 30 non-text words were matched in word frequency to the in-text words. Participants' RTs were recorded for each of the 60 words.

Finally, after completing the memory task, participants completed the text comprehension test which was comprised of sixteen MCQs (eight from each source) and eight SAQs (four from each source). Participants were then thanked for their participation and emailed a £15 voucher.

7.4 Results

7.4.1 Effects of Writing Conditions on Participants' Subjective Knowledge, Text Comprehension, and Word-Recognition Reaction Time

7.4.1.1 Analytic Approach

To explore the effects of writing activities (synthesis essay writing and free recall) on subjective knowledge and text comprehension, data was analysed using the statistical package IBM SPSS (Version 29.0.2.0 for Macintosh). All data were initially screened using Explore to assess compliance with normal distribution. With the exception of pre-reading knowledge ratings, data were reasonably normally distributed. As expected, pre-reading knowledge ratings were skewed to the right. No adjustment was made for this, given the purpose of this measure was for a baseline check (it was expected that most participants would say they have limited knowledge of the topics at the beginning of the experiment).

To analyse the effects of writing on subjective knowledge, a repeated measures ANOVA was conducted with follow-up t-tests. To analyse the effects of writing on text comprehension, four mixed ANOVAs (MCQs overall, direct MCQs, inferential MCQs, and SAQs) were conducted with writing condition (synthesis and free recall) as a between-subjects factor and source text topic (behaviour therapy and psychoanalysis) as a within-subjects factor.

Finally, to investigate the effects of writing on word-recognition RT, data was analysed using Bayesian Hierarchical Drift Diffusion Modelling (HDDM) in Python (Version 2.7.18) using the HDDM library (Wiecki et al., 2022). The group RT distribution reasonably followed an ex-Gaussian distribution and therefore was suitable for this analysis. As pointed out by Wiecki, Sofer and Frank (2013), the HDDM is also suitable, and can be useful, when there are relatively few trials across conditions and participants. Drift-diffusion modelling is a well-established method of analysing RT data.

The drift-diffusion model considers response time and response accuracy simultaneously to explain cognitive processes behind the distributions (Ratcliff, 1978; Ratcliff, Gomez & McKoon, 2004). It is based on the evidence accumulation process in two-choice forced decision tasks, whereby noisy information representative of one decision is accumulated until a threshold is reached and the decision is made (Ratcliff, Gomez & McKoon, 2004; Wiecki, Sofer & Frrank, 2013). The HDDM includes four main parameters ((1) the starting point of the process, z; (2) the threshold at which the decision is reached, a; (3) the drift rate, v, which is the mean rate of approach to the threshold; and (4) non-decision processes, t (see Figure 14 in Paper 3).

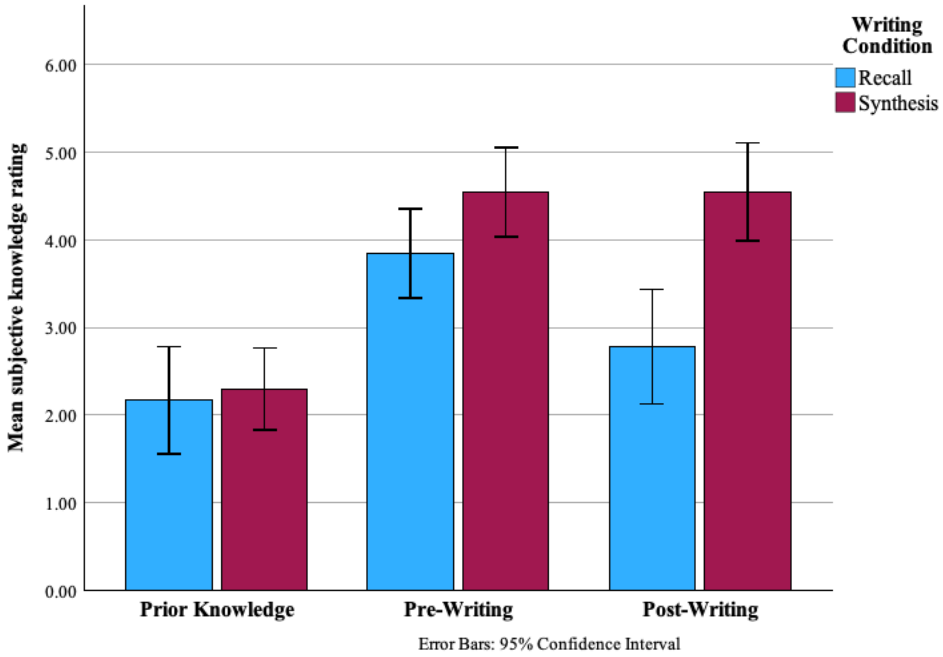
8 Bayesian HDDMs were fitted to the RT and accuracy distributions exploring all combinations of 'a', 'v' and 't' to estimate these parameters for the two conditions (synthesis and free recall), as well as the simplest model where the parameters do not vary by condition. The 'z' parameter was not included in the models because there was no substantial variation across participants, and therefore this avoids overfitting the models. The chain length of each model was 2000 and the number of burn-in iterations was 20 (as recommended by Wiecki, Sofer and Frank (2013)). Chains were generated with no thinning. The deviance information criterion (DIC) was used to select the winning model and model convergence was assessed by calculating the Gelman-Rubin statistic ($\hat{R}$) (for the winning model the chain length of each model was 5000 and the number of burn-in iterations was 20 to allow the model to converge and generate $\hat{R}$). Parameter estimates of the winning model are reported at group and participant level in Appendix I.5.

7.4.1.2 Subjective Knowledge

Descriptive statistics (means and standard deviations) for participants' subjective knowledge scores (prior knowledge, pre-writing, and post-writing) across conditions (synthesis essay and free recall), and by subscale (organisation and understanding) are available in Table 25 in Appendix I.2.

As demonstrated in Figure 17, participants' subjective knowledge ratings increased after reading the source text in both conditions (as would be expected). However, after writing about the sources, subjective knowledge ratings for participants in the free recall condition decreased, with minimal to no change in ratings for those in the synthesis essay condition. Subjective knowledge ratings on both the organisation and understanding subscales followed similar trends.

Figure 17. Mean subjective knowledge ratings at each time point (prior knowledge, pre-writing, and post-writing) by writing condition.



To assess the significance of the effects of writing on participants' subjective knowledge ratings, a repeated measures ANOVA was conducted. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2_{(2)} = 12.14$, $p = 0.002$), therefore Greenhouse-Geisser corrected tests are reported here ($\varepsilon = 0.78$). There was a significant main effect of subjective knowledge ratings ($F_{(1.56, 59.39)} = 49.57$, $p < 0.001$, $\eta_p^2 = 0.57$), as well as a significant interaction between subjective knowledge ratings and writing condition ($F_{(1.56,59.39)} = 8.34$, $p = 0.002$, $\eta_p^2 = 0.18$).

Follow-up paired sample t-tests indicated that overall, there was a significant *increase* in subjective knowledge ratings from prior knowledge to pre-writing ($t_{(39)} = -9.74$, $p < 0.001$, $d = -1.73$), and from prior knowledge to post-writing ($t_{(39)} = -5.16$, $p < 0.001$, $d = -1.05$), but a significant *decrease* from pre-writing to post-writing ($t_{(39)} = 3.04$, $p = 0.002$, $d = 0.39$).

Synthesis condition. In the synthesis essay condition, paired sample t-tests revealed that there was a significant increase in subjective knowledge ratings from prior knowledge to pre-writing ($t_{(19)} = -7.24$, $p < 0.001$, $d = -2.15$), and from prior knowledge to post-writing ($t_{(19)} = -6.04$, $p < 0.001$, $d = -2.04$). There was no significant difference between pre-writing and post-writing knowledge ratings in the synthesis essay condition ($t_{(19)} = -0.03$, $p = 0.489$, $d = 0$).

Free recall condition. In the free recall writing condition, paired sample t-tests revealed that there was a significant increase in subjective knowledge ratings from prior knowledge to pre-writing ($t_{(19)} = -6.75$, $p < 0.001$, $d = -1.39$), and from prior knowledge to post-writing ($t_{(19)} = -1.87$, $p = 0.038$, $d = -0.45$). In comparison to the synthesis essay condition, where there were no

differences in subjective knowledge ratings from pre-writing to post-writing, participants subjective knowledge ratings in the free recall condition significantly decreased from pre-writing to post-writing ($t_{(19)} = 3.98$, $p < 0.001$, $d = 0.85$).

An independent samples t-test also revealed that there was a significant difference in post-writing subjective knowledge ratings for the synthesis essay writing versus free recall condition ($t_{(38)} = -4.29$, $p < 0.001$, $d = 1.36$).

7.4.1.3 Text Comprehension

Descriptive statistics for participants' scores (as percentage correct) on the MCQ and SAQ text comprehension test across conditions and source texts are displayed in Table 26 in Appendix I.3.

MCQs. A mixed ANOVA revealed a significant main effect of source text topic on participants' total score on the MCQs ($F_{(1, 38)} = 4.678$, $p = 0.037$, $\eta_p^2 = 0.11$), indicating that participants performed better on the Psychoanalysis text comprehension test questions than the Behaviour Therapy questions. However, there was no significant interaction between the source text topic and writing condition ($F_{(1, 38)} = 0.058$, $p = 0.811$, $\eta_p^2 = 0.002$). There was also no significant effect of writing condition ($F_{(1, 38)} = 2.750$, $p = 0.105$, $\eta_p^2 = 0.067$), indicating participants in the synthesis and free recall conditions had similar performance overall on the multiple-choice text comprehension questions.

On the direct MCQs, there was no significant effect of source text topic ($F_{(1, 38)} = 3.128$, $p = 0.085$, $\eta_p^2 = 0.076$) nor a significant interaction between source text topic and writing condition ($F_{(1, 38)} = 1.267$, $p = 0.267$, $\eta_p^2 = 0.032$). Participants also performed similarly on the direct MCQs across writing conditions, as there was no significant effect of writing condition ($F_{(1, 38)} = 0.532$, $p = 0.470$, $\eta_p^2 = 0.014$).

Finally, on inferential MCQs, there was no significant effect of source text topic ($F_{(1, 38)} = 0.987$, $p = 0.327$, $\eta_p^2 = 0.025$) nor a significant interaction between source text topic and writing condition ($F_{(1, 38)} = 0.504$, $p = 0.482$, $\eta_p^2 = 0.013$). However, there was a significant effect of writing condition ($F_{(1, 38)} = 4.967$, $p = 0.032$, $\eta_p^2 = 0.116$), indicating that participants in the free recall writing condition received higher scores on the inferential MCQs than participants in the synthesis writing condition.

Short Answer Questions (SAQs). On the SAQs, a mixed ANOVA revealed a significant main effect of source text topic ($F_{(1, 38)} = 16.397$, $p < 0.001$, $\eta_p^2 = 0.301$), indicating that participants had better performance on the Psychoanalysis SAQs than the Behaviour Therapy SAQs. There was no significant interaction between source text topic and writing condition ($F_{(1, 38)} = 1.293$, $p =$

0.263, $\eta_p^2 = 0.033$), as well as no significant effect of writing condition ($F_{(1, 38)} = 0.967$, $p = 0.332$, $\eta_p^2 = 0.025$) suggesting participants in both the synthesis and free recall writing conditions performed similarly on the SAQs.

7.4.1.4 Word-Recognition Reaction Time

Descriptive statistics of RTs in the synthesis and free recall conditions for correct and incorrect responses are shown in Table 27 in Appendix I.4.

Table 13 presents the 15 HDDMs that were fitted to the data. Using DIC for model comparison, the winning model was Model 1 in which non-decision time (t) was a free parameter (i.e., allowed to vary by condition). Overall model convergence of Model 1 was good with all $\hat{R}$ values below 1.1.

Table 13. Model comparison: Results of HDDMs.

Model	Parameters		DIC
	Free	Fixed	
1	t	a, v	**3132.108**
2	v	a, t	3132.206
3	a	v, t	3135.948
4	t, v	a	3133.448
5	t, a	v	3132.178
6	v, a	t	3135.337
7	t, v, a	-	3133.651
8	-	t, v, a	3132.184

For the winning model, Table 28 in Appendix I.5 presents the parameter estimates at both group and subject levels. The results show no strong evidence to support a meaningful difference between the two conditions; this is also demonstrated in Figure 18 where the posterior distributions of 't' estimates for the writing conditions overlap. This implies that the non-decision time is similar across conditions. Posterior predictive checks indicated the model fairly reproduced the observed data (Table 14).

Figure 18. Posterior distributions of *t* estimates.

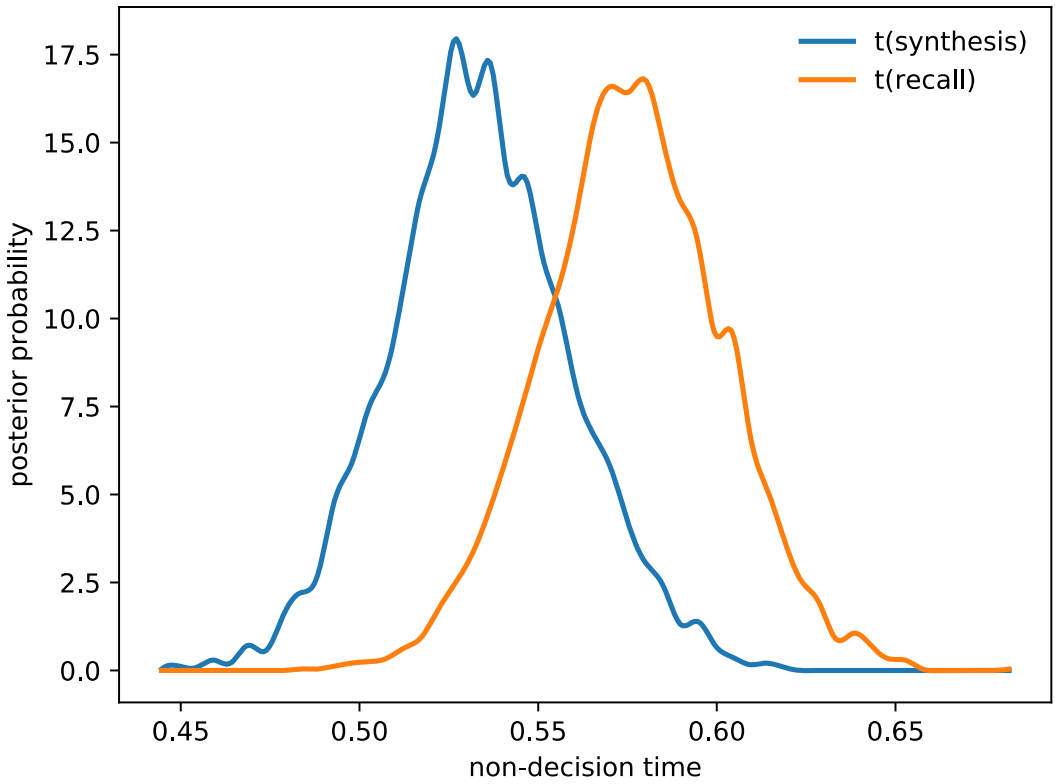


Table 14. Posterior predictive checks.

	Observed	Model Generated	SD	SEM	MSE	Credible
Accuracy	0.761	0.770	0.076	0.0001	0.006	True
RT (Upper Bound)	1.067	1.121	0.145	0.0028	0.024	True
RT (Lower Bound)	-1.208	-1.120	0.191	0.0077	0.044	True

Note. Upper Bound = Correct Responses; Lower Bound = Incorrect Responses. Credible (True) = in the 95% credible interval. Observed and model-generated values are collapsed across conditions.

7.4.2 Relationships Between Subjective Knowledge, Text Comprehension, and Word-Recognition Reaction Time

7.4.2.1 Analytic Approach

To explore the relationship between participants' subjective knowledge, their text comprehension, and their word-recognition RT, Pearson bivariate correlations were calculated between the following variables:

1. Subjective knowledge ratings post-writing
2. Total MCQ scores
3. Direct MCQ scores
4. Inferential MCQ scores
5. SAQ scores
6. *t* parameter estimates from the winning HDDM model

7.4.2.2 Correlations

Pearson bivariate correlations demonstrating the relationships between the measures listed above are reported in Table 15. As would be expected, there were highly significant strong positive correlations between the text comprehension measures (MCQ total score, MCQ direct score, MCQ inferential score, and SAQ score). There was also a highly significant moderate positive correlation between participants' subjective knowledge ratings and their score on the SAQs, indicating that participants' who had higher subjective knowledge ratings post-writing score better on the SAQ questions. There were no significant relationships between the *t* parameter estimates from the HDDM and participants' text comprehension scores or their subjective knowledge ratings.

Table 15. Pearson bivariate correlations for subjective knowledge ratings post-writing, text comprehension scores (MCQs and SAQs), and word-recognition RT (as represented by the *t* parameter of the HDDM).

Variable	Correlations					
	1	2	3	4	5	6
1. Subjective knowledge (post-writing)	-					
2. MCQ total	0.109	-				
3. MCQ direct	0.213	0.884**	-			
4. MCQ inferential	-0.016	0.892**	0.577**	-		
5. SAQ	0.454**	0.728**	0.719**	0.577**	-	
6. HDDM *t* parameter estimate	0.021	-0.020	-0.022	-0.013	0.048	-

Note. $n = 40$, * $p < 0.05$, ** $p < 0.01$, (one-tailed).

7.4.3 Effects of Writing Process Measures (Global Linearity and Sentence Production) on Participants' Subjective Knowledge Development (Pre- to Post-Writing) Across the Writing Conditions (Synthesis vs. Free Recall)

7.4.3.1 Analytic Approach

To verify the two keystroke measure scales, Global Linearity and Sentence Production, Principal Component Analysis (PCA) with Varimax rotation was conducted on the eleven keystroke measures, as in the two previous studies using these scales (Baaijen & Galbraith, 2018; Hall, 2023). Factor scores for Global Linearity and Sentence Production were subsequently calculated for each participant. The analysis was conducted in R (Version 4.4.2) with the Psych package (Revelle, 2024).

Multiple regression was conducted in SPSS to assess the effects of the independent variables (writing process measures (Global Linearity and Sentence Production) and writing conditions) on the dependent variable (subjective knowledge change pre- to post-writing). Variables were entered in blocks. Main effects were entered in Block 1, two-way interactions in Block 2, and the three-way interaction in Block 3. Regression models were simplified by progressively removing non-significant terms, starting with the three-way interaction. The models were assessed for adherence to the assumptions of normality and homoscedasticity of residuals and there was no evidence these were violated. The presence of influential cases was checked using Cook's distance, centred leverage, standardised DfBeta and covariance ratios of which distributions were acceptable.

7.4.3.2 Principal Component Analysis

Sampling adequacy was verified using the Kaiser-Meyer-Olkin test (KMO = 0.63), which was above the acceptable limit of 0.5 (Field, 2013). PCA was also an appropriate method as there was collinearity between the variables, checked using Bartlett's Test of Sphericity which was highly significant ($\chi^2(55) = 294.578$, $p < 0.001$). The results of the PCA are shown in Table 16, with loadings of 0.3 and greater highlighted in bold as meaningful correlations between the variable and the component (Tabachnick & Fidell, 2012). The results of the PCA replicate those found in both previous studies using the measures (Baaijen & Galbraith, 2018; Hall, 2023) with one exception: the text modification index (TMI) did not strongly load onto either of the components. This is likely to be because during writing, participants only made very minimal revisions, and the TMI is a more global revision measure. For comparison, the mean TMI in the current study (mean = 0.83; standard deviation = 0.07) was lower than the mean TMI in Baaijen and Galbraith's (2018) work (mean = 1.28; standard deviation = 0.16).

Table 16. Principal Component Analysis (Varimax Rotation) for Two-Factor Solution.

	Component 1: Global Linearity	**Component 2: Sentence Production**
Sentence Linearity Index	**0.87**	0.04
Percentage of I-Bursts	**-0.92**	0.11
Percentage of Time Spent on Events	**-0.50**	0.22
Percentage of Linear Transitions Between Sentences	**0.48**	**0.31**
Number of Production Cycles	**-0.87**	-0.21
Percentage of Linear Transitions Between Words	**0.71**	-0.07
Percentage of R-Bursts	**0.32**	**-0.91**
Percentage of Words Produced in P-Bursts	0.26	**0.86**
Percentage of >2 Second Pauses Between Words	-0.13	**0.79**
Text Modification Index	0.10	0.17
Mean Pause Duration Between Sentences	-0.07	**0.79**
Eigenvalues	3.55	3.05
% Variance	0.32	0.28
α	0.83	0.79

7.4.3.3 Relationships Between Writing Processes (Global Linearity and Sentence Production) and Change in Subjective Knowledge Pre- to Post-Writing

Descriptive statistics and the correlations between the writing processes and change in subjective knowledge are shown in Table 17. Bivariate correlations are mostly non-significant and weak between the variables, with one noteworthy exception: there was a significant moderate positive correlation between change in knowledge and writing condition (as explained by the findings in Section 7.4.1.2).

