

## University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Hall, S. (2023) "Estimating the Effects of Writing Beliefs, Writing Processes and Drafting Strategies on the Development of Subjective Understanding ", University of Southampton, Southampton Education School, PhD Thesis, pagination.

Data: Hall, S. (2023) Estimating the Effects of Writing Beliefs, Writing Processes and Drafting Strategies on the Development of Subjective Understanding.

<https://doi.org/10.5258/SOTON/D2523>.



**University of Southampton**

Faculty of Social Sciences

Southampton Education School

**Estimating the Effects of Writing Beliefs, Writing Processes and Drafting Strategies on  
the Development of Subjective Understanding**

by

Sophie M. Hall

ORCID ID 0000-0002-5318-6721

Thesis for the degree of Doctorate of Philosophy

March 2023

### **Abstract**

The problem-solving accounts of writing (Flower & Hayes, 1981; Bereiter & Scardamalia, 1987) are the prominent underpinning theories within writing research. These models assume that writing is a problem-solving activity - the writer combines planning, translation and revision processes to achieve their rhetorical goals, and expert writing is assumed to be a knowledge transforming process in which rhetorical problem-solving leads to both higher quality text and the development of understanding.

The dual-process model of writing (Galbraith, 2009; Galbraith & Baaijen, 2018), accepts the idea that writing is a problem-solving process, but claims in addition that spontaneous text production also leads to the development of understanding. These two different processes have conflicting effects on text quality, which the dual process model claims can be resolved by using a revision drafting strategy. The essential contrast, therefore, between the problem-solving models and the dual-process model is over whether spontaneous text production is simply a knowledge telling process (as in the problem-solving models) or is an active knowledge constituting process (as in the dual-process model).

Only two published studies thus far have examined whether the dual-process view is an appropriate depiction of the components involved in writing in relation to the development of subjective understanding. Therefore, this project attempted to further investigate the model's role in writing to learn by carrying out three studies: Study 1 examined individual differences in writing beliefs; study 2 described the development of reproducible measures of writing processes identified from keystrokes; study 3 used the methods developed in these two studies to carry out an experimental investigation of the effect of drafting strategies, writing beliefs and writing processes on the development of subjective understanding.

Study 1 used exploratory structural equation modelling to test whether a newly developed Writing Beliefs Inventory (WBI; Galbraith & Baaijen, in preparation) represented the five different types of writing beliefs (transmissional, transactional, planning, revision and audience) it claimed to, and to evaluate whether the structure of these beliefs was the same for writers with and without

dyslexia. Responses to the WBI were collected from 493 university students without dyslexia and 68 students with dyslexia. The findings of the study indicated that a five-factor WBI structure was upheld across both groups of students. The results also suggested that dyslexic writers may prioritise planning, rhetorical goals and reporting accurate information within writing more than developing their implicit understanding of the writing topic.

Study 2 (N = 48) was a methodological study concerned with identifying reproducible methods analysing keystroke logs. It first presented a transparent framework for analysing keystroke data relating to writing processes. It then explored whether experimental manipulation of the writing process, to control how spontaneously individuals typed, could elicit differences in pause structure between the two groups (demonstrated with Gaussian Mixture Modelling), but these results were inconclusive. Finally, the study evaluated the reproducibility of two scales developed by Baaijen and Galbraith (2018) designed to measure writing processes corresponding to components of the dual-process model (global revision and sentence production). This analysis revealed that, with the exception of one indicator that had been analysed slightly differently, the two composite measures were reproduced, with very similar loadings, indicating the measures' high reliability.

Finally, Study 3 combined the tools and methods examined and developed throughout Studies 1 and 2 to investigate whether writing beliefs, writing processes, and drafting strategies impacted the development of subjective understanding around a specific topic in university students' essay writing. In an experimental task, participants (N=61) were assigned to either a synthetic planning with rough drafting or an outline planning with final drafting condition. The participants were required to plan and write an argumentative essay about a current affairs topic, but the two conditions were designed to either elicit or inhibit discovery processes. The results indicated that the conditions did not lead to variations in how much participants self-rated their development of subjective understanding during writing, and a clear impact of writing beliefs on subjective understanding was not observed. However, increases in subjective understanding were linked to more spontaneous sentence production processes. These findings imply that although the experimental manipulation did not differentiate the sentence production and global revision processes used across both conditions, spontaneous sentence

production was still important for increasing subjective understanding, aligning with the components of the knowledge constitution system in the dual process model.

The projects' findings as a whole provide preliminary support for the dual-process model's explanation of subjective understanding development over the course of writing, indicating that an increase in understanding is related to active knowledge constituting processes and spontaneous sentence production, rather than basic knowledge telling processes. However, the project is limited in the amount of evidence it produces about the role of writing beliefs and drafting strategies on subjective understanding development, and hence future research should focus on these areas.

*Keywords:* Writing beliefs, writing processes, keystroke analysis, writing to learn, dual-process model, open science

## Table of Contents

<b>Abstract.....</b>	<b>2</b>
<b>Table of Contents .....</b>	<b>5</b>
<b>List of Tables .....</b>	<b>9</b>
<b>List of Figures.....</b>	<b>11</b>
<b>List of Accompanying Materials.....</b>	<b>12</b>
<b>Acknowledgements .....</b>	<b>14</b>
<b>Definitions and Abbreviations .....</b>	<b>15</b>
<b>Estimating the Effects of Writing Beliefs, Writing Processes and Drafting Strategies on the Development of Subjective Understanding .....</b>	<b>17</b>
Research Aims and Contributions .....	21
Research Questions and Overview of Research Methodology and Methods.....	25
A Note on Language.....	28
Thesis Structure .....	28
<b>1 Literature Review .....</b>	<b>31</b>
An Introduction to Writing Theory .....	31
An Overview of the Classical Problem-Solving Cognitive Perspectives of Writing .....	33
Romantic Perspectives of Writing .....	44
The Dual Process Model of Writing (Galbraith, 1999, 2009; Galbraith & Baaijen, 2018).....	46
The strengths and limitations of the Baaijen and Galbraith (2018) and Baaijen, Galbraith and de Glopper (2014) studies in their support for the dual-process model .....	66
<b>2 Methodology .....</b>	<b>71</b>
Methodological Background .....	71

Participants and Sampling .....	77
Instruments and Measures .....	79
Procedures .....	85
Ethical Considerations and the Impact of Covid-19.....	89
Introduction to Study Chapters.....	90
<b>3 Study 1: Using Exploratory Structural Equation Modelling to Examine the Writing Beliefs of UK University Students with and without Dyslexia .....</b>	<b>92</b>
Introduction .....	92
Background.....	93
Methods .....	100
Results .....	103
Discussion.....	111
<b>4 Study 2: A Transparent Methodology for the Analysis of Keystroke Data in Relation to Underlying Writing Processes. ....</b>	<b>118</b>
Introduction .....	118
Background.....	119
Methods .....	129
Results .....	145
Discussion.....	160
<b>5 Study 3: Exploring the Impact of Writing Beliefs, Writing Processes and Drafting Strategy on Subjective Understanding Development .....</b>	<b>166</b>
Introduction .....	166
Background.....	167
Methods .....	174

Results .....	178
Discussion.....	188
<b>6 General Discussion.....</b>	<b>194</b>
Introduction to General Discussion .....	194
A Brief Summary of the Results from the Three Studies.....	194
Discussion of Findings in Relation to the Dual Process View of Subjective Understanding Development .....	196
Strengths and Limitations of the Project .....	202
Real-World Implications .....	205
Conclusion.....	208
<b>References.....</b>	<b>211</b>
<b>Appendix A Writing Beliefs Inventory .....</b>	<b>222</b>
A.1 Writing Beliefs Online Questionnaire.....	222
A.2 Writing Beliefs Items and Respective Hypothesised Factors .....	229
<b>Appendix B Understanding Scale.....</b>	<b>232</b>
<b>Appendix C – First and Second Ideas Lists.....</b>	<b>234</b>
<b>Appendix D – First and Second Ratings of Importance.....</b>	<b>236</b>
<b>Appendix E – Degree of Correspondence.....</b>	<b>237</b>
<b>Appendix F – Writing Task Instructions.....</b>	<b>240</b>
<b>Appendix G – Study 1 Ethical Approval .....</b>	<b>242</b>
<b>Appendix H – Study 2/3 Ethical Approval .....</b>	<b>243</b>
<b>Appendix I – Data Management Plan .....</b>	<b>244</b>
<b>Appendix J – Additional Results Study 1 .....</b>	<b>247</b>
<b>Appendix K – Calculation Framework for Pauses, Bursts and Revisions .....</b>	<b>253</b>

<b>Appendix L – Full Mixture Model Results Study 2 .....</b>	<b>263</b>
<b>Appendix M - Additional Study 3 Model Outputs.....</b>	<b>273</b>

### List of Tables

<b>Table 1</b> WBI Statements with High Covariances, Identified through Modification Indices .....	107
<b>Table 2</b> Goodness of Fit Indices for the Three- and Five-Factor WBI ESEM and CFA Models for Non-Dyslexic Data .....	108
<b>Table 3</b> Invariance Testing Results for the 39-Item Five-Factor ESEM on the Students with and without Dyslexia .....	110
<b>Table 4</b> Invariance Testing Results for the 39-Item Five-Factor ESEM on the Students with and without Dyslexia .....	113
<b>Table 5</b> Conceptual Definitions of Keystroke Features .....	132
<b>Table 6</b> Sentence Production and Global Linearity Measures .....	141
<b>Table 7</b> Number of Components for the Best Fitting Models and Number/Percentage of Participants (Hall et al., 2022, p. 20) .....	147
<b>Table 8</b> Summary of Mixture Model Components Within Words for Outline and Synthetic Conditions .....	149
<b>Table 9</b> Summary of Mixture Model Components Between Words for Outline and Synthetic Conditions .....	151
<b>Table 10</b> Summary of Mixture Model Components for Sub-Sentences in the Outline and Synthetic Conditions .....	153
<b>Table 11</b> Summary of Mixture Model Components for Between-Sentences in the Outline and Synthetic Conditions .....	155
<b>Table 12</b> Summary of Principal Component Analysis with Varimax Rotation for 2 Factor Solution in the Current Study and Baaijen and Galbraith (2018) Study .....	157
<b>Table 13</b> Summary of Goodness of Fit statistics for 1-3 Component Solutions, Before and After Item Removal from Galbraith et al. (2023) .....	176

<b>Table 14</b> Loadings for 2-Factor Solution using Maximum Likelihood Estimation with Geomin Oblique Rotation from Galbraith et al. (2023).....	177
<b>Table 15</b> Means, Standard Deviations and Bivariate Correlations of All Variables in the Understanding Change Analysis.....	179
<b>Table 16</b> Simplified Regression of knowledge on Sentence Production and Global Revision .....	183
<b>Table 17</b> Simplified Regression of Organisation on Sentence Production and Global Revision .....	183
<b>Table 18</b> Equivalent Replication of Baaijen et al. (2014) Analysis of knowledge Regressed on Writing Beliefs.....	186
<b>Table 19</b> Equivalent Replication of Baaijen et al. (2014) analysis of knowledge regressed on writing Beliefs.....	187
<b>Table 20</b> Factor Mean Differences using the Scalar Model when the No Dyslexia Group is Constrained to Zero.....	247
<b>Table 21</b> 39-item WBI 5-factor ESEM Loadings for Students Without Dyslexia .....	248
<b>Table 22</b> Descriptive Statistics for Each Linear Within-Word GMM .....	263
<b>Table 23</b> Descriptive Statistics for Each Linear Between-Word GMM .....	266
<b>Table 24</b> Descriptive Statistics for Each Linear Sub-Sentence GMM.....	269
<b>Table 25</b> Descriptive Statistics for Each Linear Between-Sentence GMM .....	271

### List of Figures

<b>Figure 1</b> Original Hayes-Flower (1980) Model as Shown in Hayes (2012, p.370).....	34
<b>Figure 2</b> Hayes' (2012, p.371) Cognitive Model of Writing.....	35
<b>Figure 3</b> Relationship Between Sentence Production and Change in Understanding as a Function of Type of Planning (Baaijen & Galbraith, 2018, p.212) .....	66
<b>Figure 4</b> Simplified Path Diagrams of the Tested ESEM and CFA Models.....	105
<b>Figure 5</b> Example of a Non-Regular Sentence Termination Keystroke Combination (Hall et al., 2022, p.14) .....	137
<b>Figure 6</b> Example GMMS for the Log Within-Word Pause Times for Outline and Synthetic Conditions .....	149
<b>Figure 7</b> Example Mixture Models for the Log Between-Word Pause Times Across the Outline and Synthetic Conditions .....	150
<b>Figure 8</b> Example Mixture Models for the Log Sub-sentence Pause Times Across the Outline and Synthetic Conditions .....	152
<b>Figure 9</b> Example Mixture Models for the Log Between-Sentence Pause Times Across the Outline and Synthetic Conditions .....	155
<b>Figure 10</b> Proportion of Change in Understanding Over the Course of Writing by Condition .....	184
<b>Figure 11</b> Proportion of Change in Organisation Over the Course of Writing by Condition .....	185

### **List of Accompanying Materials**

Supplementary materials licensed under CC BY 4.0 can be found at: <https://osf.io/q59cj/>

These include:

- README file
- Study 1
  - MPlus script for exploratory structural equation modelling and invariance testing
- Study 2
  - Gaussian Mixture Modelling R scripts
  - Global Revision and Sentence Production measures macros and R script
- Study 3
  - R script for MANCOVA and regression analyses

Dataset licenced under CC BY-NC-SA 4.0 can be found at

<https://doi.org/10.5258/SOTON/D2523>. This includes:

- READ ME file
- Data for Studies 1, 2 and 3
- Copies of participant information sheets and consent forms

▪ Research Thesis: Declaration of Authorship

Print name: SOPHIE HALL

Title of thesis: Estimating the Effects of Writing Beliefs, Writing Processes and Drafting Strategies on the Development of Subjective Understanding

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 but expands on work published as:

Hall, S., Baaijen, V. M., & Galbraith, D. (2022). Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing*.  
<https://doi.org/10.1007/s11145-022-10284-4>

Signature: .....Date: 21.03.23

### Acknowledgements

I could not have undertaken this journey without my supervisors. David, I cannot thank you enough for your guidance not only throughout my PhD, but also my undergraduate and master's degrees – 8 years in total! Here's to hopefully many more years of collaboration. James, your stats advice and career chats will forever be appreciated! Thank you for your support over the last four years.

My thanks also go to the SCDTP for funding this work, the Southampton Education School for being a wonderful place to work in, full of excellent academics, and my fellow PhD students (especially Philly, Amy and Pip – the best ReproducibiliTea team ever!).

Finally, to Mum, Joe, Charlie, Becky, Emma and Rufus – thank you for being the best family anyone could ever wish for.

### Permissions

Figures 1 and 2 in this thesis have been reproduced with permission from SAGE Publications under their thesis reuse terms.

Figure 3 in this thesis has been reproduced with permission of the authors of the original journal article (David Galbraith and Veerle Baaijen) and under the Creative Commons Licence.

Figure 5 and Table 7 in this thesis has been reproduced with permission of the authors of the original journal article (Sophie Hall, Veerle Baaijen and David Galbraith) and under the Creative Commons Licence.

Table 13 and 14 in this thesis have been reproduced with permission from Taylor & Francis Group.

### Definitions and Abbreviations

Term	Definition
Cognitive model of writing	Models that theorise the mental processes involved in writing.
Discovery	The process of developing a deeper subjective understanding about a topic through the act of writing.
Dual process model of writing	A cognitive model that views writing as a combination of knowledge transforming and knowledge constituting processes.
Final draft	A polished, high-quality finished text product.
Knowledge constituting	The process in which ideas are synthesised. New content is generated over an iterative process of feeding spontaneously formulated utterances (shaped by the writer's semantic memory) back into the writer's existing schemas until they sufficiently capture the individual's understanding of the writing topic.
Knowledge telling	Taking ideas from long-term memory through a direct retrieval process and translating them into written sentences. No manipulation or reorganisation of content occurs throughout this process.
Knowledge transforming	A reflective process in which existing ideas are manipulated in working memory using content retrieved from episodic memory in a way that suits the needs of the writing task.
Outline plan	A hierarchically structured plan of ideas for a piece of writing.
Planning (as an activity)	An activity that takes place prior to a writing task which is aimed at capturing an individual's initial ideas for their piece of writing.
Planning (as a cognitive writing process)	The act of generating and organising ideas in relation to the rhetorical goals and/or the individual's own goals for a writing task.
Problem-solving/classical perspectives of writing	A theoretical perspective that views writing as a problem-solving process.
Revision/reviewing (as an activity)	Reviewing, editing and/or redrafting a piece of writing <i>after</i> producing an initial first draft.
Revision/reviewing (as a cognitive writing process)	Reading and editing a text throughout the process of writing.
Rhetorical goals	Explicit aims informed by the purpose and audience of the writing task.
Romantic perspectives of writing	A theoretical perspective that emphasises the importance of spontaneous text production to elicit discovery during writing.
Rough draft	A text product which is not in its final state. It captures the writer's ideas as they unfold, but does not require consideration of general writing conventions such as organisation, grammar and spelling.

---

Semantic knowledge	Commonly referred to as general knowledge. One's knowledge about concepts and relationships between those concepts.
Subjective understanding	The writer's personal interpretation of a topic.
Synthetic plan	A plan consisting of a single sentence which summarises the writer's overall goal for a piece of writing.
The writing process	The act of combining individual writing processes to create a text.
Writing beliefs	An individual's views about how one should engage in the writing process, and what the priority goals of writing are.
Writing processes	Individual cognitive processes that form the overall writing process.
Writing strategy	An approach or method used to compose a text.
Writing-to-learn	A writing activity which helps an individual develop their understanding of a topic.

---



---

Abbreviation	Term
CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis
ESEM	Exploratory Structural Equation Modelling
MANCOVA	Multivariate Analysis of Covariance
WBI	Writing Beliefs Inventory

---

### **Estimating the Effects of Writing Beliefs, Writing Processes and Drafting Strategies on the Development of Subjective Understanding**

Theory indicates that writing can be used as a tool for learning, as well as for effectively communicating thoughts and ideas to an audience (Galbraith & Baaijen, 2018). Traditionally, writing has been viewed as a problem-solving process. The writer combines planning, translating, and revision processes in a way that most effectively meet the rhetorical goals of the task at hand (Flower & Hayes, 1981). Under this school of thought, learning, or ‘discovery’, is deemed a by-product of the writing process. That is, problem-solving accounts of writing have typically assumed that in learning to write more expertly, through controlled problem-solving processes, individuals incidentally get better at developing their knowledge and understanding of the topics they write about (Flower & Hayes, 1981; Bereiter and Scardamalia, 1987).

This thesis takes a different view about how discovery is enabled during writing, focusing instead on the dual-process model of writing (Galbraith, 1999; Galbraith, 2009; Galbraith and Baaijen, 2018). Specifically, the dual-process account stipulates that discovery is attributed to the combined effect of two conflicting processes: knowledge constitution and knowledge transformation. *Knowledge constitution* is the process in which ideas are synthesised. New content is generated over an iterative process of feeding spontaneously formulated utterances (shaped by the writer’s semantic memory) back into the writer’s existing schemas until they sufficiently capture the individual’s understanding of the writing topic. *Knowledge transformation* is a reflective process in which existing ideas are manipulated in working memory using content retrieved from episodic memory in a way that suits the needs of the writing task. Both knowledge constitution and knowledge transformation are necessary for developing understanding through writing. However, the conflicting nature of the knowledge constituting processes that guide the writer’s implicit understanding of a topic, and the knowledge transformation processes involved in the explicit organisation of content to satisfactorily meet the goals of the writing task necessitates that writers must learn to balance the two systems. In doing so, the result is a final text product that accurately reflects the writer’s understanding of a topic whilst being clear and accessible to the reader.

A component of writing that may influence how an individual balances the knowledge constitution and knowledge transformation systems of the dual-process model is writing beliefs. Writing beliefs are an individual's views about how one should approach and engage in the writing process, and what the priority goals of writing are. White and Bruning's (2005) research indicated that beliefs about writing could be categorised into two distinct factors: transmissional and transactional beliefs. Writers with high transmissional beliefs are thought to have low affective and cognitive engagement with writing activities. They believe that the main purpose of writing is to report accurate information from authoritative sources to the reader. On the other hand, writers with high transactional beliefs have high levels of cognitive and affective engagement in writing. They tend to argue that the purpose of writing is to develop one's own understanding of the subject at hand. White and Bruning (2005) discovered that writing beliefs were associated with writing outcomes. Specifically, students with high transactional writing beliefs produced better quality texts in terms of organisation, idea-content development, writing conventions, voice, and general quality. In contrast, students with high transmissional beliefs performed poorly on all these measures.

Later work from Sanders-Reio et al. (2014) also found that recursive process beliefs, which monitor the extent to which writers think that revising and creating multiple drafts of a text is an important goal for writing, were also salient for improving text quality. Additionally, they found that beliefs about audience orientation (i.e. the extent to which individuals think that it is important to adapt the style of their writing to suit the needs of the reader), were positively associated with writing performance.

Further work from Baaijen et al. (2014) established that writing beliefs were linked to text quality *and* subjective understanding development in university students' essay writing. However, they found that the type of planning that the students engaged in before writing affected these relationships. Specifically, they found that writers with high-transactional beliefs wrote significantly better-quality texts than writers with low-transactional beliefs when assigned to a synthetic pre-planning strategy (i.e. the participants were not allowed to create explicit plans for their texts, only a single sentence summing up their main overall goal). Writers with low-transactional beliefs were better at writing good quality texts when they were assigned to the outline planning strategy (i.e. they

were told to create a hierarchically organised plan prior to composing their essays). For high-transactional writers, there was no difference in text quality between the two conditions, so, in general, writers with high-transactional beliefs wrote better texts than writers with low-transactional beliefs.

These findings support the assumptions of the dual process model because they imply outlining can be advantageous for helping low-transactional writers organise their thoughts prior to text production (through a knowledge-transforming approach), which ultimately improves the quality of their texts. However, for high-transactional writers, type of planning was not important, most likely because their goals for writing were guided by their tendency to believe that the purpose of writing is to develop personal understanding, and so they did not deem text quality a priority.

Baaijen et al. (2014) also found that writing beliefs moderated the relationship between planning strategy and change in subjective understanding. In general, writers with high-transactional beliefs developed their subjective understanding around the topic well, regardless of the planning condition they were assigned to. However, writers with high-transactional *and* high-transmissional beliefs developed their subjective understanding of the writing topic the most when they were in the synthetic condition. In fact, high-transactional and high-transmissional writers developed their understanding of the writing topic more than any other group when they were told to synthetically plan. Conversely, they were the worst group in terms of subjective understanding development when asked to create an outline plan.

These findings also support the assumptions of the dual process model, because they indicate that when writers are both guided by explicit, rhetorical task-related goals (high-transmissional beliefs) but also want to develop their personal understanding around the writing topic (high-transactional beliefs), a synthetic planning strategy, in which there is little rhetorical guidance, allows them to do this because it promotes knowledge constitution. When they are assigned to a rigid outline planning strategy, though, designed to inhibit knowledge constitution processes, they struggle to develop their understanding.

In the same study (reported in a different paper) Baaijen and Galbraith (2018) used keystroke data collected throughout the essay writing activity to reveal that planning type also interacted with

writing processes to affect text quality and subjective understanding development. They looked at two types of writing processes:

- 1) Global linearity – The extent to which authors revise the global structure of their texts. High levels of linearity indicated that texts were written linearly, whereas low levels indicated a lot of revision.
- 2) Sentence production – How controlled or spontaneously an author wrote their text. More spontaneous writers produced shorter bursts of language, with shorter pauses between them. Controlled authors produced longer bursts with longer pauses between them.

Writers in the synthetic condition wrote better quality essays when they revised their text's global structure during writing but produced sentences in a linear and carefully controlled manner. The quality of texts written in the outline group was also positively correlated with global linearity and controlled sentence production, but these relationships were small and non-significant. As a whole then, Baaijen and Galbraith (2018) showed that the relationship between planning and text quality was complicated - high text quality was possible when ideas were *not* organised in a plan before writing, but in order for this to happen, the writer had to have careful control over the production of their sentences, and they needed to revise the global structure of their texts during the course of writing for high text quality to be achieved.

Baaijen and Galbraith (2018) also found that planning type was related to how much the writer improved their subjective understanding of the topic throughout the writing task. Specifically, subjective understanding development was at its best in the synthetic condition when writers produced sentences spontaneously. In contrast, understanding development in the outline planning condition was shallow and practically non-existent, and the relationship between sentence production processes and understanding development, though non-significant, worked in the opposite direction. These results support the assumptions of the dual-process model, because they imply that subjective understanding development is most optimum when writers do not adhere to a pre-planned outline, but rather develop their thoughts freely and spontaneously.

The findings from Baaijen et al. (2014) and Baaijen and Galbraith (2018) are important because they reveal that pre-planning in an outline format is not universally beneficial for all writers (though

that is often assumed to be the case, particularly in the teaching of academic writing). Depending on writing beliefs and writing processes used by the writer, outline and synthetic planning strategies' effects on text quality and subjective understanding can vary. The findings also indicate that the ability to produce a good quality text whilst going through a discovery process is mainly dependent on how the individual engages with writing (i.e., what their beliefs about writing are and the transcription processes they use).