It is also worth noting that there were no significant correlations between either of the writing process measures (global linearity and sentence production) and writing condition (synthesis or free recall). This implies that the writing activity participants undertook had no direct effect on the degree to which the writing processes were carried out. There were also no significant correlations between the writing process measures and change in participants' subjective knowledge, which contrasts previous research (Baaijen & Galbraith, 2018).

Table 17. Means, standard deviations (SD), and bivariate correlations for all variables in the regression model.

			Correlations			
	Mean	SD	1	2	3	4
1. Writing condition[1]						
2. Pre-Writing Knowledge	4.20	1.13	0.31*			
3. Change in Knowledge	-0.53	1.10	0.49**	-0.02		
4. Global Linearity	0.00	1.00	0.05	0.06	0.10	
5. Sentence Production	0.00	1.00	-0.14	-0.48**	-0.09	0.00

Note. $n = 40$, $^{*}p < 0.05$, $^{**}p < 0.01$ (one-tailed). [1]Dummy coded, free recall = 0, synthesis = 1.

Change in knowledge was regressed on writing condition, global linearity, and sentence production, controlling for initial pre-writing ratings of subjective knowledge and including the two- and three-way interactions between the variables. Regression models were simplified by progressively removing non-significant terms, starting with the three-way interaction. The final model is shown in Table 18, with significant predictors highlighted in bold.

This analysis showed that the only significant predictor of participants' change in subjective knowledge from pre- to post-writing was the writing condition (synthesis or free recall), as expected by the effects outlined in Section 7.4.1.2. Neither of the writing process measures (global linearity or sentence production) predict change in participants' subjective knowledge.

Table 18. Final regression model of predictors for change in subjective knowledge (pre- to post-writing).

	Model			
	B	SE	β	p
Constant	-0.055	0.682		0.937
Pre-Writing Knowledge	-0.255	0.165	-0.263	0.130
Writing Condition	1.195	0.326	0.549	< 0.001
Global Linearity	0.095	0.157	0.086	0.550
Sentence Production	-0.154	0.179	-0.140	0.395

$R^2 = 0.30$, Adjusted $R^2 = 0.21$, $F_{(4, 35)} = 3.65$, $p = 0.014$.

7.5 Discussion

The overarching goal of this paper was to explore the cognitive processes in writing-to-learn in the context of closed-book, multiple-source based writing activities. After reading two source texts, participants wrote either a synthesis essay or free recall about the sources, and the learning effects were assessed on three measures: text comprehension, subjective knowledge, and RT on a word-recognition memory task. Participants' keystrokes were logged whilst they were writing to assess how two writing process measures, global linearity and sentence production, related to participants' subjective knowledge development from pre- to post-writing across the different writing conditions. It was predicted that the learning effects and the writing process would reflect the synthesis writing task as involving a knowledge-transforming process (conceptualised by the dual-process model; Galbraith & Baaijen, 2018), and the free recall writing task as involving a knowledge-telling process.

The most important finding in relation to these predictions is that in the free recall writing condition, participants' subjective knowledge ratings significantly decreased from pre- to post-writing, whereas in the synthesis writing condition, participants' subjective knowledge ratings stayed the same. In the free recall condition, participants were instructed to knowledge-tell: they had to retrieve information from memory and translate it into written text. No integration of the two texts was required. As a result, these participants experienced an 'illusion of explanatory depth' (Rozenblit & Keil, 2002). After attempting to retrieve information from memory and write it down, participants realised they knew less than before they started writing. Although knowledge-telling is not associated with a *development* in knowledge, it is still a valuable learning process for identifying knowledge gaps.

Decreases in subjective knowledge were not observed in the synthesis writing condition. This demonstrates that the synthesis task allowed writers to move beyond simple knowledge-telling. As the synthesis task required participants to integrate the two texts (comparing and contrasting them), the writing process became less about translating retrieved information and more about connecting material with pre-existing knowledge (elaboration) and creating a new structure by re-organising material (organisation). This reflects the knowledge-transforming process of the dual-process model (Galbraith & Baaijen, 2018) as the retrieval of information from episodic memory which is evaluated and organised in line with the writer's goals for their synthesis text. Previous research (Baaijen & Galbraith, 2018) has suggested the knowledge-transforming process is not associated with *developments* in knowledge; this is reflected here since synthesis writers' subjective knowledge did not increase, it stayed the same. This also implies that source-based writing from multiple sources is likely to be limited to knowledge-transforming rather than knowledge-constituting.

Although the differences in subjective knowledge ratings suggest the writing process was different for the synthesis and free recall conditions, these differences were not reflected by the global linearity and sentence production measures extracted from participants' keystroke logs. It is important to note, however, that global linearity and sentence production were closely reproduced from the same keystroke measures (apart from the text modification index), demonstrating the reliability of these scales. Such scales are valuable for relating keystroke logs to cognitive processes and have seldom been used in other writing research which often analyses pauses, bursts, and revisions separately. The reproduction of these scales also reinforces the importance of transparency and open science principles to allow other researchers to closely follow the same rules to analyse keystrokes.

The writing condition participants undertook did not affect the degree to which the two writing processes were carried out, nor did the writing processes predict change in participants' subjective knowledge. This suggests that the use of these keystroke measures in the current experiment did not reflect writing processes in the same way as in previous research (Baaijen & Galbraith, 2018). Baaijen and Galbraith developed the global linearity and sentence production scales from keystroke logs where participants wrote texts from their own pre-existing knowledge. In comparison, the measures were used here in source-based tasks. Since keystroke measures must be contextualised, it is likely that global linearity and sentence production do not have the same meaning in source-based writing. Future research should be carried out to develop measures that can be derived from keystroke logs that can be used to assess differences in the writing process of source-based activities. Future studies could also combine the synthesis task with outline versus synthetic planning (as in Baaijen & Galbraith's (2018) study) to compare the writing process.

With regards to text comprehension, synthesis and free recall participants performed equally well overall on both MCQs and SAQs. As the writing activities were completed in a closed-book environment, participants in all conditions engaged in retrieval practice to some extent. As shown by research on the testing effect (e.g., Glover, 1989; Butler, 2010; Roediger & Karpicke, 2006a; Karpicke, 2012), retrieval practice is a highly valuable tool for enhancing the retention of information (Karpicke, Butler & Roediger, 2009), and is an underlying process in writing-to-learn from sources in closed-book environments (Arnold et al., 2017; Papers 1 and 2). One exception is that participants in the free recall condition performed significantly better on inferential MCQs than participants in the synthesis writing condition. A possible explanation for this is that the synthesis task was less focused on remembering information from the sources and therefore they did not perform as well on these questions. From this explanation it would also be expected that free recall participants would also perform better on direct (factual) MCQs, but although their mean scores were higher (see Appendix I.3), this was not a significant difference.

Subjective knowledge ratings post-writing were significantly moderately correlated with participants' scores on the SAQs. This was also found in Papers 1 and 2, demonstrating the reliability of this finding. This is important since it shows subjective knowledge ratings are somewhat related to a more objective measure which has not been explored in any previous writing-to-learn research. Accordingly, it is recommended that future studies incorporate the subjective knowledge rating scale (Galbraith et al., 2023) as a valid measure of learning. It is possible that SAQ scores correlate more strongly with subjective knowledge ratings than MCQ scores because they allow participants an enhanced opportunity to express their knowledge since the questions are open-ended. Therefore, in research and practice, it is important that MCQs are not the only measure of learning or assessment.

Finally, this study was the first experiment with an extended writing task to use RT on a word-recognition task as a measure of learning. Previous studies (Silva & Limongi, 2019; Paper 3) have shown this measure to be reliable and valid in comparing the long-term memory consolidation of written versus spoken summaries. Here, however, no differences were found on the RT measure between the synthesis and free recall conditions, and RT, represented by the *t* parameter of the drift-diffusion model, did not correlate with text comprehension scores or subjective knowledge ratings. It is possible that writing a synthesis or a free recall led to similar long-term memory consolidation of the source text and therefore differences in RT would not be expected. However, it is also likely that the measure needs further development to be incorporated into such experiments. For example, in the current experiment, participants responded to a relatively small amount of target words (60 words) compared to in Silva and Limongi's (2019) study and in Paper 3 (240 words across eight experimental blocks). A larger number of RTs may be required to detect small differences across conditions. It is important that future studies to continue to explore this measure as it has valuable implications for investigating writing-to-learn from a cognitive neuroscience perspective.

7.5.1 Limitations

This paper offers novel insights into the cognitive processes in writing-to-learn from multiple sources in closed-book environments. It has considered several different measures of learning from different writing activities and the relationships between them, which has not been done in previous research. However, the study is not without its limitations.

The sample size was relatively small (40 participants) and was predominantly female. The statistical power for detecting small effects between conditions (e.g., for text comprehension scores and RT) was likely to be low. Future studies should endeavour to recruit larger samples from a wider variety of backgrounds. This is especially the case for writing research in which

gender differences have been reported (e.g., Bruning & Kauffman, 2016; Cordeiro, Castro & Limpo, 2018).

The use of keystroke logging, as opposed to other methods such as think-aloud protocols, is an advantage since it is unobtrusive and takes place in a familiar environment to participants (Microsoft Word). The use of two composite scales (global linearity and sentence production) combining pauses, bursts, and revisions was also a strength over previous research which has investigated these features separately. However, as acknowledged by Baaijen and Galbraith (2018), although cognitive processes relating to the dual-process model can be inferred from these measures, keystrokes do not offer direct evidence of underlying cognitive processes. As demonstrated in this study, keystrokes are also contextualised, and such measures may not be suitable for different types of writing (e.g., source-based versus writing without sources). There are also specific limitations associated with the eleven measures forming the two scales. For example, the high-level burst coding (P-, R-, and I-Bursts) does not account for instances where two types of burst occur at the same time, such as an I-Burst which ends in a pause or a revision.

These results are also based on writers constructing texts on a single topic (different modes of therapy for the treatment of neuroses). Some participants (e.g., those studying psychology) may have been more familiar with such a topic than other participants. Future research is necessary to explore the findings across more diverse topics and domains. The effects of instruction should also be investigated in future studies. Here, participants were not made aware of the writing activity they would be completing before they read the source texts. It is possible that knowing the upcoming writing task in advance of reading may change the processes participants' use to read the texts.

This experiment also did not consider potential individual differences across participants. Baaijen and Galbraith's (2018) study considered self-monitoring (the degree that learners focus their writing on rhetorical or dispositional goals), and although they found limited evidence that this moderated the findings, they emphasised that further research is needed on how self-monitoring can affect the writing process. Similarly, participants' writing beliefs may moderate the extent to which they use different writing strategies (e.g., Baaijen, Galbraith & de Glopper, 2014); this has not been explored previously in source-based writing. Finally, future research should also investigate the results in different contexts, such as for handwriting rather than typing which has been found to influence writing fluency and revision processes (e.g., van Waes & Schellens, 2003; Kellogg, 1994).

7.6 Conclusion

In conclusion, this paper advances knowledge of writing-to-learn by examining the cognitive processes in closed-book, multiple source-based writing tasks. The crucial finding was that when participants are instructed to knowledge-tell (through a free recall task), they experience a decrease in their subjective knowledge from pre- to post-writing, however when they engage in a synthesis writing task, they do not experience this decrease. This provides some support for the dual-process model in that synthesis writers appeared to engage with a knowledge-transforming process. Importantly, this study also replicated previous findings (Papers 1 and 2) that subjective knowledge ratings are significantly moderately positively correlated with scores on short-answer text comprehension questions. It also replicated writing process measures derived from participants' keystrokes (global linearity and sentence production; Baaijen & Galbraith, 2018), although these did not distinguish between writing conditions, suggesting alternative measures should be constructed for source-based writing research. Finally, this experiment represented an initial attempt to use a RT measure on a word-recognition memory task for exploring learning from an extended writing task. The limited findings highlight the need for methodological refinement of this measure in future research.

Chapter 8 General Discussion and Conclusion

The central goal of this thesis was to explore the cognitive processes underlying the effects of closed-book source-based writing on learning through the lens of the dual-process model of writing (Galbraith & Baaijen, 2018). A second goal was to compare different measures of writing-to-learn, including both a response accuracy measure (text comprehension), a response time measure (long-term memory consolidation measured by reaction time (RT) on a word-recognition task), and a measure of participants' subjective knowledge. These goals were achieved through four independent, but related, empirical papers.

This chapter will summarise the results of the four empirical papers. It will then bring these results together to discuss the overall findings of this thesis in relation to the theory and previous literature on writing-to-learn. It will explain the strengths and limitations of the thesis overall and the individual papers. Finally, it will consider implications of the work for educational practice, and make recommendations for future research.

8.1 Summary of Results

Paper 1 (Chapter 4) explored the effects of different planning strategies (outline and synthetic) and writing (essay and free recall) tasks on text comprehension and subjective knowledge. This built on previous research by Baaijen and Galbraith (2018) and Arnold et al. (2017) to explore the claims of the dual-process model in the context of source-based writing, and to assess the relationship between the text comprehension and the subjective knowledge measures. Whilst the results showed there were no effects of either planning strategies or writing task on text comprehension scores and subjective knowledge ratings, there was clear evidence that participants' subjective knowledge ratings decreased from pre- to post-writing (an 'illusion of explanatory depth'; Rozenblit & Keil, 2002).

It was therefore concluded that writing from a single source did not involve the development of new knowledge. As such single source-based writing involves knowledge-telling through retrieval practice, rather than knowledge-transforming or knowledge-constituting (Galbraith & Baaijen, 2018). Contrary to previous claims that knowledge-telling has no role in the writer's development of knowledge (Bereiter & Scardamalia, 1987; Galbraith & Baaijen, 2018), this paper suggested that an important role of knowledge-telling is to help the writer monitor their own knowledge, i.e., to identify knowledge gaps. Finally, participants' subjective knowledge ratings were moderately correlated to their text comprehension scores, suggesting they are related measures but may capture different dimensions of learning. The strongest correlations were with participants' short-answer question (SAQ) scores, suggesting these are the best predictor

of participants' perceived knowledge. This is a notable finding as it validates the subjective knowledge rating scale (Galbraith et al., 2023) as a valid measure of learning.

Paper 2 (Chapter 5) extended the findings of Paper 1 by generalising the results to a new source text topic as well as a comparison to non-writing control conditions. Experiment 1 of Paper 2 was a methodological check of a new set of text comprehension questions for use in Experiment 2. In Experiment 2, two single source-based activities were compared (essay and free recall) in different formats (written or mental (thinking)) on participants' text comprehension and subjective knowledge. This was to establish whether the effects found in Paper 1 are specific to *writing*, or also true of general cognitive processing involving retrieval. Results showed that participants who wrote an essay or free recall performed better on SAQs than participants who thought about writing an essay or free recall. This demonstrates the important and unique value of writing in specifically enabling this effect of retrieval practice in a way that thinking does not. Even though their SAQ text comprehension performance was better, participants in the written conditions experienced a larger decrease in their subjective knowledge from pre- to post-writing than participants in the thinking conditions. Taken together, these results indicate the powerful role of writing for the retention of information as well as metacognitive monitoring (realising knowledge gaps). Notably, Paper 2 (Experiment 2) also replicated the result from Paper 1 that subjective knowledge ratings were significantly correlated with SAQ scores and factual MCQ scores.

Much like previous research on writing-to-learn, Papers 1 and 2 measured learning through a response accuracy measure (a text comprehension test) as well as a measure of participants' subjective knowledge. Paper 3 (Chapter 6) evaluates a new methodological tool for measuring long-term memory consolidation from source-based writing, as a replication and extension of work by Silva and Limongi (2019). A direct replication (Experiment 1) and a partial replication to control for a possible orthographic priming effect (Experiment 2) reproduced the findings from Silva and Limongi's study. Participants RTs on a word-recognition task were faster after writing a summary than after speaking a summary. This was explained by the non-decision parameter of the drift-diffusion model (reflecting encoding and/or motor processes). These results reinforce the idea of *writing*-to-learn (in comparison to speaking) with a novel measure and highlight the importance of the external representation and visual feedback of written text production.

Paper 4 (Chapter 7) integrates Papers 1 to 3 by bringing together the three measures of learning (text comprehension, subjective knowledge, and RT) in a multiple source-based writing task. Participants' performance on these measures was compared for a synthesis writing task (comparing and contrasting two text passages) versus a free recall writing task (writing down everything one could remember from two text passages). Whilst writing the synthesis or free

recall, participants keystrokes were logged to provide further insight into their writing process. The most striking finding was that subjective knowledge ratings decreased pre- to post-writing for participants in the free recall condition, but not for participants in the synthesis writing condition. It was therefore concluded that a multiple source-based synthesis writing task reflects the knowledge-transforming process of the dual-process model (Galbraith & Baaijen, 2018) to a greater extent than single source-based writing tasks. Synthesis writing allowed participants to move beyond retrieval practice and knowledge-telling through requiring the writer to create new ideas during text production. This was further supported by the finding that participants in the free recall condition performed better on inferential multiple-choice questions (MCQs) than participants in the synthesis condition, suggesting the synthesis task was less about remembering material from the source texts. Despite this, the writing activities had no direct effect on the degree to which writing process measures (global linearity and sentence production), extracted from participants' keystroke logs, were carried out. These measures were previously developed from research where participants write from their own knowledge (Baaijen & Galbraith, 2018); it is likely they may not have the same meaning for source-based writing. In replication of the findings from Papers 1 and 2, participants' subjective knowledge ratings were positively correlated with their SAQ scores. Finally, an initial attempt to incorporate RT as a measure of long-term memory consolidation highlighted that further work is needed to use the measure in studies on more complex writing.

8.2 Discussion of the Key Findings Across Papers

Taken together, the findings from these four papers offer novel insights into the cognitive processes in writing-to-learn. Theoretically, they highlight the distinct mechanisms involved when writing from sources in a closed-book setting in comparison to writing from sources in an open-book setting and writing from existing knowledge in the absence of sources. Methodologically, for the first time in writing-to-learn research, multiple measures of learning from writing, and the relationships between them, have been considered.

8.2.1 Theoretical Insights

Central to the results is the consistent finding from Papers 1 and 2, and for free recall participants in Paper 4, that participants subjective knowledge ratings decreased from pre- to post-writing. After reading a source text, participants initially believed they had gained knowledge and understood the information within the text. However, after attempting to write about the source, they came to realise their knowledge was much lower. This is consistent with the literature on the 'illusion of explanatory depth' (Rozenblit & Keil, 2002). In this phenomenon,

after individuals attempt to write an explanation of a topic (e.g., how a device works), they rate their understanding lower than they had before they explained the topic (Mills & Keil, 2004; Keil, 2003). One advantage of experiencing a decrease in subjective knowledge after writing is for metacognitive monitoring. Even though participants had not necessarily 'learned' in terms of knowledge *gain*, learning has still taken place since participants recognise gaps in their knowledge. Previous research (e.g., Little & McDaniel, 2015; Glenberg & Epstein, 1987) has shown that students use such metacognitive monitoring to inform their future study. Therefore one could expect that if participants were provided with an opportunity to revisit the source text and then revise their texts, they would then experience an increase in their subjective knowledge. Future research could explore this prediction.

The decrease in subjective knowledge ratings experienced by participants in this thesis is in contrast to previous research assessing the claims of the dual-process model, as well as to general assumptions in writing theory that knowledge-telling does not play a role in learning (Bereiter & Scardamalia, 1987). Baaijen and Galbraith (2018) did not observe such a decrease in participants' subjective knowledge from pre- to post-writing. The difference between the studies in this thesis and Baaijen and Galbraith's work is that in the former, participants wrote from sources, whereas in the latter, participants wrote based on their own personal understanding in the absence of a source.

By way of reminder, the dual-process model (Galbraith & Baaijen, 2018) predicts that discovery of knowledge (reflected by an *increase* in subjective ratings) is related to the effective management of two processes: knowledge-transforming and knowledge-constituting. Knowledge-constituting represents the active synthesis of knowledge from semantic memory as the text is produced which facilitates a development of understanding; this was reflected in Baaijen and Galbraith's study by participants who synthetically planned and wrote spontaneously. Knowledge-transforming represents goal-directed retrieval of knowledge from episodic memory; when individuals only engage in this process (e.g., by outline planning and more controlled sentence production), their development of understanding is more limited (Baaijen & Galbraith, 2018). The knowledge-transforming process differs from the knowledge-telling process in Bereiter and Scardamalia's (1987) classic model of writing since it requires the organisation of knowledge towards rhetorical goals, whereas knowledge-telling is essentially the translation of retrieved ideas into written text. In writing-to-learn, it has been assumed that knowledge-telling has no responsibility for learning and the writer's knowledge development.

When participants wrote from single sources in a closed-book setting in Papers 1 and 2, they engaged in retrieval practice, regardless of the writing condition they were subject to (essay writing or free recall). As a result, they predominantly engaged in a knowledge-telling process.

This means closed-book single source-based writing was treated as translating retrieved information into written text, as opposed to the active construction of new ideas. When writing based on a single source, it seems that the writers main aim is to report what they remember from the text, rather than integrating their own knowledge. Therefore, it is unlikely they would engage in knowledge-constituting. Furthermore, since the source text has already provided an organised structure and argument, there is little need for the writer to engage in knowledge-transforming. This explains why writers did not experience a subjective increase in their knowledge, and in fact experienced a decrease. However, this allows for a re-interpretation of the knowledge-telling process as a cognitive strategy that does have a role in learning. That is, knowledge-telling can help a writer to consolidate their knowledge of a text and reveal knowledge gaps. Rather than simply just translating thought into text, knowledge-telling can be an important mechanism for learners to monitor their understanding. This is a key novel finding of this project.

Interestingly, the knowledge decrease was not observed in a synthesis writing task in Paper 4. This was a multiple source-based closed-book writing task in which participants were instructed to *'compare and contrast'* two text passages and '*conclude with an overall evaluation*'. In the same study, participants who were instructed to simply *'recall and write down'* everything they could remember from the two text passages did experience a decrease in their subjective knowledge. It therefore appeared that whilst free recall participants were restricted to knowledge-telling strategies, a synthesis writing prompt allowed participants to engage in knowledge-transforming. Free recall participants performed significantly better on inferential MCQs than synthesis participants. This further suggests the synthesis task was less about remembering information from the source, and more about integrating the content with participants' own existing knowledge to generate a new text.

The removal of the knowledge decrease effect for the synthesis writing task illustrates that closed-book source-based writing does not have to be restricted to knowledge-telling and retrieval practice processes. This is likely related to the instruction given to participants, and importantly their interpretation of such instruction, which encourages them to use organisation and elaboration processes in addition to retrieval. This is supportive of recent work by Nückles and colleagues (2020) on self-regulated learning journals. When writing about source material from lectures and courses, learning journals provide students with cognitive prompts to encourage organisation and elaboration processes, and metacognitive prompts to encourage students to monitor their own knowledge and comprehension. These have been found to be very effective for enhancing learning outcomes in university and high school students (Nückles et al., 2020; Hübner, Nückles & Renkl, 2010).

To summarise so far, the writing process in closed-book tasks with single sources is mostly comprised of knowledge-telling. However, knowledge-telling has been reinterpreted here as a cognitive process that can help writers consolidate their knowledge and identify knowledge gaps, through the benefits of retrieval practice. To move beyond knowledge-telling in source-based tasks, it is important to prompt learners to use elaboration and organisation processes (e.g., comparing and contrasting in a synthesis writing task). The learning benefits in such tasks are likely to be because of goal-directed knowledge-transforming.

As pointed out by Arnold et al. (2017), despite the exhaustive literature on the testing effect (e.g., Glover, 1989; Butler, 2010; Roediger & Karpicke, 2006a, 2006b; Karpicke, 2012), writing-to-learn theory has overlooked retrieval practice. The results discussed so far do suggest retrieval practice is a valuable underlying processes in the effects of writing on learning. In spite of this, the results of Papers 2 and 3 suggest there must still be something inherently unique about writing, beyond retrieval practice, in source-based tasks. In Paper 2, writing activities (essay and free recall) were compared to mental equivalents where participants were instructed to think about how they would write an essay or free recall text. In Paper 3, writing summaries was compared to speaking summaries. In both papers, the non-writing tasks (thinking and speaking) were also conducted in a closed-book environment, and therefore also involved retrieval practice. Yet, the findings showed that writing was superior for text comprehension (Paper 2) and long-term memory consolidation (Paper 3), as well as for metacognitive monitoring, since participants who wrote experienced bigger decreases in their subjective knowledge than participants who thought about writing (Paper 2).

One possible explanation is that writing provides an external representation of knowledge in a way that thinking or speaking does not. By having a physical record of their knowledge, thoughts, and ideas retrieved so far, writing facilitates reflection and self-regulation. Nückles et al. (2020) recognised this for supporting metacognitive monitoring in learning journals. In addition, studies exploring the permanency of the written trace by comparing writing with invisible writing provide support for this explanation. For example, writing produced better recall of information compared to invisible writing, dictating and speaking in Grabowski's (2007) experiments. The 'external store' of writing can support also individuals to engage in higher-level writing processes (Olive & Passerault, 2012; Olive & Piolat, 2002; Berninger & Winn, 2006). Future research could investigate this explanation by adding an invisible writing condition in the basic experimental design of reading a source text, writing an essay or free recall text about it, and completing learning outcome measures. One such study is currently being undertaken by Galbraith and colleagues within the paradigm of Silva and Limongi's (2019) experiment. Participants read short source texts, produce a summary with normal typing or invisible typing, and complete the word-recognition RT task. Early exploration of the data, however, has

suggested minimal differences between the writing and invisible writing conditions. This suggests the external representation of writing is not the only explanation of the writing-superiority effect.

Another possibility is that writing requires individuals to elaborate on retrieved information and translate it into a structured text. This is likely to involve more complex cognitive processing than speaking or thinking, which requires individuals to retrieve information but not necessarily to elaborate on or organise it. This explanation would provide support for the knowledge-transforming process of the dual-process model (Galbraith & Baaijen, 2018) which is assumed to reflect the goal-directed organisation of retrieved knowledge. Future research could explore this explanation directly by comparing writing and speaking or thinking in closed- vs. open-book source-based tasks. As the retrieval element is removed in an open-book task (where learners have access to source materials), this may allow writers to have more capacity to engage with organisation and elaboration processes. Existing research supports this notion. Arnold et al. (2021) found that whilst participants performed equally well on a text comprehension test following open- or closed-book essay writing, those in the open-book condition benefited from organisation and elaboration processes (generative learning) reflected by the production of better-quality essays. Waldeyer et al. (2020) also found generative learning benefits for an open-book source-based activity versus a retrieval practice benefit for a closed-book source-based activity.

One important question for future research to explore is the extent to which source-based tasks can elicit knowledge-constituting (Galbraith & Baaijen, 2018), and therefore developments of knowledge. The insights from the research presented in this thesis suggest that writing from single sources is restricted to knowledge-telling, and that a synthesis task involving writing from multiple sources involves knowledge-transforming. It is important to investigate whether source-based tasks completely rule out the possibility of knowledge-constituting. A future study could explore this by manipulating planning (outline versus synthetic) in a source-based synthesis writing task.

Up to this point, the theoretical contributions of this thesis have been discussed. Attention now shifts to discussion of the methodological contributions.

8.2.2 Measuring Learning Outcomes

Writing-to-learn research has traditionally been dominated by recall and comprehension measures of learning (Ackerman, 1993). Baaijen and Galbraith (2018; Galbraith et al., 2023) used subjective knowledge ratings to measure participants' perceived knowledge discovery. Prior to this thesis, it was unknown how these subjective knowledge ratings were related to text

comprehension performance. The results from Papers 1, 2, and 4 indicate that participants' subjective knowledge is moderately correlated with their text comprehension performance. In Paper 1, both MCQ scores and SAQ scores were correlated with subjective knowledge ratings. In Paper 2, direct (factual) MCQ scores and SAQ scores were correlated with subjective knowledge ratings. Finally, in Paper 4, only SAQ scores were correlated with subjective knowledge ratings.

There is clear evidence that participants' subjective knowledge ratings are related to objective measures of learning. Therefore, the subjective knowledge rating scale (Galbraith et al., 2023) is a valid measure. SAQs were significantly positively correlated with subjective knowledge ratings consistently throughout this thesis. This suggests they provide the best opportunity for participants to demonstrate their knowledge of a source text. MCQs may reflect a more superficial measure of learning (e.g., the ability to remember a fact). Furthermore, text comprehension was generally not influenced by different writing activities (e.g., no differences between essay and free recall). This can be explained by the fact that these activities all involved retrieval practice, which can help to answer the questions on the text comprehension test.

Second, and more importantly, these results reflect that learning from writing has different dimensions. This was because, for the first time in writing-to-learn research, multiple different measures of learning, and how they are related, were considered. Papers 1 and 2 established the relationship between text comprehension (Arnold et al., 2017) and subjective knowledge (Baaijen & Galbraith, 2018) measures; Paper 3 established the reliability of a new methodological tool (Silva & Limongi, 2019) for measuring long-term memory consolidation of writing-to-learn; and finally Paper 4 brought together these three measures.

What can be concluded from this is that learning from source-based writing can be assessed under two dimensions: remembering, or consolidating, information (text comprehension and RT), and understanding information (subjective knowledge). This may also reflect the distinction between episodic and semantic memory. Text comprehension and RT could be considered as capturing episodic memory, and subjective knowledge as capturing semantic memory. A notable implication of this is that it may help to explain contradictory results, for example in research on summary writing. Several studies have reported advantageous effects of summarising for learning measured by a text comprehension test (e.g., Doctorow, Wittrock & Marks, 1978; Thiede & Anderson, 2003; Gil et al., 2010). Despite these findings, Spirgel and Delaney (2016) and Dunlosky et al. (2013) concluded that summarising is not an effective tool for learning. It therefore seems likely that when learning is considered as the ability to remember information, summarising can be considered as a beneficial learning tool, however when

learning is considered as the ability to understand information, summarising is not considered as a beneficial learning tool.

8.3 Recommendations for Future Research

As outlined in the previous section, future research could adapt the research designs from this thesis to evaluate possible explanations of the results. For example, comparing closed- and open-book source-based writing and non-writing tasks, as well as tasks with and without visual feedback. In addition, another important consideration for future research is individual differences which may affect the relationships between writing and learning. This is important for making well-informed recommendations for educational settings with a range of learners.

8.3.1 Individual Differences

This thesis considered the range of outcome measures of learning from writing (text comprehension, subjective knowledge, and RT), which has not been done previously in writing-to-learn research. To maintain this focus, individual differences which might moderate the relationship between writing and learning were not included. Furthermore, this project was heavily influenced in its early stages by the Covid-19 pandemic (outlined in Chapter 3). Decisions on the measures to include in experiments were made based on the ability that the data could be collected online. However, it is important for future research to explore individual differences (variations in the abilities of learners which may affect the effectiveness of different cognitive strategies; Arnold et al. (2017)). These might include, but are not limited to, individual differences in writing beliefs and working memory capacity, which are discussed next. Future studies may also consider writing motivation, interest, and self-efficacy, as well as individual differences in reading skill.

8.3.1.1 Writing Beliefs

Participants' beliefs about writing may moderate the relationship between writing and learning. This is reflected in a handful of studies using the Writing Beliefs Inventory (White & Bruning, 2005), which distinguishes between transmissional and transactional writers. Writers with high transmissional beliefs view writing as a way to transfer information or report content from authoritative sources, whilst writers with high transactional beliefs view writing as a way to develop their understanding by integrating their thinking (White & Bruning, 2005; Galbraith & Baaijen, 2018). White and Bruning (2005) found undergraduates with high transactional beliefs tended to have more cognitive and emotional engagement with their emerging texts, viewing writing as self-expression, and were more likely to write for pleasure than high transmissional

writers. This also meant high transactional writers produced better quality texts. These findings were extended by Baaijen, Galbraith & de Glopper (2014) who found the effectiveness of different writing strategies depended on participants' writing beliefs. Those with high transactional beliefs developed their understanding more through writing through knowledge-constituting, whereas those with low transactional beliefs did not develop their understanding since their writing process was mostly comprised of translating knowledge into text. Similarly, Sanders-Reio et al. (2014) suggested transmissional beliefs can be viewed as 'knowledge-telling' whereas transactional beliefs can be viewed as 'knowledge-transforming'.

The findings of this thesis suggested that closed-book source-based writing, in particular writing from single sources, is largely a knowledge-telling process. Therefore, it could be implied that this task is more accustomed to writers with high transmissional beliefs. In addition, overall, writing from single sources tended to decrease participants' subjective knowledge pre- to post-writing, because knowledge-telling helped them realise information that they do not know. Writing beliefs could moderate this effect. Whilst writers with high transmissional beliefs may stick to knowledge-telling and report the information from sources directly, writers with high transactional beliefs may transform knowledge from the sources to a greater extent through integrating it with their own thoughts and ideas. This may then be reflected in their knowledge ratings, where high transmissional writers experience a larger decrease in knowledge pre- to post-writing than high transactional writers, who may be able to move away from knowledge-telling to develop their understanding through writing.

8.3.1.2 Working Memory

An individual's working memory capacity may also moderate the relationship between writing and learning. Previous research has suggested that a learner's working memory can affect both reading comprehension (e.g., Daneman & Carpenter, 1980; Carretti et al., 2009) and writing (e.g., McCutchen, 2000; Kellogg, 1996; Kellogg et al., 2013; Olive 2004; Swanson & Berninger, 1996). Working memory is important in the writing process because it provides a store for temporary thoughts, ideas, and information during text production (Olive, 2022). An individual's limited working memory capacity can restrict the extent to which cognitive strategies are carried out during writing, which can subsequently affect writing quality and fluency (Ransdell & Levy, 1996). Written text production engages all components of working memory (Olive, 2022). The executive function of working memory is said to be fundamental to writing (Olive, 2012) since writers have to monitor multiple processes simultaneously. In addition, Galbraith et al. (2009) found the spatial component of working memory to have an important role in the development of thoughts and ideas during planning processes. In Chenoweth and Hayes' (2003) study, participants who were subject to articulatory suppression during writing made more surface-

level mistakes (e.g., typos, spelling and grammar) and perceived the task to be more challenging, as well as changes to the moment-by-moment construction of their texts.

Arnold et al. (2017) considered the moderating role of working memory capacity in their study comparing retrieval and non-retrieval writing activities. They predicted that working memory would influence retrieval activities in particular, and could also influence the extent that participants engaged with elaboration and organisation processes when writing. However, results showed no effects of working memory capacity. Nonetheless, previous research has found that individuals with high working memory capacity are more capable of strategic encoding processes and less susceptible to interference during retrieval from episodic memory than individuals with low working memory capacity (Unsworth, Brewer & Spillers, 2012; Unsworth & Spillers, 2010). Since both retrieval and encoding processes were found to be important for the superiority effect of writing in this project, individual differences in working memory capacity are an important consideration for future research. The dual-process model (Galbraith & Baaijen, 2018) would hypothesise that the knowledge-transforming process would be most affected by working memory capacity, since it is responsible for the reflective evaluation of thought towards rhetorical goals. Therefore, one could predict working memory capacity to moderate the learning outcomes particularly in synthesis writing tasks.

8.4 Strengths and Limitations

The following section will elaborate on the strength of this research in enhancing theoretical and methodological knowledge of writing-to-learn from sources. It will also acknowledge the limitations associated with a largely homogenous sample of female undergraduates across all studies, as well as consider the strengths and limitations associated with online experimental data collection in Papers 1 and 2.

8.4.1 Enhancing Theoretical and Methodological Knowledge of Writing-to-Learn

A strength of this research is that it has enhanced current knowledge of the underlying cognitive processes involved in the effects of closed-book source-based writing on learning. Previous writing-to-learn theory claims that knowledge-telling does not play a role in the development of understanding; instead, development of understanding was argued to be associated with knowledge-transforming and knowledge-constituting (Bereiter & Scardamalia, 1987; Baaijen & Galbraith, 2018). However, the results of this thesis indicate an important function of knowledge-telling for learning; that is, to help learners identify gaps in their knowledge.

A further strength of this research is that it considers multiple different measures of learning from source-based writing, and how they relate to one another. Previous writing-to-learn research has not made such an explicit distinction between the different dimensions of learning that writing can influence. The text comprehension test was used to assess participants' factual and inferential knowledge about content from the text passages they read. Recall and comprehension measures are dominant in writing-to-learn (e.g., Ackerman, 1993; Gunel, Hand & McDermott, 2009; Gingerich et al., 2014) and source-based writing research (e.g., Thiede & Anderson, 2003; Gil et al., 2010; Wiley & Voss, 1999). Although an ecologically valid measure due to their frequent use in educational settings, text comprehension tests can reduce the meaning of learning to how well students have retained information and do not necessarily measure the extent that a student has grasped sound knowledge and understanding of the topic. To capture participants' own perceptions of their knowledge and understanding, a Subjective Knowledge Rating Scale (Galbraith et al., 2023) was also used to assess learning in this thesis. Papers 1 and 2 of this thesis represent the first studies to consider the relationship between participants' subjective knowledge ratings and their text comprehension scores, and showed that these measures are related to some extent.

In addition to the text comprehension and subjective knowledge measures, this thesis also established a novel measure of learning from writing by replicating and extending Silva and Limongi's (2019) work. Measuring RT on an episodic memory word-recognition task was a reliable way to assess long-term memory consolidation from writing. Although work is still needed to explore how the measure can be incorporated in studies with more complex and extended writing tasks, the methodological contribution provided in this thesis is particularly valuable in enhancing the writing-to-learn field by initiating a connection between writing-to-learn and cognitive neuroscience.

8.4.2 Limitations of the Sample

One of the main limitations across the four papers in this thesis is the sample of participants. In the first three papers, the samples consisted predominantly of female undergraduate psychology students. In Paper 4, participants were recruited from a wider variety of subjects, but the sample was still mostly comprised of female undergraduates from the social sciences. This is common in psychological research due to the ease of accessing these participants (e.g., through university research participation schemes) and is relevant for this thesis in studying writing in higher education. However, it does restrict the generalisability of the findings to broader populations.

When recruiting participants through the psychology research participation scheme, it was expected that most participants would be women. This is because the majority of undergraduate students studying psychology in the UK are female (85% at Russell Group universities) (Johnson et al., 2020). Nevertheless, this is a noteworthy limitation for research on writing because gender differences have been reported in previous research. For example, females are reported to have higher writing self-efficacy (e.g., Bruning & Kauffman, 2016) and higher writing skill (e.g., Cordeiro, Castro & Limpo, 2018; Villalón et al., 2015) than males. In their study on synthesis writing, Vandermeulen et al. (2020) found female participants felt more positively towards writing and were more emotionally involved with text production than male participants, even though females found writing more cognitively demanding. Noting these gender differences, future research could consider the extent to which the findings of this thesis are limited to a female population.

Even though this thesis was focused students' writing in higher education, it would be interesting to conduct further research to establish whether the findings extend to younger groups, such as in primary and secondary school children. Source-based writing is not an activity restricted to university settings; it is used throughout all stages of education. A key finding of this thesis was that when producing a source-based text, particularly from a single source, writers tend to engage in a knowledge-telling process. According to Bereiter and Scardamalia (1987), the knowledge-telling process is used frequently by younger, novice writers. Therefore one might expect similar results to be found in younger students. It can also be expected that children would also report decreases in their subjective knowledge pre- to post-writing when engaging in knowledge-telling. The illusion of explanatory depth has been found in children as well as adults (Mills & Keil, 2004).

8.4.3 Online Experimental Research

The forced online data collection for Papers 1 and 2 as a result of the Covid-19 pandemic reflects both a strength and a limitation of this research. As a strength, it offers valuable insights into conducting experimental research designs remotely as an alternative to laboratory experiments. A defining feature of experimental methods is the high control, especially in experiments which require precise measurements (Cohen, Manion & Morrison, 2018). As such, experimental research has rarely been conducted online.

In general, online research is more convenient for both researchers and participants (Clifford & Jerit, 2014). Participants are able to take part in online research using their own computer, in their own home, at a time most suitable to them (Stoner, Felix & Blank, 2022). Online data collection also provides the opportunity to recruit a larger sample of participants from a wider

variety of demographic backgrounds. This is reflected in the high number of people who volunteered to participate in the studies reported in Paper 1 (229) and Paper 2 (195 for Experiment 2). However, many of these participants did not complete the full study, and therefore the number of participants included in analyses is much lower (78 in Paper 1; 116 in Paper 2, Experiment 2). Since data collection for Paper 1 and Paper 2 (Experiment 2) was conducted in two sessions, two-days apart, it was anticipated that there would be some dropout, even if data were collected in-person. However, online data collection was likely to exacerbate this dropout due to the lack of direct interaction between the participant and the researcher. One method of mitigating this in future online experimental studies could be to conduct the experiment through a video call with the participant to allow them to complete the study under the guide of the researcher. On the other hand however, this may make recruitment of participants more challenging as the study has to be completed at a convenient time for both the participant and the researcher.