To date, however, the Baaijen et al. (2014) and Baaijen and Galbraith (2018) studies are the only two studies that have explicitly examined how drafting strategies and writing beliefs may interact to impact text-quality and subjective understanding development through a dual-process lens. As such, this project further tests the dual-process model's assumptions. Firstly, it does this by addressing some of the gaps presented in previous research that has looked at the role of drafting strategy, writing beliefs and writing processes on subjective understanding (these gaps will be briefly introduced in the next section of the chapter). Secondly, the project attempts to partially replicate the Baaijen et al. (2014) and Baaijen and Galbraith (2018) studies with a new sample of university students, to investigate whether their findings can be upheld beyond the remit of the original study.

## **Research Aims and Contributions**

The project first focuses on writing beliefs, and testing the suitability of a tool to examine them. This is because although previous research has indicated that writing beliefs are linked to text quality and subjective understanding development, these studies rely on measures that may not capture the full breadth and type of beliefs that exist. Previous writing beliefs inventories (White & Bruning, 2005; Sanders-Reio et al., 2014) have examined transactional, transmissional, revision and audience beliefs, but these scales fail to capture any beliefs about planning. Planning is an integral part of the writing process, and is likely linked to both text quality and subjective understanding development (Baaijen & Galbraith, 2018). Hence, it is important to understand how one's views about planning might be implicated in how they engage with planning processes during writing, and the impact those views might have on writing outcomes.

As such, the first study in this project aims to validate a new writing beliefs inventory (WBI; Galbraith & Baaijen, in preparation), which has a theoretical five-factor structure consisting of transmissional, transactional, revision, audience *and* planning beliefs in a sample of UK university students.

The second aim within Study 1 was to test whether the five-factor structure of the WBI could be generalised to a sample of university students with dyslexia. This is because dyslexic students commonly report academic writing as one of the most challenging tasks they encounter whilst at university, particularly in terms of being able to organise and express their ideas in writing (Mortimore & Crozier, 2006). Hence, by examining the writing beliefs of university students with dyslexia, we can test whether i) the WBI is an appropriate tool for examining their writing beliefs and ii) whether the types of beliefs that dyslexic students display differ from students without dyslexia, and whether this may be implicated in their writing outcomes.

At this point, it is salient to mention that the original overall aim of this PhD project was to compare dyslexic and non-dyslexic writers, and to assess how the process of writing and development of understanding may have been affected by disruptions to the language production process. However, due to the impact of Covid-19 on the timeline of the project, it was unviable to obtain enough data from dyslexic university students in order to draw such a comparison. Hence, Study 1 (which was conducted pre-Covid), is the only study in this project that compares students with and without dyslexia. However, the findings of that comparison are still integral to understanding how writing beliefs are implicated within the development of subjective understanding over the course of a writing task. Any discrepancies between the two groups' beliefs could help partially explain why dyslexic students appear to struggle more with academic writing than those without (beyond the typical word-level and spelling issues linked to poor writing outcomes in dyslexic students; Galbraith et al., 2012; Sumner & Connelly, 2020; Wengelin, 2007).

The second study in this PhD project focuses on creating a transparent methodology for analysing keystrokes in relation to writing processes. Keystroke logging, a tool for collecting real-time keyboard and mouse inputs throughout a writing task, is fast becoming a popular technique for investigating the writing process. However, virtually all previous research has investigated *individual*

features of text production (i.e. pauses, bursts, or revisions). In order to get a rich and coherent picture of the processes that individuals use during writing, though, it is important to consider these measures collectively (Baaijen, Galbraith & de Glopper, 2012). Baaijen & Galbraith (2018) created two composite scales using pauses, bursts and revisions to capture sentence production processes and global revision (corresponding to the knowledge constitution and knowledge transforming components of the dual-process model respectively). Hence, the primary aim with Study 2 was to create a transparent and reproducible way of calculating the measures for these two scales, and to assess whether overall global revision and sentence production components from Baaijen and Galbraith's (2018) measures could be replicated. The secondary aim of Study 2 was to investigate whether experimental manipulation to the drafting strategy participants used might make it easier to determine how the structure of pauses within texts could relate to underlying processes. Mixture modelling (McLachlan & Peel, 2001) has been used in previous studies to try and enhance the interpretation of pauses at different intervals of text production (i.e., within words, between words, between sub-sentences, between sentences and between paragraphs). In this study, mixture modelling was used to investigate whether experimental manipulation of the writing task (i.e. writers being assigned to different planning and drafting scenarios) had an impact on the structure of pauses observed during writing between conditions. Such findings would make it easier to map pauses onto underlying controlled or spontaneous sentence production processes, which could give further insight into how these components align with the dual process model of writing.

Finally, study 3 aimed to use the newly developed WBI (Galbraith and Baaijen, in preparation) examined in study 1, with the sentence production and global linearity measures investigated in study 2 to look at how writing beliefs and writing processes related to subjective understanding development during an experimental writing task. This study used methods similar to Baaijen et al. (2014) and Baaijen and Galbraith (2018), but with two major differences. Firstly, a multi-item scale was used to capture subjective understanding change over the course of writing. This is because the original research relied on a single item scale. However, Galbraith, Peters, Hall and Baaijen (2023) argue that a multi-item scale may be more appropriate for capturing subjective understanding in relation to how knowledge develops during writing and the organisation of that knowledge (reflecting

the components of the knowledge constituting and transforming systems within the dual process model).

Secondly, the study extends the writing task manipulation used in the original research. In the previous investigation, participants were asked to spend five minutes either creating an outline plan (a hierarchically structured plan of ideas for their essays about a specific topic) or a synthetic plan (a single sentence summing up their overall opinion on a topic), before typing an essay in 30 minutes. However, both conditions were required to write a final version of an essay, essentially meaning that text products across conditions were supposed to be of the same high quality. As such, this may have limited the amount that the participants could develop their understanding of the writing topic because, under the assumptions of the dual-process model, the knowledge transforming processes involved in ensuring that the final product met the rhetorical goals of the writing task would have conflicted with the extent to which the writers could engage with producing new ideas (the knowledge constituting system). Hence, the current project extended the manipulation in the original study by requiring the two conditions to engage in different drafting techniques. Specifically, writers in the outline condition were required to write final versions of an essay (i.e. a polished, high-quality end product). However, writers in the synthetic condition were only required to write rough drafts (i.e. an end product which captured the writer's ideas as they unfolded, but did not require consideration of general writing conventions such as organisation, grammar and spelling). In theory, this should have given the writers in the synthetic condition more of a chance to engage with the knowledge constitution component of the dual process model, and hence they should have been able to develop their understanding of the writing topic more than those in the outline condition.

Nevertheless, the extent to which writers engage with the knowledge constituting and knowledge transforming systems, regardless of the conditions they are assigned to, will likely be influenced by their beliefs about writing and the underlying writing processes that they typically engage in. As such, in line with the original investigation, this research also considered how writing beliefs were implicated in subjective understanding development during an essay writing activity. In contrast to Baaijen et al. (2014) though, the current research used the newly developed WBI (Galbraith and Baaijen, in preparation) examined in Study 1, as opposed to the White and Bruning

(2005) inventory, so it could look more in depth at the role of planning beliefs on subjective understanding. In line with Baaijen and Galbraith (2018), the study also examined the influence of global revision and sentence production processes on subjective understanding by using the composite measures examined in Study 2. The purpose of this was to see whether the previous research findings could be upheld in a new sample of students.

This thesis provides an important contribution to the field of writing research in two ways. Primarily, it builds on the existing evidence for the dual-process model, highlighting how writing beliefs, drafting strategies and writing processes may influence the extent to which writers engage in the knowledge constituting and knowledge transforming systems, and subsequently the effect that this can have on subjective understanding development throughout a writing task.

Secondly, the project uses methods of open science. Study 2 addresses many issues relating to reproducibility that are prominent in keystroke research, and so the study provides a framework that other researchers can use to more transparently conduct keystroke research in the future. To further aid the transparency of this project, the scripts and datasets used throughout have been made openly available (link to these supplementary materials on p.11).

Finally, in terms of a contribution beyond research, the evidence presented throughout this thesis has the potential to impact academic writing teaching practice. Most notably, findings from Study 1 indicate that consideration of students' writing beliefs may be important for improving writing outcomes. Study 3 demonstrates that when the aim is to use writing as a tool for learning, encouraging students to write more spontaneously might be beneficial for maximising the extent to which they develop their subjective understanding around a given topic. A full discussion of the real-world implications of the findings from this project can be found in Chapter 6 (pp.205 – 208).

## **Research Questions and Overview of Research Methodology and Methods**

Study 1 utilised a survey design with an online questionnaire featuring the WBI (Galbraith & Baaijen, in preparation). Exploratory structural equation modelling (ESEM; Asparouhov & Muthén, 2009; Marsh et al., 2009) was used to examine whether the WBI had a hypothesised five-factor

structure demonstrating five different types of writing belief: transmissional, transactional, planning, revision and audience. Invariance testing was then used to investigate whether the WBI structure was consistent across UK university students with and without dyslexia, and whether the two groups displayed varying levels of the five writing beliefs. Hence, the research questions for Study 1 were:

1. Concerning the Galbraith and Baaijen (in preparation) Writing Beliefs Inventory:
  - a. To what extent does it validly capture the five underlying components of writing beliefs that it was designed to capture?
  - b. To what extent does it demonstrate measurement invariance across items, factors, and UK university students with and without dyslexia?

Study 2 attempted to create a reproducible and transparent framework for analysing keystroke components (pauses, bursts and revisions) in relation to writing processes (global revision and sentence production). Study 2 also examined whether experimental manipulation of planning and drafting strategy (i.e. synthetic planning with rough drafting or outline planning with final drafting) led to differences in how pauses were distributed at different locations within the text (i.e., within-word, between-word, between-sub-sentence and between-sentence). Gaussian Mixture Modelling (GMM; McLachlan & Peel, 2000) was used to model the pause data. The results give an indication as to whether the participants in the rough draft condition engaged with deeper, reflective writing processes than those in the final drafting condition (as the dual process model would predict), and whether this could be observed with pause analysis. Finally, Study 2 used the keystroke measures developed in the reproducible framework to try to replicate the two writing process measures (sentence production and global linearity) identified by Baaijen and Galbraith (2018), using principal components analysis (PCA). The research questions for Study 2 were:

- 2a. How do we create a reproducible method for defining, identifying, and calculating pauses, bursts, and revisions from keystroke data?

2b. Do Gaussian mixture models reveal differences in the durations and mixing proportions of pauses throughout text composition in the outline and synthetic drafting strategies?

2c. Can the Baaijen and Galbraith (2018) two-component composite measures of global linearity and sentence production be reproduced on a new sample of university students' essay writing keystroke data?

Studies 2 and 3 use the same experimental dataset. UK university students ( $N = 61$ ) were assigned to one of the two drafting conditions described above. They were asked to type an essay in 30 minutes on a current affairs topic (either veganism or social media), during which time their keystrokes were logged. Before typing their essays, they completed the WBI (Galbraith and Baaijen, in preparation) and a pre-test multi-item subjective understanding scale. After essay writing, they filled in the understanding scale for a second time.

For Study 3, the difference between the pre- and post-test subjective understanding scores was used to examine whether subjective understanding developed over the course of writing, and if this differed between the two writing conditions. The results from the WBI (Galbraith and Baaijen, in preparation) were used to assess whether writers' preconceived beliefs about writing were associated with their changes in subjective understanding. Additionally, the keystroke data collected throughout the writing task was used to examine whether global revision and sentence production processes were related to subjective understanding during writing, and whether these findings supported or opposed the assumptions of the dual-process model. It was predicted that spontaneous sentence production should have been associated with increases in subjective understanding, along with higher levels of global revision. Research Question 3 was as follows:

3. Do drafting conditions, writing beliefs and writing processes affect the development of subjective understanding during essay writing?

## **A Note on Language**

Some of the research in this project focuses on dyslexic university students. As such, it is important to discuss the terminology about dyslexia used throughout the thesis. Terminology around special educational needs, disabilities and learning differences has been debated for some time. Specifically, there is tension over using person-first language or identity first language (Collier, 2012). Person-first language refers to the individual first (e.g., a person with dyslexia). Identity-first language refers to their label first (e.g., a dyslexic person).

In previous years, person-first language has been advocated for, particularly within government and health care settings because of the idea that disability, education needs and learning differences only make up part of an individual's identity. Hence, the idea was that people should be referred to as exactly that: people (Gernsbacher, 2017). However, in more recent years, there has been a push for identity first language, because disability advocates see their labels as being a big part of their identity, and so putting their disability to the wayside can be interpreted as accentuating stigma (Collier, 2012; Gernsbacher, 2017).

In dyslexia research, person-first versus identity-first language is still debated, mainly because there is variation in preferred terminology amongst the dyslexic community (Evans, 2014; Laffar-Smith, 2017; McCall et al., 2020). Hence, it is recommended to use a combination of person- and identity-first language in research contexts, as there is no clear preference in the disability community on which to use (Dunn & Andrews, 2015). Therefore, in line with Dunn and Andrew's (2015) recommendation, both person-first and identity-first language is used throughout this thesis.

## **Thesis Structure**

The following section outlines the structure of this thesis.

### ***Chapter 1: Literature Review***

The literature review introduces the cognitive approach to writing theory, focusing on the assumptions of the problem-solving and dual-process accounts of writing, and what these models say about two typically linked writing outcomes: text quality and subjective understanding development. In the latter section of the chapter, a critique of the research relating to these two models is presented, specifically focusing on studies that explore the relationships between writing beliefs, writing strategies and text composition processes with text quality and understanding development.

### ***Chapter 2: Methodology***

This chapter discusses the project's methodological philosophy and research design. It also gives an overview of the participants, instruments, measures, and procedures for each of the three studies. Finally, it discusses the ethical considerations for the project, including project adaptations due to the Covid-19 outbreak in 2020.

### ***Chapter 3: Study 1***

This chapter presents the first study within this project, which was concerned with using ESEM to examine the structure of writing beliefs captured by the WBI (Galbraith and Baaijen, in preparation) in UK university students with and without dyslexia.

### ***Chapter 4: Study 2***

This chapter presents study 2, which attempted to identify a reproducible framework for analysing keystrokes in relation to underlying writing processes. The study used this framework to examine a technique for improving the accuracy of mapping pause data onto writing processes (mixture modelling with experimentally manipulated planning and drafting strategies). Finally, this study attempted to replicate the global linearity and sentence production writing process measures developed by Baaijen and Galbraith (2018).

***Chapter 5: Study 3***

This chapter presents study 3, which examined how drafting strategies, writing beliefs and writing processes affected change in subjective understanding during an essay writing task for UK university students.

***Chapter 6: General Discussion***

This chapter brings together the findings of all three studies and discusses whether they support the dual-process model's view of how subjective understanding develops over the course of writing. The chapter also discusses the strengths and weaknesses of the project as a whole, and highlights some recommendations for future research. Finally, the chapter describes the real-world implications of the project's results.

## **1 Literature Review**

The first half of this chapter will talk broadly about the various theories of writing developed over the years, before establishing why this project situates itself under the umbrella of cognitive writing theories. It will discuss the main aims of cognitive perspectives on writing, focusing on two schools of thought defined by Galbraith and Baaijen (2018) as the classical and romantic models of writing. In the discussion of these perspectives, particular attention is given to how they explain variations in two fundamental writing outcomes – text quality and development of understanding. The strengths and weaknesses of each perspective will be presented, highlighting how neither presents a comprehensive view of the writing process. The chapter then introduces the dual-process model of writing (Galbraith, 1999; Galbraith, 2009; Galbraith & Baaijen, 2018) as an addendum to the classical view of writing, and argues that the addition of this model to pre-existing theory helps to provide a more holistic view of the writing process in relation to text quality and changes in the writer's subjective understanding.

The second half of this chapter then explores the strengths and weaknesses of the dual process model, drawing on evidence in support of the model and highlighting several of the model's limitations. These include the lack of discourse about how individual differences in approaches to writing can influence text quality and change in subjective understanding in the model, and about how text production processes map onto the models' knowledge transforming and knowledge constituting systems.

The chapter ends with an explanation of the project's aims, which seek to address some of the challenges and limitations set out throughout the literature review.

### ***An Introduction to Writing Theory***

When writing was first established as an area of study, research focused on how to improve the quality of written text, particularly in school-aged children, and so the emphasis was on practical writing instruction (Foster, 1983; Nystrand, 2006; Rose, 1985). However, as the field has grown,

theory has become more of a central focus, and research is now often steered towards understanding *how* writing happens as a mental process, as well as a physical activity.

The general consensus among scholars now is that writing is not just a way of communicating, but also a tool for improving cognition, manipulating and reorganising knowledge from long-term memory and, most interestingly, a method for creating new knowledge and understanding (Leggette, Rutherford, Dunsford, & Costello, 2015). This notion of knowledge creation has typically been referred to as ‘discovery’, and implies that inside and outside of the writer’s mind, there is something waiting to be discovered that will improve the writer’s understanding around a topic. But, this term has been controversial. For example, Flower and Hayes (1980), argue that writers don’t find meaning, they make meaning, and so a better term would be ‘invention’. This instead highlights the deliberate problem-solving involved in developing content. Others, though, argue that discovery, or knowledge creation, is a spontaneous and self-reflective rather than deliberate process (Britton et al., 1975) that is optimised when rhetorical, problem-based constraints are removed from the writing task at hand.

Accordingly, as researchers began to develop models of writing, reflection on those outcomes beyond effective communication (e.g., discovery or invention) were further considered and debated, albeit to varying extents. This resulted in a series of overarching paradigms in writing research, which vary in the way they characterise the processes and goals used in writing. One example is the cognitive perspective, which ultimately views writing as a goal directed problem-solving process carried out within a limited capacity mental system, and over time, has adopted Baddeley and Hitch’s (1974) working memory model as a characterisation of its limited capacity system. Another is the sociocultural perspective. Driven by Vygotsky’s (1978) work on language development, this framework argues that an individual’s writing ability and goals are shaped by their surrounding social and cultural contexts, both consciously and unconsciously (Mohammadzadeh, Ahour & Saeidi, 2020). Finally, the social-cognitive perspective of writing, first developed by Flower (1994), is broadly based on Bandura’s (1977) social learning theory. It bridges the gap between the cognitive and socio-cultural perspectives, and posits that writing and an individual’s ability to write are influenced by their internal cognitive processes as well as their societal and cultural contexts. This perspective implies that societal and cultural contexts actively *inform* the development of writing processes.

Flower's (1994) socio-cognitive perspective offers the most complete model of writing out of the three described here. Nevertheless, it is the cognitive perspective that provides the framework for this PhD, not because the author disagrees with the notions of the socio-cognitive theory, but because the cognitive perspective remains the most dominant within the field. Flower and Hayes' cognitive model of writing has, indeed, been substantially developed since its original 1981 form, and remains the most cited model to date. But, importantly, it has provided the basis for many important pieces of writing research that inform much of the work conducted throughout this project, such as Bereiter & Scardamalia's (1987) work on novice and expert writers, and Kellogg's (1988; 1990; 1994) studies on working memory and drafting strategies in writing. It is for these reasons that this project is underpinned by the cognitive perspective of writing.

However, as will become apparent throughout the rest of the chapter, the cognitive paradigm in and of itself does not present a fully-comprehensive and holistic model of writing. As such, the thesis also draws on elements of the sociocultural and social cognitive perspectives of writing, with the goal of identifying important, but noticeably overlooked, factors contributing to some of the outcomes of writing described at the beginning of this section – development of understanding and effective communication (text quality). We start with a description of Flower and Hayes' (1981) cognitive model of writing, and an explanation of how it has been updated over time.

### ***An Overview of the Classical Problem-Solving Cognitive Perspectives of Writing***

Flower and Hayes' (1981) model aligns with the classical, cognitive theoretical perspective, which defines writing as a problem-solving task; writers utilise different components of the writing process to satisfy rhetorical goals (i.e., the goals of the task and the writer's own goals).

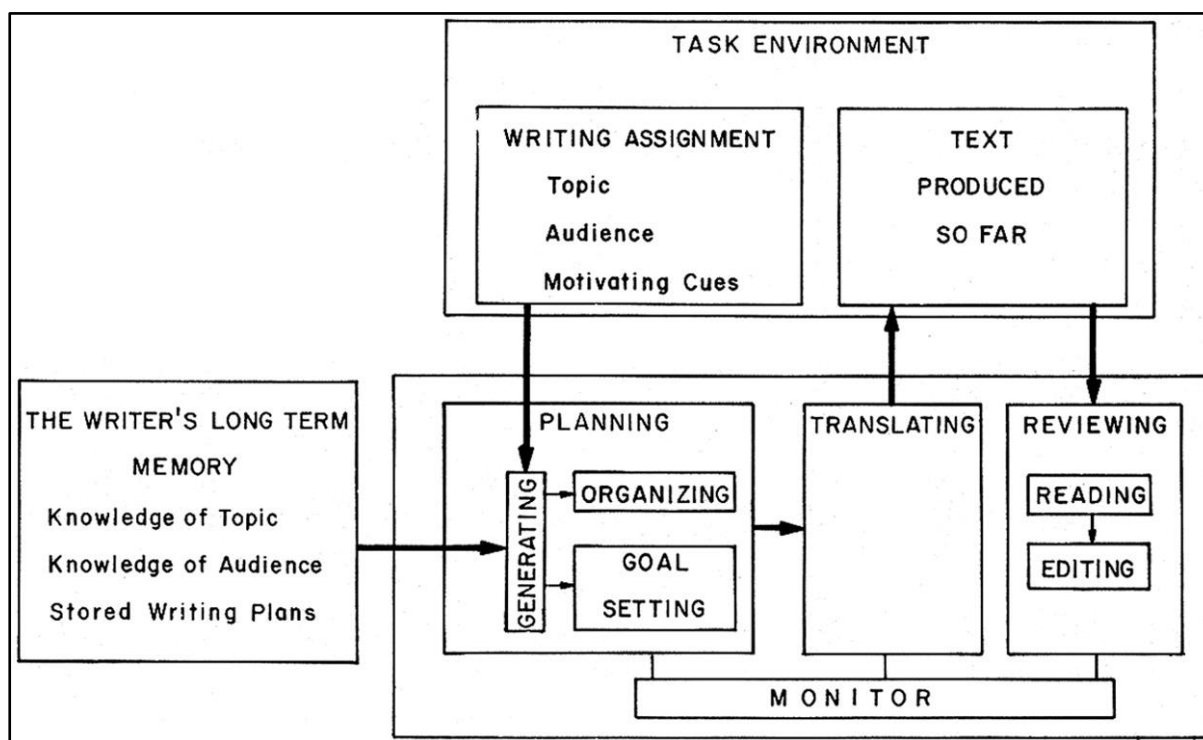
**Planning, translation and reviewing.** This model splits the writing process into three fundamental sub-processes: planning, translation, and reviewing. Planning is further broken down into three sub-categories: generating, organisation and goal setting. The model suggests that the writer generates ideas by retrieving information from long-term memory and the task environment. The organising component of planning then arranges thoughts to satisfy the third element of planning, goal

setting. Goal setting specifies the aims towards which the generating and organising processes are directed.

After planning, translation refers to the writer putting those ideas into readable text. Finally, reviewing involves reading and editing written text. The monitor controls and shifts attention to coordinate these fundamental processes. Thus, planning, translation, and reviewing do not necessarily happen linearly - these processes can occur interchangeably and simultaneously.

**Figure 1**

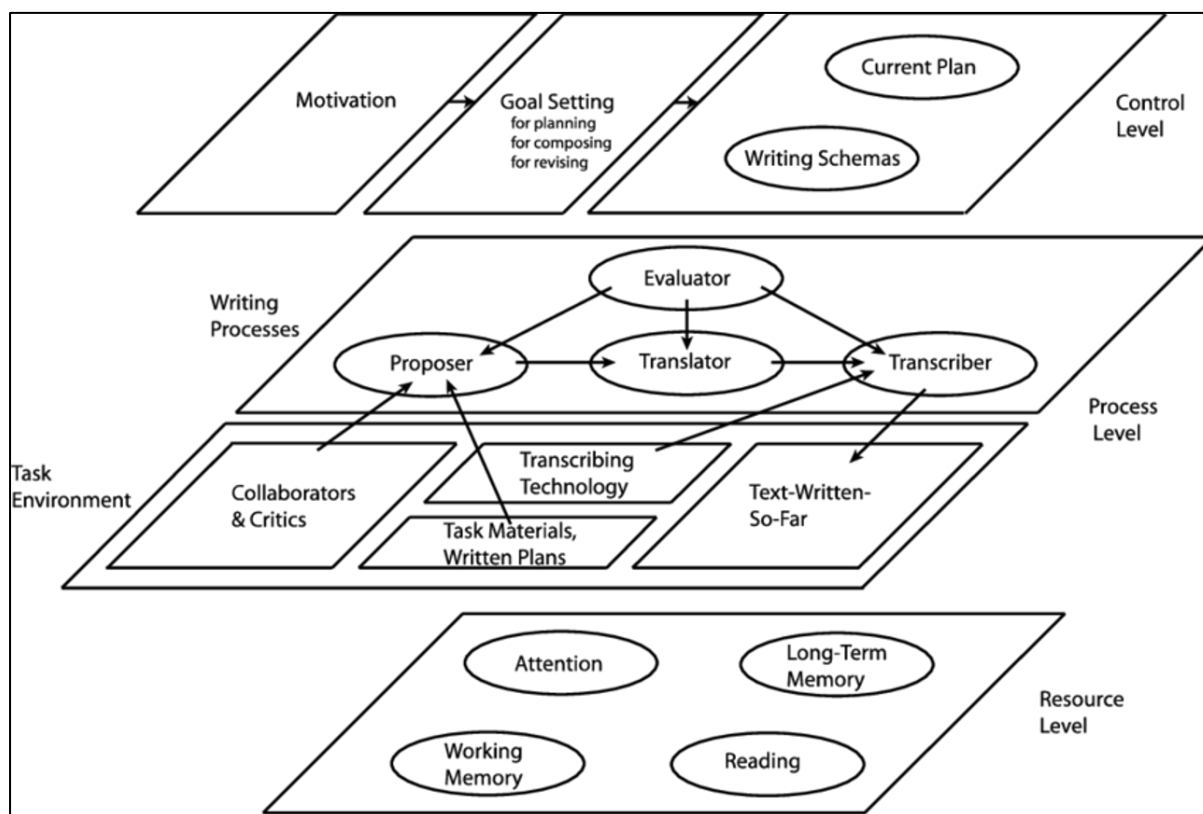
*Original Hayes-Flower (1980) Model as Shown in Hayes (2012, p.370). Reproduced with Permission of the Rights Holder.*



**An Updated View of the Problem-Solving Account of Writing.** Whilst the terms 'planning', 'translation' and 'reviewing' (sometimes referred to as 'revision') are still commonly used in writing research, the original Flower & Hayes (1981) model of writing has been revised over the years. In the most recent version from Hayes (2012; see Figure 2) some elements of the model have been adapted, and there are also some new additions.