Aside from the strengths and challenges associated with participant recruitment, the most notable limitation of online data collection for the current project is the lack of experimental control. Research comparing online research to laboratory research suggests participants experience significantly more distraction in online settings (e.g., Clifford & Jerit, 2014). During the reading and writing activities, it is possible that participants were not focused on the task for the full amount of time. This mostly has implications for Experiment 2 in Paper 2, in particular the two thinking (mental) conditions. As participants did not have an explicit physical task to be completing (as is the case for participants allocated to the two writing conditions), it is possible they experienced enhanced levels of distraction. This could be an alternative explanation as to why participants in the thinking conditions performed worse on the SAQs than participants in the written conditions.

The final insight from remote data collection is the importance of the tools or platforms used to host experiments online. At the time of data collection for Paper 1, iSurvey was the only available tool provided by the university that could be used to collect data remotely. Prior to the pandemic, this tool was largely used for simple survey and questionnaire research. However, it had the functionality available for online experiments in randomly assigning participants to conditions in which they could complete different tasks. Although the platform guided participants to spend five minutes planning and fifteen minutes writing, there was variation amongst participants in the time they spent on these activities. This limitation was reduced in Experiment 2 of Paper 2 because the university was able to provide access to Qualtrics, which is a more reliable platform for conducting experimental research as it prevents participants from moving onto the next aspect of the procedure before the time has elapsed. That said, it does not eliminate the distractions participants could have experienced through not being in a room with

an experimenter. Although it reduces the convenience for both the researcher and participant, future studies should consider the use of a video call to increase experimental control in online research.

8.5 Implications for Educational Practice

Although much of this thesis has focused on enhancing theoretical understanding of writing-to-learn from sources, the findings also carry valuable implications for educational practice. In particular are the benefits that closed-book source-based writing can have during study sessions for students in higher-education.

Writing from sources in a closed-book setting can be used as a useful study tool for students to engage in retrieval practice. Even though they may largely engage in knowledge-telling processes, this type of writing can still be useful for helping students to consolidate their knowledge, but more importantly to help them monitor their learning to identify gaps in their knowledge. This can enable students to take appropriate next steps when studying, such as identifying material that they need to revise again or providing an indication of when they are ready to be tested. From the point of view for teachers and other educators, writing from sources is a good way to assess how well students have learned a text. However, it is important that both objective tests, as well as how students perceive their own knowledge development, are considered when making such evaluations. The results here demonstrated that text comprehension tests may only capture one dimension of learning, and students' own assessments of their knowledge should also be considered in writing-to-learn assignments. This can assist with identifying particular areas where students need more targeted interventions.

To move beyond retrieval practice and knowledge-telling in source-based writing, educators could design instruction that prompts the synthesis of knowledge, including both cognitive processes (elaboration and organisation) as well as metacognitive processes (monitoring one's own comprehension). One way of doing this could be through learning journals (Nückles et al., 2020), but also through synthesis-based tasks which require students to compare and contrast information from multiple sources. These tasks can make source-based writing a very productive tool for learning. Synthesis writing tasks should be recommended as a tool for students to monitor their own learning from sources, but also encourage more elaborate writing processes which support learning.

8.6 Conclusion

Writing-to-learn has represented a shift in writing research from the *product* to the *process*. Whilst it is generally accepted that writing can enhance learning, the underlying cognitive processes of how this takes place are still under exploration. The dual-process model (Galbraith & Baaijen, 2018) places the focus on the writer's development of knowledge, which takes place through knowledge-transforming and knowledge-constituting. However, the claims of this model were limited to studies where writers produce texts based on their own knowledge. This thesis extends research on the dual-process model to a different type of writing common in educational settings: closed-book source-based writing.

The consistent finding across papers that participants experience a decrease in their subjective knowledge after writing about a source text demonstrates the role of retrieval practice in writing-to-learn. Engaging in retrieval through writing can not only help learners to remember information from sources but also help them to monitor their own learning and identify knowledge gaps. Therefore, contradictory to the assumptions of previous writing-to-learn theory, the knowledge-telling process can play a role in knowledge development. However, it is not just retrieval that is important in closed-book source-based writing; comparisons with non-writing retrieval activities (thinking and speaking) also showed favourable effects of writing. Furthermore, the decrease in subjective knowledge was not observed when participants undertook a synthesis writing task, implying these participants moved beyond simple knowledge-telling.

This thesis also considered different measures of learning from writing, and assessed how they relate to one another, which has not been done in previous writing-to-learn research. By including a range of outcome measures, this project demonstrates that there are multiple dimensions of learning from writing. Participants' subjective knowledge was significantly, although moderately, positively related to their scores on text comprehension tests (in particular to scores on SAQs). This highlights the value of measuring both objective test performance and participants subjective knowledge in writing-to-learn studies. By replicating the findings of Silva and Limongi (2019), a novel measure of long-term memory consolidation from writing was established (RT), and Paper 4 made a first attempt at implementing this measure in a more complex experimental design.

Appendix A Subjective Knowledge Rating Scale (Galbraith et al., 2023)

Based on the source text passage you read…

On a scale of 1 to 7, where 1 = very little and 7 = a great deal, please rate:

Item	**Subscale**
How much you feel you know about the topic	Understanding
How well you understand the topic	Understanding
How organised your thoughts about the topic are	Organisation
How well you could explain the topic	Understanding & Organisation
How clear your thoughts about the topic are	Organisation
How clear your interpretation of the topic is	Understanding
How coherent your thoughts about the topic are	Organisation
How structured your thoughts about the topic are	Organisation
How much you can make sense of the topic's issues	Understanding & Organisation
How clear the relationships between your ideas about the topic are	Organisation
How well you comprehend the topic's issues	Understanding
How well-ordered your thoughts about the topic are	Organisation

Appendix B Source Texts and Text Comprehension Test Questions (Papers 1, 2, and 4)

B.1 Solar Activity Source Text

Solar Activity

The clearest visible sign of solar activity are sunspots, which are visible spots on the sun that are darker than the surrounding area. The existence of sunspots has been known for a long time, since the largest ones can be seen with the naked eye by looking at the Sun through a suitably dense layer of fog. More precise observations became available beginning in the 17th century, when Galileo started to use the telescope for astronomical observations.

A sunspot looks like a ragged hole in the solar surface. By looking at spots near the edge of the sun's surface, called the solar disc, it can be seen that the spots are slightly depressed with respect to the rest of the surface. The surface temperature in a sunspot is about 1500 degree Kelvin below that of its surroundings, which explains the dark color of the spots. Akin to magnetic storms, spots erupt where various magnetic forces pierce the surface of the Sun.

The diameter of a typical sunspot is about 10,000 km, and its lifetime is from a few days to several months, depending on its size. The larger spots are more likely to be long-lived. Sunspots often occur in pairs or in larger groups. By following the motions of the spots, the period of rotation of the Sun can be determined.

The variations in the number of sunspots have been followed for almost 250 years. Evidently, the number of sunspots varies with an average period of 11 years. The actual period may be between 7 and 17 years. In the past decades, it has been about 10.5 years. Usually, the activity rises to its maximum in about 3 - 4 years and then falls off slightly more slowly. The period was first noted by Samuel Heinrich Schwabe in 1843. These variations in the number of sunspots have been fairly regular since the beginning of the 18th century. However, in the 17th century, there were long intervals when there were essentially no spots at all. This inactive period is called the Maunder minimum. The mechanism behind these irregular variations in solar activity is not yet understood.

Measured in tesla, magnetic fields in sunspots may be as large as .45 tesla. (For comparison, the magnetic field of the Earth is .03 milliteslas (mT).) The strong magnetic field inhibits heat transfer, which explains the lower temperature of the spots.

The periodic variation in the number of sunspots reflects a variation in the general solar magnetic field. As the cycle advances, the spots move closer to the sun's equator. Specifically, spots of the next cycle begin to appear while those of the old one are still present near the equator. This characteristic pattern in which the spots appear is known as the butterfly diagram. Spots belonging to the new cycle have a magnetic charge, called polarity, opposite to that of the old ones. Spots in the opposite hemispheres also have opposite polarity. Since the field is thus reversed between consecutive 11 year cycles, the complete period of solar magnetic activity is two such cycles.

The Sun shows several other types of surface activity: faculae and plages, prominences, and flares. The faculae and plages are local bright regions, and usually occur where new sunspots are forming and disappear when the spots disappear. Apparently, they are caused by the enhanced heating of the chromosphere, a layer of the sun's atmosphere, in strong magnetic fields. The areas surrounding sunspots are brighter than the normal surface of the Sun because of these two forms of solar activity.

The prominences are among the most spectacular solar phenomena. They are glowing gas masses in the corona—the plasma atmosphere of the Sun, easily observed near the edge of the Sun. The temperature of prominences is about 10,000 - 20,000 degree Kelvin. There are several types of prominences; a rare type of prominence called an eruptive prominence throws gas violently outwards. If eruptions like these are aimed at Earth, they can disturb the magnetosphere, the planet's magnetic field, with dramatic consequences. Past eruptions have knocked out satellites, wrecked television reception and caused power surges and blackouts.

The flare outbursts are among the most violent forms of solar activity. They appear as bright flashes, lasting from one second to just under an hour. In the flares, a large amount of energy stored in the magnetic field is suddenly released. The detailed mechanism is not yet known. Flares can be observed at all wavelengths. The hard X-ray emission of the Sun may increase hundredfold during a flare. The flares give rise to disturbances on Earth. The X-rays cause changes, which affect short-wave radio communications. Direct radio emission may disturb operation of radars and other devices. When they enter the Earth's magnetic field a few days after the outburst, the flare's particles give rise to strong auroras, glowing phenomena in the sky seen at polar latitudes.

B.2 Solar Activity Multiple-Choice Questions

*Note: correct answers are in **bold**.* (d) = direct/factual question. (i) = inferential question.

1. (d) All of the following are true about sunspots EXCEPT:
    a. They are slightly depressed with respect to the rest of the Sun's surface.

   b. **Once they are formed, they remain on the Sun's surface permanently.**
   c. The typical diameter of a sunspot is 10,000 km.
   d. They often occur in pairs or in larger groups.

2. (i) Sunspots can have the same polarity when:
   a. One belongs to a new cycle and the other belongs to the old cycle.
   b. They are in opposite hemispheres.
   c. They form as a pair
   d. Sunspots never have the same polarity.

3. (d) Prominences (glowing masses) occur in the:
   a. Corona (the plasma atmosphere of the Sun).
   b. Photosphere
   c. Chromospheres
   d. Sunspots on the solar surface.

4. (i) If the temperature surrounding a sunspot was 6,000 degrees Kelvin, what could be the approximate temperature of the sunspot?
   a. 8,500 K
   b. 4,500 K
   c. 0 K
   d. -1,500 K

5. (d) All of the following are correct about the *Maunder minimum* EXCEPT:
   a. It was a period in which there were essentially no sunspots.
   b. It occurred during the 17th century.
   c. It was an unusual deviation from the pattern of 11 year cycles of varying sunspot number.
   d. Sunspots occurred during this period but were notably smaller than typical sunspots.

6. (i) Based on their visual characteristics, with regards to temperature, the relationship between that of prominences and sunspots would best be described as which of the following:
   a. Prominences are lower in temperature than sunspots
   b. Prominences are the same temperature as sunspots
   c. Prominences are higher in temperature than sunspots
   d. It not clear what the relationship between prominences and sunspots is regarding their temperatures

7. (d) What is the size of the magnetic fields in sunspots?
   a. The magnetic fields are of negligible size.
   b. The magnetic fields are too large to measure.
   c. .03 mT.
   d. **.45 tesla.**

8. (I) If temporary difficulties with short-wave radio communications were being experienced on Earth, which solar activity might you expect to be occurring?
   a. A peak in the number of sunspots
   b. A solar flare
   c. An eruptive prominence
   d. A plage

B.3 Solar Activity Short-Answer Questions and Scoring Criteria

Pretend you are an amateur astronomer who likes to watch solar activity from your backyard.

Address the following problems—

Question 1. Strange things have been happening on Earth and the mayor of your town has come to you for help. She knows that eruptive prominences and solar flares are both types of violent solar activity and suspects one of these is causing all of the trouble. She wants you to describe how they are different and how we might be able to tell which of these is causing the problems on Earth. What are characteristics of each solar activity that could help you distinguish between them (describe 1 for each)? You begin to suspect eruptive prominences are the culprit. What strange activity might have led you to this conclusion (list 2)?

Score /4.

- 2 marks for outlining the differences between prominences and solar flares
- 1 mark for each 'strange activity' from eruptive prominences (2 marks available)

Possible answers:

- Eruptive prominences: Glowing gas masses in the corona; easily observed near the edge of the sun; temperature 10,000 - 20,000 degree Kelvin; throws gas violently outwards
- Solar flares: Bright flashes lasting from one second to just under an hour; can be observed at all wavelengths; hard x-ray emission of sun may increase hundredfold during flare; give rise to disturbances; affect short-wave radio communication
- Strange activity from eruptive prominences: Disturbance to magnetosphere – knocked out satellites; wrecked TV reception; power surges and blackouts

Question 2. As part of your hobby you take careful notes of what you observe, on a daily basis. Today you are observing the sun, and you compare your observations with those you included in your research notes two years ago. If the solar cycle has advanced, what two changes should you see today, as compared to what you noted two years ago?

Score /2.

- 1 mark for each change identified

Possible answers:

- Spots move closer to the sun's equator; spots of the next cycle begin to appear while those of the old one are still present near the equator (pattern known as butterfly diagram); spots belonging to new cycle have a magnetic charge (polarity), opposite to that of the old ones; faculae and plages usually appearing when new sunspots form and disappear when sunspots disappear

Question 3. You want to show a friend solar activity in the sky, but you do not have access to a telescope at the moment. Which of these solar activities (sunspots, faculae, eruptive prominences, solar flares) would they be most likely to be able to see? Please give two reasons to explain your answer.

Score /3.

- 1 mark for correctly naming the solar activity (sunspots)
- 1 mark per correct reason (2 marks available)

Possible answers:

- Sunspots – visible spots on the sun that are darker than surrounding area; existence known for a long time since largest ones can be seen with the naked eye by looking at the sun through a dense layer of fog
- Possibly marks for solar flares – can be observed at all wavelengths; when enter Earth's magnetic field a few days after outburst, the flare's particles give rise to strong auroras, glowing phenomena in the sky seen at polar latitudes

Question 4. Your younger brother was born on the first year of a solar activity period. At age 6, he's become quite the solar enthusiast, and has recently become fascinated with faculae. He's begged you to take him with you solar viewing the next time there's an abundance of faculae. Assuming this is an average period of solar magnetic activity, calculate how old your brother will be the next time there's an abundance of this type of solar activity. How do you know this?

Score /4.

- 1 mark for identifying correct age
- 3 marks for explanation

Possible answers:

- Abundance of faculae occurs when new sunspots are forming
- Number of sunspots varies with an average period of 11 years
- Activity rises to a maximum in about 3-4 years

B.4 Deserts and Savannas Source Text

Deserts and Savannas

Deserts and semideserts by most accountings cover about a third of the land on Earth. There is no uncomplicated way to define a desert. One of the simplest definitions is that a desert

receives less than about 25 cm (10 inches) of precipitation per year, but this is an imperfect standard because there are regions that receive much more than that but are undoubtedly deserts. Following the lead of Imanuel Noy-Meir, most biologists prefer a definition that emphasises two attributes of deserts that are of extreme importance for plants and animals. First, a desert is a place where precipitation is so low that availability of water exerts a dominant controlling effect on biological processes. Second, when precipitation occurs in deserts, it comes in infrequent, largely unpredictable events. Rains in deserts are highly irregular and unreliable in both time and space. One year may bring 5 times (or even 20 times) the rain of another. One 30-km2 area may be drenched during a storm while a nearby area of similar size receives nothing, because rains in deserts are usually produced by isolated storms rather than broad fronts. Large herbivores must often conduct their lives in ways that – more than anything – are opportunistic, taking advantage of rain or moisture whenever and wherever it occurs.

The classic look of the dry savannas is of endless plains of grass with trees dotted here and there. Taking a big-picture view of Earth, dry savannas are in many cases neighbouring environments to deserts. Gradients of moisture often occur on continental scales of space, with moist forests in some regions, deserts in others, and dry savannas between the two – in intermediate regions where there is more moisture than in the deserts, but far less than in the forests. In dry savannas – as in deserts – water is a dominant controlling factor for biological processes. This is true in a somewhat different way than deserts, however. Savannas usually have discrete rainy seasons interrupted by discrete rainless seasons. Because of the rainless season each year, although on an annual basis savannas receive more rain than deserts do and receive it more predictably, the plants and animals living in savannas must often endure profound drought for long lengths of time. The Serengeti plains of East Africa are examples of this second sort of water-controlled ecosystem. Routinely during the long dry season each year, no rain falls for 4-6 months; thus streams dry up, and soils become so parched that the grasses turn crisp and brown.

Deserts and dry savannas exist for several reasons. The single most important cause of desert and semidesert conditions on a planetary scale is the global pattern of air movements, whereby air warmed at the equator rises to high altitudes and displaces air at those altitudes in such a way that high-altitude air descends to Earth's surface somewhere else. Air at high altitudes, partly because it is cold, contains little moisture, even when saturated with water vapour. Thus, in regions where high-altitude air tends consistently to fall to low altitude, the land can become parched. The present global pattern is for high-altitude air to descend in two bands encircling the globe at latitudes roughly 30° north and 30° south of the equator. Most of the world's great deserts – including the largest, the Sahara – are products of this process. Another common but more localised cause of desert or savanna conditions is rain-shadowing caused by highlands. If

a region's only reliable source of moisture is winds blowing in from the ocean, and if intervening highlands force the winds to rise – so that the air cools and its moisture condenses to form rain or snow – the winds may have little moisture left by the time they blow into the region of interest. Near Los Angeles, for example, the coastal mountains force the prevailing winds blowing east from the Pacific Ocean to rise, causing mountain rain. Seen from above, the land still farther east, the Mojave Desert, looks much like a dry, sandy-coloured shadow cast by the mountains.

The existence and extent of deserts and dry savannas often depend on nothing more permanent than the gossamer movements of air. Accordingly, over the long reaches of geological time, deserts and savannas have come and gone as patterns of air movement have changed. The modern Sahara Desert, for example, has existed for less than 6000 years. Nonetheless, desert conditions have come and gone in North Africa for at least 7 million years. Currently the Sahara Desert is tending to expand north and south. Thus, for example, a recent consensus report by climate scientists concluded that the southern reaches of Europe (e.g., Italy and southern Spain) are likely to become far drier over the upcoming decades.

Despite the variability just discussed, deserts and savannas have probably always existed somewhere. Thus, they have long presented plants and animals with special challenges, and they have long presented opportunities for species that are able to make do with little water and tap the meagre water resources available.

B.5 Deserts and Savannas Multiple-Choice Questions

*Note: correct answers are in **bold**.* (d) = direct/factual question. (i) = inferential question.

1. (d) All of the following are true about deserts, except:
   a. Availability of water exerts a dominant controlling effect on biological processes
   b. Rains in deserts are usually produced by isolated storms
   c. Deserts cover about a third of the land on Earth
   d. A desert must be over 30° Celsius to be defined as a desert

2. (i) With regards to rainfall, the relationship between deserts and savannas would best be described by which of the following:
   a. Rainfall is more unpredictable in deserts than in savannas
   b. Rainfall is more unpredictable in savannas than in deserts
   c. Deserts receive less rain than savannas
   d. Deserts and savannas receive roughly the same amount of rain

3. (d) The most important cause of desert conditions is:
   a. Sandstorms
   b. Global pattern of air movements
   c. Extreme temperatures
   d. Drought

4. (i) If the quantity of rain declined over an extended period of time, which of the following would be the likely sequence of environmental change?
   a. Forest, desert, savanna
   b. Desert, savanna, forest

c. Savanna, forest, desert
d. Forest, savanna, desert

5. (d) During the long dry season in The Serengeti plains, no rain falls for what period?
   a. 3 years
   b. 2 weeks
   c. 4-6 months
   d. 5-7 days

6. (i) If the orbit of the Earth changed, so that the equator was 10° north of its current position, between roughly what latitudes would high-altitude air descend (relative to the current position of the equator)?
   a. 40° north, 20° south
   b. 10° north, 30° south
   c. 20° north, 40° south
   d. 50° north, 10° south

7. (d) The modern Sahara Desert has existed for how long?
   a. The modern Sahara Desert has always existed
   b. Less than 6000 years
   c. 6 million years
   d. We don't know how long the modern Sahara Desert has existed

8. (i) The addax antelope, native to the Sahara Desert, rarely need to drink since they are able to get most of the water they need from the plants they eat. Why is this a good adaptation to desert conditions?
   a. There is an abundance of plants in the desert that the addax antelope can eat
   b. Animals in the desert can only get water from plants
   c. Rain is unpredictable in the desert
   d. There are long periods of drought in the desert

B.6 Deserts and Savannas Short-Answer Questions and Scoring Criteria

Question 1. The Atacama Desert in Chile is the driest nonpolar desert in the world. The most arid region of the Atacama Desert is situated between two mountain chains, the Andes and the Chilean Coast Range. How could this mountainous environment have led to the formation of the Atacama Desert?

Score /4.

- 1 mark for naming the correct process (rain shadowing)
- 3 marks for the explanation

Possible answers:

- Most likely cause is rain shadowing caused by highlands/mountains
- Only reliable source of moisture is winds blowing in from the ocean and the maintain ranges force the winds to rise, so the air cools and moisture condenses to form rain or snow

- This leaves little moisture by the time the winds reach the region of the Atacama Desert

Question 2. The conditions in deserts and savannas present challenges for the plants and animals that live there. For example, baobab trees have adapted to be able to store water between the bark and meat of the tree that they can sip on during a drought. Based on this characteristic, is the baobab tree found in deserts or savannas? Why?

Score /3.

- 1 mark for correct answer (savannas)
- 2 marks for explanation

Possible answers:

- Savannas have discrete rainy seasons interrupted by discrete rainless seasons
- Rainless seasons mean plants and animals must endure long periods of drought
- Baobab tree has adapted by storing water during rainy season so they can cope during drought

Question 3. Provide two reasons why the effect of global patterns of air movements on high altitude air lead to deserts and dry savannas.

Score /2.

- 1 mark for each correct reason (2 marks available)

Possible answers:

- Hot air rises from equator to high altitudes where cold air at those altitudes is displaced and air descends to earth somewhere else
- Air at high altitudes is cold and contains little moisture – where high altitude air falls to low altitude the land can become parched

Question 4. Compared to other deserts, The Sonoran Desert in Arizona can receive a relatively high amount of precipitation per year, at up to 15 inches. Why would it be problematic to define it as a desert on this fact alone? What other two attributes of deserts enable the Sonoran Desert to be defined as a desert?

Score /4.

- 1 mark for the simple definition of a desert
- 1 mark for the reason why this is problematic in the case of the Sonoran Desert

- 1 mark for each attribute that enable the Sonoran Desert to be defined as a desert (2 marks available)

Possible answers:

- Simple definition of a desert: less than 25cm / 10 inches of rain
- Problematic definition because the Sonoran Desert receives more rain than this
- Attributes of a desert: precipitation is so low that availability of water exerts dominant controlling effect on biological processes; when precipitation occurs, it is infrequent and unpredictable (unreliable in both time and space)

B.7 Psychoanalysis Source Text

Psychoanalysis

The general theoretical basis of psychoanalysis is Freud's personality theory. The model of personality assumed in this theory is based on conflict. Three personality instances are supposed to collide with each other: id, ego, and super-ego. Id represents the realm of impulses and needs, its energetic aspect supposedly consisting of an overall striving for pleasure called "libidinous energy". Super-ego stands for the rules and prohibitions imposed by society. It aims at restricting the power of Id or, respectively, tries to direct the libidinous energy of id into channels that are in keeping with social standards. Ego is responsible for the interaction with the external world and for conscious behaviour control. Its nature is determined – among other things – by the needs coming into consciousness from the id and by the social standards emanating from super-ego. Ego tries to create a balance between the two opposing sides. If, in the course of psychological development very strong libidinous impulses are confronted with unrelenting prohibitions on the part of super-ego, ego may fail to manage its mediating function. In this case, the needs threatening to overthrow the balance are repressed from consciousness. They may, however, continue to influence behaviour by reappearing in the guise of other needs.

The basic therapeutic principle of psychoanalysis consists in removing the psychological complex at the bottom of any neurosis. This is done via the therapist who helps bring back into consciousness the original early experiences repressed into the unconscious and to relive them. Due to the importance attached to experiences of early childhood, psychoanalysis often concentrates on very remote events of the biography. The re-activation of suppressed experiences is supposed to offer a possibility of repeating the originally unsuccessful conflict of ego with id and with super-ego in a more constructive manner thus leading to the resolution of the psychological complex. Removal of the unconscious origins leads at the same time to the disappearance of the problematic neurotic behaviours.

Neurotic behaviour according to psychoanalysis is the result of repression. Its roots usually reach back into early childhood, during which "traumatic" (i.e., psychologically wounding) experiences occurred necessitating the repression of libidinous impulses and thus also increasing susceptibility for later traumatic experiences. The consequence of repression is the so-called psychological complex residing in the subconscious and being the actual cause of the neurotic disorder. This complex constitutes the pool of energy responsible for the symptomatic behaviour and the persistence of the neurosis. For this reason, psychoanalysis aims at removing this psychological complex.

Psychoanalysis considers neurotic problem behaviour as a symptom of a deeper unconscious psychological complex. For psychoanalysis symptoms represent the visible discharges of some hidden unconscious causes. It is only by removing these causes that the symptoms may be brought to extinction. According to psychoanalysis, a therapy that merely attempts to remove the symptoms may be quite successful for a limited period of time. But because the actual causes have not been removed relapses occur after a while or it leads to "symptom shifts", that is, the appearance of new symptoms caused by the same psychological complex.

Psychoanalysis belongs in the main to a hermeneutic interpretive tradition of science. It attempts to interpret the verbal statements of the patient on his/her experiences and phantasies as a way of understanding the developmental history of the neurotic disorder in retrospect. Psychoanalysis does not evaluate itself by a concept of science based on experiment and prediction. Because its objective is to treat the overall personality (that means taking into account a greater number of interacting factors) and because the psychoanalytical treatment usually lasts over several years, individual therapy goals can scarcely be determined in advance. For this reason, too, it is difficult to objectively evaluate therapeutic success.

B.8 Psychoanalysis Multiple-Choice Questions

*Note: Correct answers are in **bold**.*

Direct/Factual Multiple-Choice Questions

1. What is the basic therapeutic principle of psychoanalysis?
    a) Treating symptoms with medication
    b) Removing the psychological complex at the bottom of any neurosis
    c) Teaching coping mechanisms
    d) Cognitive restructuring

2. According to psychoanalysis, neurotic behaviour is the result of what?
    a) Repression
    b) Environmental stimuli
    c) Genetic predisposition
    d) Symptom shifts

3. How long does psychoanalytical treatment usually last?
    a) A few weeks
    b) Several years
    c) A lifetime
    d) One or two days

4. In Freud's personality theory, what is the model based on?
    a) Alignment with social norms
    b) Self-actualisation
    c) Conflict
    d) Cognition and learning

Inferential Multiple-Choice Questions

1. Which of the following is the best description of the Ego in Freud's personality theory?
    a) The Ego is the source of our impulses and needs
    b) The Ego is responsible for enforcing social norms and rules
    c) The Ego is the mediator between the Id and the Super-ego
    d) The Ego is the unconscious part of the mind

2. Which of the following would a psychoanalyst aim to encourage in the process of treating a patient with neurosis?
    a) The patient recalling emotionally charged experiences from early childhood
    b) The patient describing their recent day-to-day experiences
    c) The patient recalling only their happiest memories
    d) The patient describing their imagined future self

3. Why does psychoanalysis emphasise addressing the unconscious causes of neurotic behaviour?
    a) Symptoms are often temporary and may disappear on their own
    b) Treating only symptoms can lead to relapses or new symptoms appearing
    c) Symptoms are not indicative of deeper psychological issues
    d) Unconscious causes are easier to address than overt symptoms

4. According to the principles of psychoanalysis, which statement best characterises the role of the psychological complex in neurosis?
    a) It represents a repository of unresolved conflicts and distressing memories that continue to influence the individual's behaviour
    b) It serves as a defence mechanism that protects the conscious mind from unbearable anxiety or guilt
    c) It manifests as a physical or biological abnormality that directly causes neurotic symptoms
    d) It represents a learned pattern of behaviour that persists due to reinforcement or habit

B.9 Psychoanalysis Short-Answer Questions and Scoring Criteria

Question 1. A patient has been diagnosed with generalised anxiety disorder. According to the principles of psychoanalysis, how could the patient be treated?

Score /4.

- 2 marks for treatment principles of psychoanalysis

- 2 marks for application to generalised anxiety disorder

Possible answers:

- Removing psychological complex at the bottom of the disorder. Therapist helps bring back into consciousness original early experiences repressed into unconscious and to relive them.
- Concentrating on very remote events of biography.
- Reactivating suppressed experiences supposed to offer possibility of repeating originally unsuccessful conflict of EGO with ID and with SUPER-EGO in a more constructive manner.
- Removal of unconscious origins.
- Therapy targets overall personality, taking into account a greater number of interacting factors, and treatment usually lasts several years.
- Relate this to the specific example of generalised anxiety disorder. e.g., could suggest therapist brings back experiences of when the patient felt anxiety.

Question 2. Describe the three personality instances in personality theory and how these are interrelated.

Score /4.

- 2 marks for defining the personality instances (ID, EGO, SUPEREGO)
- 2 marks for explaining how they are interrelated

Possible answers:

- Three instances are ID, EGO and SUPER-EGO, include brief description of each.
    - ID – impulses and needs, energetic aspect consisting of overall striving for pleasure ('libidinous energy')
    - Super-ego – rules and prohibitions imposed by society; aims at restricting power of ID or tries to direct libidinous energy of ID into channels keeping within social standards
    - Ego – interaction with external world and conscious behaviour control; nature determined by needs coming into consciousness from ID and by social standards from super-ego
- EGO nature is determined by needs coming into consciousness from ID and social standards from SUPER-EGO. EGO tries to create balance between the two opposing sides.

Question 3. Based on psychoanalysis theory, explain the distinction between causes and symptoms and their role in maintaining neuroses.

Score /4.

- 2 marks for distinction between causes and symptoms
- 2 marks for role in maintaining neuroses

Possible answers:

- Symptoms represent viable discharges of some hidden unconscious causes and by removing these causes, symptoms may be brought to extinction
- Therapy that merely attempts to remove symptoms may be successful for a limited amount of time, but because the actual causes have not been removed relapses occur after a while or it leads to 'symptom shifts' (appearance of new symptoms caused by same psychological complex)

Question 4. What relation does psychoanalysis have to traditional experimental science, and how does this affect its evaluation of neurotic disorders?

Score /3.

Possible answers:

- Psychoanalysis mainly belongs to a hermeneutic interpretive tradition of science
- Psychoanalysis does not evaluate itself by concept of science based on experiment and prediction
- Individual therapy goals can rarely be determined in advance – because objective of psychoanalysis is to therapy the overall personality and because treatment usually lasts over several years
- Difficult to objectively evaluate therapeutic success

B.10 Behaviour Therapy Source Text

Behaviour Therapy

The general theoretical basis of behaviour therapy is constituted by behaviouristic learning theories. In these theories learning is considered as a process, by which the behaviour of an organism (humans, animals) changes as a function of an environmental situation. They proceed on the assumption that behaviour is determined by environmental cues: connections are supposed to exist between the stimuli of the environment (S) and the response of the organism

(R). Learning occurs through changing these S-R connections. This change may be induced either by classical or operant conditioning. Classical conditioning presupposes the existence of some reflex-like S-R connections. Another stimulus S+ which so far has been “neutral” is presented together with stimulus S. After several repetitions S+, too, can elicit the response R. In other words, there has been created a new connection S+-R. In operant conditioning the organism finds itself in a new situation (i.e., stimulus condition S), which it does not consider optimal. The organism reacts to S by trying out several responses from its behavioural repertory. If one response R has a successful outcome, this increases the probability that in the future the response R will occur under condition S.

The fundamental principle of treatment in behaviour therapy is to resolve engrained inadequate S-R connections and to replace them by more adequate ones with the help of classical or operant conditioning. In the case of behaviours conflicting with the social standards of the culture in question, the therapist has the function of destroying the S-R connections at the basis of these behaviours and encouraging alternative behaviours that cause the extinction of the previous problem behaviour. If a desirable behaviour is merely lacking in specific situations, he will have to develop corresponding ways of acting for this situation step-by-step. Treatment is confined to the behavioural habits exhibited by the patient in the present and dispenses with a far-reaching analysis of earlier experiences.

A neurotic disorder is considered by behaviour therapy as a maladapted behaviour that has its origins in a learning process. Behaviour therapy proceeds on the assumption that “neurotic” behaviour is basically acquired along the same principles as “normal” behaviour and that both kinds of behaviour are amenable to change via learning. Accordingly, the principles of learning theory may offer an explanation for the conditions of origin as well as suggestions for the corresponding therapy methods. Therefore, behaviour therapy may be considered as the application of the principles of classical and operant conditioning from learning theory to the realm of neurotic disorders.

According to behaviour therapy, neurotic problem behaviour does not represent a symptom based on some deeper unconscious cause in the psyche of man. A neurotic symptom is simply considered as a maladapted behavioural habit resulting from erroneous learning. Behaviour therapy does not know of any neurosis “behind” the symptom, merely the neurotic symptom itself: if the symptom is removed, the neurosis has disappeared, too. “Therapy of the symptom” of this kind results in continuous improvement without danger of relapse or return of the “same” neurosis in the guise of other symptoms, that is, of symptom shifts.

Behaviour therapy lies in the line of a scientific tradition based on experiment and prediction and is rather sceptical towards hermeneutic interpretative approaches. On the basis of its

fundamental learning theories, behaviour therapy, may by means of deduction, comes to statements that are empirically testable. Because behaviour therapy confines itself to the extinction of problem behaviour and, hence, a well-defined objective, and because, furthermore, the therapy does not take very long, the goals of the therapy are easy to determine beforehand. This also creates the possibility of assessing successfulness with a fair degree of objectivity.

B.11 Behaviour Therapy Multiple-Choice Questions

Direct/Factual Multiple-Choice Questions

1. According to behaviour therapy, how is behaviour determined?
   a) Environmental cues
   b) Free will
   c) Psyche of an individual
   d) Genetic make-up of an individual

2. What is the fundamental principle of treatment in behaviour therapy?
   a) To identify early childhood trauma
   b) To analyse the unconscious
   c) To resolve inadequate stimulus-response connections
   d) To investigate family history

3. What does behaviour therapy's 'therapy of the symptom' result in?
   a) Increasing the risk of relapse
   b) Continuous improvement without danger of relapse or symptom shifts
   c) Emergence of new symptoms
   d) Temporary release of symptoms followed by symptom shifts

4. What is behaviour therapy sceptical towards?
   a) Philosophical traditions
   b) Scientific tradition
   c) Reward based learning
   d) Hermeneutic interpretative approaches

Inferential Multiple-Choice Questions

5. Which of the following is the best explanation for why behaviour therapy is effective?
   a) Because it helps people identify and understand the root causes of their problems
   b) Because it helps people to change their maladaptive behaviours through conditioning
   c) Because it provides people with support and encouragement to cope with their problems
   d) Because it helps people to develop a deeper understanding of their unconscious mind

6. According to behaviour theories, what is the most likely implication of the assumption that neurotic behaviour is acquired along the same principles as normal behaviour?
   a) Neurotic behaviour is a sign of a serious mental illness that requires medication
   b) Neurotic behaviour is a reflection of a person's personality and cannot be changed
   c) Neurotic behaviour can be treated using the same principles of learning that are used to treat new behaviours

d) Neurotic behaviour is a result of a person's early childhood experiences

7. What is the main difference between classical and operant conditioning?
   a) Classical conditioning involves pairing a stimulus with a response, while operant conditioning involves pairing a response with a consequence
   b) Classical conditioning is more effective for long-term learning, while operant conditioning is more effective for short-term learning
   c) Classical conditioning involves learning a new skill, while operant conditioning involves unlearning an old skill
   d) Classical conditioning is a type of associative learning, while operant conditioning is a type of cognitive learning

8. According to behaviouristic learning theories, why might neurotic problem behaviour be considered maladaptive?
   a) Neurotic problem behaviour makes individuals feel unhappy
   b) Neurotic problem behaviour makes individuals physically unwell
   c) Neurotic problem behaviour reflects an unhealthy brain
   d) Neurotic problem behaviour conflicts with cultural norms

B.12 Behaviour Therapy Short-Answer Questions and Scoring Criteria

Question 1. A patient has a phobia of spiders. According to the principles of behaviour therapy, how could the patient be treated?

Score /4.

- 2 marks for treatment principles of behaviour therapy
- 2 marks for application to phobia of spiders

Possible answers:

- Principle of therapy is to resolve engrained inadequate S-R connections and replace them by more adequate ones with the help of classical or operant conditioning
- Therapist has function of destroying S-R connections at basis of these behaviours and encouraging alternative behaviours that cause extinction of previous problem behaviour
- 'Normal' and 'neurotic' behaviour are amenable to change via learning
- Example of how phobia could be 'cured' through classical or operant conditioning

Question 2. How does behaviour theory view symptoms and what is the implication of this for treatment?

Score /4.

- 2 marks for behaviour therapy view of symptoms
- 2 marks for implications for treatment

Possible answers:

- Neurotic problem behaviour does not represent a symptom based on some deeper unconscious cause – a neurotic symptom is simply considered as a maladapted behavioural habit resulting from erroneous learning
- If symptom is removed, the neurosis has disappeared to
- 'Therapy of the symptom'. -continuous improvement without danger of relapse or return of 'same' neurosis in guise of symptom shifts

Question 3. Explain the two ways in which learning occurs according to behaviour theories.

Score /4.

- 2 marks for each explanation of classical and operant conditioning

Possible answers:

- Behaviour determined by environmental cues – connections exist between environmental stimuli (S) and response of the organism (R) (S-R connections)
- Classical conditioning presents a neutral stimulus alongside stimulus S and eventually the neutral stimulus elicits the response
- In operant conditioning, the organism responds to a new S by trying several responses, and if one has a successful outcome, this increases probably that in future the response R will occur under condition S

Question 4. What relation does behaviour therapy have to traditional experimental science, and how does this affect its evaluation of neurotic disorders?

Score /3.

Possible answers:

- Scientific tradition based on experiment and prediction
- Behaviour therapy comes to statements that are empirically testable
- Goals of therapy easy to determine beforehand and assessment of success can be made with fair degree of objectivity

Appendix C Source Texts and Memory Task Words (Paper 3)

C.1 Text 1 – Bats

Bats have fascinating characteristics related to their genetics, way of life, and survival skills. In terms of genetics, all existing bat species belong to the animal kingdom of mammals, but with the additional and unique ability of flight. In terms of their way of life, they are animals that sleep hanging – upside down – from tree branches, or rock protrusions in caves. Within the variety of existing bat species, for example, insect-eating bats are surprisingly beneficial to livestock and arable farming. Insect-eating bats feed on parasites which damage harvests and spread disease to farm animals. All bats are capable of hearing insects walk, or the beating of their wings. However, the most surprising survival skill of these flying mammals is echolocation (a system of movement and reference to bodies which detects the spread of sound waves). Bats in full flight emit very high frequency sounds, which hit bodies in their path and rebound to the bats, marking the location of their prey. Echolocation in bats is like a kind of radar, which prevents them from colliding with trees and rocks in the natural environment because, despite having eyes, bats cannot see.

In-Text Memory Task Words	Non-Text Memory Task Words
bats	vampire
characteristics	typology
mammals	viviparous
sound	boom
prey	loot
farm	car
system	airplane
waves	radio
location	computer
harvests	bike
animals	book
insects	glasses
pests	aspirin

In-Text Memory Task Words	Non-Text Memory Task Words
bodies	banana
radar	fruit

C.2 Text 2 – Jargon

The study of jargon (or slang) – which refers to vocabulary and specific expressions used by members of a particular social group, such as lawyers, teenagers, teachers, or builders – proves to be beneficial in establishing differences in the dynamics of social relationships, and in the use of language adapted to certain communicational contexts. In this sense, the main benefit of using jargon is to facilitate a particular system of communication that is only explicit amongst members of the group. As such, there is no need to explain the use of a term or complete expression, since both have been learned at an opportune moment. A second benefit of the use of jargon lies in the bonds established between members of the group. These bonds are sometimes so strong that they alienate people unconnected to the group's activities or the jargon that it uses. Additionally, jargon allows groups to maintain an identity, and project an image of what they do. Finally, the use of jargon gives group members a sense of belonging and strengthens their self-esteem through the simple fact of feeling accepted as equals by their peers.