**Figure 2**

*Hayes' (2012, p.371) Cognitive Model of Writing. Reproduced with Permission of the Rights Holder.*



**The relationship between planning and revision processes with writing outcomes.** The first thing to note in Hayes (2012) is that the model now views planning and reviewing/revision as independent writing activities, rather than basic components of the writing process, because these activities themselves often utilise the act of writing. For example, planning can involve individuals creating a written outline to inform a text in advance of composing said text (e.g. writing a plan for an essay). Similarly, revision of text may involve redacting existing text and then adding to or replacing it with new text (e.g. rewriting a sentence).

Having planning and revision recognised as independent writing activities in themselves within the problem-solving account of writing is important because how the writer engages with these activities has an implication on their outcomes. That is, the way the writer chooses to draft their text has an impact on how they are able to meet the rhetorical goals of the writing task. For example,

planning can be utilised as an activity that is separated from the rest of the general writing process, completed *before* the writing of the main text. Revision can be utilised as an activity that is separated from the rest of the general writing task, to be completed *after* writing a draft of the text.

In fact, it has become standard for writers to create explicit plans for their texts before writing in many instances. For example, students are commonly encouraged to make outline plans for university essays before drafting their texts. Similarly, revision has been utilised as a task that is separated from the main writing process. For example, students may be encouraged to write rough drafts of essays before systematically going back through their texts and editing them into a finalised version. It is important, then, to understand how systematic compartmentalisation of the writing process might have an impact on writing outcomes.

Kellogg (1988) argued that writing is a very cognitively demanding task due to the coordination of several processes. He suggested that coordinating multiple writing processes could result in the writer's cognitive overload - where an individual's working memory capacity is exceeded. Thus, their ability to complete a task successfully is hindered (Sweller, 1994). Therefore, researchers have investigated methods for breaking down the writing process and combining it in different ways to reduce cognitive load. For example, Kellogg (1988) tested the utility of outline drafting versus rough drafting. For outline drafting, the writer separated the planning and translation components of writing by generating their ideas in a plan before composing their texts. Kellogg thought that this would help reduce the attention that writers gave to planning during the actual writing task, instead allowing them to concentrate on translation and revision processes. For rough drafting, the writers focused directly on translation without being concerned about how well expressed their texts were. They did not create any form of a plan before writing but instead wrote freely. After completing their rough drafts, though, the writers spent time reviewing and revising their texts. Rough drafting, hence, separated the processes of translation and revision.

This study showed that whilst outline planning had a positive effect on text quality, rough drafting had no impact on quality. The results imply that outlining strategies help improve text quality because planning is separated from the translation process used during text production, reducing cognitive load throughout. These findings are interesting, however, because they imply that separating

planning from translation has a better effect on text quality than separating translation from revision. However, both techniques should reduce cognitive load in theory, so it is difficult to understand why this discrepancy occurs. Galbraith and Torrance (2004) suggested that one reason could be that the crucial factor in reducing cognitive load throughout writing is separating the organisation component of writing from other processes. Whilst outline planning does this (because the writer is forced to think of ideas for their text up front) a rough drafting strategy does not. Galbraith and Torrance (2004) also showed that in some instances, revision-based drafting strategies could have a beneficial impact on text quality (i.e. when used in an iterative drafting context. That is, the writer creates multiple drafts of a text, but without sight of each previous draft. This prompts higher levels of global reorganisation of ideas contained in the initial text in subsequent drafts). Nevertheless, an outline planning strategy was still seen to have the most positive impact on quality.

Given that the research presented here tends to indicate that an outline planning strategy is more beneficial than a rough/revision drafting strategy in improving text quality, one might wonder why anybody would ever use a revision strategy. Indeed, some evidence implies that matching the writing strategy that an individual employs during a writing task with their preferred writing strategy (whether that be a planning or revision-based strategy) may be beneficial for improving outcomes related to learning. This work will be discussed in detail later in the chapter (see pages 55 - 59).

Maintaining focus on Kellogg's work on planning strategies for now, though, Kellogg (1990) crucially demonstrated that even differentiation in the way that writers used planning strategies alone could be seen to have an impact on text quality. Specifically, he observed that when writers created *hierarchically* organised outline plans before writing, (i.e. they made an outline plan in which their ideas were ordered, corresponding to the order of the ideas presented in the final text), this had an even bigger impact on text quality than when using a general outline planning drafting technique. This suggests that how a writer chooses to *organise* their ideas is perhaps just as important as formulating ideas in advance of text production for improving overall quality.

**The control-level: the relationship between dispositional and externally driven processes with writing outcomes.** Kellogg's (1990) research indicated that organisation was an important aspect in the planning component of writing. Organisation is indeed also seen as a central component

of the problem-solving account of writing in the form of goal setting, which may help to explain why it has a positive impact on text quality. Planning and revision processes are directed by goal-setting, and goal-setting differs between writers dependent on what they consider to be the main objective of the writing task. Objectives can be intrinsically or extrinsically motivated. For example, some individuals may prioritise clear communication for *external* audiences as a primary goal within their texts, whereas others may consider it important to develop their *implicit* understanding of the writing topic.

Motivation within itself is an important control-level component within the Hayes (2012) model of writing because it drives an individual's willingness to write. Lack of motivation has a negative impact on writing engagement, which subsequently affects how successfully an individual can carry out a writing task. Along with motivation and goal setting, the control level of the model includes the writer's current plan for writing and their writing schemas. Schemas are the writer's internalised depictions of what specific tasks should look like. Although they may share similarities, a writing schema for an essay will inherently be different to a writing schema for a newspaper article. However, writing schemas are malleable. That is, as an individual becomes more experienced with writing, their schemas will be updated to reflect their new understanding of what a certain writing task should consist of. Consequently, writing schemas influence how the writer carries out the writing process in order to satisfy rhetorical goals.

As summarised here, many of the components in the control-level tier of the Hayes (2012) cognitive model are associated with the writer's personal approach to writing, and so the relationships model components have with specific writing outcomes will vary considerably between writers due to individual differences. As will become clearer later in this chapter, the components of the control-tier are likely to play a significant role in an individual's *writing beliefs* - an important socio-cognitive concept examined throughout this thesis, which capture an individual's thoughts about how writing should be approached.

**The process-tier: the proposer, translator and transcriber mechanisms of writing.** For now, though, we move to the writing process level of the Hayes (2012) model. There are now four components that form the 'mechanical' process of writing. First, the proposer puts forward the ideas

the writer wants to express. These ideas are formulated through demanding cognitive processes (i.e. planning of new content and evaluation of previous content to inform new ideas). Next, the translator takes these ideas and converts them into linguistic strings. Finally, the transcriber expresses these linguistic strings in the visual form of written script. Every stage of the sequence is reviewed by an evaluating component (similar to the monitor), allowing the writer to revise and edit their text where necessary.

These four mechanisms are expressed somewhat in the original model (Flower & Hayes, 1981) but have been elaborated in the most recent version from Hayes (2012). For example, the ‘planning’ component from the original model has now been replaced with the ‘proposer’. Ideas from the proposer are influenced by a range of factors, including explicitly written plans, internally composed mental plans stored in short term memory, and elements of the task environment (e.g., critical feedback from readers, ideas from collaborators, and materials used to inform the content of the text product). The ‘translator’ and ‘transcriber’ essentially mirror the ‘translation’ component of the original Flower and Hayes (1981) model, but as two separate processes rather than one. Finally, the evaluator monitors how the proposer, translator and transcriber are integrated with one another. The way in which these components are balanced has an effect on how the writer plans for new text and revises pre-existing content.

Another important component of the process tier is the task environment which is comprised of external factors that influence how the basic writing process is conducted. As already mentioned, this includes the influence of collaborators, critics and task materials on the proposer, but also the influence of transcribing technology and previously written text on the transcriber.

**The resource-tier: basic cognitive functions and their roles within writing.** Finally, the lowest tier in the model is the resource level. This tier incorporates attention, working memory, long-term memory, and reading. These are the base resources required for the act of writing to take place. As in the original model, attention (i.e. selectively processing information) is divided across all of the basic writing mechanisms to varying degrees and working memory temporarily stores and manipulates information from short-term and long-term memory so that it can be utilised during writing. However, limits on individuals’ working memory capacity ultimately affect how successfully

one can carry out a writing task. Finally, reading is used to support the writer in reviewing their texts, and long-term memory is the store from which writers can retrieve ideas for their texts.

**The strengths and weaknesses of the Flower and Hayes models of writing.** To summarise, the general notion consistent throughout all versions of the Flower and Hayes model of writing (Flower & Hayes, 1980, 1981; Hayes, 2012; Hayes & Flower, 1986) is that writing is a problem-solving activity. Hence, writers try to combine the components of the basic writing process (which are influenced by the control level, the task environment and the resource level) in a way that most effectively meet the rhetorical goals of the task.

As well as the evidence on planning, organisation and cognitive load from Kellogg (1988;1994;1990), many other studies indeed suggest the existence of the systems within the model. For example, research from Hayes & Chenoweth (2006) indicates that writing processes compete for cognitive resources. They found that the transcription process was slowed down when verbal working memory capacity was reduced, indicating that lack of working memory resources negatively impacted the writer's ability to transcribe text. Research also shows the importance of motivation in successfully carrying out writing. Hayes, Schriver, Hill, and Hatch (1999) found that college students assessed as 'basic' writers were less engaged in an activity supposed to improve their writing skills than the average honours student. These results imply that low motivation could be linked to poorer writing.

Perhaps most importantly, and in line with the notion that writing is a problem-solving activity, evidence implies that the way in which the writer combines the various components of the Hayes (2012) model of writing throughout the duration of a writing task has an impact on text quality. Breetvelt, van den Bergh and Rijlaarsdam (1994) found that during a think-aloud writing task, the effects of several key components of the Hayes model on text quality, including goal setting, generating ideas, transcribing, evaluation and revision were not consistent over the duration of the writing task. For example, when ninth-graders were goal setting in the first part of the writing task, there was no relationship between these processes and text quality. However, during the second part of the writing task, goal setting had a positive relationship with text quality. During the final part, rereading (reviewing) processes had a positive effect on text quality, but this relationship was not

observed in the first and second stages of the writing task. These findings demonstrate that in order to produce a high-quality text which effectively communicates the writer's goals, the individual must combine components of writing in a specific way at different stages of the task. This further demonstrates the problem-solving nature of writing, because the success of the final product is dependent on how the writer has combined key components throughout the writing process, and this can vary dependent on the stage of writing that an individual is in.

Yet, one problematic feature of the Flower and Hayes model prevents it from providing a comprehensive view of the writing process. That is, it does not adequately describe how knowledge and understanding develops over the course of writing, and how writers discover new ideas through the means of writing. The classical, problem-solving perspective assumes that in learning to write more expertly, through controlled problem-solving processes, individuals get better at developing their understanding of the topics they are writing about. Under this view, discovery is a by-product of the writing process (and, subsequently, any explanation of discovery within the model is vague). However, this seems to contradict a claim made in the paper *The Cognition of Discovery* from Flower and Hayes' (1980, p.22) which implies that discovery should be considered a central feature of the writing process:

“The mythology of discovery doesn't warn the writer that he or she must often build or create new concepts out of the raw material of experience; nor does it tell the writer how to do it. And yet, this act of creating ideas, not finding them, is at the heart of significant writing.”

One could argue that the Flower and Hayes model of writing was not intended to model processes of discovery, but rather describe the basic process of putting thoughts into words, and the factors that contribute to that process. But, in Flower and Hayes' own words, discovery is at the “heart of significant writing”, consequently implying that discovery is (or should be) a major contributing component within their model.

In fact, Flower and Hayes (1980) contend that ‘discovery’ is actually a process of ‘invention’. New ideas have to be constructed through deliberate problem-solving directed towards rhetorical goals, rather than through a spontaneous and vague process of ‘finding’. Invention happens because, through the process of writing, individuals adapt their existing ideas to rhetorical goals, and

communicative goals force writers to re-evaluate and reorganise (transform) their knowledge. Hence, this is evidence that Flower and Hayes' (1981) model was actually set up to explain how discovery, invention, or knowledge transformation occurs, not as a *feature* of the writing process, but *as* the writing process in and of itself.

What Flower and Hayes neglected in the formulation of their model was explicitly testing whether writing led to a transformation of knowledge (i.e. whether it led to learning). Although their model predicts that writing should lead to learning, they do not specify under which conditions this would happen. They describe writing as a process of discovery, but this is not the same as modelling the *effects* of writing on the writer's change in knowledge and understanding. Hence, consideration of when writing might lead to more or less knowledge change is not discussed within the Flower and Hayes (1981) model, not because discovery is not a part of the model, but rather because it is such a general feature of the model that it is not treated as variable.

However, whilst Flower and Hayes, in the formulation of their model, did not test whether writing led to discovery, they do briefly allude to which mechanisms might be involved in this process in another paper (Flower & Hayes, 1984). Specifically, they suggest that semantic knowledge (general knowledge) can be expressed in line with the Collins and Quillian (1969) symbolic network theory, which describes knowledge representation as a network of hierarchically organised concepts (nodes) and relationships (labels). When combined through a serial search, the nodes and labels allow the writer to access prepositions about concepts and their relationships. Galbraith and Baaijen (2018, p.241) give the following as an example:

“To verify a proposition about whether robins can fly, activation spreads from the ‘robin’ node up to the ‘bird’ node and then to the ‘fly’ attribute of ‘bird’.”

Although Flower and Hayes do not explicitly discuss this in their 1984 paper, under a symbolic network theory of knowledge representation, the assumption would be that during the act of writing, the writer accesses their semantic knowledge through a process of hierarchical activation between nodes. However, because each node represents a concept in a *pre-established* semantic network, changes to knowledge can only involve the *reorganisation* of concepts existing within that network, rather than the creation of any new concepts. Arguably, this reorganisation could lead to the writer

combining concepts in different ways, which could result in a newly ‘discovered’ or ‘invented’ understanding of the writing topic at hand. To stress, though, this is not really explained in the Flower and Hayes model, so further clarification is needed.

To summarise, then, the Flower and Hayes (1981) cognitive model, and its further iterations, indeed characterise writing intrinsically as a process of discovery or knowledge transformation. But, because this is such a general assumption, they do not specify ways in which this might vary between writers, nor do they explicitly test the conditions under which knowledge does or does not change.

Consequently, and in response to the lack of an explanation of knowledge change in the problem-solving account of writing, Bereiter and Scardamalia (1987) introduced the knowledge telling and knowledge transforming framework, which combines the notions of knowledge retrieval from the Flower and Hayes (1981) model and those of knowledge development in Flower and Hayes (1984).

They explained the importance of considering knowledge representation in a writing model framework by emphasising that any writing task can present new challenges to the writer, leading them to question the knowledge they already hold in relation to the task. Accordingly, this may result in the writer needing to refine their thoughts to either clarify their knowledge or make it more appropriate to the goals of the writing task. Bereiter and Scardamalia suggested that individuals may take two approaches to representing knowledge through writing: knowledge telling and/or knowledge transforming. Knowledge telling involves the writer simply writing down what they already know. If an individual defines writing as a knowledge telling process, then they will not develop their understanding. Less experienced writers, for example young children, would tend to use a knowledge telling process. On the other hand, knowledge transforming involves the writer manipulating pre-existing knowledge to better suit the needs of the task’s communicative, rhetorical goals. When writers use a knowledge transforming process, their understanding around the task may, as a side effect, change, and this results in discovery (i.e. because, through the reworking of content to better suit the needs of the task, they make links not previously activated between concepts, furthering their understanding of the writing topic). More experienced writers would tend to use a knowledge transforming process.

The mechanisms of the knowledge transforming approach in writing can be explained by Flower and Hayes' (1984) symbolic network explanation of the representation of semantic knowledge. Hence, Bereiter and Scardamalia's knowledge telling/transforming model can be interpreted as an addition to the Flower and Hayes (1981) model, which makes the classical, problem-solving account of writing as a whole more holistic.

However, one fundamental flaw remains with the classical view, which is that even with the addition of Bereiter and Scardamalia's (1987) model, the problem-solving school of thought still does not adequately explain how writers produce *new* knowledge. Instead, and as explained already, there is a reliance on pre-existing knowledge being manipulated to meet rhetorical goals to improve understanding, but that is all that happens - a *reworking* of the content that already exists in an explicit semantic network. This may help to *clarify* a writer's understanding of a topic by activating links between concepts in a writer's semantic knowledge network, but one could argue that this process does not equate to discovery because that network of knowledge is already explicitly pre-defined.

### ***Romantic Perspectives of Writing***

Whilst the classical view of writing, as it stands, does not contain an adequate explanation of how writers develop new knowledge during a writing task, there is a perspective that puts the development of knowledge and understanding at the heart of its theories - the romantic view of writing. Most notably brought into the limelight by Britton et al. (1975), these models emphasise the importance of discovery within the writing process. In contrast to the classical models, they suggest that *spontaneous* production of text (i.e. text production with no prior planning and no reworking of text) is, by definition, a process of discovery. This directly contradicts Bereiter and Scardamalia's (1987) claims, which suggest that spontaneous text production is just a process of knowledge telling.

Britton and his colleagues argued that spontaneous text production is best facilitated when the writer focuses less on traditional writing conventions, such as spelling and grammar, because they interrupt the spontaneous language production process. Instead, they argued that spontaneous expression in writing is a fundamental, yet often neglected, part of the writing process and if writers

wish to be successful, they need to 'adapt [their] inventiveness and continue to rely on it rather than switching to some other mode of operating' (Britton et al., 1982, p.140). In essence, Britton was suggesting the freer an individual is to spontaneously produce text, the more likely they are to develop and discover new thoughts and understanding relating to a topic.

Yet, there are two major issues with the romantic perspective of writing. The first is that although it suggests writing is a process of discovery, it does not present a clear model for the mechanisms involved in that process. The second is that there is a flaw in the pragmatic logic of the romantic perspective: the assumption that the act of text production needs to be completely spontaneous and void of constraints (e.g. spelling, syntax and grammar) if the end product is to capture discovery at its most effective. This is problematic because unplanned, unrevised writing could result in a completely incomprehensible final product. Instead, the text would just reflect the structure of the writer's spontaneous thoughts as they occurred. Consequently, this implies that because spontaneous text production is the vehicle for discovery, only a revision-based drafting strategy would be compatible with both developing the writer's understanding and then, through revision and reorganisation, developing a coherent text. In fact, Britton and colleagues concede that a good, developed writing process should involve "hearing an inner voice dictating forms of the written language appropriate to the task in hand" (Britton et al., 1982, p.140), implying a recognition that successful writing cannot rely on a *completely* spontaneous process and will require adherence to general written conventions at some level. However, they do not go as far as to explain exactly where the balance should lie between complying to writing conventions and producing spontaneous text, and whether this is a process that should take place throughout the planning, drafting or revision of a piece of writing.

The classical and romantic perspectives of writing are indeed oppositional. The classical view treats writing as a systematic, problem-solving process, highlighting the need for constant evaluation throughout to produce effective pieces of written communication that meet rhetorical goals. However, it does not adequately explain how the act of discovery and the writer's change in understanding happens. The romantic perspective underlines the importance of discovery through writing, explaining that traditional conventions of writing need to be relaxed so that the writer can express themselves

spontaneously, resulting in maximum levels of discovery. Yet, the romantic account fails to explain the mechanisms involved in discovery and also how the writer organises content to meet the rhetorical goals of the task.

### **The Dual Process Model of Writing (Galbraith, 1999, 2009; Galbraith & Baaijen, 2018)**

In response to the lack of a clear explanation of knowledge representation and discovery in writing in both the classical and romantic views, Galbraith created the dual process model of writing (Galbraith, 1999, 2009; Galbraith & Baaijen, 2018). This model splits the act of writing into two systems. The first, knowledge-transforming, accepts the views of the classical perspective and implies that the writer draws content from long-term memory in a direct retrieval process. This means that the writer takes pre-existing ideas from their long-term memory and translates them into a written format. These ideas might be manipulated in working memory using a knowledge-transforming approach (Bereiter & Scardamalia, 1987) before transcription to meet the rhetorical goals of the writing task. However, Galbraith (2009) does not regard this alone as a discovery process, because it relies on explicit knowledge and hence by definition, no new ideas are created, only potentially manipulated and reorganised.

The second process of content creation in this model is based on discovery. The knowledge-constituting system relies on synthesising content rather than retrieval from long-term memory. Similar to Flower and Hayes' (1984) comments on semantic knowledge representation, this process relies on mechanisms described by a network-based view of knowledge representation. The fundamental difference, however, is that Galbraith uses the parallel distributed processing (PDP) model (McClelland and Rogers, 2003), rather than the symbolic network of Collins and Quillian (1969). In this model, propositions are synthesised, rather than already existing in an explicit network. Information gets processed by activating nodes and labels in parallel across layers of sub-conceptual units. The first layers consist of input units about items (e.g. 'daisy', 'robin', 'canary') and/or relations (e.g., 'is a', 'can', 'has'). Then, there is a layer of hidden units which combines the input unit

information. The final layer is the output units which correspond to all of the possible combinations of propositions that can be true of the input items and relations (e.g., 'grow', 'move', 'fly').

As illustrated above, the PDP model stores the same information as the Collins and Quillian (1969) network. What differs is the way in which the information is processed. The concepts in the network are not represented by an explicit, hierarchical network. They are instead momentary patterns of activation across all layers of representation units. Once the processing is completed, the activation disappears. Propositions about concepts are created on the basis of how strong the fleeting activations between units are, and this results in the output layer of the network, meaning that, unlike the Collins and Quillian (1969) network, knowledge is created incidentally and implicitly, rather than being formed through retrieval from an explicit semantic network stored in memory.

In the knowledge constituting system, the recognition of a new piece of knowledge is only made once expressed in explicit language. That idea is then fed back into the knowledge representation system to further inform the writer's disposition toward the topic (Galbraith & Baaijen, 2018). This process then repeats itself through a feedback loop.

It is essential to stress here that the sub-conceptual units making up the semantic knowledge network are not in themselves categorised as parts of knowledge, but instead, they are fragments of an individual's past experiences. Only when the fragments are synthesised according to the connections between them (i.e. through spontaneous sentence production) are they recognised as explicit ideas and, hence, new coherent knowledge. Like the romantic models of writing, the focus of the knowledge-constituting system is the invention of new ideas and spontaneous self-expression based on the writer's dispositional experiences.

How the writer balances the knowledge constituting and knowledge transforming systems determines whether the writing process is carried out successfully. Galbraith (2009, p.23) explains why this balance is necessary:

"The two processes are both necessary for effective writing. The knowledge-transforming system and the associated manipulation of mental models in working memory are required in order to explicitly organize the text and ensure that it satisfies rhetorical goals...The knowledge-constituting process synthesizes content activated

by the knowledge-retrieval process into explicit, connected sentences, and in the course of this may produce novel content that clarifies the writer's understanding...

The joint result of these two processes is the creation of a coherent knowledge object, which satisfies rhetorical goals."

Suppose the writer uses more of a knowledge-transforming approach to writing. In that case, the processes they use in writing centre around organisation to satisfy the writer's rhetorical goals, but they do not develop their understanding of the topic. Suppose instead that the writer uses more of a knowledge-constituting approach. In that case, the synthesis of ideas happens spontaneously, which means that the writer has to relax writing constraints to produce ideas freely. This spontaneity leads to an increased level of understanding about the topic, but the writer still needs to explicitly organise those ideas in comprehensible text production. As such, when the writer follows up that knowledge constitution process with a knowledge transforming approach, it enables them to organise their newly synthesised ideas to meet the task's rhetorical goals. If the writer masters the balance and interaction between knowledge-transforming and knowledge-constituting, the resulting text should capture the individual's implicit understanding of the topic whilst being able to exist independently from them.