In-Text Memory Task Words	Non-Text Memory Task Words
vocabulary	speech
use	routine
beneficial	helps
study	thesis
identity	crowd
jargon	pencil
group	draft
expressions	notebook
teenagers	frame
people	screen
lawyers	keyboard
system	cards
image	alcohol
sense	medicine

In-Text Memory Task Words	Non-Text Memory Task Words
self-esteem	ointment

C.3 Text 3 – Forgetting

What is forgetting, and how does it happen? The definition of forgetting implies an interference with information as it enters our memory. Forgetting, therefore, is an inability to recover the information we need, at the moment that we need it. According to psychological research, forgetting can happen in two ways: proactively and retroactively. Proactive interference inhibits our ability to recall new information. For example, if you have to set a new facebook password annually, you might have difficulty recalling it in the first few months, because the previous password still holds validity in your mind. That is to say, you keep recalling the information of your old password when you attempt to log in to your account every day. However, in retroactive interference, the opposite happens: new information inhibits our ability to remember old information. But in both cases, we forget because the new and old information compete to establish themselves in our system of memory. So, when we try to remember any kind of information, we experience a conflict - or interference - commonly called forgetting.

In-Text Memory Task Words	Non-Text Memory Task Words
forgetting	neglect
interference	cross
information	greetings
ability	expertise
research	search
system	specialists
recall	fossils
cases	ideas
password	dinosaur
account	birds
conflict	evidence
moment	series
facebook	mould
mind	fowl
memory	key

C.4 Text 4 – Theory of Crime

For many years, researchers of various disciplines have debated the reasons why a person commits crime. Today, therefore, there are many theories accounting for the possible causes of criminal conduct. The primary, and most popular theory is that of sociocultural origin, which assumes that the celebration of violence in our culture motivates people to adopt violent behaviour. This thesis argues that television and cinema are responsible for celebrating forms of violence, and that this primarily influences young people. The second theory holds that crime is a consequence of social and economic inequalities. However, if social classes did not exist, this theory would disappear because the poor would not steal from – or kill – the rich in order to obtain coveted objects. A third explanation relates criminal actions to upbringing and family life. Children that receive a poor and inadequate family education could end up becoming criminals. In the early stages of a child's life, parents are role models for their children to follow. The final theory emphasises biological factors. According to this theory, both genetics and severe malnutrition play an important part. Researchers have talked of certain differences in brain structure between criminals and people who do not commit crime. Children of homicidal parents could inherit this gene. Severe malnutrition could also alter the development of the human brain. What is certain is that there is no absolute evidence that proves or disproves any of these theories. Our reasoning can lead us to believe that any one of them is valid.

In-Text Memory Task Words	**Non-Text Memory Task Words**
theory	answer
crime	offense
causes	root
evidence	proof
popular	public
people	release
behaviour	science
violence	philosophy
culture	mathematics
upbringing	women
children	reading
parents	writing
classes	house
genetics	achievement

In-Text Memory Task Words	Non-Text Memory Task Words
brain	horse

C.5 Text 5 – Mummification

In order to mummify a cadaver, the ancient Egyptians first washed the body in water from the Nile river. Next, they removed all bodily organs except the heart. It was thought that the heart hosted a person's soul. After removing the organs, the Egyptians bathed the body's cavities in wine to perfume it, and filled it with any available waste material that would help maintain its shape as a human figure. The cadaver, which they covered in salt, was then placed under the sun for 35 or 40 days in order to dry out. After this time, the whole process was repeated, replacing bandages, salts, and filling material. Next, the Egyptians closed the incisions with bandages soaked in a new rubbery liquid, derived from a plant which grew near the waters of the Nile. The embalming process was thus complete. Then, a mask (either a likeness of the dead person or of an Egyptian deity) was placed on the head of the cadaver. Finally, the mummy – buried together with food and personal effects – was placed inside a hand-carved sarcophagus. The Egyptians believed that in the next life, the mummy would be able to use all the wealth it had accumulated.

In-Text Memory Task Words	Non-Text Memory Task Words
mummy	dispossession
incisions	wounds
bandages	ties
organs	members
sarcophagus	tomb
egyptian	key
body	handbag
material	backpack
figure	board
cadaver	school
objects	dance
filling	play
mask	beach
heart	mountain
plant	snow

C.6 Text 6 – The Great Wall of China

The Great Wall of China is a monumental work of civil engineering, loaded with history and renowned as the world's largest tourist attraction. With a length of 4500 kilometres, the height of five men, and a width capable of accommodating six horses running side by side without bumping into each other, the Great Wall of China is a construction so big that it can be seen from space. In terms of the details of its construction, the wall was built in four parts. Each part is made from the same material of compressed earth and stones. The Great Wall of China follows mountain slopes with a gradient of 70 degrees. For this reason, walking along the wall is very difficult due to the unevenness of the steps. Today, The Great Wall of China is a tourist attraction with 2500 years of history, which attracts many visitors. Local residents say that tourists are keen to make the huge effort of climbing the wall in order to understand the majesty of its size, and the history that it represents.

In-Text Memory Task Words	**Non-Text Memory Task Words**
wall	enclosed
construction	foundation
attraction	fascination
tourists	traveller
mountain	peaks
work	printer
world	paper
horses	door
stones	street
slopes	glass
history	napkin
earth	fork
steps	meal
gradient	building
width	elevator

C.7 Text 7 – Cosmetic Surgery

Cosmetic surgery presents itself these days as an easy procedure, painless and without any risk to the health of men and women. To such an extent that it is now very common to associate cosmetic surgery with basic procedures like face lifts, figure-shaping massages, or the tattoing

of eyebrows, eyes and lips. However, cosmetic surgery comes with risks. These include adverse reactions to anaesthetic, excessive bleeding, and postoperative infections. Muscles and nerves that are the target of a surgical procedure could be damaged during the surgery, and so the patient could be at risk of suffering paralysis in these areas. There is also the possibility that the patient's own body could reject the implant that is being inserted (for example, a breast or gluteus prosthesis). Even if the patient recovers satisfactorily from the surgery, there is a risk that the results are not pleasing. In this case, it is very likely that additional surgery will be required. Finally, a reasonably successful outcome to the cosmetic surgery could even encourage the person to undergo further cosmetic operations, due to the positive reinforcement of their improved self-image. What is certain is that undertaking cosmetic surgery to achieve a more contoured nose, enlarged breasts, or other body part, does not ensure happiness, good friendships, or true love.

In-Text Memory Task Words	**Non-Text Memory Task Words**
surgery	amputation
risks	conflict
results	referral
implant	substitutes
infections	contagious
lips	elephant
prosthesis	dynamite
nose	chequebook
friendships	geometry
massages	telephone
tattooing	tortoise
paralysis	president
bleeding	congress
happiness	university
health	kitchen

C.8 Text 8 – Post-Traumatic Stress

After experiencing a traumatic event, some people may continue reliving this experience time after time. Any stimulus could cause the person to relive the horror they experienced. For example, the person could experience a flashback, that is, as if they were watching their own horror film, but as the protagonist, instead of just sitting in the front row. During a flashback, the

person could feel an uncontrolled anxiety, that culminates in fainting, which is the body's own method of physiological protection against imminent emotional and mental damage. This entire collection of symptoms is a mood disorder described as post-traumatic stress. Its description dates back to the comprehensive treatment (medical and psychological) that war veterans received. However, nowadays, any person could suffer from it, for example: victims of rape, child abuse, violent crime, or industrial accidents, like the recent tragedy in the Amuay oil refinery in 2012. Other causes of post-traumatic stress for a person could be natural disasters like earthquakes, tsunamis, tornados, fires and, in general, accidents involving means of transport.

In-Text Memory Task Words	**Non-Text Memory Task Words**
stress	anguish
event	fact
horror	startle
symptoms	signals
protagonist	star
experience	lettuce
stimulus	turkey
person	pig
collection	ball
flashback	street
disorder	truck
mental	seller
victims	pharmacy
abuse	market
tragedy	tomato

C.9 Practice Text – Pets

Physical contact with a pet is beneficial to elderly people, having been clinically proven to extend life expectancy. Recent clinical studies have shown that walking a dog lowers the blood pressure of elderly people, who are very likely to be retired and living alone. Owners will often talk to their pets, and this communication helps them keep mentally active. Pet ownership entails a complete routine of activities (feeding, walks, bathing, medical assistance), which demand the owner's complete attention and memory. We also know that a pet depends upon the care provided by its owner for its survival, and that being needed by another living creature

gives life a purpose. Finally, owning a pet encourages the formation of relationships with other people in the same position of responsibility for the wellbeing of their pets.

In-Text Memory Task Words	**Non-Text Memory Task Words**
pet	animals
contact	union
care	guidance
purpose	hope
routine	automatic
owner	colours
feeding	pasta
walking	teeth
creature	shampoo
position	product
elderly	client
dog	sale
pressure	culture
people	society
bathing	apple

Appendix D Memory Task Words (Paper 4)

Psychoanalysis

In-Text Memory Task Words	Non-Text Memory Task Words
traumatic	gallery
instances	domain
society	health
needs	lady
conflict	victory
repression	window
impulses	creation
energy	director
personality	complicated
consciousness	universal
appearance	exhibition
problem	times
therapeutic	sister
fantasies	ambition
libidinous	detail

Behaviour Therapy

In-Text Memory Task Words	Non-Text Memory Task Words
maladapted	characters
process	evidence
humans	muscles
cues	visit
situation	date
operant	soldier
reflex	dog
danger	pair
interpretative	unknown
response	television

In-Text Memory Task Words	Non-Text Memory Task Words
connections	organic
classical	tension
culture	primary
erroneous	partner
application	majority

Appendix E Instructions for Writing Activities

E.1 Essay and Free Recall Instructions (Papers 1 and 2)

Essay

Solar Activity

"Write a draft essay describing the different types of solar activity, including their properties, their relationships with one another, and their effects on Earth. Be clear, detailed, and thorough, so that a high-school student who has not read the text could understand. Your essay should have an introduction and a clear thesis, and you should make sure to back up your points with supporting details. You will have 15 minutes to write your essay."

Deserts and Savannas

"Write a draft essay describing deserts and savannas, including their properties, their relationships with one another, and their effects on plants and animals. Be clear, detailed, and thorough, so that a high-school student who has not read the text could understand. Your essay should have an introduction and a clear thesis, and you should make sure to back up your points with supporting details. You will have 15 minutes to write your essay."

Free Recall

"Recall and write down as close to word for word as possible, everything you can from the text passage you read. If you can't remember the exact words, just try to express roughly what the text said. You will have 15 minutes."

Mental Essay

Solar Activity

"Think about how you would write a draft essay describing the different types of solar activity, including their properties, their relationships with one another, and their effects on Earth. You will have 15 minutes to think about your essay. Please continue to look at the screen whilst you are thinking. At one-minute intervals, you will be prompted on screen to write down the first word that comes to mind, and you will have 15 seconds to write the word."

Deserts and Savannas

“Think about how you would write a draft essay describing deserts and savannas, including their properties, their relationships with one another, and their effects on plants and animals. You will have 15 minutes to think about your essay. Please continue to look at the screen whilst you are thinking. At one-minute intervals, you will be prompted on screen to write down the first word that comes to mind, and you will have 15 seconds to write the word.”

Mental Recall

“Think about recalling, as close to word for word as possible, everything you can from the text passage you read. If you can't remember the exact words, just try to think roughly what the text said. Please continue to look at the screen whilst you are thinking. At one-minute intervals, you will be prompted on screen to write down the first word that comes to mind, and you will have 15 seconds to write the word.”

E.2 Synthesis and Free Recall Instructions (Paper 4)

Synthesis

“Write a draft essay comparing and contrasting the two different modes of therapy for neuroses (psychoanalysis and behaviour therapy) and conclude with an overall evaluation.”

Free Recall

“Recall and write down, as close to word for word as possible, everything you can remember from the text passages you read.”

Appendix F Ethics Approval

F.1 Paper 1

Approved by Faculty Ethics Committee - ERGO II 53981.A1

E ERGOII <ERGOII@soton.ac.uk>
To: Peters A.
Wednesday 1 July 2020 at 17:17

Approved by Faculty Ethics Committee - ERGO II 53981.A1

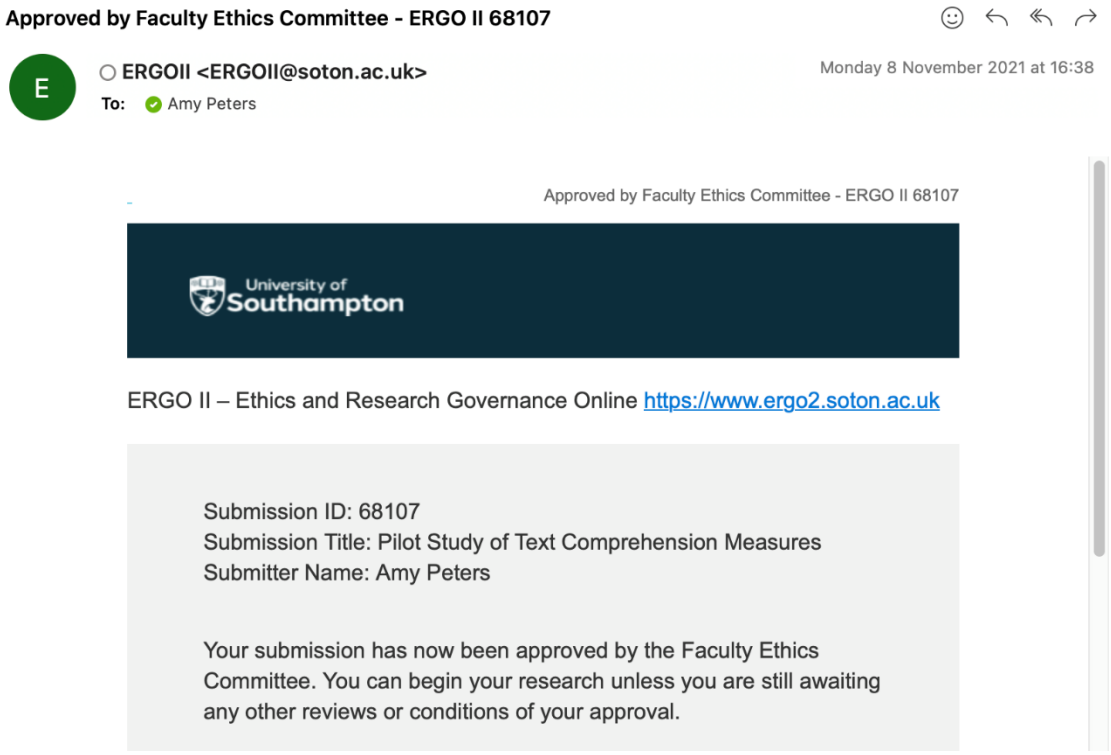
University of Southampton

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 53981.A1
Submission Title: Exploring the learning processes involved during writing. (Amendment 1)
Submitter Name: Amy Peters

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

F.2 Paper 2, Experiment 1



Approved by Faculty Ethics Committee - ERGO II 68107

E ERGOII <ERGOII@soton.ac.uk>
To: Amy Peters
Monday 8 November 2021 at 16:38

Approved by Faculty Ethics Committee - ERGO II 68107

University of Southampton

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 68107
Submission Title: Pilot Study of Text Comprehension Measures
Submitter Name: Amy Peters

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

F.3 Paper 2, Experiment 2

Approved by Faculty Ethics Committee - ERGO II 70438

E ERGOII <ERGOII@soton.ac.uk>
To: Amy Peters

Monday 24 January 2022 at 15:06

Approved by Faculty Ethics Committee - ERGO II 70438

University of Southampton

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 70438
Submission Title: Exploring the learning processes involved during writing: follow-up study.
Submitter Name: Amy Peters

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

F.4 Paper 3

Approved by Faculty Ethics Committee - ERGO II 70416

E ERGOII <ERGOII@soton.ac.uk>
To: Amy Peters

Monday 24 January 2022 at 12:28

Approved by Faculty Ethics Committee - ERGO II 70416

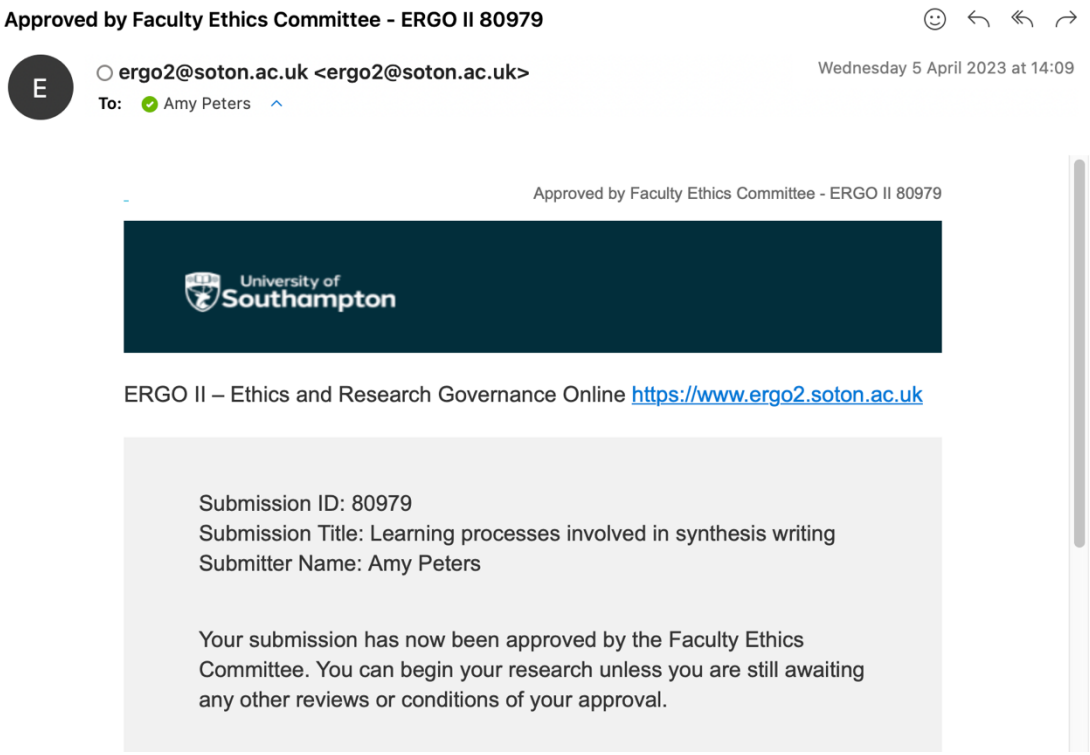
University of Southampton

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 70416
Submission Title: Using a mental-chronometry measure to explore the effects of summarising on reading comprehension
Submitter Name: Amy Peters

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

F.5 Paper 4



Approved by Faculty Ethics Committee - ERGO II 80979

E

ergo2@soton.ac.uk <ergo2@soton.ac.uk>

To: Amy Peters

Wednesday 5 April 2023 at 14:09

Approved by Faculty Ethics Committee - ERGO II 80979

University of Southampton

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 80979
Submission Title: Learning processes involved in synthesis writing
Submitter Name: Amy Peters

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

Appendix G Supplementary Materials – Paper 1

G.1 Descriptive Statistics – Text Comprehension

Table 19. Means (and standard deviations) of text comprehension test scores by planning and writing condition (Paper 1).

			Writing condition	
			Essay	Free Recall
Planning condition	Outline	Total MCQ	5.00 (1.87)	4.54 (1.63)
		Direct MCQ	2.85 (0.99)	2.54 (0.99)
		Inferential MCQ	2.15 (1.21)	2.00 (1.17)
		Short answer	4.42 (2.68)	3.73 (2.12)
	Synthetic	Total MCQ	5.00 (1.84)	4.91 (1.41)
		Direct MCQ	3.00 (1.17)	2.86 (1.17)
		Inferential MCQ	2.00 (1.17)	2.05 (0.79)
		Short answer	4.71 (2.01)	3.89 (2.48)

G.2 Descriptive Statistics – Subjective Knowledge

Table 20. Means (and standard deviations) of subjective knowledge ratings by planning and writing condition (Paper 1).

			Writing condition	
			Essay	Free Recall
Planning condition	Outline	Pre-writing knowledge	4.17 (0.93)	3.44 (1.02)
		Post-writing knowledge	3.19 (0.95)	2.70 (1.08)
		Two-days later knowledge	3.01 (0.98)	2.14 (0.84)
	Synthetic	Pre-writing knowledge	3.87 (1.14)	3.61 (1.17)
		Post-writing knowledge	3.20 (1.19)	3.25 (1.29)
		Two-days later knowledge	2.72 (0.95)	2.87 (1.30)

Table 21. Means (and standard deviations) of subjective knowledge ratings by knowledge subscale (Paper 1).

	Type of Knowledge		
	Overall	Organisation	Understanding
Pre-writing	3.70 (1.09)	3.65 (1.15)	3.78 (1.13)
Post-writing	3.04 (1.15)	3.05 (1.18)	3.00 (1.14)
After two-days	2.62 (1.07)	2.54 (1.05)	2.73 (1.12)

Appendix H Supplementary Materials – Paper 2

H.1 Descriptive Statistics – Text Comprehension – Experiment 1

Table 22. Means (and standard deviations) of text comprehension test scores by text and condition (Paper 2, Experiment 1).

				Condition		
			Overall	No text	Text absent	Text present
Text	Overall	MCQ	4.48 (1.94)	3.05 (1.55)	4.63 (1.65)	5.66 (1.61)
		SAQ	4.31 (3.09)	1.79 (1.37)	3.80 (2.40)	6.95 (2.63)
	Solar activity	MCQ	4.29 (2.06)	2.72 (1.61)	4.44 (1.79)	5.56 (1.69)
		SAQ	4.63 (3.45)	1.37 (1.66)	4.31 (2.86)	7.21 (2.73)
	Deserts and savannas	MCQ	4.69 (1.78)	3.40 (1.43)	4.82 (1.51)	5.77 (1.54)
		SAQ	3.99 (2.65)	2.14 (0.99)	3.23 (1.64)	6.62 (2.54)

H.2 Descriptive Statistics – Text Comprehension – Experiment 2

Table 23. Means (and standard deviations) of text comprehension scores (percentage correct) by text and condition (Paper 2, Experiment 2).

			Condition			
			Written essay	Mental essay	Written recall	Mental recall
Text	Overall	MCQ	55.00 (16.54)	50.86 (19.74)	52.42 (21.02)	54.44 (17.84)
		MCQ Direct	66.00 (23.80)	58.62 (27.77)	59.68 (26.36)	65.32 (27.14)
		MCQ Inferential	44.00 (14.93)	43.10 (23.99)	45.16 (29.87)	43.55 (20.38)
		SAQ	27.46 (12.81)	18.44 (14.30)	27.11 (16.98)	23.01 (12.19)
	Solar Activity	MCQ	50.00 (18.54)	52.78 (22.91)	51.79 (24.93)	56.25 (19.36)
		MCQ Direct	56.82 (25.23)	58.33 (30.92)	55.36 (26.27)	68.75 (26.61)
		MCQ Inferential	43.18 (16.17)	47.22 (28.30)	48.21 (31.72)	43.75 (26.61)
		SAQ	22.90 (11.16)	16.45 (12.61)	24.72 (16.05)	22.00 (11.98)
	Deserts & Savannas	MCQ	58.93 (14.23)	47.73 (13.48)	52.94 (17.97)	52.50 (16.50)
		MCQ Direct	73.21 (20.72)	59.09 (23.11)	63.23 (26.69)	61.67 (28.14)
		MCQ Inferential	44.64 (14.47)	36.36 (13.06)	42.65 (29.00)	43.33 (11.44)
		SAQ	31.04 (13.26)	21.68 (16.86)	29.07 (17.95)	24.10 (12.75)

H.3 Descriptive Statistics – Subjective Knowledge – Experiment 2

Table 24. Means (and standard deviations) of participants' subjective knowledge ratings at each time point by text and 'writing' condition (Paper 2, Experiment 2).

			Condition			
		Subjective Knowledge	Written essay	Mental essay	Written recall	Mental recall
Text	Solar Activity	Pre-Writing	2.79 (0.59)	3.33 (1.13)	3.88 (0.79)	3.78 (0.91)
		Post-Writing	2.25 (1.02)	2.96 (1.11)	2.56 (1.16)	3.25 (0.89)
		Two-Days Later	1.73 (0.67)	2.35 (0.95)	1.95 (1.06)	2.56 (0.89)
	Deserts & Savannas	Pre-Writing	3.99 (0.99)	3.91 (0.87)	3.14 (1.22)	3.75 (1.24)
		Post-Writing	3.05 (1.35)	3.91 (1.17)	2.39 (1.22)	3.24 (1.27)
		Two-Days Later	2.82 (0.70)	2.76 (1.11)	2.04 (1.01)	2.94 (1.25)

Appendix I Supplementary Materials – Paper 4

I.1 Subject Background of Participants

Subject Area	Subject	Number of Participants	Total
Social Sciences	Education and Psychology	13	28
	Education	6	
	Psychology	2	
	Criminology and Psychology	2	
	Criminology	1	
	Accounting and Finance	1	
	Economics and Finance	1	
	Marketing	1	
	Law	1	
Engineering	Engineering	3	3
Medicine and Nursing	Nursing	2	3
	Medicine	1	
Physical Sciences	Biochemistry	1	3
	Biology	1	
	Environmental Science	1	
Humanities	History	1	3
	Modern History and Politics	1	
	English and Politics	1	

I.2 Descriptive Statistics – Subjective Knowledge

Table 25. Means (and standard deviations) for subjective knowledge ratings at each time point by writing condition and knowledge rating subscale (Paper 4).

		Writing Condition		
		Overall	Synthesis	Free Recall
Prior Knowledge	Overall	2.23 (1.15)	2.30 (1.00)	2.17 (1.31)
	Organisation	2.21 (1.13)	2.25 (0.96)	2.17 (1.31)
	Understanding	2.28 (1.21)	2.35 (1.07)	2.21 (1.37)
Pre-Writing	Overall	4.20 (1.13)	4.55 (1.09)	3.85 (1.09)
	Organisation	4.13 (1.15)	4.44 (1.11)	3.82 (1.12)
	Understanding	4.32 (1.17)	4.72 (1.13)	3.92 (1.10)
Post-Writing	Overall	3.67 (1.57)	4.55 (1.20)	2.78 (1.40)
	Organisation	3.63 (1.63)	4.54 (1.23)	2.71 (1.47)
	Understanding	3.73 (1.52)	4.60 (1.18)	2.87 (1.33)

I.3 Descriptive Statistics – Text Comprehension

Table 26. Means (and standard deviations) of text comprehension scores (percentage correct) by writing condition and source text (Paper 4).

			Writing condition		
			Overall	Synthesis	Free Recall
Text	Overall	Overall MCQ	76.88 (17.05)	72.50 (18.18)	81.25 (15.04)
		Direct MCQ	74.06 (18.86)	71.88 (18.97)	76.25 (18.98)
		Inferential MCQ	79.69 (19.55)	73.13 (20.79)	86.25 (16.17)
		SAQ	37.98 (17.81)	40.75 (18.84)	35.21 (16.73)
	Behaviour Therapy	Overall MCQ	74.06 (21.81)	69.38 (21.64)	78.75 (21.50)
		Direct MCQ	70.63 (25.25)	66.25 (23.33)	75.00 (26.90)
		Inferential MCQ	77.50 (25.82)	72.50 (27.98)	82.50 (23.08)
		SAQ	33.75 (21.08)	35.33 (21.40)	32.17 (21.19)
	Psychoanalysis	Overall MCQ	79.69 (15.42)	75.63 (16.46)	83.75 (13.51)
		Direct MCQ	77.50 (19.45)	77.50 (19.70)	77.50 (19.70)
		Inferential MCQ	81.88 (21.92)	73.75 (22.18)	90.00 (18.85)
		SAQ	42.21 (16.68)	46.17 (17.95)	38.25 (14.68)

I.4 Descriptive Statistics – Reaction Time

Table 27. Means (and standard deviations) of reaction times (ms) by condition and accuracy (Paper 4).

		Response Accuracy		
		Overall	Correct	Incorrect
Condition	Overall	1100.89 (355.54)	1067.33 (343.72)	1207.53 (371.49)
	Synthesis	1068.99 (351.02)	1036.06 (332.23)	1174.54 (387.76)
	Free Recall	1131.85 (357.34)	1097.80 (352.11)	1239.14 (353.14)

I.5 HDDM Parameter Estimates

Table 28. Parameter estimates of the winning HDDM at group and participant level (Paper 4).

Parameter	a		v		t (recall)		t (synthesis)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Group	1.600	0.031	0.767	0.040	0.576	0.025	0.534	0.025
Participant 1	1.628	0.106	0.908	0.144	-	-	0.722	0.029
Participant 2	1.498	0.097	0.833	0.133	0.621	0.022	-	-
Participant 3	1.714	0.110	0.891	0.135	-	-	0.578	0.029
Participant 4	1.636	0.088	0.738	0.113	0.529	0.020	-	-
Participant 5	1.570	0.092	0.722	0.126	-	-	0.731	0.026
Participant 6	1.648	0.090	0.666	0.119	0.611	0.024	-	-
Participant 7	1.228	0.098	0.888	0.146	-	-	0.556	0.017
Participant 8	1.569	0.080	0.769	0.113	0.482	0.018	-	-
Participant 9	1.624	0.099	0.810	0.128	-	-	0.741	0.029
Participant 10	1.529	0.082	0.669	0.122	0.623	0.021	-	-

Parameter	a		v		t (recall)		t (synthesis)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 11	1.591	0.097	0.798	0.123	-	-	0.678	0.027
Participant 12	1.652	0.095	0.696	0.123	0.586	0.022	-	-
Participant 13	1.731	0.102	0.769	0.116	-	-	0.582	0.023
Participant 14	1.577	0.093	0.874	0.136	0.622	0.021	-	-
Participant 15	1.609	0.090	0.660	0.124	-	-	0.590	0.022
Participant 16	1.620	0.087	0.670	0.125	0.519	0.024	-	-
Participant 17	1.633	0.093	0.754	0.118	-	-	0.582	0.244
Participant 18	1.658	0.091	0.740	0.111	0.589	0.024	-	-
Participant 19	1.566	0.081	0.660	0.121	-	-	0.488	0.019
Participant 20	1.653	0.098	0.789	0.123	0.577	0.026	-	-
Participant 21	1.694	0.114	0.894	0.146	-	-	0.428	0.027
Participant 22	1.686	0.093	0.726	0.116	0.497	0.023	-	-
Participant 23	1.487	0.085	0.758	0.122	-	-	0.474	0.017
Participant 24	1.761	0.118	0.910	0.137	0.643	0.031	-	-

Parameter	a		v		t (recall)		t (synthesis)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 25	1.630	0.091	0.880	0.129	-	-	0.507	0.021
Participant 26	1.539	0.089	0.729	0.123	0.584	0.024	-	-
Participant 27	1.612	0.093	0.751	0.121	-	-	0.579	0.023
Participant 28	1.736	0.119	0.767	0.125	0.641	0.035	-	-
Participant 29	1.380	0.082	0.762	0.124	-	-	0.381	0.016
Participant 30	1.651	0.092	0.712	0.115	0.530	0.023	-	-
Participant 31	1.674	0.084	0.614	0.122	-	-	0.253	0.013
Participant 32	1.545	0.083	0.710	0.116	0.494	0.020	-	-
Participant 33	1.487	0.087	0.792	0.123	-	-	0.528	0.019
Participant 34	1.673	0.091	0.845	0.118	0.474	0.021	-	-
Participant 35	1.603	0.087	0.764	0.120	-	-	0.444	0.022
Participant 36	1.602	0.081	0.603	0.130	0.514	0.018	-	-
Participant 37	1.668	0.100	0.779	0.120	-	-	0.545	0.028
Participant 38	1.589	0.089	0.663	0.127	-	-	0.503	0.022

Parameter	a		v		t (recall)		t (synthesis)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Participant 39	1.516	0.080	0.774	0.120	0.468	0.016	-	-
Participant 40	1.494	0.091	0.947	0.152	0.776	0.022	-	-

List of References

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science*, *2*(3), 153–157. https://doi.org/10.1111/j.1467-9280.1991.tb00123.x

Ackerman, J. M. (1993). The promise of writing to learn. *Written Communication*, *10*(3), 334–370. https://doi.org/10.1177/0741088393010003002

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, *17*, 286–303. https://doi.org/10.1016/j.learninstruc.2007.02.004

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Kluwer Academic Publishers.

Alves, R. A., Castro, S. L., & Olive, T. (2008). Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill. *International Journal of Psychology, 43*(6), 969-979. https://doi.org/10.1080/00207590701398951

Anders, R., Riès, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57–73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 802–814. https://doi.org/10.1037/0278-7393.15.5.802

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, *54*(4), 577–596.

Arnold, K. M., Umanath, S., Thio, K., Reilly, W. B., McDaniel, M. A., & Marsh, E. J. (2017). Understanding the cognitive processes involved in writing to learn. *Journal of Experimental Psychology: Applied*, *23*(2), 115–127. https://doi.org/10.1037/xap0000119

Arnold, K. M., Eliseev, E. Drew., Stone, A. R., McDaniel, M. A., & Marsh, E. J. (2021). Two routes to the same place: Learning from quick closed-book essays versus open-book essays. *Journal of Cognitive Psychology*. https://doi.org/10.1080/20445911.2021.1903011

Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (2016). A diffusion model analysis of episodic recognition in preclinical individuals with a family history for Alzheimer's disease: The adult children study. *Neuropsychology*, *30*(2), 225–238. https://doi.org/10.1037/neu0000222

Atasoy, S. (2013). Effect of writing-to-learn strategy on undergraduates' conceptual understanding of electrostatics. *Asia-Pacific Education Researcher*, *22*(4), 593–602. https://doi.org/10.1007/s40299-013-0062-4

Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction*, *36*(3), 199–223. https://doi.org/10.1080/07370008.2018.1456431

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, *29*(3), 246–277. https://doi.org/10.1177/0741088312451108

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2014). Effects of writing beliefs and planning on writing performance. *Learning and Instruction*, *33*, 81–91. https://doi.org/10.1016/j.learninstruc.2014.04.001

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Balota, D. A., Yap, M. J., Cortese, M. J., Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59*(4), 495-523. https://doi.org/10.1016/j.jml.2007.10.004

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20*(3), 160-166. https://doi.org/10.1177/0963721411408885

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29–58. https://doi.org/10.3102/00346543074001029

Barenberg, J., & Dutke, S. (2019). Testing and metacognition: retrieval practise effects on metacognitive monitoring in learning from text. *Memory*, *27*(3), 269–279. https://doi.org/10.1080/09658211.2018.1506481

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of Multiple Source Use* (pp. 99–116). Routledge.

Beattie, V., Collins, B., & McInnes, B. (1997). Deep and surface learning: a simple or simplistic dichotomy? *Accounting Education*, *6*(1), 1–12.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68.

Bereiter, C. (1980). Development in writing. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 73–93). Lawrence Erlbaum Associates.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Routledge.

Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 96–114). The Guilford Press.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123–144). Lawrence Erlbaum Associates.

Bjork, R. A. (1998). Assessing our own competence: Heuristics and illusions. In *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). MIT Press.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*(3), 849–858. https://doi.org/10.1037/a0035934

Bourdin, B., & Fayol, M. (2002). Even in adults, written production is still more costly than oral production. *International Journal of Psychology*, *37*(4), 219–227. https://doi.org/10.1080/00207590244000070

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities*. London: Macmillan.

Brown, A. L., & Day, J. D. (1983). Macrorules for summarising texts: The development of expertise. *Journal of Verbal Learning and Verbal Behaviour*, *22*, 1–14. https://doi.org/10.1016/S0022-5371(83)80002-4

Brown, A. L., Day, J. D., & Jones, R. S. (1983). The development of plans for summarizing texts. *Child Development, 54*(4), 968-979. https://doi.org/10.2307/1129901

Bruning, R. H., & Kauffman, D. F. (2016). Self-efficacy beliefs and motivation in writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 160–173). The Guilford Press.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*(5), 279–283. https://doi.org/10.1177/0963721412452728

Carretti, B., Borella, E., Cornoldi, C., & de Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246–251. https://doi.org/10.1016/j.lindif.2008.10.002

Castells, N., Minguela, M., & Nadal, E. (2022). Writing a synthesis versus reading: strategies involved and impact on comprehension. *Reading and Writing*, *36*, 849–880. https://doi.org/10.1007/s11145-022-10341-y

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, *20*(1), 99–118. https://doi.org/10.1177/0741088303253572

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182. https://doi.org/10.1016/0364-0213(89)90002-5

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477. https://doi.org/10.1016/0364-0213(94)90016-7

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*(2), 120–131.

Cooper, R. P., & Peebles, D. (2015). Beyond single-level accounts: The role of cognitive architectures in cognitive scientific explanation. *Topics in Cognitive Science*, *7*(2), 243–258. https://doi.org/10.1111/tops.12132

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *The Journal of the Learning Sciences*, *6*(4), 347–365. https://doi.org/10.1207/s15327809jls0604_1

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Conijn, R., Roeser, J. & van Zaanen, M. Understanding the keystroke log: the effect of writing task on keystroke features. *Reading and Writing, 32*, 2353-2374. https://doi.org/10.1007/s11145-019-09953-8

Cordeiro, C., Castro, S. L., & Limpo, T. (2018). Examining potential sources of gender differences in writing: The role of handwriting fluency and self-efficacy beliefs. *Written Communication, 35*(4), 448-473. https://doi.org/10.1177/0741088318788843

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, *10*, 1–8. https://doi.org/10.1080/15434303.2011.622016

Cumming, A., Lai, C., & Cho, H. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes*, *23*, 47–58. https://doi.org/10.1016/j.jeap.2016.06.002

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Delaney, Y. A. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, *7*(3), 140–150. https://doi.org/10.1016/j.jeap.2008.04.001

Dinsmore, D. L., & Alexander, P. A. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review*, *24*, 499–567. https://doi.org/10.1007/s10648-012-9198-7

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 887–915. https://doi.org/10.1037/xlm0000352

Doctorow, M., Wittrock, M. C., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology*, *70*(2), 109–118.

Drabick, D. A. G., Weisberg, R., Paul, L., & Bubier, J. L. (2007). Keeping it short and sweet: Brief, ungraded writing assignments facilitate learning. *Teaching of Psychology*, *34*(3), 172–176. https://doi.org/10.1080/00986280701498558

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, *28*(2), 122–128.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data*. MIT Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3

Eysenck, M. W., & Keane, M. T. (2015). *Cognitive psychology: A student's handbook* (7th ed.). Psychology Press. https://doi.org/10.4324/9781315778006

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London: SAGE Publications.

Fiorella, L., & Mayer, R. E. (2015). Eight ways to promote generative learning. *Educational Psychology Review*, *28*, 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, *35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 31–50). Lawrence Erlbaum Associates.

Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600

Friend, R. (2002). Summing it up. *The Science Teacher*, *69*(4), 40–43.

Fry, S. W., & Villagomez, A. (2012). Writing to learn: Benefits and limitations. *College Teaching*, *60*, 170–175. https://doi.org/10.1080/87567555.2012.697081

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, *21*, 45–72.

Galbraith, D. (1996). Self-monitoring, discovery through writing and individual differences in drafting strategy. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, Models, and Methodology in Writing Research* (pp. 121–141). Amsterdam University Press.

Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 139–160). Amsterdam University Press.

Galbraith, D., Ford, S., Walker, G., & Ford, J. (2005). The contribution of different components of working memory to knowledge transformation during writing. *L1-Educational Studies in Language and Literature*, *5*, 113–145. https://doi.org/10.1007/s10674-005-0119-2

Galbraith, D., Hallam, J., Olive, T., & le Bigot, N. (2009). The role of different components of working memory in writing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*(31), 3028–3033.

Galbraith, D., & Baaijen, V. M. (2018). The work of writing: Raiding the inarticulate. *Educational Psychologist*, *53*(4), 238–257. https://doi.org/10.1080/00461520.2018.1505515

Galbraith, D., & Baaijen, V. M. (in press). Managing spontaneity: A dual-process perspective on creative writing.

Galbraith, D., Peters, A., Hall, S., & Baaijen, V. (2023). Measuring discovery through writing. In R. Horowitz (Ed.), *The Routledge International Handbook of Research on Writing* (2nd ed., pp. 573–584). Routledge.