Imperatively, Galbraith & Baaijen (2018) suggest that the dual process model of writing should not be considered an opposing theory to the Flower and Hayes model (Flower & Hayes, 1981; Hayes, 2012; Hayes & Flower, 1986), but rather an explanation of how new knowledge is generated and subjective understanding is developed, which can be seen as an addendum to their model. The dual process model helps provide the problem-solving account of writing with a comprehensive explanation of knowledge generating and retrieval processes in writing by indicating that there is conflict between reflective, controlled thought (problem solving/knowledge transforming) and spontaneous discovery (knowledge constitution) during text production, which the writer must learn to balance when producing a text.

***Strengths and Weaknesses of the Dual Process Model***

Whilst the dual process model provides a cohesive explanation of discovery, which was distinctly lacking in other cognitive models, there are some issues with the model as a whole. For example, it relies on the notion of feedback loops for taking newly synthesised ideas and relaying them to the writer's dispositional knowledge to further inform the synthesis of more new ideas. However, there is currently no existing literature in favour of a feedback loop system for knowledge synthesis, and so this part of the model remains wholly theoretical. Moreover, the model claims that content generation is a central feature of the writing process, through a combination of knowledge constituting and knowledge transforming processes. Yet, it says very little about how content generation actually maps onto text production. That is, the model is not detailed about the specific processes involved in sentence production. Similarly, there is an issue with Galbraith's (2009) claim, which is that knowledge transforming is not a process of discovery, because it relies on explicit knowledge and hence by definition, no new ideas are created, only potentially manipulated and reorganised. However, as will become clear in the review of literature supporting the dual process model, some evidence (Baaijen and Galbraith, 2018) clearly indicates that knowledge transforming is at least linked with increased ratings of knowledge (global linearity). So, this claim needs to be modified to accept that both knowledge transforming and knowledge constituting can change understanding, but in different ways. Finally, the dual process model has been far less researched than the Flower and Hayes (Flower & Hayes, 1980, 1981; Hayes, 2012; Hayes & Flower, 1986) models of writing, and so empirical evidence for the model is somewhat lacking. This, however, is not to say that no evidence supporting the model exists.

For example, some studies have provided evidence of the knowledge transforming system. To reiterate, the knowledge transforming system relies on writers drawing information from long-term memory through direct retrieval or the reorganisation of pre-existing content. As discussed at the beginning of this chapter, Kellogg (1990) found that outline drafting strategies, which were designed to enable individuals to write more reflectively (i.e. by reducing cognitive load), increased the quality of the texts they produced. Specifically, students who created ordered idea lists corresponding to the

order of their final texts wrote better quality essays than students who used no pre-planning activity, or just wrote their ideas down in an unordered list. The knowledge transforming system of the dual process model would indeed predict better quality after using a hierarchical planning technique, like the one employed in this study, because the writer would have been able to reorganise their pre-existing knowledge to better meet the rhetorical goals of the task. In the same way, Kellogg's findings suggest that effective writing involves organising ideas in line with the aims of the task.

The dual process model would also predict that, if the knowledge-transforming model was correct, organised planning could also lead to an improvement in the writer's subjective understanding of the writing topic, if it were used in combination with knowledge-constitution. This is because it would enable the writer to transform their thoughts more effectively. Whilst Kellogg (1990) did not examine the impact of hierarchical planning strategies on subjective understanding development, evidence from Klein et al. (2017) alludes to the notion that the organisation of ideas to satisfy rhetorical goals does indeed have an impact on change in understanding during writing.

In their study, child writers were asked to produce texts persuading readers about how they should classify different animals. The participants were randomly assigned to one of three conditions. In one group, the children received a rhetorical prompt before writing, explicitly instructing them to incorporate rhetorical goals into their texts. These prompts encouraged them to provide reasons for their classifications, come up with another potential classification and reasons for it, and provide a counter-argument for their second alternative classification. The children who received a rhetorical prompt before writing achieved better scores on an objective test of their knowledge about animal classifications after writing *and* produced better quality texts than children who received either content prompts (i.e. prompts about the *type* of content they could consider writing) or a control prompt (i.e. standard instructions for the task).

Whilst this study did not involve any explicit planning activity, the findings indicate that organising writing on the basis of rhetorical goals can further a writer's subjective understanding of a topic, as well as improving their writing quality. Hypothetically, this is perhaps because throughout the process of reorganising knowledge to suit the requirements of the task, the writers were able to synthesise new propositions about these concepts that they had not considered before, supporting the

idea that knowledge constitution used in combination with knowledge-transforming increases a writer's understanding of the writing topic.

Kellogg's (1990) and Klein et al.'s. (2017) findings are consistent with the knowledge transforming system of the dual process model. They suggest that generating and organising content is associated with an increase in subjective understanding when directed towards rhetorical goals and that subjective understanding is also increased when a writing strategy is used that enables writers to focus on reflective problem solving, free from the requirement to combine this simultaneously with expressing thought in fully formed text.

However, neither study *directly* tests the knowledge constituting component of the model, so empirical support for this component is lacking. In fact, another major criticism of the dual process model is that there is little evidence for the knowledge constituting system. It is far less well studied than the knowledge transforming system.

However, Galbraith (1992) produced some evidence consistent with the underlying theory of the knowledge-constituting system. In this study, he examined writers who were classified as either high-self monitors (i.e. external, rhetorical clues guided their behaviour) or low-self monitors (i.e. behaviour was guided by their dispositional beliefs). The writers were assigned to either a full-text essay condition, or a making notes in preparation to write an essay condition. The results indicated that high-self monitors produced more ideas after writing notes than low-self monitors. This was expected, given that the writing of high self-monitors is thought to be directed towards rhetorical goals, and so change in understanding will be activated by rhetorical problem solving for these writers (knowledge transforming). The low-self monitors, however, produced more ideas when they wrote a full text, rather than writing notes. Galbraith argued that low-self monitors created more ideas when writing essays in comparison to writing notes because their essay writing reflected a 'dispositional spelling out' (Galbraith, 1999, p.140) – that is, the writers' subjective understanding developed as they spontaneously articulated their thoughts during text production (knowledge constitution), and this was better elicited in an extended writing task than a fragmented note taking task.

Later work from Galbraith (1999) extended on the original study by introducing two different planning scenarios. Low and high self-monitors were asked to write essays about a given topic, but

before doing so, they were assigned to one of three planning strategies, either i) unplanned, in which writers just wrote down their thoughts as they occurred, without being concerned about the essay's organisation, ii) synthetically planned, in which writers had five minutes after making an initial list of ideas to write a single sentence summing up their main response to the writing topic, which they could use to inform their essay, or iii) outline planned, in which participants were given five minutes to construct an organised outline plan for their essays after making an initial list of ideas. For the synthetic and outline conditions, participants were told that their essays should be well-organised, but that they did not have to worry about how well the text was expressed.

The most important finding from this study was that for low self-monitors writing under synthetic planning guidelines, subjective understanding development was related to the number of new ideas they produced throughout writing their essays. The synthetic condition was the only condition for which this relationship was observed, and so these findings support the knowledge-constituting system's assumptions because they suggest that new ideas and, consequently, development in subjective understanding were the result of spontaneous, disposition-driven text production.

A later study by Galbraith, Torrance and Hallam (2006) extended the design of Galbraith (1999) by requiring the synthetic and outline conditions to also used different writing strategies. Specifically, the synthetic condition was required to write a rough draft of an essay, in which they had to write down their thoughts as they occurred, but not worry about how well organised or expressed the final texts were. On the other hand, the writers in the outline condition were told to write a well-structured essay, in which they communicated their ideas as clearly as they could to the reader, but without worrying too much about the mechanical features of text composition (e.g. spelling; note that half of the unplanned control group were told to write rough drafts, and half were told to write well-structured drafts). The study yielded similar findings to Galbraith (1999) in that for the synthetic/rough drafting condition, low-self monitors produced more new ideas than high self-monitors. Again, this implies that disposition-guided text production is an active knowledge-constituting process, rather than a knowledge telling process, and that the process is inhibited when rhetorical goals (i.e. the prompts for the outline/well-structured essay condition) are brought into the

picture. Additionally, in the planned essay conditions where rhetorical constraints were raised, high self-monitors produced more ideas than low self-monitors, which was to be expected given that high self-monitors are driven by rhetorical goals.

As a whole, then, the findings from Galbraith (1992;1999) and Galbraith et al., (2006) provide evidence in line with the knowledge transforming and knowledge constituting systems of the dual process model. Evidence for the knowledge transforming system comes from high-self monitors producing more ideas during note taking and writing under outline planning contexts because their processes of discovery are driven by rhetorical goals. On the other hand, low self-monitors produce more ideas under synthetic planning/rough drafting contexts because discovery is enhanced by disposition-driven spontaneous text production.

However, there are some important points worth considering that mean that the current evidence for the dual process model is not as robust as it could be. The first point is that participants bring their own definitions of writing into the writing task, and these will influence how the writer chooses to carry out the task, regardless of which condition they have been assigned to. In fact, such individual differences between writers may moderate any effect that condition has on text quality and change in subjective understanding. Galbraith and his colleagues (1992; 1999; 2006) went some way to looking at this by investigating individual differences between low and high self-monitors. The issue, however, is that self-monitoring is not directly related to writing processes, and so it would be better to use an individual difference measure that is related to writing.

The second point worth considering is that these studies say little about how the processes involved in writing composition relate to the two systems within the dual-process model. That is, they say nothing about what engaging with knowledge transforming looks like in comparison to engaging with knowledge constituting. As already mentioned, the dual-process model predicts that engaging in these two systems would have different effects on how individuals write sentences (spontaneous or controlled) and form the global structure of their texts (linearly or with high levels of revision). However, in the studies described here, relationship between engagement in the two systems and writing processes is only *inferred*, not formally investigated.

Consequently, in the second half of this literature review, we look at the impact that differences in approaches to the writing process may have on text quality and subjective understanding. We then explore how the processes involved in text composition may map onto the knowledge constituting and knowledge transforming components of the dual process model.

### ***Individual Differences in Approaches to Writing: Writing Beliefs and Writing Strategies***

When introducing the Hayes (2012) model of writing, we discussed how individual differences may lead to discrepancies in how writers approach their writing tasks and what they perceive to be the key objectives of these tasks. Indeed, research on individual differences in writing is well established. Yet, in the evidence supporting the dual process model, there is limited discussion about how these individual differences may influence the extent to which writers engage with the knowledge telling and knowledge transforming systems, and the implications this could have on text quality and subjective understanding development outcomes.

**Research on writing beliefs.** One of the main areas in which individual differences in writing has been examined is in research on writing beliefs. This work is underpinned by Bandura's (1986) social learning theory, which states that individuals gain knowledge and understanding through vicarious learning, which then informs their behaviour. Yet, an individual's knowledge and understanding may not be evident in their behaviour alone and extraneous, motivational factors, for example beliefs, may affect if and how an individual exhibits their knowledge.

Before social learning theory was incorporated into writing research, it had an influence on reading research. Schraw & Bruning (1996) discovered that individuals held implicit beliefs about reading and these were implicated in how they engaged with texts. They found that thoughts about reading could be categorised into two factors independent from one another – transmissional and transactional beliefs. Readers with high-transmissional beliefs thought that text interpretation was transmitted to the reader from the text/author. They also focused on the information explicitly stated in the text, rather than developing an understanding through reading between the lines. People with high-transactional beliefs, however, developed their understanding of a text via a transaction between

the text, the reader, and the author. This was constructed not only with the information explicitly available in the text, but also through their personal inferences about the content. As these factors were found to be independent of each other, some readers were observed to have simultaneously high and low levels of transactional and transmissional beliefs. Schraw and Bruning (1996) found that readers who had high-transactional beliefs were generally better at engaging emotionally and cognitively with texts, leading to a deeper understanding of what they were reading. Conversely, readers who had high-transmissional beliefs engaged more objectively with their texts, meaning their interpretations of extracts was limited to what they ‘factually’ understood from them.

Following on from this work, White & Bruning (2005) tried to apply the transactional versus transmissional beliefs structure to writing. Similar to how the two domains related to reading, writers with high-transmissional beliefs had lower affective and cognitive engagement with writing activities, and treated writing as a process of translating their pre-existing knowledge directly into text (example item: *“One of my writing goals is to have to make as few changes as possible”*). Conversely, writers with high-transactional beliefs had high affective and cognitive engagement in writing, and treated writing as a transaction with the text in which their understanding developed and was elaborated as they produced it (example item: *“Writing helps me understand better what I am writing about”*). The researchers also found that writers with high-transactional beliefs produced better quality texts than writers with low-transactional beliefs in terms of organisation, idea-content development, conventions, voice and overall quality.

While White & Bruning (2005) were the first to discuss writing beliefs explicitly, their transmissional and transactional factors are reminiscent of Bereiter & Scardamalia's (1987) knowledge telling versus knowledge transforming approaches to writing. Knowledge telling essentially maps onto transmissional beliefs because they both reflect a limited engagement within text production and relatively direct relations between knowledge in long-term memory and what is expressed in the text. Knowledge transforming and transactional beliefs, however, both reflect higher levels of engagement within text production, and a belief that understanding evolves as the text is produced.

Work on writing beliefs was extended by Sanders-Reio et al. (2014). They examined transactional, transmissional, recursive process and audience orientation beliefs within writing. Recursive process beliefs capture attitudes towards revising and rewriting text versus producing it in a purely linear fashion. Although Sanders-Reio et al. (2014) examined recursive process beliefs independently of any other belief type, it is necessary to note that White and Bruning (2005) also included recursive process beliefs as part of the transactional factor in their scale. However, audience orientation was an entirely novel concept introduced by Sanders-Reio and colleagues. Audience orientation refers to the idea that expert writers adapt their writing to suit their readers' needs. This idea is in line with the problem-solving model and the dual process model - both highlight the need for a coherent knowledge object that can exist independently from the writer so that the audience can understand it without assistance.

Sanders-Reio and colleagues' (2014) scale was administered to 738 undergraduates and was used to predict their grades on a class paper. The researchers found that the students' writing beliefs did indeed predict variance in their paper scores. Audience orientation was the strongest positive predictor of paper grade, while transmissional beliefs were the strongest negative predictor. Recursive process was also a significant predictor of paper score but this was less strong. Conversely to White & Bruning (2005), the researchers found that transactional beliefs did not relate to writing performance, even though they were still positively associated with writing self-efficacy and enjoyment of writing.

These results could indicate that transactional beliefs may encourage positive feelings towards writing, rather than directly relating to writing performance, motivating writers to keep on task productively for longer. It could also be that transactional beliefs are more about developing understanding, rather than producing high-quality texts, which is why they were not directly related to writing quality.

Importantly, though, both the White and Bruning (2005) and Sanders-Reio et al. (2014) studies do not consider two important things. Firstly, although Sanders-Reio et al. (2014) extend the work of White and Bruning (2005) to include beliefs about revision, they do not include any beliefs about planning. Yet, under the classical view of writing, planning would be considered an equally

important strategy as revision in helping the writer achieve their own goals and the goals of the task at hand. To date, no writing beliefs scale examines planning as a beliefs concept.

Secondly, neither White and Bruning (2005) or Sanders-Reio et al. (2014) examined the type of writing strategy that participants used to produce their text. However, it is very possible that writing beliefs might interact with the type of writing strategies that individuals employ throughout a writing task to have varying effects on outcomes such as text quality and subjective understanding development. In order to understand *why* it may be important to consider the role of writing strategies and writing beliefs collectively, it is necessary to delve deeper into the research that has been conducted on writing strategies.

**Research on writing strategies.** Along with writing beliefs, another main area in which individual differences can play a role in how writers approach the writing process is in the writing strategies they employ throughout a task. Writing strategies can be defined as the methods writers use to plan, compose, and revise text, amongst other writing related activities (Torrance et al., 2000 in Penuelez, 2012, p. 83). Similar to the work conducted on writing beliefs, researchers have found that individuals tend to gravitate towards specific writing strategy preferences.

Torrance, Thomas and & Robinson (1994) found that students participating in social science degrees could be split into three distinct groups in terms of their writing strategy preferences: planners, who planned extensively for a writing task, but then made very few revisions; revisers, who developed the content and the structure of their texts through the use of extensive revision, rather than advance planning; and mixed-strategy writers, who used both advance planning and extensive revision to compose texts. Torrance and his colleagues found that the students' preferences towards specific writing strategies had implications on their writing productivity and the extent to which they found writing difficult. Planners were observed to be more productive than revisers and mixed-strategy writers. Additionally, planners and revisers found writing to be of a similar difficulty to one another, with mixed-strategy writers finding writing most difficult. The findings suggested that a strong preference towards a *specific* strategy, rather than using a mix of strategies, could potentially be more beneficial for writers.

Subsequently, researchers became interested in how certain writing strategies could relate to specific writing outcomes. Kellogg (1988;1990;1994) had already explored how using a strategy that broke the process of writing into planning followed by drafting was seen to lead to better text quality outcomes, and more so than using an approach that divided the drafting phase from the revision phase of writing. However, Galbraith and Torrance (2004) criticised the version of the revision strategy tested by Kellogg, because the writers were not given detailed instruction about how to write using a revision strategy. Instead, they were just instructed to read through their first draft, highlight the main changes they would like to make to it, and then sum up their overall point. As such, Galbraith and Torrance (2004) recommended that future studies should continue to investigate the utility of revision-based strategies by incorporating more detailed instructions about how to create such drafts (note that *preliminary* evidence from this study indicated that a multiple-drafting strategy, in which writers create not just one, but multiple redrafts of text, starting with a spontaneous, unorganised draft and ending with a rhetorically appropriate final version, may help to develop the writer's understanding as well as produce a good quality text).

Generally, though, more recent work on writing strategies has focused predominantly on two aspects: whether the utility of writing strategies is related to a writer's preference for that strategy and, furthermore, whether various writing strategies lead to differences in writing-to-learn related outcomes, not just text quality. Much of this work has been conducted by Kieft and colleagues.

For example, Kieft, Rijlaarsdam and van den Bergh (2008) hypothesised that writing using an individual's preferred strategy could have one of two effects. In line with Kellogg's (1988) assumptions, they implied that if a writing task was matched to an individual's preferred writing strategy, this could reduce their cognitive load which could subsequently improve their discovery outcomes. Alternatively, Kieft and colleagues suggested that individuals could benefit from learning how to use strategies they do not usually gravitate to, as this could help to supplement their skills in these techniques, which could also ultimately improve writing-to-learn outcomes.

Consequently, these researchers conducted a study in which high school students were asked to write argumentative texts about short stories they had read. The participants were assigned to one of two short writing courses, which they had completed prior to writing their texts. One group

participated in a planning strategy course, whilst the other group took part in a revision strategy course. Both courses highlighted the importance of discovery in writing, but the planning strategy prompted idea generation through written plans developed *prior* to essay writing, whereas the revising strategy promoted the use of spontaneous, free-writing to develop ideas, and then revising and re-writing the text to produce a more polished version of it. The participants also filled in a questionnaire which was designed to measure their writing strategy preference (either for a planning or revision strategy).

The findings of the study indicated that for students with a preference for the revising strategy, the revising condition resulted in a better argumentative literary interpretation text. These findings are somewhat in line with the assumptions of the dual process model, because they suggest that writers with a strong preference for revision *after* spontaneous text production do indeed discover more when they use a revision strategy, implying that they have engaged with the knowledge constituting system followed by the knowledge transforming system. However, there was no correlation between preference for planning strategy in the planning condition and its effects on literary interpretations, implying that the matching of writing strategy preference to strategy used may only be effective when in reference to a revision strategy.

Interestingly, the study's authors emphasise that the lack of effect observed in combining a planning strategy preference with using a planning-based condition on participant's literary interpretations could be due to flaws with the scale they used to measure the writers' strategy preferences. They imply that the students who reported a preference for the planning strategy on the questionnaire items could have been doing this due to social desirability (because in writing instruction, it is commonplace for students to be taught to plan before composing a text), rather than because it was the strategy technique they would usually use.

However, it could just be the case that the match-hypothesis for writing strategies (i.e. aligning a writer's strategy preference with the strategy they actually use) is not a reliable method for improving writing outcomes. This does not mean, though, that strategy preferences are not important in predicting writing outcomes, particularly in relation to understanding development goals. As already discussed, a study by Torrance et al. (1994) found that when writers had a strong preference

towards a single, rather than mixed, strategy, this improved intermediary writing outcomes, like productivity. In a similar vein, Kieft, Rijlaarsdam, Galbraith & van den Bergh (2011) found that it is perhaps the *intensity* of an individual's preference towards a writing strategy that impacts knowledge development outcomes, rather than matching the type of strategy used with their preference.

In this study, different writing instruction courses produced varying effects on high school students' writing skill in a literary interpretation task. Crucially, though, this relationship was dependent on the *extent* to which students used their preferred writing strategies. They found that students who had a *strong* preference for planning or revision had better writing skill in literary interpretation tasks when they received writing instruction based on a planning strategy. On the other hand, students with a *weak* preference for planning or revising were better off when they received instruction based on a revising strategy. The findings suggest that using a planning strategy may be better for writers who already tend to plan and/or revise, whereas using a revision strategy may help to improve writing performance in those who have an underdeveloped writing strategy, (i.e. they neither plan nor revise during writing).

Again, though, the authors note that the items they used to capture writers' preferences towards a planning or revision strategy (taken from Janssen and Overmaat's (1990) process scale) could have been flawed. They state that the revision scale they employed may have measured the extent to which writers monitored their text during writing, rather than how much they revised their texts.

It is clear, then, that the evidence on writers' individual differences in writing strategy preferences, and how these are related to outcomes in writing-to-learn tasks is mixed. Crucially, though, this could be due to the items being used to measure strategy preferences in these studies not being fit for purpose. Ultimately, this emphasises the need for further work on measures that can accurately capture an individual's strategy preference. Sanders-Reio et al.'s (2014) writing beliefs scale does this to an extent, because it incorporates revision beliefs, which capture a writer's tendency to edit and revise during text composition. Still, this scale has not been used to examine the relationships between revision tendencies and using a revision-based strategy on writing outcomes.

Moreover, and as discussed already, the measure does not incorporate anything about planning beliefs.

Another limitation with the current work on writing strategies is the way that it deals with examining writing-to-learn outcomes. Both studies from Kieft and colleagues (2008; 2011) assessed literary interpretation as a proxy measure for the development of knowledge and understanding through the process of writing. However, they did not *directly* assess discovery in writing. Yet we know that discovery is an integral component of writing-to-learn.

Galbraith, Torrance and Hallam's (2006) study, which was discussed earlier in the chapter (see page 49) did attempt to directly assess the role of planning and revision strategies on knowledge development by monitoring the number of new ideas writers produced before and after an essay writing activity by either being assigned to an outline or synthetic planning condition (which are essentially planning or revision writing strategies). However, this study did not examine how an individual's preferences/beliefs about how writing should be approached may affect the utility of the *actual* strategy they use on understanding development.

To the knowledge of the author of this thesis, to date, the only piece of research that has looked at the interactional effect of writers' beliefs about approaches to writing and the strategies they actually employ on text quality and understanding development is that of Baaijen, Galbraith and de Glopper (2014). In this study, university students were asked to write a newspaper article about a current affairs topic. The participants were randomly assigned to either an outline planning strategy or synthetic planning strategy condition. As in Galbraith (1999), writers in the outline condition were given five minutes to construct an organised outline plan for their essays after making an initial list of ideas. Writers in the synthetic condition were given five minutes to write a single sentence summing up their main response to the writing topic. The students across both conditions were then given 30 minutes to type their articles (during which time their keystrokes were logged). Prior to planning and writing the articles, the participants filled in the White and Bruning (2005) Writing Beliefs Inventory. They were also asked to rate on a seven-point scale how much they thought they knew about the writing topic. They were then asked to complete the same scale at the end of the task, and the difference between these scores was taken as a measure of subjective understanding development.

The articles were also assessed in terms of text quality. Two independent judges evaluated the articles on a 9-point scale based on a range of criteria, including text fluency and whether the participants showed evidence of discussing issues rather than just listing topical facts.

The results showed that writing beliefs did indeed moderate the relationship that the writing strategy the participants used had on text quality and development of subjective knowledge. Writers who had high transactional beliefs produced good quality texts regardless of the condition they were assigned to. In line with the dual process model's assumptions, this suggests that high affective and cognitive engagement, and a proneness to revise during text production, is beneficial in achieving clear and effective communication to the reader because the writer can create new content through knowledge constitution and then revise that content to make it reader friendly, using knowledge transforming.

On the other hand, for writers with low transactional beliefs, who were less cognitively and affectively engaged in writing, pre-planning was beneficial for improving their text quality. This was likely because the act of putting ideas into an organised plan elicited a knowledge transforming approach. Hence, these writers could draw items from long-term memory and reorganise them to meet the rhetorical goals of the task.

These findings also demonstrate that an individual's beliefs about how they should approach a writing task could be important for determining how much their subjective understanding of a topic changes during writing. Writers with high-transactional beliefs generally did better at developing their knowledge in comparison to low-transactional writers, implying that high cognitive and affective engagement in writing is beneficial for eliciting a knowledge-constituting process. Significantly though, writers with high-transmissional *and* high-transactional beliefs developed their understanding the most when they used the synthetic planning strategy. They did not develop their subjective understanding well when using outline planning. According to Baaijen et al. (2014), the likely reason for this effect was that high-transactional and high-transmissional beliefs interacted with one another, meaning that outlining inhibited these writers' knowledge-constituting processes (as the writers felt constrained to keep to the ideas in their written plans).

Baaijen et al. (2014) provide evidence about the interactional effect that writing beliefs and writing strategies can have on text quality and subjective understanding development outcomes, and their findings support the assumptions of the dual process model of writing. Yet, their study is the only piece of research to date which provides this evidence. Moreover, there are limitations to this evidence, which will be explored later in the chapter. However, in a brief interlude from the evidence on writing beliefs and writing strategies, it is salient to explain how this study addressed one of the most significant criticisms of the dual process model to date.

This particular issue was highlighted earlier in the chapter in the discussion of the dual process model. Specifically, it was noted that studies providing evidence for the model explain very little about how the processes involved in writing composition actually map onto the knowledge constituting and knowledge transforming systems. Yet, the dual-process model predicts that engaging in these two systems should have varying effects on how individuals write sentences (spontaneous or controlled) and how they form the global structure of their texts (linearly or with high levels of revision).

So, as well as looking at the effects of writing beliefs combined with writing strategies on text quality and understanding development, Baaijen and colleagues examined whether sentence production and global composition measures could be related to the two systems of the dual-process model. These findings were reported in a later paper (Baaijen & Galbraith, 2018), albeit taken from the same investigation, and the next section of this chapter describes what those findings revealed.

### ***Writing Processes and their Relationships with Components of the Dual Process Model***

In Baaijen and Galbraith (2018), the authors were interested in how linearly (i.e. little revision) or recursively (i.e. much revision) individuals created their texts. Similar measures of the writing process had been used in previous research to investigate how linearly or non-linearly children produced poems, and these measures could be related to the quality of their final poems (Groenendijk, Janssen, Rijlaarsdam, & van den Bergh, 2008). This previous work, however, only examined the relationship between the linearity and non-linearity of text production with text quality, because

subjective understanding development was not a goal within this specific writing task. However, under the assumptions of the dual-process model, knowledge transforming processes would be associated with carefully controlled sentence production and high level of revision to the text's global structure. Such processes would reflect the individual's manipulation of pre-existing knowledge to meet the task's rhetorical goals. On the other hand, knowledge-constitution would be related to spontaneous sentence production and little global revision, reflecting the author shaping their ideas at the point of utterance.

Baaijen and Galbraith (2018) used the keystroke data collected throughout essay writing to form two composite measures of writing processes used during text composition. The first, global linearity, examined the extent to which writers edited the global structure of their texts throughout writing. Writers who scored highly on this measure were said to produce their texts linearly. That is, they composed their texts in a sentence-by-sentence linear format, with little editing of the overall structure. On the other hand, the writers who had low scores edited the structure of their texts consistently throughout writing. The sentence production measure captured how controlled writers were in their composition of sentences. Writers with high-scores on this measure produced very controlled sentences (long bursts of text production and long pauses between sentences). Writers with low scores on this measure created very spontaneous sentences (short bursts of text production with short pauses in between).

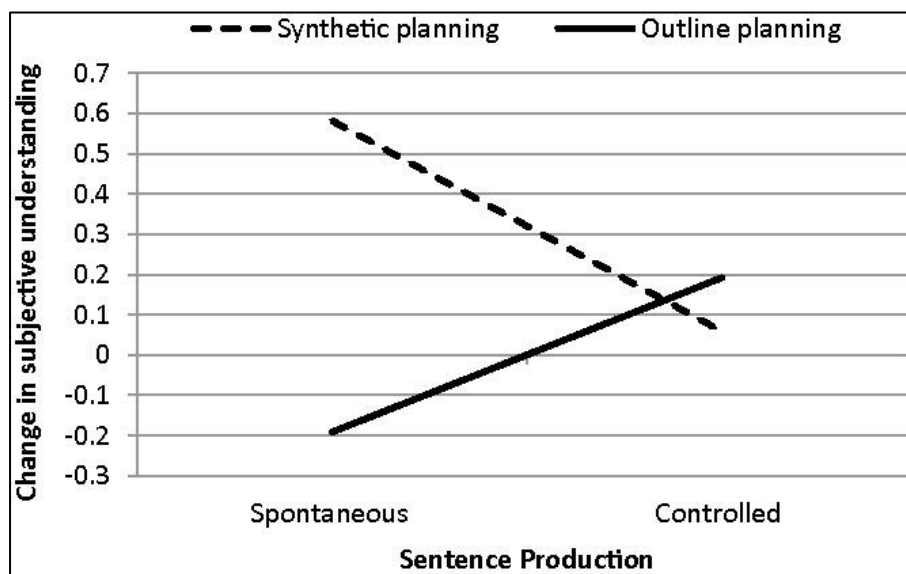
Analysis revealed that the relationship between drafting condition and text quality was largely dependent on the type of composition processes the writers were using. For example, writers in the synthetic condition wrote better quality essays when revising their text's global structure during writing but when producing sentences in a linear and carefully controlled manner. These findings imply that using a synthetic drafting strategy results in better quality writing when writers do not compose text spontaneously. This supports the dual process model's assumptions because it implies that if writers are not given a chance to organise their thoughts in a plan prior to writing, they will only produce a good quality text if they use a controlled approach to sentence production, rather than producing sentences at the point of utterance.

Alternatively, the quality of texts written in the outline group was generally high, but this was not correlated with global linearity and controlled sentence production, indicating that text quality depended primarily on the quality of the outline plans themselves (although this was not assessed). Again, this alludes to the notions of the knowledge transforming system, which suggests that when the writer has organised their thoughts before putting them into written words, this will aid with text quality.

Baaijen and Galbraith (2018) also found that the planning strategy writers used impacted their change in understanding about the writing topic before and after writing. Change in subjective understanding was best for the synthetic condition when the writers produced spontaneous sentences rather than controlled sentences. This directly supports the dual process model because it suggests that discovery was best enhanced when writers could produce text at the point of utterance, rather than being bound by traditional writing conventions. In contrast, however, change in subjective understanding in the outline planning condition was practically non-existent, which implies that although outlining may be beneficial for improving text quality, it could limit the extent to which an individual is able to engage with knowledge constitution during text production.

**Figure 3**

*Relationship Between Sentence Production and Change in Understanding as a Function of Type of Planning (Baaijen & Galbraith, 2018, p.212). Reproduced with Permission of the Rights Holder.*



### **The strengths and limitations of the Baaijen and Galbraith (2018) and Baaijen, Galbraith and de Glopper (2014) studies in their support for the dual-process model**

Overall, the results from Baaijen and Galbraith (2018) support the idea that synthetically planned spontaneous text production, rather than controlled text production, is associated with discovery and that this is independent of the effects of more global planning directed to rhetorical goals. This further reinforces the need for a *dual* process model of writing, which highlights the balance between effective communication and development of knowledge. The findings also suggest that global linearity and sentence production are indeed differentially related to text quality. Furthermore, the evidence from Baaijen et al. (2014) suggests that writing beliefs can moderate these effects, possibly because the writing of individuals with high transactional beliefs is relatively more oriented towards dispositional goals, whereas writers with low transactional beliefs direct their writing towards rhetorical goals.

However, there are some important limitations within this research, meaning that there are weaknesses in the support it provides for the dual process model. For example, Baaijen et al. (2014) used the White and Bruning (2005) Writing Beliefs Inventory, which only considers transactional, transmissional and revision beliefs (as part of the transactional factor). Yet, as highlighted in the evidence on writing strategy preferences earlier in the chapter, it is also imperative to look at the role that a writer's approach to planning may have on their writing outcomes. The evidence from Baaijen et al. (2014) and Baaijen and Galbraith (2018) clearly indicates that the type of planning employed impacts text quality and subjective understanding development. Hence, one would expect the beliefs an individual has about planning preferences to be an important contributor towards the effectiveness of the strategy they are instructed to use. Indeed, the dual process model would predict that those with strong beliefs about using a planning strategy in writing should engage less in the knowledge constituting system than those with a preference for using a revision strategy instead. However, Baaijen et al. (2014) does not provide evidence about this assumption because the White and Bruning (2005) scale is not comprehensive enough to examine the effects of planning beliefs.

Therefore, the first aim of this PhD project was to attempt to validate a new Writing Beliefs Inventory (WBI; Galbraith and Baaijen, in preparation) which included beliefs about planning. Later in the thesis, the potential impact of writing beliefs, including those relating to planning, on subjective understanding development is investigated, so establishing a comprehensive measure of writing beliefs was an important precursor to that evaluation.

Additionally, study 1 of the PhD project examined whether the new WBI could reliably and validly measure writing beliefs in the same way across university students with and without dyslexia. There are two reasons why this study sought to validate the new WBI in two separate groups of writers. Firstly, and as will become clear in the Study 1 chapter, all pre-existing writing beliefs measures have been examined in very homogenous student groups, so their utility in examining the beliefs of writers who do not fit into these groups is unknown. Hence, the study demonstrates the need to test writing belief scales in different groups of individuals. Secondly, the study specifically examines students with and without dyslexia as separate groups because people with dyslexia have specific difficulties with transcription (spelling) and translation processes (Rose, 2009), which could

influence the approaches they use for writing differently to those without dyslexia. As such, it is important to understand whether writing beliefs can conceptually be presented in the same way for individuals with and without dyslexia, or whether this might differ *because* of the specific difficulties that dyslexic writers may experience.

Originally, this PhD project intended to utilise the data from students with and without dyslexia throughout all studies in the thesis, and the overall goal was to examine how the two groups might have differed in their engagement with the processes involved in writing. However, due to the Covid-19 pandemic, the data collection time attributed to the project was significantly reduced, and so it was difficult to obtain a statistically reasonable number of participants with dyslexia to take part in each phase of the project. Hence, Study 1 is now the only study in the thesis that considers differences between students with and without dyslexia.

Another issue with the evidence presented by Baaijen and colleagues (2014; 2018) suggests a limit to the robustness of the support it provides for the dual process model in terms of processes. The authors use composite measures they had created to examine writing processes through keystroke data, and how these processes may relate to the systems of the dual process model. But, their research is currently the only evidence to use these measures, so we do not yet know if these measures will produce similar findings in other studies. As such, it is important to assess whether these measures can be used to reliably identify the same writing processes in a new sample of writers, and, subsequently, whether they show the same relationships with text quality and understanding development. As the authors provide novel findings about the relationships between writing processes with text quality and understanding development, it is vital to test whether the methods used within their research are transparent and accurate, so that the work can be reproduced.

In Study 2 of the PhD project, therefore, a reproducible framework for aligning keystroke data with cognitive processes in writing was created. This ultimately results in replicating the two composite measures of writing processes initially identified by Baaijen et al. (2012) and then tested and used in Baaijen and Galbraith (2018) but provides a comprehensive, step-by-step guide for accurately obtaining the variables used to calculate them. The study also used mixture modelling to explore whether experimental manipulation of the drafting process could lead to differences in the

structure of pauses that individuals make throughout writing. This was conducted to help with mapping pauses onto underlying writing processes and, in the context of this project, to see how pauses may be related to the two systems of the dual process model.

The design of the experimental conditions used in Baaijen et al. (2014) and Baaijen and Galbraith (2018) was also somewhat limited, and hence the studies do not provide a comprehensive view of how writing beliefs, writing processes and writing strategies may interact to affect text quality and understanding development. Although there were two different planning conditions, all participants were required to draft their texts in the same way. Specifically, they were told to write a well-formed essay. This may have stunted the full extent to which some of the writers could develop their subjective understanding, because text production may have been directed to the task's rhetorical goals (i.e. producing a polished final product), rather than developing their own ideas. Moreover, under the assumptions of the dual-process model, if transactional writing beliefs are largely about prioritising the development of subjective understanding throughout the act of writing, one would expect their relationship with knowledge development to be better if a revision writing strategy (in which writers write a draft of a text without pre-planning, but then spend considerable time editing it afterwards) was used for high transactional writers (because outline planning reduces the development of understanding). However, the original investigation did not test this hypothesis, so it is limited in the evidence it provides for the mechanisms of the knowledge constituting and knowledge transforming systems.

As such, Study 3 of the PhD project aims to semi replicate the Baaijen and Galbraith (2018) and Baaijen et al. (2014) studies. In contrast, in this project, the experimental manipulation is extended by requiring participants in the synthetic and outline planning conditions to use different drafting techniques as well as planning techniques. This means that the *whole* writing strategy participants use is manipulated, rather than just the planning component. Writers in the synthetic condition were only required to write rough drafts of an essay, which should have theoretically extended the amount of subjective understanding they could develop throughout the course of their writing task. On the other hand, writers in the outline condition were required to write a final, polished version of an essay (using the same set up as the original study). It was expected, much in the

same way as the original study, that subjective understanding development would have been stunted in the outline planning/final drafting condition because of an increased need to focus on rhetorical goals (i.e. producing a well-formed text). However, in the synthetic/rough drafting condition, where rhetorical goals were relaxed, it was expected that change in subjective understanding would have been maximised.

To summarise, then, this PhD project sought to provide further evidence about whether the claims of the dual process model could be upheld by elaborating on the research design used by Baaijen and colleagues (2014; 2018) in a way that created a stronger contrast between planned and rhetorically guided writing with spontaneous rough drafts. It measures the effects of these conditions on writing processes and the development of understanding and also assesses how writing beliefs (crucially, beliefs about planning amongst other types) and writing processes may be implicated in understanding development. To achieve this, the project first assessed the validity of the Galbraith and Baaijen (in preparation) writing beliefs inventory and compared how it generalises to writers with dyslexia. This helped to establish how valid the inventory was before it was used in the experimental study and, importantly, adds to our understanding of how dyslexia might affect writing beliefs. Then, the replicability of the Baaijen and Galbraith (2018) measures of writing processes was examined, so that they could be used to investigate the role of writing processes in subjective understanding development throughout the experimental study. The project concludes with the experimental study, which used the writing beliefs measure from Study 1, writing process measures from Study 2, and the extended manipulation to writing strategy (including changes to drafting technique as well as planning technique) to assess the impact that these features could have on how much a writer can develop their understanding around a given topic during a writing task.

## 2 Methodology

This chapter explains the overall methodological approach for the PhD project. It first reintroduces the research questions from the introduction chapter, and explains how the research in this project has been designed to fulfil those questions. It also describes why a quantitative, post-positivistic approach has been chosen as the underpinning research paradigm for this project. The chapter then moves on to give a detailed description of the sampling techniques, instruments, measures and procedures used in each of the studies. Finally, the chapter discusses the ethical issues presented by the research, and how they were addressed. This chapter does not go into detail about the analytical methods throughout the project. Instead, a detailed description of each study's analytical choices is presented in the separate study chapters.

### Methodological Background

#### *Recap of Research Questions and Methodological Choices*

In this project, three studies were conducted. Study 1 addressed the following research questions:

1. Concerning the Galbraith and Baaijen (in preparation) Writing Beliefs Inventory:
  - a. To what extent does it validly capture the five underlying components of writing beliefs that it was designed to capture?
  - b. To what extent does it demonstrate measurement invariance across items, factors, and UK university students with and without dyslexia?

Study 1 utilised a survey design with an online questionnaire featuring the WBI. Survey methods were used because a large amount of data could be collected in a relatively short period of time (Cohen et al., 2017), which was appropriate for the limited timescale of this PhD project. Surveys

have also traditionally been used to test the accuracy and validity of newly designed instruments (Cohen et al., 2017). As the overall aim of Study 1 was to validate the five-factor structure of the Writing Beliefs Inventory and test whether it was reliable for usage amongst multiple groups of writers, a survey was a well-grounded choice of research design to address this goal.

To be able to make statistically robust inferences about whether the Writing Beliefs Inventory captured the underlying components it was designed to, and to demonstrate whether the inventory was indeed suitable for usage amongst students with and without dyslexia, it was essential to obtain a large sample of participants. Survey methods made this practicable because an online questionnaire could be used to collect the data. Online questionnaires are easily accessible, which means the study's data could be obtained from a range of individuals within a wide target demographic (Cohen et al., 2017). To generalise the findings to the broader population of UK university students, the aim was to recruit participants from several UK universities. So, using an online questionnaire was efficient for collecting data from students outside of the University of Southampton.

Studies 2 and 3 used the same dataset, which had been created using an experimental research design involving keystroke analysis. For Study 2, the research questions were as follows:

2a. How do we create a reproducible method for defining, identifying, and calculating pauses, bursts, and revisions from keystroke data?

2b. Do Gaussian mixture models reveal differences in the durations and mixing proportions of pauses throughout text composition in the outline and synthetic drafting strategies?

2c. Can the Baaijen and Galbraith (2018) two-component composite measures of global linearity and sentence production be reproduced on a new sample of university students' essay writing keystroke data?

Because Study 2 attempted to create a reproducible and transparent framework for analysing keystroke components (pauses, bursts and revisions) in relation to writing processes (global revision

and sentence production), it was necessary to have a large sample of keystroke data to base this framework upon. Thus, the pragmatic decision was made to use the keystroke data collected for Study 3 also in Study 2, because it meant that fewer experiments needed to be conducted, which was important given the limited time available for data collection within the PhD timeline. Additionally, Studies 2 and 3 utilised the same experimental manipulation, albeit to examine different outcomes, so it made sense to use the same datasets across both of the studies. Finally, Study 2 focused on utilising the keystroke measures developed in the reproducible framework to try to replicate the two writing process measures (sentence production and global revision) identified by Baaijen and Galbraith (2018). The primary aim of this replication, of course, was to test whether sentence production and global revision were reproducible concepts. However, in running this replication, the results of Study 2 also provided evidence as to whether the composite measures were an appropriate tool for examining the relationship between writing processes and subjective understanding change in Study 3.

Study 3 addressed Research Question 3:

2. Do drafting conditions, writing beliefs and writing processes affect the development of subjective understanding during essay writing?

Experimental research methods were appropriate for Study 3 because they enabled the investigation of the effects that writing beliefs, writing processes and drafting strategy could have on change in subjective understanding over the course of writing, but under carefully controlled conditions. Participants were randomly allocated to their drafting condition, meaning that any differences observed in self-reported change in understanding between the two groups would most likely be due to how the writing task was manipulated, rather than confounding variables (note that a caveat to this, non-probability sampling, is discussed in the next section).

### ***Philosophical Approach***

From the descriptions of each of the three studies, it is clear to see that the methods within this project would traditionally align with the positivist research paradigm because quantitative, empirical observation is used as a means to measure behaviour (Beck, 1979).

However, in reality, quantitative tools can only *estimate* human behaviours because they are being used to try to make inferences about non-observable variables. As with any research tool that has ever existed, all scientific measurements produce error (Trochim, 2020). As a result, we can never truly accurately measure cognitive concepts, and we can never wholly understand their relationships with one another. As such, a post-positivist epistemology is applied throughout this project. Post-positivism recognises that objective knowledge does not exist, but instead, knowledge is situated within social and cultural contexts (Cohen et al., 2017). This means that all observation methods produce only a *representation* of knowledge, rather than a completely accurate depiction.

Quantitative methods were chosen to conduct this research because of the following advantages set out by Goertzen (2017):

- Findings are generalisable to a specific population;
- Datasets can be large, and findings can be representative of a population;
- Documentation regarding the research framework and methods can be shared and replicated;
- Standardised approaches allow for studies to be replicated over time.

As can be seen, a significant attraction to the use of quantitative methods is that research findings may be generalised from a sample to a population. However, it is essential to remember that, based on the assumptions of inferential statistics, such generalisations should only be conducted when the data is collected using random sampling methods. The reality of the current research project is that it would have been challenging to use genuinely random sampling methods because this project had a limited amount of time for data to be collected. Also, to maintain high ethical standards, all participants in this study were recruited voluntarily and attempting to recruit volunteers through random sampling methods would have been extremely difficult. Thus, this research uses non-

probability sampling, and hence traditional sample-to-population inferences should technically not be applied.

However, a truly random sample of participants would be challenging to acquire in any project because, even when enacted with a tremendous amount of precision, “random sampling seldom results in random samples” (Polit & Beck, 2010, pp.1453). This is perhaps best illustrated with research conducted with people (like in the current project). The vast number of between- and within-person extraneous variables in a single sample of participants in one study is likely to mean that the sample would be biased because most individuals display one or more of the same extraneous variables. Thus, one could argue that generalising from a sample to a population is appropriate in very few instances.

As such, throughout this project, Firestone’s (1993) model of generalisability is used as a framework for drawing inferences from data to the population of UK university students. His models of analytic generalisation and case-to-case transfer are applied. Analytic generalisation refers to distinguishing and highlighting relevant results to all participants. While this is more commonly used to generalise qualitative data, this ideal underpins quantitative analysis (Polit & Beck, 2010).

Case-to-case transfer puts the work of transferability onto the reader rather than the researcher. In this model, it is implied that the researcher’s job is to provide detailed evidence through their analysis, but it is the *reader’s* responsibility to decide whether these findings can be extrapolated to other scenarios.

In short, it is not uncommon for quantitative research to violate the assumption of probability sampling when concluding from sample data to a target population, but very few researchers address this violation. Therefore, by being transparent about the issue now, it is recommended that the reader of this thesis use the evidence presented to make their own decisions about whether the project’s findings can be generalised to the population of UK university students. Moreover, much of the work presented in this thesis is exploratory. Where this is the case, it is explicitly stressed that the research only gives a preliminary insight into the concepts investigated rather than giving any confirmatory conclusions. In such instances, it is highly recommended that the reader does not consider the findings generalisable.

It is also necessary to mention that the philosophical principles of open science underpin the methodological choices made throughout this project. Fecher & Friesike (2014) describe ‘open science’ as an umbrella term used to capture research issues relating to the accessibility of knowledge creation and interpretation for researchers and the public. Specifically, in this project, four open science concepts are focused on: reliability, reproducibility, validity and transparency.

Reliability is the extent to which research findings are consistent when the concepts being measured are tested over multiple instances (Kirk et al., 1986). Reliability is an important component of open science because high reliability indicates that research findings are more likely to be reproducible. Hence, reliability and reproducibility are two open science concepts that go hand in hand. Reproducibility, as a concept, does not have a strict definition. Goodman et al. (2016) suggest that ‘reproducibility’ should be broken into three sub-terms: methods reproducibility, results reproducibility and inferential reproducibility. Methods reproducibility refers to examining whether authors provide enough detail to ensure that other researchers can repeat their study procedures. Results reproducibility refers to when the same results can be obtained in a study that is independent of the original. The methods of the two studies do not have to be exactly the same but should be very similar. Inferential reproducibility refers to when the same conclusions are drawn from separate researchers using the same data, or across replications of a study (i.e. different datasets are used but the same conceptual findings exist).

In this project, reliability and reproducibility are examined in multiple ways. Firstly, the project looks at whether the WBI has a reliable 5-factor latent structure, by testing whether that structure can be replicated in two groups of participants. Reproducing the same results in both samples would imply that the WBI produces consistent patterns in categorising writing beliefs across separate groups of university students. This would provide support for using the WBI as a writing beliefs measure in future studies (including Study 3 within the current project). The project also examines reliability by attempting to reproduce the Baaijen and Galbraith (2018) global linearity and sentence production writing process measures. Suppose consistent findings can be found with how the relevant components of these two scales load onto each other. In that case, there will be evidence to suggest that the original study is both methods and results reproducible.

Validity refers to when scores from a tool exhibit empirical patterns consistent with the theoretical construct being examined (Shrout & Lane, 2012). In terms of open science, validity is important because research findings should represent what is actually happening in the real world - not just in sample data. An overarching aim within this project was to validate the WBI for usage in university student writers using ESEM (Asparouhov & Muthén, 2009). ESEM has been described as a more valid factor analysis technique than traditional confirmatory factor analysis because it allows cross-loadings within models (Tóth-Király et al., 2017). Research indicates that allowing items to load onto multiple latent constructs gives a more realistic interpretation of the data because, in real life, items are likely to represent multiple theoretical constructs. If ESEM provides a better model fit than CFA for the WBI within this project, those findings will imply that cross-loadings are more appropriate for modelling the WBI data. Hence, a more accurate technique would have been utilised.

Transparency is the need to be explicitly clear about the methods and analytical decisions used in a study (Given, 2008). The benefit of transparency is that it allows other researchers to fully scrutinise the methods used within a piece of research. Transparency underpins all methodological and analytical reporting in this project. However, it is especially highlighted in Study 2, which focuses on developing a reproducible set of procedures for examining writing processes.

## **Participants and Sampling**

The participants in Study 1 were UK university students with and without dyslexia. Participants were primarily recruited through the University of Southampton Psychology Research Participation Scheme. Most participants were undergraduate students from the University of Southampton and received course credits for participating in the project. This was also the case for Studies 2 and 3 (both these studies used data collected from the same sample of participants). Students recruited through the Southampton Research Participation Scheme for Study 1 received two credits on completion. Participants for Study 2/3 recruited through the Southampton Research Participation Scheme received 48 credits on completion. Although most participants who took part in the project were University of Southampton students, the project was open to all students who attended a UK

university and had English as a first language because the aim was to produce findings that would be generalisable to the wider UK student population. However, in the end, only a small percentage of the students involved in the study were not from the University of Southampton. The implications of this will be elaborated on in the general discussion chapter.

For statistical analysis to be considered robust, high statistical power is necessary. However, judging statistical power in studies that use latent modelling (like Study 1) can be difficult due to many conflicting accounts of adequate sample size. For Study 1, target sample size was 200 participants (with dyslexia  $n = 100$ , without dyslexia  $n = 100$ ). This was based on the assumption that, for multi-group modelling, there should be 100 participants per group (Kline, 2005). The actual number of participants obtained for Study 1 was 612. However, only 69 of the total sample of students were dyslexic. Since the dyslexic sample was approximately ten percent of the total sample, the number of dyslexic participants in the study was approximately proportionally representative of the number of people estimated to have dyslexia in the UK, (1 in 10; National Health Service, 2020). Additionally, many guidelines for multi-group modelling imply that the numbers of participants with dyslexia in the study were statistically adequate. These include recommendations from Wang & Wang (2012) who state that sample sizes should be between 5-20 participants per latent factor, depending on how many observed variables the latent factors have. The models tested in Study 1 have between 8 and 19 observed variables per latent factor and so meet the expectations set out by this definition. Other guidelines suggest that a sample size of 50 (per group) is adequate if each latent factor has upwards of 6 observed variables (Boomsma, 1985; Marsh & Hau, 1999), again implying that Study 1 was adequately powered.