Galbraith, D., Baaijen, V., Peters, A., & Hall, S. M. (2025). Writing to learn. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (Third Edition). Guilford Press.

Galbraith, D., van Waes, L., & Torrance, M. (2007). Introduction. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and Cognition: Research and Applications* (pp. 1–10). Emerald Group Publishing Limited.

Gelati, C., Galvan, N., & Boscolo, P. (2014). Summary writing as a tool for improving the comprehension of expository texts: An intervention study in a primary school. In P. D.

Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Eds.), *Writing as a learning activity* (pp. 191–216). Brill.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*(3), 157-173. https://doi.org/10.1016/j.cedpsych.2009.11.002

Gingerich, K. J., Bugg, J. M., Doe, S. R., Rowland, C. A., Richards, T. L., Tompkins, S. A., & McDaniel, M. A. (2014). Active processing via write-to-learn assignments: Learning and retention benefits in introductory psychology. *Teaching of Psychology*, *41*(4), 303–308. https://doi.org/10.1177/0098628314549701

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702–718. https://doi.org/10.1037/0278-7393.11.1-4.702

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, *15*(1), 84–93. https://doi.org/10.3758/BF03197714

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602. https://doi.org/10.3758/BF03202442

Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*(3), 392–399.

Gould, J. D. (1980). Experiments on composing letters: Some facts, some myths, and some observations. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 97–127). Lawrence Erlbaum Associates.

Grabowski, J. (1996). Writing and speaking: Common grounds and differences toward a regulation theory of written language production. In M. C. Levy & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 73–91). Lawrence Erbaum Associates.

Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In G. Rijlaarsdam, M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and Cognition: Research and Applications* (Vol. 20, pp. 165–179). Elsevier.

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, *81*(4), 710–744. https://doi.org/10.17763/haer.81.4.t2k0m13756113566

Graham, S., Kiuhara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, *90*(2), 179–226. https://doi.org/10.3102/0034654320914744

Graham S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. https://doi.org/10.1111/tops.12142

Gunel, M., Hand, B., & McDermott, M. A. (2009). Writing for different audiences: Effects on high-school students' conceptual understanding of biology. *Learning and Instruction*, *19*, 354–367. https://doi.org/10.1016/j.learninstruc.2008.07.001

Hall, S. M. (2023). *Estimating the effects of writing beliefs, writing processes and drafting strategies on the development of subjective understanding*. [Doctoral thesis, University of Southampton].

Hall, S. M., Baaijen, V. M., & Galbraith, D. (2024). Constructing theoretically informed measures of pause duration in experimentally manipulated writing . *Reading and Writing*, *37*, 329–357. https://doi.org/10.1007/s11145-022-10284-4

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 1–27). Lawrence Erlbaum Associates.

Hayes, J. R. (2006). New directions in writing theory. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 28–40). The Guilford Press.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hayes, J. R., & Flower, L. S. (1980). Identifying the organisation of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3–30). Lawrence Erlbaum Associates.

Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist*, *41*(10), 1106–1113.

Hebert, M., Gillespie, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing*, *26*, 111–138. https://doi.org/10.1007/s11145-012-9386-3

Hebert, M., Graham, S., Rigby-Wills, H., & Ganson, K. (2014). Effects of note-taking and extended writing on expository text comprehension: Who benefits? *Learning Disabilities: A Contemporary Journal*, *12*(1), 43–68.

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 157–183. https://doi.org/10.1037/xlm0000819

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*(4), 473–493. https://doi.org/10.3102/00346543056004473

Higham, P. A., Fastrich, G. M., Potts, R., Murayama, K., Pickering, J. S., & Hadwin, J. A. (2023). Spaced retrieval practice: Can restudying trump retrieval? *Educational Psychology Review*, *35*(98), 1–41. https://doi.org/10.1007/s10648-023-09809-2

Higham, P. A., Zengel, B., Bartlett, L., & Hadwin, J. A. (2021). The benefits of successive relearning on multiple learning outcomes. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000693

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, *69*, 151–164. https://doi.org/10.1016/j.jml.2013.03.002

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heijltjes, A., & van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology*, *44*, 95–106. https://doi.org/10.1016/j.cedpsych.2016.02.005

Hudson, P. T. W., & Bergman, M. W. (1985). Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. *Journal of Memory and Language*, *24*(1), 46–58. https://doi.org/10.1016/0749-596X(85)90015-4

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001Getrightsandcontent

Janssen, D., van Waes, L., & van den Bergh, H. (1996). Effects of thinking aloud on writing processes. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 233–250). Lawrence Erlbaum Associates.

Johnson, J., Madill, A., Koutsopoulou, G. Z., Brown, C., & Harris, R. (2020, July 24). *Tackling gender imbalance in psychology*. The Psychologist (The British Psychological Society). https://www.bps.org.uk/psychologist/tackling-gender-imbalance-psychology

Johnson, D. R., Murphy, M. P., & Messer, R. M. (2016). Reflecting on explanatory ability: A mechanism for detecting gaps in causal knowledge. *Journal of Experimental Psychology: General, 145*(5), 573–588. https://doi.org/10.1037/xge0000161

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*(4), 469–486. https://doi.org/10.1037/a0017341

Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, *21*(3), 157–163. https://doi.org/10.1177/0963721412443552

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English, 20*(2), 121-140. https://doi.org/10.58680/rte198615612

Keil, F. C. (2003). Folkscience: Coarse intepretations of a complex reality. *Trends in Cognitive Sciences*, *7*(8), 368–373. https://doi.org/10.1016/S1364-6613(03)00158-X

Keil, F. C., & Wilson, R. A. (2000). *Explanation and cognition*. MIT Press.

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 355–365.

Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, *103*(3), 327–342.

Kellogg, R. T. (1994). *The psychology of writing*. Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 57–71). Lawrence Erlbaum Associates.

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research*, *5*(2), 159–190. https://doi.org/10.17239/jowr-2013.05.02.1

Kiewra, K. A. (2019). Soaring to successful synthesis writing. *Journal of Writing Research*, *11*(1), 163–209. https://doi.org/10.17239/jowr-2019.11.01.06

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*(3), 203–270.

Klein, P. D. (2004). Constructing scientific explanations through writing. *Instructional Science*, *32*, 191–231. https://doi.org/10.1023/B:TRUC.0000024189.74263.bd

Klein, P. D., & Boscolo, P. (2016). Trends in research on writing as a learning activity. *Journal of Writing Research*, *7*(3), 311–350. https://doi.org/10.17239/jowr-2016.07.3.01

Klein, P. D., Boscolo, P., Gelati, C., & Kirkpatrick, L. C. (2014). Introduction: New directions in writing as a learning activity. In P. Klein, P. Boscolo, L. Kirkpatrick, & C. Gelati (Eds.), *Writing as a Learning Activity* (pp. 1–14). Brill.

Klein, P. D., Haug, K. N., & Arcon, N. (2017). The effects of rhetorical and content subgoals on writing and learning. *The Journal of Experimental Education*, *85*(2), 291–308.

Klein, P. D., Piacente-Cimini, S., & Williams, L. A. (2007). The role of writing in learning from analogies. *Learning and Instruction*, *17*(6), 595–611. https://doi.org/10.1016/j.learninstruc.2007.09.006

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311–333. https://doi.org/10.1037/0096-3445.124.3.311

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences, 20*(7), 512-534.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora

and megastudies. *Quarterly Journal of Experimental Psychology*, *66*(3), 563–580. https://doi.org/10.1080/17470218.2012.658820

Leech, G., Rayson, P., Wilson, A. (2001). *Word Frequencies in Written and Spoken English*. London: Routledge.

Le Bigot, L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research*, *39*(4), 445-470. https://doi.org/10.1080/10862960701675317

Le Bigot, N., Passerault, J.-M., & Olive, T. (2008). Memory for words location in writing. *Psychological Research*, *73*, 89–97. https://doi.org/10.1007/s00426-008-0135-9

Leggette, H. R., Rutherford, T., Dunsford, D., & Costello, L. (2015). A review and evaluation of prominent theories of writing. *Journal of Applied Communications*, *99*(3). https://doi.org/10.4148/1051-0834.1056

Leijten, M., & van Waes, L. (2006). Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment. In K. Sullivan & E. Lindgren (Eds.), *Computer Key-Stroke Logging and Writing: Methods and Applications* (pp. 73–93). Elsevier.

Leijten, M., & van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyse and visualise writing processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Leijen, M., van Waes, L., & Ransdell, S. (2010). Correcting text production errors: Isolating the effects of writing mode from error span, input mode, and lexicality. *Written Communication*, *27*(2), 189–227. https://doi.org/10.1177/0741088309359139

Levy, C. M., Marek, J. P., & Lea, J. (1996). Concurrent and retrospective protocols in writing research. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, Models, and Methodology in Writing Research* (pp. 542–556). Amsterdam University Press.

Lindgren, E., Westum, A., Outakoski, H., & Sullivan, K. (2019). Revising at the Leading Edge: Shaping Ideas or Clearing Up Noise. In E. Lindgren & K. Sullivan (Eds.), *Observing Writing* (pp. 346–365). Brill.

Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition*, *43*, 85–98. https://doi.org/10.3758/s13421-014-0453-7

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.

Markman, E. M. (1979). Realising that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, *50*(3), 643–655. https://doi.org/10.2307/1128929

Mateos, M., Solé, I., Martín, E., Cuevas, I., Miras, M., & Castells, N. (2014). Writing a synthesis from multiple sources as a learning activity. In P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Eds.), *Writing as a learning activity* (pp. 169–190). Brill.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*(1), 13–23. https://doi.org/10.1207/S15326985EP3501_3

McGinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, *27*(3), 226–248. https://doi.org/10.2307/747793

Meade, G., Grainger, J., & Declerck, M. (2021). Friend or foe? Flankers reverse the direction of orthographic neighbourhood effects. *Language, Cognition and Neuroscience*, *36*(5), 535–542. https://doi.org/10.1080/23273798.2020.1849750

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2021). Word frequency effects on judgments of learning: More than just beliefs. *The Journal of General Psychology*, *148*(2), 124–148. https://doi.org/10.1080/00221309.2019.1706073

Meyer, D., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234.

Mills, C. M., & Keil, F. C. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology*, *87*, 1–32. https://doi.org/10.1016/j.jecp.2003.09.003

Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The value of using tests in education as tools for learning - not just for assessment. *Educational Psychology Review*, *35*(89), 1–21. https://doi.org/10.1007/s10648-023-09808-3

Myers, C. E., Interian, A., & Moustafa, A. A. (2022). A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1039172

Nelson, N., & King, J. R. (2022). Discourse synthesis: Textual transformations in writing from sources. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10243-5

Newell, G. E. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English, 18*(3), 265-287. https://doi.org/10.58680/rte198415670

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524. https://doi.org/10.1016/j.tics.2013.08.003

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, *19*(3), 259–271. https://doi.org/10.1016/j.learninstruc.2008.05.002

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimise cognitive load in self-regulated learning. *Educational Psychology Review*, *32*, 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

Olive, T. (2004). Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist*, *9*(1), 32–42. https://doi.org/10.1027/1016-9040.9.1.32

Olive, T. (2010). Methods, techniques, and tools for the on-line study of the writing process. In N. L. Mertens (Ed.), *Writing: Processes, Tools and Techniques* (pp. 1–18). Nova Science Publishers, Inc.

Olive, T. (2012). Working memory in writing. In V. W. Berninger (Ed.), *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology* (pp. 485–503). Psychology Press. https://doi.org/10.4324/9780203805312

Olive, T. (2022). How do novice and skilled writers engage working memory? In J. W. Schwieter & Z. (Edward) Wen (Eds.), *The Cambridge Handbook of Working Memory and Language* (pp. 504–526). Cambridge University Press.

Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. *European Journal of Cognitive Psychology, 21*(5), 758-785. https://doi.org/10.1080/09541440802079850

Olive, T., & Passerault, J.-M. (2012). The visuospatial dimension of writing. *Written Communication*, *29*(3), 326–344. https://doi.org/10.1177/0741088312451111

Olive, T., & Piolat, A. (2002). Suppressing visual feedback in written composition: Effects on processing demands and coordination of the writing processes. *International Journal of Psychology*, *37*(4), 209–218. https://doi.org/10.1080/00207590244000089

Olson, D. R. (1996). *The world on paper: The conceptual and cognitive implications of writing and reading*. Cambridge University Press.

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2011). Complementary learning systems. *Cognitive Science, 38*(6), 1229-1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

Pan, S. C., Dunlosky, J., Xu, K. M., & Ouwehand, K. (2024). Emerging and future directions in test-enhanced learning research. *Educational Psychology Review*, *36*(20), 1–16. https://doi.org/10.1007/s10648-024-09857-2

Piolat, A. (2007). Effects of note-taking and working-memory span on cognitive effort and recall performance. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and Cognition: Research and Applications* (pp. 109–124). Emerald Group Publishing Limited.

Primor, L., & Katzir, T. (2018). Measuring multiple text integration: A review. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02294

Pugh, K. R., Frost, S. J., Sandak, R., Gillis, M., Moore, D., Jenner, A. R., & Mencl, W. E. (2006). What does reading have to tell us about writing? Preliminary questions and methodological challenges in examining the neurobiological foundations of writing and writing disabilities. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 433–448). The Guilford Press.

Ransdell, S., & Levy, C. M. (1996). Working memory constraints on writing quality and fluency. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 93–105). Lawrence Erlbaum Associates.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Rivard, L. P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education*, *88*(3), 420–442. https://doi.org/10.1002/sce.10114

Rivard, L. P., & Straw, S. B. (2000). The effect of talk and writing on learning science: An exploratory study. *Science Education*, *84*(5), 566–593. https://doi.org/10.1002/1098-237X(200009)84:5<566::AID-SCE2>3.0.CO;2-U

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roelle, J., Endres, T., Abel, R., Obergassel, N., Nückles, M., & Renkl, A. (2023). Happy together? On the relationship between research on retrieval practice and generative learning using the case of follow-up learning tasks. *Educational Psychology Review*, *35*(102), 1–27. https://doi.org/10.1007/s10648-023-09810-9

Rouet, J.-F., Favart, M., Gaonac'h, D., & Lacroix, N. (1996). Writing from multiple documents: Argumentation strategies in novice and expert history students. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, Models, and Methodology in Writing Research* (pp. 44–60). Amsterdam University Press.

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*(6), 759–769. https://doi.org/10.3758/BF03202637

Säljö, R. (1979). Learning about learning. *Higher Education, 8*, 443-451. https://doi.org/10.1007/BF01680533

Sanders-Reio, J., Alexander, P. A., Reio, T. G., & Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance? *Learning and Instruction*, *33*, 1–11. https://doi.org/10.1016/j.learninstruc.2014.02.001

Schnotz, W. (1984). Comparative instructional text organisation. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 53–81). Lawrence Erlbaum Associates, Inc.

Schumacher, G. M., & Gradwohl Nash, J. (1991). Conceptualizing and measuring knowledge change due to writing. *Research in the Teaching of English*, *25*(1), 67–96.

Segev-Miller, R. (2004). Writing from sources: The effect of explicit instruction on college students' processes and products. *L1-Educational Studies in Language and Literature*, *4*, 5–33. https://doi.org/10.1023/B:ESLL.0000033847.00732.af

Silva, A. M., & Limongi, R. (2019). Writing to learn increases long-term memory consolidation: A mental-chronometry and computational modelling study of "epistemic writing." *Journal of Writing Research*, *11*(1), 211–243. https://doi.org/10.17239/jowr-2019.11.01.07

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80.

Smagorinsky, P. (1998). Thinking and speech and protocol analysis. *Mind, Culture, and Activity*, *5*(3), 157–177. https://doi.org/10.1207/s15327884mca0503_2

Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating information: An analysis of the processes involved and the products generated in a written synthesis task. *Written Communication*, *30*(1), 63–90. https://doi.org/10.1177/0741088312466532

Spelman Miller, K., Lindgren, E., & Sullivan, K. P. H. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly*, *42*(3), 433–454. https://doi.org/10.1002/j.1545-7249.2008.tb00140.x

Spelman Miller, K., Sullivan, K. P. H. (2006). Keystroke logging: An introduction. In K. Sullivan, & E. Lindgren (Eds.), *Computer Keystroke Logging and Writing* (pp. 1-9). Elsevier.

Spirgel, A. S., & Delaney, P. F. (2016). Does writing summaries improve memory for text? *Educational Psychology Review*, *28*, 171–196. https://doi.org/10.1007/s10648-014-9290-2

Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*(2), 256-287. https://doi.org/10.1177/0741088390007002004

Spivey, N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*(1), 7-26.

Stoner, J. L., Felix, R., & Blank, A. S. (2022). Best practices for implementing experimental research methods. *International Journal of Consumer Studies*, *47*, 1579–1595. https://doi.org/10.1111/ijcs.12878

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology*, *63*(2), 358–385.

Tabachnick, B. G., & Fidell, L. S. (2012). Using multivariate statistics (6th ed.). New York: Harper and Row. https://doi.org/10.1037/022267

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2021). Age differences in diffusion model parameters: a meta-analysis. *Psychological Research*, *85*, 2012–2021. https://doi.org/10.1007/s00426-020-01371-8

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, *28*, 129–160. https://doi.org/10.1016/S0361-476X(02)00011-5

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*(4), 331–362. https://doi.org/10.1080/01638530902959927

Tulving, E., & Craik, F. I. M. (2000). *The Oxford Handbook of Memory*. Oxford University Press.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory* , *67*(1), 1–16. https://doi.org/10.1016/j.jml.2011.12.005

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review*, *17*(2), 200–205. https://doi.org/10.3758/PBR.17.2.200

Vandermeulen, N., de Maeyer, S., van Steendam, E., Lesterhuis, M., van den Bergh, H., & Rijlaarsdam, G. (2020). Mapping synthesis writing in various levels of Dutch upper-secondary education: A national baseline study on text quality, writing process and students' perspectives on writing. *Pedagogische Studiën*, *97*(3), 187–236.

Vandermeulen, N., van Steendam, E., & Rijlaarsdam, G. (2022). Introduction to the special issue on synthesis tasks: where reading and writing meet. *Reading and Writing*. https://doi.org/10.1007/s11145-022-10394-z

van Waes, L., & Schellens, P. J. (2003). Writing profiles: the effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics*, *35*(6), 829–853. https://doi.org/10.1016/S0378-2166(02)00121-2

Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: what is their role in writing performance? *Educational Psychology, 35*(6), 653-674. https://doi.org/10.1080/01443410.2013.836157

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. https://doi.org/10.3758/BF03196893

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*(3), 333–362. https://doi.org/10.1207/s1532690xci2203_3

Waldeyer, J., Heitmann, S., Moning, J., & Roelle, J. (2020). Can generative learning tasks be optimised by incorporation of retrieval practice? *Journal of Applied Research in Memory and Cognition*, *9*(3), 355–369. https://doi.org/10.1016/j.jarmac.2020.05.001

Weigle, S. C., Yang, W., & Montee, M. (2013). Exploring reading processes in an academic reading test using short-answer questions. *Language Assessment Quarterly*, *10*(1), 28–48. https://doi.org/10.1080/15434303.2012.750660

Wengelin, Å., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, *41*(2), 337–351. https://doi.org/10.3758/BRM.41.2.337

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475–482. https://doi.org/10.1007/BF03395630

White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, *30*, 166–189. https://doi.org/10.1016/j.cedpsych.2004.07.002

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*(14), 1–10. https://doi.org/10.3389/fninf.2013.00014

Wiecki, T. V., Sofer, I., Pedersen, M. L., Fengler, A., Govindarajan, L., Bera, K., & Frank, M. J. (2022). *Introduction – HDDM 0.9.8 documentation*. https://hddm.readthedocs.io/en/latest/

Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Eds.), *Writing as a learning activity* (pp. 120–148). Brill.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301–311. https://doi.org/10.1037/0022-0663.91.2.301

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single- versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1324–1344.

Young, R., and Sullivan, P. (1984). Why write? A reconsideration. In R. J. Conners, L. S. Ede, & A. A. Lunsford (Eds.), *Essays on classical rhetoric and modern discourse* (pp. 215-225). Southern Illinois University Press.

Yu, G. (2013). The use of summarisation tasks: Some lexical and conceptual analyses. *Language Assessment Quarterly*, *10*, 96–109. https://doi.org/10.1080/15434303.2012.750659

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organisational processes. *Memory & Cognition*, *38*(8), 995–1008. https://doi.org/10.3758/MC.38.8.995