For Study 2, power analysis was not conducted because the aim of Study 2 was not to make robust statistical inferences, but rather to develop a set of methods for the analysis of keystrokes which, in turn, could be used to collect robust statistical data about the processes used in participants' writing.

The statistical power of Study 3 is likely low due to not being able to collect data from enough participants during the Covid 19 outbreak in 2020. When the project's data collection started again in October 2020, approximately 6-months of allocated data collection time had passed. To

negate issues related to low sample size, Study 3 was adapted to be more exploratory. That is, although the original aim was to make generalisable inferences about how writing beliefs, writing processes and drafting strategies could affect subjective understanding throughout writing, it is recommended that readers interpret the findings of Study 3 as *indicators* of the potential relationships that may exist between these variables, rather than confirmatory findings.

Moreover, the statistical analyses conducted in Study 3 were simplified. The original aim was to use structural equation modelling to test how the relationships between drafting strategy, writing beliefs and writing processes could impact change in subjective understanding throughout writing. Additionally, a range of pre-test measures for each participant were collected to be used as controls for individual differences in aspects such as working memory functioning, typing proficiency and spelling ability. Again though, due to small sample size, the data from these measures were omitted in the statistical analyses, and the analyses themselves were simplified (linear regressions instead of structural equation models), to reduce the number of model parameters. To be completely transparent, the final sample size for the study is low even for the simplified analyses that were conducted, which is further reason to interpret the results as exploratory and in need of following up with replication studies.

## **Instruments and Measures**

### ***Study 1***

Study 1 was a survey which used an online questionnaire hosted on the [isurvey.soton.ac.uk](http://isurvey.soton.ac.uk) website. The questionnaire consisted of the following sections:

**Demographic Questions.** Participants were asked their birthdate, gender, university, degree title, and year of study. This was to help validate that they were indeed university students and therefore part of the target population. The participants were also asked whether English was their first language so that any students who did not have English as a first language could be removed from the final sample. Finally, participants were also asked to state whether they had received a

dyslexia diagnosis. This information was used to split the sample into two sub-samples: a group with dyslexia and without dyslexia.

**Galbraith and Baaijen (in preparation) Writing Beliefs Inventory.** This Writing Beliefs Inventory (WBI) consisted of 43 questions that were presented to participants in random order. The statements were designed to reflect respondents' preconceived attitudes towards writing. Participants gave their responses on a 5-point Likert scale, with the points labelled 'Strongly Agree', 'Agree', 'Neutral', 'Disagree', and 'Strongly Disagree'. The Galbraith and Baaijen (in preparation) inventory was designed to represent five different types of writing beliefs:

1. Planning beliefs – how much a writer plans prior to writing and uses their plan during writing;
2. Revision beliefs - a writer's proneness to editing and reorganising text throughout writing (referred to in previous scales as recursive process beliefs);
3. Audience beliefs - The extent to which the writer keeps the needs of the audience in mind throughout writing;
4. Transmissional beliefs - The extent to which the writer believes the main purpose of writing is about reporting accurate information from authoritative sources, reflecting limited cognitive and emotional engagement throughout the writing process;
5. Transactional beliefs - the extent to which the writer believes the main purpose of writing is to develop their personal understanding of the writing topic throughout writing, reflecting a higher level of cognitive and emotional engagement during the writing process.

The Galbraith and Baaijen (in preparation) WBI is largely based on the previous writing beliefs questionnaires from White & Bruning (2005) and Sanders-Reio et al. (2014) and uses a similar selection of items from the transactional, transmissional, recursive (revision) and audience dimensions. However, the number of items is reduced to ensure equal coverage of the five domains. Additionally, it introduces a new set of items representing planning beliefs.

The reason for using this new questionnaire is because it introduces items about planning. Since this project investigates how planning plays a role in understanding development via writing, it

is important to use a scale that incorporates planning beliefs. Also, and as discussed in the literature review, previous writing beliefs scales have not included any items about planning, so this updated inventory addresses this limitation.

The full questionnaire can be seen in Appendix A (pp.222 – 231).

### ***Studies 2 and 3***

The tools used in study 2 were as follows:

**Keystroke Datasets.** The data used in Study 2 were collected as part of the experiment designed for Study 3. Specifically, 48 keystroke logging datasets were used to develop measures of pauses, bursts and revisions. These measures were used in Study 3 to make inferences about the relationship between writing processes and drafting strategies with change in subjective understanding.

The datasets were created using Inputlog Version 8 (Leijten & Van Waes, 2013). Participants who had their data collected pre the outbreak of Covid-19 completed the keystroke logging writing task on University laptops supplied to them by the researcher, which were identical for all participants. Participants who had their data collected post Covid-19 outbreak completed the keystroke logging task on their personal computers.

The keystroke datasets included an array of variables containing information about real-time computer and mouse keyboard inputs produced by participants during the writing task in Study 3. Each keystroke log recorded every key and mouse input, the start and end time of each input, the action time of an input (the length of time the input lasted for) and the pause (empty) time between inputs. The writing task used to create these keystroke datasets is described in the instruments and measures section for Study 3.

### ***Study 3***

The tools used in Study 3 were as follows:

**Writing Beliefs Inventory (Galbraith and Baaijen, in preparation).** Participants who took part in the study pre-Covid-19 outbreak completed a paper-based version of the 43-item Galbraith and Baaijen (in preparation) WBI. Participants who took part in the study after the Covid-19 outbreak completed the questionnaire online at [isurvey.soton.ac.uk](https://survey.soton.ac.uk). The results from these questionnaires were used to examine the relationship between writing beliefs and knowledge development in the essay writing of UK university students.

**Essay Questions.** The participants were randomly assigned to one of two essay questions:

1. Does social media do more harm than good?
2. Should we all become vegans?

These questions were chosen because of their argumentative nature. For university exams and coursework assignments, students are often asked to argue for or against a question. However, due to the sheer length of this study, it would be impractical to ask participants to revise a topic prior to writing an essay. These two questions were chosen because they cover current affairs frequently discussed in the news, and so it was assumed that participants would have some level of opinion and understanding about them.

The participants were given 30 minutes to type their essays on a laptop. All participants who took part in the study prior to the covid-19 outbreak used the same style of university laptop, supplied to them as part of the research. Students who took part in the study remotely, after the outbreak, used their own personal laptops to write their essays.

**Subjective Understanding Scale.** Participants completed a 12-item scale of subjective understanding developed for the purposes of this project. The scale was administered twice to participants throughout the experiment - once before participants started writing their essays and once after they finished. The differences in the subjective understanding scale ratings prior to and post essay writing were used to estimate how participant's subjective understanding of the topic changed over the course of the writing session. Participants who took part in the study before the Covid-19

outbreak conducted these tasks on paper, whereas participants who took part in the study remotely completed the task on a computer.

In previous studies looking at the development of understanding through writing, researchers have used single-item Likert scales to assess writers' knowledge change (Baaijen et al., 2014; Baaijen & Galbraith, 2018). However, a potential issue with this method is that a single item may not be sensitive enough to measure a concept as complex as subjective understanding.

A new scale of understanding was developed for this study to overcome this issue. More items have been used to create a more reliable measure of each participants' development of subjective understanding through writing.

Originally, a 10-item version of this scale was tested with volunteer students from the University of Southampton (n=42). The internal reliability of the scale was very high ( $\alpha = .959-.967$ ), and principal component analysis (PCA) reliably showed a single factor structure that had been deemed to represent subjective understanding.

Since testing the scale with PCA, additional items have been added to make a 12-item version. These include *"how clear the relationships between your ideas about the topic are"*, and *"how well-ordered your thoughts about the topic are"*. These items have been added in an attempt to capture more information about participants' organisation of thoughts about the writing topic. Due to time constraints, the factor structure of the finalised scale was tested after data collection of this study, in a separate investigation conducted by Galbraith et al. (2023). This revealed that a reduced 8 item version of the understanding scale had a two-component structure, which the authors hypothesised to represent subjective organisation and how much writers feel they know about the topic as separate but related constructs. A copy of the understanding scale used in this study, as participants saw it, can be found in Appendix B (pp. 232 – 233), and discussion of the finalised subject understanding scale can be found in Chapter 5 (pp. 175 – 178).

**Idea Lists, Importance Ratings and Idea Correspondence Ratings.** Before the essay writing task began, participants were asked to list the ideas that came to mind about their assigned writing topic. They were then asked to rate their ideas in terms of importance on a 5-point Likert

scale, where 1 = a minor point and 5 = a major point. These two tasks were repeated at the end of the writing task.

Finally, participants were asked to compare their first list of ideas with their second list. To do this, they were told to look at each of the ideas in their second list and search for the corresponding idea(s) in their first list. If an idea was present, they were told to write it down on an idea comparison worksheet, otherwise, they were asked to leave the relevant box on the sheet blank. Finally, participants were asked to rate the similarity of the paired ideas on a 6-point Likert scale, where 1 = an identical point and 6 = no corresponding point.

The idea correspondence ratings, importance ratings and difference in subjective understanding scale were used as indicators of subjective understanding change over the course of the writing task. However, in the final analysis, only the difference in subjective understanding ratings was used to estimate understanding change. Participants who took part in the study pre-Covid-19 outbreak conducted these tasks on paper, whereas participants who took part in the study post-Covid-19 outbreak completed the task on a computer. Copies of the worksheets for these measures can be found in Appendices C – E (pp. 233 – 238).

**Keystroke Logging.** During the 30- minute essay writing task, participants had their computer keystrokes logged for statistical analysis using the Inputlog version 8 software. Inputlog is a widely used keystroke logging tool that allows for a detailed breakdown of real-time information about computer-based text input (Leijten & Van Waes, 2013). Participants who had their data collected pre Covid-19 outbreak completed the keystroke logging writing task on laptops supplied to them by the researcher, which were identical for all participants. Participants who had their data collected post Covid-19 outbreak completed the keystroke logging task on their personal computers.

**Drafting Instructions.** Participants allocated to the synthetic planning/rough drafting condition were instructed to write a single sentence summing up their overall ideas for the essay on a piece of paper in the 5 minutes prior to the start of the essay writing. This meant that they were not allowed to create a full outline plan to inform their essays. Instead, for the essay writing component of the experiment, the students were instructed to write a rough draft of an essay about their topic, in which they did not have to worry about spelling or organisation.

Participants allocated to the outline planning/final draft condition were instructed to make a structured outline 5 minutes before writing their essays. For the essay writing component, they were told to write a final version of an essay, taking into account spelling and organisation.

Whether participants took part in the study pre or post Covid-19 Outbreak, the planning tasks in both conditions were completed on paper. The essays were typed into word documents whilst Inputlog tracked all keyboard and mouse inputs in the background.

A copy of the full writing instructions for both conditions can be found in Appendix F (pp.240-241).

## **Procedures**

### ***Study 1***

Participants completed all the demographic questions, written assessment grade questions, and WBI questions (in that order) all within one online questionnaire. The WBI items were presented in a random order to reduce the likelihood of order effects. Before receiving the questions, all participants were shown a consent screen, which detailed information about what the questionnaire involved and what the results were being used for. To give consent, participants were asked to tick a consent box at the end of the screen, which would proceed onto the questionnaire.

The questionnaire took approximately 5 – 15 minutes to complete. All responses were recorded anonymously. However, at the end of the questionnaire, the option was given for respondents to leave their contact details (name and email address) if they were interested in taking part in Study 2. Participants studying BSc Psychology or a joint honours course with BSc Psychology at the University of Southampton were directed to the questionnaire via [e-folio.soton.ac.uk](mailto:e-folio.soton.ac.uk) and received two course credits on completion. Most participants completed the questionnaire online, but 21 completed the questionnaire on paper during a BSc Education seminar. All questions were the same in the paper-based version, but the WBI items were not presented in random order.

### ***Studies 2 and 3***

Participants who took part in this study prior to the Covid-19 outbreak did their study sessions in classrooms at the University of Southampton. Only the researcher and participants were present in the classroom at the time of the study. This meant that there was minimal noise disruption. The study was split into two 1 ½ hour sessions for each participant. The first session was always conducted on a one-to-one basis between the researcher and a participant but the second session could be conducted with up to four participants attending at any one time. Larger numbers of participants in the second session was beneficial in terms of time efficiency (i.e. more participants completed the experiment in a shorter amount of time) but limiting the numbers to four meant that it was still manageable for the researcher to deal with any problems that might arise (e.g. technical issues).

Participants who took part in the study after the Covid-19 outbreak did their study session remotely via a Microsoft Teams call with the researcher. This meant that in each session, only one participant could complete the experiment at a time.

#### **Session 1:**

1. Participants were emailed a participant information sheet, which detailed information about the study. Those who did the study in person were then asked to reread the information sheet at the beginning of their sessions and then sign a consent form. Those who did the study remotely were emailed the consent form and asked to return an electronically signed version to the researcher via email.
2. For the in-person version of the study, participants were first asked to fill in a paper-based version of the WBI (Galbraith and Baaijen, in preparation). They also completed some other measures to capture individual differences in spelling, typing and working memory proficiency, which were not used in the final data analysis due to small sample size and needing to reduce model parameters. For the remote version of the study, participants were emailed an online version of the WBI, which they were asked to complete before attending their first session. During session 1, they

were introduced to Inputlog Version 8 (the keystroke logging program) and were guided through how to install the software onto their personal computers.

3. At the end of Session 1 (both in-person and remote), the researcher thanked the participant for their time and arranged an appointment to conduct Session 2.

### **Session 2:**

1. At the beginning of the session, participants were introduced to one of two essay questions, either “should we all become vegans?” or “does social media do more harm than good?”
2. Participants were then given 10 minutes to list all ideas they could think of relating to their essay question on a designated worksheet. This was completed by hand for the in-person version of the study, and in Microsoft Word for the remote version of the study (this was how all worksheets were completed, unless stated otherwise).
3. On a designated worksheet, participants were then asked to rate their ideas in terms of importance on a 5-point Likert scale, where 1 = a minor point and 5 = a major point.
4. After, participants filled in the subjective understanding scale to rate how much they felt they understood about their essay topic.
5. Participants were then given the instructions for the writing task. These corresponded to either the outline planning with full drafting condition, or the synthetic planning with rough drafting condition.
6. Both conditions then had 5 minutes to complete their planning activity - either creating a hierarchically organised plan for their essay (outline plan with full draft condition), or writing a single sentence summing up their overall opinion about the writing topic (synthetic plan with rough draft condition). This task was completed on a blank piece of paper, regardless of whether the participant did the study remotely or in person.
7. For the in-person version of the study, participants wrote for 30 minutes on laptops supplied to them by the researcher. For the remote version of the study, participants wrote using their own laptops. During the writing time, Inputlog version 8 was

logging all computer keyboard and mouse inputs. Throughout the writing session, participants were allowed to refer to their plans. Participants in the outline/final draft condition were instructed to write a final version of an essay, whereas participants in the synthetic/rough draft condition were told to write a rough draft of an essay.

8. Once the 30 minutes was over, the participants rated their understanding of the writing topic on the subjective understanding scale again, taking into consideration how they felt after they had finished writing.
9. Participants then had 10 minutes to write a new list of ideas about their essay question.
10. Once again, on a designated worksheet, they were then asked to rate their new ideas in terms of importance on a 5-point Likert scale, where 1 = a minor point and 5 = a major point.
11. Afterwards, the participants had to compare the first list of ideas to the list of ideas they created after writing. They did this on a designated worksheet, which required them to take each idea in their second list and search for a corresponding idea(s) in their first list. If an idea matched up, they were told to write it down on the worksheet. Finally, they were asked to rate the similarity of the paired ideas on a 6-point Likert scale, where 1 = identical point, 5 = vaguely similar, and 6 = no corresponding point.
12. After completing the experiment, participants who completed the study remotely were asked to email all relevant study files to the researcher. Participants who completed the study for BSc Psychology course credits were assigned 48 credits. All participants were also allowed to enter into a prize draw to win a £100 Amazon Voucher.

## **Ethical Considerations and the Impact of Covid-19**

### ***Ethical Considerations Before the Covid-19 Pandemic***

Study 1 was approved by the University of Southampton's ethics and research governing body, ERGO, on 28th February 2019 (see Appendix G, p.243). Data collection took place between 4th March 2019 and 10th February 2020. Study 2 was approved by the University of Southampton ethics board (ERGO) on the 25th May 2018 (see Appendix H p.244) - data collection took place from June 2018 to March 2020, before the UK went into lockdown because of the Covid-19 pandemic. At this point, the University of Southampton mandated all face-to-face data collection to cease.

Until the outbreak of Covid-19, the primary ethical considerations for this project were participant confidentiality, anonymity, and participant recruitment. Participant confidentiality and anonymity were paramount because the project involved participants disclosing potentially sensitive information about their dyslexia status. Participants who had dyslexia and took part in the project disclosed this information at their discretion. The data obtained from each participant were anonymised after collection using unique identity numbers, which replaced names on the documents completed by participants. Therefore, only the researcher knew the identity of the participants involved in the study. All participants remained anonymous throughout write-ups of the research and the participation of students involved in the project was kept confidential (i.e. the researcher did not disclose who participated in the research to anyone).

Recruitment of participants with dyslexia also presented ethical issues. Due to confidentiality issues, it would have been unethical to obtain any contact details of students with dyslexia through university enabling services. Therefore, all advertising material for this project stated that people interested in participating should contact the researcher, rather than the researcher contacting them. Furthermore, it was asked that in their emails, potential participants stated whether they had a dyslexia diagnosis or not. This way, only participants who were willing to disclose information about their dyslexia status were recruited for the study.

The data management plan for this project (Appendix I pp.244 - 246) explains, in greater detail, the ethical considerations about how data were handled in this project.

### ***Ethical Considerations During the Covid-19 Pandemic***

Due to the Covid-19 outbreak, face-to-face data collection for the project was suspended in March 2020. However, later that year, the University of Southampton allowed researchers to continue with remote data collection. This meant that ethical approval needed to be re-obtained for an online version of the study from the University's ERGO team.

The first challenge of adapting studies 2 and 3 so that they could be conducted remotely was finding the most appropriate way of running the research sessions. Microsoft Teams video calls were used to host each session with participants. Microsoft Teams was chosen because it is the University's approved video conferencing software. During the video calls, participants were told to sit in a quiet place with minimal distractions to semi-emulate the controlled conditions of the face-to-face version of the study so that both versions resulted in data collected in similar environmental setups.

The second challenge in adapting the study was finding a way to collect keystroke data remotely. Although Inputlog will only record keyboard and mouse inputs when the program is open, sensitive data can still be elicited if the participant does not close down Inputlog properly after a writing session. Thus, when participants were running Inputlog from their own computers, they needed to be adequately trained in how to use the software. As such, an Inputlog training session was added into the experiment during session 1 for all participants who took part in the research remotely. This reduced the likelihood of data leakage due to the participants misusing the software. In addition, after the study, participants were reminded to uninstall Inputlog to prevent accidental keystroke logging outside of the research sessions.

### **Introduction to Study Chapters**

The studies described in this methodology chapter will now be presented in three self-contained study chapters. Each chapter starts with an introduction and background section, providing a rationale

for the work which draws on the literature presented in Chapter 1, but in greater detail. A methods section is then presented for each study. Because this methodology chapter gives an overview of each study's research design and procedures, the methods sections within the three study chapters briefly recaps what has been described here, but focuses predominantly on describing the analytical techniques used in each of the studies. Finally, the results and a discussion of the findings are presented in each chapter.

### **3 Study 1: Using Exploratory Structural Equation Modelling to Examine the Writing Beliefs of UK University Students with and without Dyslexia**

#### **Introduction**

Research indicates that writing beliefs are linked to writing outcomes such as text quality and subjective understanding development (Baaijen et al., 2014; Sanders-Reio et al., 2014; White & Bruning, 2005). However, most published writing beliefs studies focus only on university students without any form of reported writing difficulty (e.g. dyslexia), so it is difficult to know whether the findings from these investigations can be upheld with dyslexic writers. Also, the current measures that exist are limited in the scope of writing beliefs that they cover. In particular, existing measures such as the White and Bruning (2005) and Sanders-Reio et al. (2014) questionnaires do not account for how beliefs about planning can impact writing outcomes, meaning that current evidence might not provide a comprehensive picture of the role of writing beliefs in the writing process.

This chapter starts with a background discussion of the limitations with the existing evidence base on writing beliefs, drawing particular focus to how existing measures do not necessarily examine the full breadth of writing beliefs that could impact important writing outcomes. It also highlights how existing writing beliefs measures have only been examined in populations of writers without any form of writing difficulty, and why that is problematic if we are to consider writing beliefs as an integral influence on outcomes such as text quality and knowledge development.

The chapter then introduces Study 1 of the project, a survey study that examined whether the Baaijen and Galbraith (in preparation) Writing Beliefs Inventory (WBI) appears to be a suitable measure for investigating the writing beliefs of UK university students with and without dyslexia. The study firstly examined the structure of the WBI using Exploratory Structural Equation Modelling (ESEM) in university students without dyslexia. It then used invariance testing to see whether that structure could be upheld in students with dyslexia. Finally, the two groups of students' latent factor means were compared to see whether there were any discrepancies in how the two groups gravitated towards different beliefs. The chapter concludes with a discussion of the results, highlighting whether

the WBI was observed to validly capture five different types of writing beliefs, and whether this was consistent across the students with and without dyslexia. The discussion also describes how differences in the latent means between the students with and without dyslexia might partially explain why these two groups are often observed to have varying writing outcomes.

## **Background**

Chapter 1 (p.51) introduced the concept of writing beliefs, and how research has indicated that they may play a role in both text quality and subjective knowledge development (Baaijen et al., 2014; Sanders-Reio et al., 2014; White & Bruning, 2005). It is assumed that the reader is now familiar with the concept of writing beliefs, and so this background section will instead focus on the limitations of the existing writing beliefs research.

### ***Limitations of Research on Writing Beliefs***

The first limitation to highlight, in terms of existing writing beliefs literature, is that the relationship between transactional beliefs and writing performance is somewhat unclear. White and Bruning (2005) found that high levels of transactional beliefs were associated with better text quality. However, Sanders-Reio et al. (2014) found no relationship between transactional beliefs and students' text quality. Yet, it is important to note that Sanders-Reio et al. (2014) explicitly examined recursive process beliefs separately from transactional beliefs, unlike White and Bruning (2005). White and Bruning (2005) argued that transactional and recursive process beliefs were not distinct concepts from one another, because they both involved the writer developing their understanding around a topic. Indeed, given that writing theory indicates that increased engagement and subjective understanding in writing at least partly involves the editing and reorganisation of content (a knowledge transforming approach; Bereiter and Scardamalia, 1987; Galbraith and Baaijen, 2018) one could presume that transactional and recursive beliefs should be examined collectively.

However, Sanders-Reio et al. (2014) found that they could be modelled as separate constructs. They also demonstrated that whilst transactional beliefs were not related to paper grade, recursive

process beliefs were positively related. As such, the differences in the relationship between transactional beliefs and text quality between the two studies could have been due to the differences in how the two groups examined recursive process beliefs. Because White and Bruning (2005) included recursive belief items as part of their transactional scale, the relationship between transactional beliefs and increased writing performance could have simply been due to the recursive process items that formed part of that factor. This would imply that transactional beliefs on their own are not linked to increased writing quality. On the other hand, whilst Sanders-Reio et al. (2014) argued that transactional and recursive process beliefs should be treated as separate latent concepts, they used an oblique rotation procedure rather than orthogonal rotation in the analysis of their questionnaire items. This highly implies that they expected transactional and recursive process beliefs to be somewhat related (and this was indeed the case).

Further work from Baaijen et al. (2014) established that transactional beliefs could be linked to increases in text quality and change in subjective understanding during university students' essay writing (note though, the type of planning that the students engaged in before writing affected these relationships). However, the study relied on the White and Bruning (2005) questionnaire, and so we do not know whether those results were due to the recursive process items making up the transactional factor. As such, the relationship between transactional beliefs and text quality is somewhat unclear.

The next important limitation to highlight with previous writing beliefs research is the fact that the existing inventories are limited in the breadth of writing beliefs they examine. White and Bruning (2005), for example, only examined transactional, transactional and recursive process beliefs (which were included as part of the transactional factor). Sanders-Reio et al. (2014) identified audience orientation beliefs as an additional set of beliefs to be examined. Audience orientation beliefs describe the extent to which individuals believe they must adapt their writing to suit the needs of their readers. Sanders-Reio et al. (2014) actually found that audience orientation beliefs were the strongest predictor of students' grades on a class paper, highlighting the necessity of including them within a writing beliefs measure.

Yet, because Baaijen and Galbraith (2014) showed that the type of planning that a writer engaged in interacted with their writing beliefs to have varying effects on text quality and subjective

understanding development, one may presume that student's beliefs about planning (specifically how they approach the planning process) might also be implicated in their writing outcomes. However, no research has formally investigated this hypothesis, because the existing writing beliefs measures do not capture planning beliefs.

One final major limitation with the existing measures of writing beliefs is that they have exclusively been validated in populations of writers without any form of writing difficulty. The White and Bruning (2005) Writing Beliefs Inventory was tested on university students, but the authors do not say whether any of the students had learning differences that could affect their writing skill, so presumably the students in this study had no obvious writing difficulties. The Sanders-Reio et al. (2014) questionnaire was also tested on university students, but again, the authors made no reference to whether any of the students might have learning difficulties affecting their writing skill.

The problem with trying to validate writing beliefs questionnaires in populations of writers solely without any form of writing difficulty is that we do not know whether those tools are appropriate for usage with students who do experience writing-related learning differences. Moreover, one might expect that if an individual has a writing difficulty, this may influence the beliefs they have about writing, which ultimately might be implicated in their writing outcomes. One particular group that are important to focus on for validating a measure of writing beliefs is that of university students with dyslexia.

### ***Dyslexia and Writing at University***

The number of dyslexic individuals enrolling on higher education courses has been increasing yearly (Higher Education Statistics Authority, 2020), implying that, regardless of the negative impacts of dyslexia, dyslexic individuals are excelling through compulsory education into university. One might suggest that this reflects how universities are becoming more accessible institutions. However, the literature implies that students with dyslexia generally leave university with lower grades than their peers without dyslexia (Sumner & Connelly, 2020). In terms of theorising potential reasons for this trend, it is hard to ignore the fact that writing is reported as one of the most challenging tasks

dyslexic students encounter at university (Hatcher, Snowling, & Griffiths, 2002; Mortimore & Crozier, 2006). To understand why this may be the case, it is important to understand what the literature says about the mechanisms of dyslexia.

Dyslexia is commonly defined as a learning difference which “affects the skills associated with accurate and fluent word reading and spelling” (Rose, 2009, p.9). Although dyslexia is thought of as a learning difference that primarily impacts reading skill, it is also implicated in writing skill. Most research indicates that dyslexic individuals’ problems with writing are primarily down to word-level issues. For example, evidence from Connelly et al. (2006) revealed that students with dyslexia made more spelling mistakes and produced generally lower quality and shorter texts in comparison to students without dyslexia. Further evidence from Wengelin (2007), which compared the writing of adults with and without dyslexia on an array of different writing tasks, demonstrated that writers with dyslexia were poorer in terms of fluency than the writers without dyslexia. They also paused more throughout writing, especially at the word-level, and made more spelling revisions. More recent findings from Sumner & Connelly (2020) showed that during a handwritten task, students with dyslexia were found to produce poorer quality texts and more spelling errors and revisions than students without dyslexia. Both groups, however, performed similarly in terms of time spent writing, amount of text written, hand writing execution, and non-spelling related revision behaviour.

These results collectively indicate that the issues that dyslexic individuals experience with writing are likely due to skills related to the decoding and encoding of words. This aligns with the most predominant theory of dyslexia, the phonological deficit hypothesis, which suggests that dyslexic individuals have poor phonemic awareness (i.e. being able to match language sounds to letters). Ultimately, this leads to issues with reading (decoding) and spelling (encoding) words. Yet, evidence from Galbraith et al. (2012) demonstrated that dyslexic writers may experience issues beyond those at the word-level, which are implicated in the quality of work they produce. In an experimental essay planning and writing task, the authors found that even when spelling and capitalisation were controlled for, dyslexic students were still rated as producing poorer quality texts than students without dyslexia. These findings imply that word-level processes only partially mediate the influence that dyslexia has on text quality. Further analysis indicated that the students with

dyslexia had issues with outline planning but that some of the writers with dyslexia had a strong positive correlation between content planning *during* writing with text quality. Essentially, this indicated that for some dyslexic students, planning during writing was better for improving text quality than planning prior to writing. It also implies that dyslexic writers may have an issue with organising content for writing tasks in formal plans, and further evidence from Mortimore and Crozier (2006) suggests that issues with organising ideas might be a prominent feature of the writing experience for dyslexic individuals. Their survey study looked at the experiences of 136 male students across 17 UK universities. They found that whilst students accessed the support provided to them by their institutions (e.g., extra time in exams, assistive technology support and access to dyslexia tutors), many of their needs were still unmet. One area lacking in support was academic writing skills, and the students notably reported having difficulties organising essays and expressing their ideas in writing. As such, Mortimore and Crozier's evidence hinted that dyslexic university students may struggle with their approach to academic writing and that university provision may not be adequate in reducing such issues.

Based on the evidence presented thus far about how dyslexia might have an impact on writing performance, there appears to be several ways in which their writing beliefs might be impacted differently to writers without dyslexia. Firstly, the evidence on word-level issues (Connelly et al., 2006; Sumner & Connelly, 2021; Wengelin, 2007) suggests that dyslexic writers may find the translation process of writing more difficult than writers without dyslexia. This could have all kinds of effects on how dyslexic writers believe writing should be approached. For example, would writers with dyslexia be more inclined to use a knowledge telling approach over a knowledge transforming/constitution approach, because it is less cognitively demanding? Consequently, would this mean that students with dyslexia would be more likely to gravitate towards higher levels of transmissional beliefs and lower levels of transactional beliefs?

Secondly, if the issues that dyslexic students experience with writing go beyond word-level problems, and also suggest that they experience issues with processes related to planning and organisation, does this mean that writers with dyslexia would have lower levels of planning beliefs than those without? Or, conversely, would dyslexic writers prioritise planning before and during

writing, because the issues they experience with the translation process might mean that they need to take a more controlled approach to writing, rather than be able to successfully write text relatively spontaneously?

To date, these theories about the impact that dyslexia could have on writing beliefs have not been tested, because the majority of the research on writing beliefs has centred on individuals without any reported form of writing difficulty.

### ***The Current Study***

In response to the limitations within the writing beliefs evidence presented in this section, this study aimed to examine the structure of a new measure of writing beliefs (the Writing Beliefs Inventory; WBI) created by Galbraith and Baaijen (in preparation) in response to previous writing beliefs measures from White and Bruning (2005) and Sanders-Reio et al. (2014). The questionnaire is composed of 43 statements about writing beliefs, and respondents are asked to rate how much they agree with each of the statements, using a 5-point Likert scale, where 1 = strongly agree and 5 = strongly disagree.

Items representing transmissional and transactional beliefs were directly drawn from the White and Bruning (2005) questionnaire. Items representing revision and audience beliefs were directly drawn from Sanders-Reio et al. (2014). Where the WBI differs from previous inventories, however, is that it also includes beliefs about planning. Specifically, these items, which were developed by Galbraith and Baaijen, represent how much a writer plans prior to writing, and uses their plan during writing. As a whole, the inventory is theorised to examine the following beliefs as five separate constructs (see Appendix A.2 for how items theoretically map onto these constructs):

6. Planning beliefs – how much a writer plans prior to writing and uses their plan during writing;
7. Revision beliefs - a writer's proneness to editing and reorganising text throughout writing (referred to in previous scales as recursive process beliefs);
8. Audience beliefs - The extent to which the writer keeps the needs of the audience in mind throughout writing;

9. Transmissional beliefs - The extent to which the writer believes the main purpose of writing is about reporting accurate information from authoritative sources, reflecting limited cognitive and emotional engagement throughout the writing process;
10. Transactional beliefs - the extent to which the writer believes the main purpose of writing is to develop their personal understanding of the writing topic throughout writing, reflecting a higher level of cognitive and emotional engagement during the writing process.

This study tested the hypothesised structure of the WBI outlined above via exploratory structural equation modelling (ESEM), and confirmatory factor analysis (CFA). First and foremost, this investigation is a validation study. Because the WBI items have been drawn from established scales, Galbraith and Baaijen (in preparation) argue that these items should represent the same latent constructs identified in the previous examination of those scales (i.e. those mentioned above). The novelty of the WBI is the addition of items relating to a hypothesised planning factor, and so the analyses presented in this study indicate whether those planning items can indeed be grouped into an additional construct, separate to the other hypothesised factors.

Consequently, the study also addresses the lack of consensus between White and Bruning (2005) and Sanders-Reio et al. (2014) about whether revision and transactional items represent the same underlying construct, or two separated (but related) constructs. It does this by comparing a five-factor model versus a three-factor model via goodness-of-fit indices. Whilst the five-factor model divides transactional and revision items into two separate factors (like Sanders-Reio and colleagues), the three-factor model groups transactional with revision beliefs (like White and Bruning). The three-factor model also groups transmissional with planning beliefs (audience beliefs were kept on their own). This is based on a hypothesis identified by Galbraith and Baaijen (in preparation), which states that if White and Bruning are correct in saying that transactional and revision beliefs should be represented in the same latent construct (because high-transactional writers are likely to revise their texts during writing as they develop their personal understanding), then transmissional and planning beliefs should also be represented in a single construct. This is because writers with high-transmissional beliefs, who believe that the main purpose of writing is to give the audience accurate information, are also likely to prioritise careful and controlled planning before text composition,

allowing them to ensure that the information they transmit to the reader is accurate. Evidence indeed indicates that hierarchical outline planning is associated with high text quality (Kellogg, 1990; Baaijen & Galbraith, 2018), implying that writers are better at communicating their thoughts to the reader when they organise their ideas prior to text composition.

After identifying which of the three- and five-factor models best fit the data from the students, the secondary aim of this study was to test whether that model could be used to examine WBI data from dyslexic students. This would provide evidence about the generalisability of the WBI beyond writers without any writing difficulties. The comparison between these two groups also enables insight into whether students with and without dyslexia display different levels of each writing belief, and whether any differences align with theories about how dyslexia impacts writing processes and outcomes.

Taking these aims into consideration the research questions for study 1 were:

Concerning the Galbraith and Baaijen (in preparation) Writing Beliefs Inventory:

- a. To what extent does it validly capture the five underlying components of writing beliefs that it was designed to capture?
- b. To what extent does it demonstrate measurement invariance across items, factors, and UK university students with and without dyslexia?

## Methods

The data from this study were collected via an online questionnaire. UK university students with and without dyslexia ( $N = 562$ ) completed the questionnaire via [isurvey.soton.ac.uk](https://survey.soton.ac.uk). The questionnaire had two sections; the first section consisted of demographic questions about the participants' gender, age, degree type (e.g., undergraduate, masters, PhD), subject area and year of study. They were also asked to confirm that they had English as a first language and whether or not they had a dyslexia diagnosis. The second part of the questionnaire consisted of the 43-item WBI (Galbraith and Baaijen, in preparation). The WBI items were presented randomly to each participant,

except for 21 participants who completed the questionnaire on paper during a BSc Education seminar. The questionnaire took approximately 5 – 15 minutes to complete, and all participants completed it anonymously. More information about the tools and procedure for this study can be read in Chapter 2 (pp.79 – 85).

### ***Brief Introduction to Exploratory Structural Equation Modelling (ESEM)***

Writing beliefs questionnaires have typically been validated using methods that could be considered flawed. For example, the White and Bruning (2005) Writing Beliefs Inventory was developed using exploratory factor analysis (EFA). One severe limitation of EFA is that it cannot be used to examine invariance amongst different groups. That is, you cannot determine whether your instrument represents the concepts being examined in the same way for multiple groups. This presents issues in determining whether the White and Bruning (2005) Writing Beliefs Inventory is suitable to use amongst writers with and without dyslexia. Sanders Reio et al. (2014) used confirmatory factor analysis (CFA) to validate their scale. CFA is traditionally seen as methodologically advanced in comparison to EFA because it allows for invariance testing and other techniques that are not applicable within the EFA framework (e.g., certain goodness-of-fit statistics and the inclusion of method factors or correlated uniquenesses; Tóth-Király et al., 2017). Yet, EFA holds advantages over CFA in many ways because it gives a more realistic representation of the data (Tóth-Király et al., 2017). EFA allows for the cross-loading of items onto multiple factors, whereas CFA constrains items onto single target factors. In real life, items on a measurement are likely to represent numerous latent constructs rather than single ideas. Hence, EFA provides a more ecologically valid examination of data. In writing beliefs research, it is essential to examine how statements on a questionnaire may represent more than one underlying writing belief, because it would go some way as to deciding whether some beliefs (e.g. transactional and recursive) should be treated as single or multiple constructs. In short, Both EFA and CFA have their strengths and limitations. Therefore, it may be more appropriate to use hybrid approaches that bridge the gap between the two traditional methods of examining questionnaire structure.

To robustly validate the structure of the Galbraith and Baaijen (in preparation) WBI, exploratory structural equation modelling was used. Exploratory structural equation modelling (ESEM; Asparouhov & Muthén, 2009; Marsh et al., 2014) is a framework developed to incorporate the ecological validity benefits of EFA (cross-loading items onto multiple factors) with the methodological advancements offered by CFA (including invariance testing, multiple goodness of fit indices, method factors and correlated uniquenesses; Tóth-Király et al., 2017).

Asparouhov et al. (2015) demonstrated with simulation studies that failing to account for cross-loadings as small as 0.1 could lead to inflated model parameter estimates. However, using ESEM reduces parameter estimates and generally improves model fit. It also provides better discriminant validity of factors (Marsh et al., 2010; Morin & Mañano, 2011). This is because ESEM works by targeting items onto their hypothesised main factor, as one would when using the CFA framework. However, those items are then cross-loaded onto all other factors within the model, with the loading being targeted, but not forced, towards 0 using the oblique rotation procedure. If that close-to-zero cross-loading provides a poor model fit with the data at hand, it will become inflated - showing the presence of cross-loadings within the model.

Another strength of ESEM over EFA is that it can be used with invariance testing (Meredith, 1993), which traditionally could only be applied within the CFA framework. Invariance testing allows researchers to formally investigate whether instruments can be used among individuals with different characteristics. This means that comparisons can be drawn between groups, and generalisations can be made about how latent concepts manifest in diverse populations (Tóth-Király et al., 2017; note that this is an important step in determining whether the WBI is suitable for usage amongst university students with and without dyslexia). Marsh's invariance testing for the ESEM framework has up to 13 steps, but the 6 steps of standard invariance testing, used in this study, are the central components of the taxonomy (Marsh et al., 2009). These are:

1. Configural invariance: ensuring that groups hold the same factor structure;
2. Weak or metric invariance: ensuring that groups have the same factor loadings;
3. Strong or scalar invariance: ensuring that item intercepts are similar for all groups;
4. Strict or residual invariance: ensuring that measurement errors are similar for all groups;

5. Variance-covariance invariance: ensuring that variances and covariances are similar across groups;
6. Latent means invariance: ensuring that latent mean scores are similar across groups.

The first four steps described here are essential for ensuring that an instrument does not create measurement bias between different users. Step 1 makes sure that groups display the same conceptual framework on an instrument. Achieving step 2 allows for the comparison of factor correlations between groups. Step 3 ensures that groups show similar item scores when the underlying latent construct is held at the same level, and step 4 allows for the comparison of manifest scores between groups. Steps 5 and 6 are unnecessary for achieving invariance because they do not test for measurement bias. However, they indicate whether group-based differences exist in the variance-covariance matrix structures or whether the means of the latent constructs measured vary between groups (note that these two steps are necessary for determining whether students with and without dyslexia have varying levels of specific writing beliefs).

To sum up, through examining whether the WBI has a three- or five-factor structure using ESEM, and determining whether this structure is the same across university students with and without dyslexia, this study will give insight into whether the WBI is an appropriate writing beliefs measure to be used in future studies.

## **Results**

### ***Participants and Data Screening***

All analysis for Study 1 was completed in Mplus version 7.4 (Muthén & Muthén, 2015). Scripts for the analyses can be found in the project's OSF repository (link on p.1). 493 university students without dyslexia and 69 participants with dyslexia completed the online questionnaire. Of these participants, the overwhelming majority were University of Southampton students. 80.8% of respondents were undergraduate students studying BSc Psychology (including joint honours Education and Psychology and Criminology and Psychology students). This was expected as these

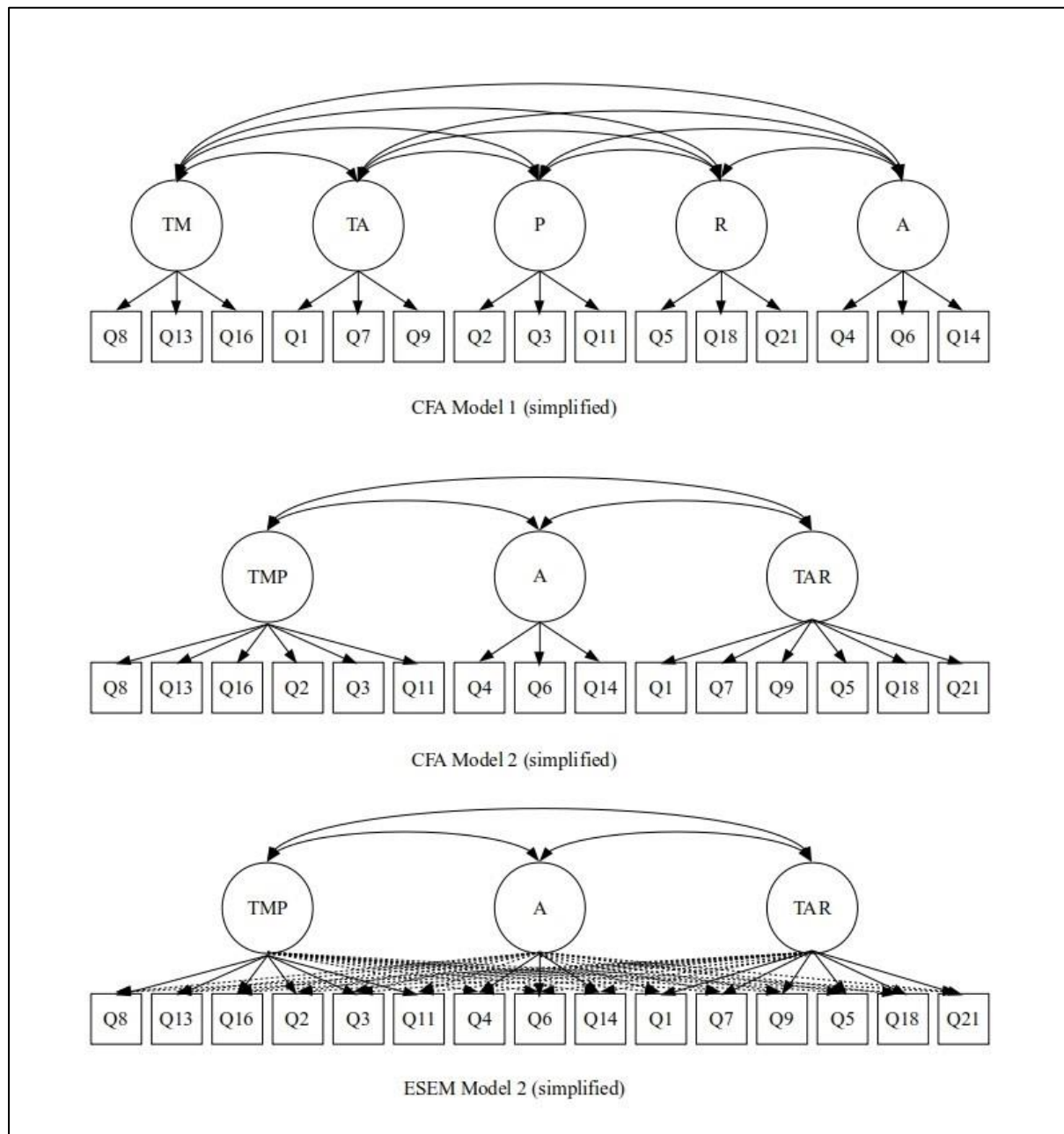
students received course credits for completing the questionnaire. 94% of respondents were enrolled in undergraduate courses. The remainder were enrolled on masters, PhD or 'other' university courses. 83.8% of respondents identified as female and 15.8% identified as male. The remaining participants chose not to disclose their gender.

The data from the WBI proportion of the questionnaire were treated as continuous because guidelines generally imply that it is acceptable to treat ordinal scales with five or more categories as interval level data (Byrne, 2013; Wang & Wang, 2012). Participant responses were screened for missing data, and 12 cases were found. One participant had four instances of missing data across their WBI responses, and eight participants had one case of missing data across their WBI responses. After considering the missing data, it was assumed that the data were missing at random. Additionally, some of the Likert scale responses for the WBI statements were skewed beyond the typical  $\pm 1$  threshold for univariate normality (Muthén & Kaplan, 1985; skewness of the WBI items ranged between -0.05 to +1.123). Hence, the robust maximum likelihood estimator was used to conduct the models. This estimator is robust to data non-normality, suitable for 5+ point Likert scales and ideal for usage with missing data (Muthén & Muthén, 2015; Rhemtulla et al., 2012).

### ***Testing the Structure of the WBI using ESEM and CFA in Students without Dyslexia***

Galbraith and Baaijen (in preparation) hypothesised that the WBI would have a five-factor structure, representing planning, revision, audience, transmissional and transactional beliefs. However, they also suggested that the WBI may be condensed into a three-factor format, representing a combination of transmissional and planning items in one factor and transactional and revision items in another, with audience beliefs being separated into their own factor.

Therefore, a five- and three-factor WBI model were compared. The study also examined whether ESEM provided a better fit than CFA regarding the five- and three-factor models. Figure 4 displays simplified versions of the tested models.

**Figure 4***Simplified Path Diagrams of the Tested ESEM and CFA Models*

*Note.* TM = transmissional, TA = transactional, P = planning, R = revision, A = Audience, TMP = transmissional and planning combined, TAR = transactional and revision combined

The top model in Figure 4 represents the first CFA tested. This model consisted of the five types of writing beliefs represented by different factors, and each of the 43 items constrained to their

target factor. The middle model in Figure 4 describes the second CFA tested, which combined transmissional and planning beliefs into one factor, and transactional and revision beliefs into another. The bottom model in Figure 4 represents one of the ESEMs, which used the three-factor design of the middle model. However, all items were allowed to cross load onto every other factor within the model (although these cross-loadings were targeted toward 0). The final model tested (not presented in the diagram due to the number of arrows making it difficult to interpret visually) uses the same ESEM design as the bottom model in Figure 4, but instead of three factors, five factors were included, with every item being allowed to load onto their target and non-target constructs. The oblique rotation procedure was used for all four models because it was anticipated that the factors would correlate.

This initial ESEM and CFA testing was only conducted with the data from students without dyslexia because the goal was to define a suitable factor structure for the WBI before testing whether it generalised to dyslexic students. The main indices used to assess model fit were the comparative fit index (CFI), the Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA). The criteria for good fit of each of these indices are as follows:

- CFI and TLI values:  $>0.90$  = adequate fit;  $>0.95$  = excellent fit.
- RMSEA values:  $<0.08$  = adequate fit;  $<0.06$  = excellent fit.

However, it is vital to note that CFI and TLI cut-offs should be interpreted with caution because of the relatively small number of applications they are based upon. Baumgartner & Homburg (1996) note that it can be very difficult to achieve the specified 0.9 thresholds with real-life data. Hence, they suggest that these thresholds should not be considered definitive, but instead, they should be considered reference points to use collectively with other fit indices.

The robust chi-square ( $\chi^2$ ) test of exact fit was also used to assess model fit. However, this test is often susceptible to minor model misspecifications and sample size. Therefore, some researchers recommend using the relative chi-square statistic ( $\chi^2 / df$ ) alongside robust  $\chi^2$  (Brookings & Bolton, 1988; Wang & Wang, 2012), which involves taking the chi-square value and dividing it by degrees of freedom (df). Generally, chi-square is used as an absolute fit index, with a low chi-square value relative to the degrees of freedom (and higher p-value) indicating better model fit. Below are the criteria for these two goodness-of-fit indices:

- $\chi^2$  values: smallest is best. P-value  $> 0.05$  = good fit but this statistic can be oversensitive to sample size.
- $\chi^2 / df < 2$  = cut off point for good fit.

***Results of Testing the Structure of the WBI Using ESEM and CFA on Students Without Dyslexia***

**Table 1**

*WBI Statements with High Covariances, Identified through Modification Indices*

Modification Indices	Statement 1	Statement 2
61.605	Writing's main purpose is to give other people information.	Writing is primarily about transmitting information.
60.737	The most important reason to write is to report what authorities think about a subject.	The key to successful writing is accurately reporting what authorities think.
49.5	Writing is often an emotional experience.	Writing is a process involving a lot of emotion.
48.094	It is important for writers to get their ideas straight before they start to write.	A good writer makes sure they know what they think before they start to write.

When initially running the models, modification indices revealed that there were four pairings of WBI statements that had very high covariances. These can be seen in Table 1. High modification indices suggest that some items may address the same questions. As such, there is redundancy in the model due to those items creating unnecessary additional model parameters. From Table 1, it can be seen that the high-covariance pairings are conceptually very similar (in fact, they could be interpreted

as the same items with slightly different wording). Hence, one within each pair was removed from the analysis, meaning that all further models were conducted on a smaller 39-item version of the WBI.

Table 2 shows the fit of the 4 ESEM and CFA models. As indicated, the five-factor ESEM provided the best fit out of all tested models, with CFI and RMSEA displaying excellent model fit (CFI = 0.958; RMSEA = 0.023, 90% CI = 0.017-0.028), robust chi-square and relative chi-square indicating good fit ( $\chi^2 = 702.1$ ,  $p > 0.05$ ;  $\chi^2 = 1.26$ ), and TLI indicating adequate fit (TLI = 0.944). None of the other tested models provided evidence of good model fit. These results indicate two things. Firstly, the presence of cross-loadings in the 5-factor ESEM suggests that items in the WBI are at least partially related to multiple underlying constructs. Hence, ESEM provides a more ecologically valid explanation of how the WBI items relate to their underlying constructs than CFA. Secondly, results imply that the 5-factor model is more valid than the 3-factor model. Hence, although cross-loadings are present within the data, the results imply that the five types of writing belief should be treated as separate, rather than grouped constructs.

**Table 2**

*Goodness of Fit Indices for the Three- and Five-Factor WBI ESEM and CFA Models for Non-Dyslexic Data*

Model	$\chi^2$ (sig/not-sig)	df	$\chi^2/\text{df}$	CFI	TLI	RMSEA	RMSEA 90% CI
1. 5-factor ESEM	702.1	556	1.26277	0.958	0.944	0.023	0.017 - 0.028
2. 5-factor CFA	1616.169 (sig)	692	2.335504	0.735	0.716	0.052	0.049 - 0.055
3. 3-factor ESEM	1245.991 (sig)	627	1.987226	0.823	0.79	0.045	0.041 - 0.048
4. 3-factor CFA	1827.33	699	2.614206	0.677	0.657	0.057	0.054 - 0.060

Follow-up analysis of factor loadings indicated that the 5-factors of the ESEM were pretty well defined - there were moderate loadings for most target factors ( $\lambda = -0.056 - 0.765$ ,  $M = 0.306$ ). However, it is important to note that some items did not load as expected. For example, a few items

did not load onto their target factor at all, or produced very weak loadings. A few significant cross-loadings were also present, although these did not undermine the five-factor structure ( $\lambda = -0.52 - 0.547$ ,  $M = 0.074$ ). Small correlations were also present between the factors ( $\lambda = -0.068 - 0.252$ ,  $M = 0.1376$ ), suggesting that the five factors identified in the analysis were weakly related. A full breakdown of the item and factor loadings for the five-factor ESEM are in Appendix J (pp. 247 – 252).

Based on this initial analysis, it appears that out of all the tested models, the five-factor ESEM provided the best fit to the data. In fact, out of the four models tested, it was the only model that provided a good fit. This demonstrates that the cross-loadings in the ESEM (as opposed to the lack of cross-loadings in the CFA) provided a superior representation of the WBI data from the non-dyslexic sample of students. However, whilst inspection of the loadings showed that most items at least moderately loaded onto their target factors, some items did not load as expected, and so a proposed reduced model, with some items removed, is presented in the discussion section of this chapter.

### ***Testing the Invariance of the Five-Factor ESEM Across University Students with and without Dyslexia***

The next step was to test whether the 5-factor WBI ESEM observed with the students who did not have dyslexia could be replicated in the students that had dyslexia. In this study, only the six main steps of invariance testing within the Marsh et al. (2009) framework were interpreted because these are the most vital for investigating measurement biases and group-based differences. Additionally, all 39-items (including those that did not load onto their target factors) were retained for this analysis, to see whether they provided similar, better or worse fit in the dyslexic group of students. Table 3 displays the invariance testing results.

**Table 3**

*Invariance Testing Results for the 39-Item Five-Factor ESEM on the Students with and without Dyslexia*

ESEM model	$\chi^2$	df	$\chi^2/df$	CFI	TLI	RMSEA 90%	
						RMSEA	CI
Reduced item configural model	1653.144*	1112	1.48664	0.882	0.843	0.042	0.037 - 0.046
reduced item weak/metric model	1865.51	1282	1.455159	0.873	0.854	0.04	0.036 - 0.044
<i>configural - weak difference</i>				-0.009	0.011	-0.002	
reduced item strong/scalar model	1924.691*	1316	1.462531	0.868	0.851	0.041	0.037 - 0.044
<i>weak-strong difference</i>				-0.005	-0.003	0.001	
reduced item strict/residual model	1999.893*	1355	1.475936	0.86	0.847	0.041	0.037 - 0.045
<i>strong-strict difference</i>				-0.008	-0.004	0	
reduced item variance-covariance model	2026.542*	1370	1.479228	0.857	0.846	0.041	0.037 - 0.045
<i>strong-variance model difference</i>				-0.003	-0.001	0	
reduced item variance-latent mean model	2074.915*	1375	1.509029	0.848	0.836	0.043	0.039 - 0.046
<i>variance - latent mean difference</i>				-0.009	-0.01	0.002	

*Note.* for  $\chi^2$ , \* $p < .05$ .

Models were determined as not achieving invariance if decreases in CFI and TLI were 0.01 or larger, and increases in RMSEA were 0.015 or larger (Chen, 2007; Cheung & Rensvold, 2002; Tóth-Király et al., 2017). Firstly, the fit of the configural model was assessed as this indicated whether both groups held the same conceptual framework. The goodness of fit statistics showed that the model had marginal fit, and this would have likely been better if the sample size of dyslexic students was larger (CFI = 0.882 = marginal fit; TLI = 0.843 = marginal fit; RMSEA = 0.042 (90% CI = 0.037-0.046) = excellent fit;  $\chi^2 = 1653.144$  ( $p < 0.05$ ) = poor fit;  $\chi^2 / df = 1.49$  = good fit). Thus, the decision was that

it was acceptable to continue investigating the invariance of the model between the two student samples. Weak/metric invariance was achieved ( $\Delta$  CFI = - 0.009;  $\Delta$  TLI = +0.011;  $\Delta$  RMSEA = - 0.02), indicating that groups had the same factor structure. Strong/scalar invariance was achieved ( $\Delta$  CFI = -0.005;  $\Delta$  TLI = -0.003 ;  $\Delta$  RMSEA = +0.001), indicating that the groups had similar item intercepts. Strict/residual invariance was achieved ( $\Delta$  CFI = -0.008;  $\Delta$  TLI = -0.004;  $\Delta$  RMSEA = 0), indicating that the groups had similar measurement errors. These results imply that measurement bias did not occur for the 39-item version of the WBI across the dyslexic and non-dyslexic student groups, as the first four levels of invariance were achieved.

To test for group-based differences, variance-covariance invariance was examined first. This was achieved ( $\Delta$  CFI = - 0.003;  $\Delta$  TLI = -0.001;  $\Delta$  RMSEA = 0), indicating that the two groups had the same variance-covariance structures. However, the last step of invariance, the latent-means model, was not achieved ( $\Delta$  CFI = - 0.009;  $\Delta$  TLI = -0.01;  $\Delta$  RMSEA = +0.002). This indicated that the two groups may have had different latent factor means. To explore this further, the five-factor ESEM was reconducted with the latent means constrained to 0 in the dyslexic group and freely estimated in the other group. This meant that the latent means of the groups could be directly compared in standard deviation units (Tóth-Király et al., 2017). The results indicated that students with dyslexia had higher latent means for the transmissional and planning factors (+0.477 standard deviations,  $p < 0.01$ ; +0.321 standard deviations,  $p < 0.05$ ) and a significantly lower latent mean for the transactional factor (-0.779 standard deviations,  $p < 0.001$ ) than the students without dyslexia. The implications of these results will be discussed in the next section of this chapter.

## Discussion

### *WBI Factor Structure*

The overall goal of this study was to test whether a newly developed writing beliefs questionnaire, the WBI, which was designed to measure beliefs about planning as well as transactional, transmissional, revision and audience beliefs, was an appropriate tool for examining writing beliefs amongst different populations. The first step towards this was investigating whether

the WBI was better represented by theoretically-driven five-or three-factor models. Comparison of ESEMs and CFAs showed that a five-factor ESEM provided the best fit in the baseline non-dyslexic data.

These results have two implications. The first is that the WBI appears to be validly represented by five separate latent constructs (revision, planning, transactional, transmissional and audience), rather than three. In line with Sanders-Reio et al. (2014), this evidence suggests that revision and transactional beliefs should be treated as distinct from one another. Similarly, the analysis implies that planning and transmissional beliefs are separate constructs.

However, the second implication is that whilst the results from this investigation do suggest that the WBI represents five separate beliefs, some individual items are representative of multiple latent factors, as evidenced by the ESEMs (in which cross-loadings were present) which provided more ecologically valid depictions of writing beliefs than the CFAs, demonstrated by better model-fit indices. Theoretically, it is appropriate to expect some items within the WBI to relate to multiple constructs due to the likelihood of some shared underlying themes between beliefs. This is evidenced by the fact that experimental studies have shown that writers with high transmissional beliefs tend to produce better quality texts when they create organised plans before writing (Baaijen et al., 2014), implying that some of the underlying mechanisms of planning and transmissional beliefs may be shared. Further inspection of the factor loadings for the transmissional and planning constructs in the WBI indeed indicated that several planning items had positive loadings with the transmissional factor, further supporting this theory.

To summarise, then, the five-factor ESEM provided best fit of all models tested in the non-dyslexic student sample. However, some items did not show a positive loading onto their target factor, suggesting that it may be worthwhile removing them in future iterations of the WBI to produce a more parsimonious scale. As such, a revised 29-item version of the five-factor WBI is proposed in Table 4, with items that have a negligible loading onto their target factor ( $<0.25$ ) removed.

**Table 4***Proposed revised 29-item version of the WBI*

Transmissional	Transactional	Revision	Planning	Audience
Writing is primarily about transmitting information.	Writing is a process involving a lot of emotion.	A primary goal of writing should be to have to make as few changes as possible. (REVERSE CODED)	The key to successful writing is to stick to one's plan.	Good writers anticipate and answer their audience's questions.
The key to successful writing is accurately reporting what authorities think.	Writers need to immerse themselves in their writing.	Writing is a process of reviewing, revising, and rethinking.	Good writing requires making a detailed outline before writing.	It's important to keep your overall purpose in mind while writing.
Writing's main purpose is to share the information in sources accurately.	Writing helps me understand better what I'm thinking about.	Good writing involves developing ideas over a series of drafts.	Good writing involves getting each sentence right before moving on to the next.	Good writers thoroughly explain their opinions and findings.
Good writers try to be objective.	Writing helps me see the complexity of ideas.	Writing requires going back over it to improve what has been written.	The key to successful writing is making a well-organised plan.	The key to good writing is conveying information clearly.
Writing should focus on the information in books and articles.	My thoughts and ideas become clearer to me as I write and rewrite.  It is important to develop a distinctive writing style.  Writing helps new ideas emerge.		Thorough planning is the most important aspect of writing.	Good writers keep their audience in mind.  Good writers adapt their message to their readers.  Good writers are reader-friendly.  It's important to select the words that suit your purpose, audience, and occasion.

Because the WBI is still in its development phase, and is currently being tested in a sample of Dutch and English students (Baaijen and Galbraith, in preparation), the revised WBI presented here is subject to change. The results from that investigation will need to be compared with the results of this study to confirm the final scale's items and underlying structure, and further testing may be required. Nevertheless, the evidence from this investigation points towards the WBI being sufficiently validly represented by the five underlying beliefs proposed by Baaijen and Galbraith (in preparation), albeit subject to further improvement by reducing the number of items representing each of those factors.

### ***WBI Invariance: Dyslexic and Non-dyslexic Students***

The study's next aim was to identify to what extent the five-factor ESEM demonstrated measurement invariance across items, factors, and UK university students with and without dyslexia. The first five models of invariance (configural through to variance-covariance) were achieved, indicating that there was invariance across the items and factors of the WBI, and between the separate groups of students with and without dyslexia. This established that the two groups had the same conceptual structure for their beliefs. However, the latent means model of invariance was not achieved, which indicated that there might be differences in the factor means between the two groups.

Further inspection of the data revealed several interesting things. Firstly, the students with dyslexia scored 0.477 standard deviations higher on the transmissional factor than the students without dyslexia ( $p < 0.05$ ). Secondly, the dyslexic students scored -0.779 standard deviations lower on the latent transactional means ( $p < 0.001$ ). These results indicate that students with dyslexia believe that reporting accurate information from authoritative sources is more important for academic writing than writers without dyslexia, and that they find it less useful as a means of developing their subjective understanding than writers without dyslexia. In fact, because the biggest variation between the dyslexic and non-dyslexic students was observed with transactional beliefs, this implies that viewing writing as a means for developing understanding may be the most important difference between the two groups. The discrepancies in transmissional and transactional means between the writers with and without dyslexia also suggests that dyslexic UK university students are less emotionally and

cognitively engaged with academic writing than those without dyslexia. Finally, the dyslexic students scored 0.321 standard deviations higher on the planning factor than those without dyslexia ( $p < 0.05$ ), suggesting that students with dyslexia prioritise planning more before and throughout writing than those without dyslexia.

The variation in latent means between the groups suggest several things about the impact that writing beliefs might have on writing outcomes. For example, the fact that the dyslexic students had a lower transactional factor mean could partially explain why in previous research, dyslexic students have reported finding it difficult to express their ideas in writing (Mortimore & Crozier, 2006) - if transactional beliefs represent a higher level of cognitive and emotional engagement in writing, and a tendency to try to develop subjective understanding over the course of writing, these writers may find expressing their ideas in written text relatively easy. The fact that the dyslexic students had a lower mean on the transactional factor in comparison to the students without dyslexia implies that they could find developing understanding during writing challenging, which might result in difficulties in organising and expressing ideas.

Moreover, the fact that the dyslexic students scored higher on the transmissional factor suggests that students with dyslexia believe that reporting accurate information is an important aspect to consider when writing, more so than the students without dyslexia. As such, these writers may only experience a limited amount of cognitive engagement within writing, because they will be primarily concerned with transferring the information they know to be true, rather than manipulating and building on that knowledge to develop their own ideas around the writing topic. Again, this might give some insight into why dyslexic writers find it difficult to organise and express their ideas in writing.

Additionally, the dyslexic students had a higher planning mean than the non-dyslexic students, suggesting a tendency to prioritise planning before and during the course of writing. As previous evidence has shown, however, whilst controlled planning is beneficial for text quality, it can stunt the extent to which writers develop their subjective understanding (Baaijen et al., 2014). As such, the fact that dyslexic students had a higher planning factor mean implies that these students may

be limited in the extent to which they develop their own ideas around a topic over the course of writing.

It is also important to consider, though, that although dyslexic students had higher means for the transactional and planning factors than the students without dyslexia, these are not necessarily negative outcomes. These results could indicate that dyslexic university students represent a population of individuals who have actually learned to cope particularly effectively with academic writing. For example, dyslexic students' high transactional beliefs might reflect trends in academic writing tuition. All students are likely to have been taught that reporting accurate information is an important outcome in academic writing. Hence dyslexic writers' high transactional beliefs may echo their consideration of this information. High planning beliefs might also represent a trend towards tutors and lecturers highlighting planning as a useful tool for improving academic writing outcomes. However, the development of these skills might have come at a cost for dyslexic students. That is, they may over-emphasise the importance of accurate reporting and planning within their own schemas, to the extent that it damages writing outcomes such as subjective understanding development. Future research should continue to investigate whether students with and without dyslexia score differently on the latent constructs represented in the WBI and how this subsequently might be related to their writing outcomes.

One final thing to consider is how the results from the latent means analysis in this study relate back to the nature of dyslexia. As highlighted previously, dyslexia is defined as a problem with the encoding and decoding of individual words. As such, the findings from this study suggest that these word-level issues might increase cognitive load during the process of writing, making it hard for dyslexic writers to engage more cognitively and affectively with writing. Subsequently, this could result in them finding it difficult to develop their understanding (as represented by their low transactional beliefs). Instead, it could force them to focus more on transmitting the information they already know (represented by their high transactional beliefs). It could also explain why dyslexic writers prioritise planning before and during writing, because they may see this as a way of mitigating some of the impact that the effects of dyslexia have on their text quality.

Hence, the potential relationships between transactional, transmissional and planning beliefs, and the outcomes of writing for dyslexic students are worth investigating in future studies because such evidence would help to deepen our understanding of why dyslexic individuals find academic writing difficult (beyond word-level issues).

To conclude, the results from this study give initial evidence to suggest that the WBI has a five-factor structure representing revision, planning, transmissional, transactional and audience beliefs. It also indicates that ESEM gives a more valid model representation than CFA because items are likely to cross-load onto multiple factors. Additionally, the results imply that the WBI does not produce measurement bias when used with dyslexic students and so can be applied to examine the writing beliefs of UK university students with and without dyslexia simultaneously.

Furthermore, comparison of the two groups' latent means revealed some important differences between students with and without dyslexia in their beliefs about writing. However, because this study does not formally assess writing outcomes, we do not really know whether these differences are positive or negative, in terms of their impact on writing in dyslexic students. As such, future research should look at how writing beliefs relate to text quality and subjective understanding in samples of dyslexic writers.

Finally, it is important to highlight that this study utilised a relatively small sample size of dyslexic students and, as such, the results presented here should only be considered preliminary. Further investigation of the WBI factor structure, the number of items that should make up the scale, and whether it is suitable for students with and without dyslexia should be conducted in order to fully validate the scale for usage in the population of UK university students.

## **4 Study 2: A Transparent Methodology for the Analysis of Keystroke Data in Relation to Underlying Writing Processes.**

### **Introduction**

This chapter starts with a background discussion about how keystroke logging can provide valuable insight into processing during writing but that typically, keystroke logging research has fallen victim to issues of reproducibility and transparency, which ultimately affects the reliability and validity of findings. In response to these issues, Study 2 is then presented, with the first aim of this methodological study being to create a reproducible framework for extracting keystroke features typically used to make inferences about writing processes (i.e. pauses, bursts and revisions).

Pauses, in particular, are considered a useful feature for making inferences about the types of writing processes individuals use. Researchers have started to use Gaussian Mixture Modelling (GMM) to examine the structure of pauses at different intervals of text production (within-word, between-word, sub-sentence, between-sentence and between-paragraph; Baaijen et al., 2012; Chenu, Pellegrino, Jisa, & Fayó, 2014; Roeser, De Maeyer, Leijten, & Van Waes, 2021; Van Waes, Leijten, Roeser, Olive & Grabowski, 2021). However, there is great difficulty in trying to align pauses with writing processes, because a single pause may be representative of *multiple* processes. As such, the second aim of this study was to use GMMs in combination with experimental manipulation of a writing task (i.e. varying planning and drafting strategies), in order to see whether this experimental manipulation made it easier to align pauses with processes.

Finally pauses, bursts and revisions are typically used in isolation when trying to interpret how keystrokes align with processes. However, because the three features are interrelated, there is an argument for using them in combination (Baaijen, Galbraith & de Glopper, 2012). As such, the final aim of this study was to try to replicate the Baaijen and Galbraith (2018) composite keystroke measures of global linearity and sentence production. Success in replicating these measures would suggest that they are reliable and reproducible, warranting their use in future studies that use keystrokes to examine writing processes.

This chapter concludes with a discussion of the results, highlighting how challenging and complex it is to create a reproducible framework for the analysis of keystrokes. This section also discusses the preliminary evidence that the study provides about how experimental manipulation of the writing process may impact the amount of time that writers spend on cognitively demanding, reflective writing components, such as planning and revision. Finally, the discussion reflects on the successfulness of the replication of the Baaijen and Galbraith (2018) global linearity and sentence production measures, what this says about the reliability and reproducibility of the measures, and whether they are appropriate measures to be used in other studies that examine writing processes through keystroke analysis.

The body of work in this chapter builds upon work first published by Hall, Baaijen and Galbraith (2022). Hall and colleagues investigated methods for constructing theoretically informed measures of pause duration in experimentally manipulated writing. This chapter extends that work by also incorporating the analysis of bursts and revisions. It also extends the mixture modelling work presented in that study by conducting similar analyses on a larger dataset.

## **Background**

Keystroke logging has become a popular tool for analysing the writing time course. Tools such as Inputlog (Leijten & Van Waes, 2013) and Scriptlog (Andersson et al., 2006) track all computer keyboard and mouse inputs during a writing session and then provide the researcher with a real-time breakdown of how an individual has composed a text. Specifically, keystroke loggers give information about action time and empty time. Action time is the time a key is pressed down for, and empty time is the time between key presses.

To try to make sense of keystroke data, the researcher needs to map the keystrokes and mouse inputs onto pauses, bursts, and revisions, which can subsequently be used to make inferences about writing processes. However, in doing this, the researcher is faced with a problem of alignment (Baaijen et al., 2012; Galbraith & Baaijen, 2019). That is, the data obtained from a keystroke log are not only representative of language production, but also other underlying writing processes such as

problem-solving and editing. Hence, trying to distinguish how keystroke features align with processes is difficult at best, and potentially impossible at worst (Baaijen et al., 2012; Schilperoord, 2001).

Moreover, many writing researchers use procedures to make inferences about how keystrokes relate to writing processes that are not transparent and reproducible. Issues of reproducibility have been a concern within the psychological sciences for a while, and there has been a push towards the use of more open and transparent practices within the field (Aarts et al., 2015; Munafò et al., 2017). Yet, as a domain, writing research is still not very transparent about the methods used to analyse keystrokes - this is both in terms of the literature and the programs used to collect keystrokes, as will be demonstrated throughout this chapter. First though, it is important to introduce some of the issues that researchers face when trying to align keystrokes with writing processes.

### *Identifying Pauses, Bursts and Revisions*

**Pauses.** In a keystroke log, pauses reflect empty time in which no text production occurs. Pauses are thought to be representative of multiple writing processes, including rereading of previously written text, planning for the next unit of text and internal revision of planned language production (Baaijen et al., 2012). Due to the numerous processes that may be reflected by pauses, it is a standard technique in writing research to analyse ‘cognitive pauses’ - pauses that fall above a specific time threshold. This is based on the notion that in the classic cognitive models of writing (Bereiter & Scardamalia, 1987; Flower & Hayes, 1981; Hayes, 2012), planning and revision are the main higher-order procedures involved in writing, and so it is important to identify the writing features representative of these longer, reflective processes. Typically, pauses above a threshold of two seconds have been identified as representative of reflective processes, whereas pauses falling below this threshold are more likely to represent automated transcription processes (Leijten & Van Waes, 2013).

There are several issues with this approach to identifying which pauses to analyse. Firstly, the standard two-second threshold is uninformed. There is not much empirical basis for the decision to use approximately two seconds as the cut-off point for what should and should not be considered a

pause (Chenu et al., 2014). Secondly, one could argue that individual differences in typing skill and cognitive processing time mean that a generalised threshold approach to identifying pauses is invalid because that threshold would vary between writers in real life. Also, using thresholds to identify pauses means that, typically, pauses below the threshold are ignored. However, pauses that fall below threshold might give considerable insight into the types of processes individuals use throughout writing. Only analysing above-threshold pauses leads to significant data loss. Finally, and perhaps most importantly, studying pauses above a threshold does not make the process of mapping those pauses onto writing processes any easier because a single above threshold pause could still be representative of underlying planning, rereading or revision processes.

Moreover, there are discrepancies in how researchers define pauses based on their associated keystrokes. Pauses can occur at the within-word, between-word, between-subsentence, between-sentence or between-paragraph units of text production. A between sentence pause, for example, is defined in Inputlog on the basis of the <FULL STOP> keypress (Leijten & Van Waes, 2013). The keystrokes associated with said pause include the last <LETTER> keypress before the <FULL STOP>, the <FULL STOP> itself, and all keypresses leading up to and including the first <LETTER> press of the new sentence. These keys are classified by Inputlog (Leijten & Van Waes, 2013) as being ‘before sentence’ or ‘after sentence’, situated around the <FULL STOP>. It is up to the researcher whether to treat the empty time between each key press associated with the between sentence transition as separate pauses, the sum of the ‘before sentence’ keys as one pause and the sum of the ‘after sentence’ keys as another pause, or the sum of all the keys associated with the between-sentence transition as a single pause.

There are issues with each of these approaches. Taking each inter-key interval as a single pause leads to the researcher obtaining a collection of, mostly, very short pauses. If using the cognitive pause approach, they will struggle to get many pauses (if any) that reach the two-second threshold. Using the ‘before sentence’ and ‘after sentence’ approach is problematic because keystroke loggers, such as Inputlog, do not distinguish between linear and non-linear text production. For example, when a writer presses the <CAPS LOCK> key to initiate a new sentence, Inputlog classifies the interval before this keypress as a before-sentence transition, regardless of whether that <CAPS LOCK> occurs

directly after the <FULL STOP> and <SPACE> keypresses, or whether that keypress occurs after a revision or movement away from the leading edge. Hence, the researchers who take the sum of the duration of keypresses and inter-key intervals that are part of the before sentence transitions and after sentence transitions separately (based on the automated output that Inputlog provides) might underestimate the amount of time that could elapse between the end of a sentence and the start of the next. On the other hand, researchers who take the sum of all the keys and inter-key intervals associated with a between sentence transition as a single pause create a very shapeless, nondescript measure unless they separate linear pauses from non-linear event transitions.

The technique of separating linear pauses from non-linear event-based transitions was identified by Baaijen et al. (2012). They emphasised the importance of separating pauses formed as part of forward text production (linear pauses) from other pauses associated with events such as revisions or insertions (non-linear events). This is because pauses that form part of forward text production are more likely to be representative of sentence production processes (Baaijen & Galbraith, 2018). The fact that a linear pause only leads to the production of new text means that the underlying processes that pause reflects are more likely to be associated with planning content. On the other hand, pauses associated with revisions or insertions are more likely to represent how recursively a writer creates the global structure of their text (Baaijen & Galbraith, 2018). To date, though, not a lot of writing research explicitly distinguishes between linear pauses and non-linear event transitions. So, it is presumed that many researchers treat all pauses, regardless of whether they are associated with forward text production or revisions/insertions, in the same way. This is problematic because the underlying processes of the two pause types are highly likely to be different.

**Bursts.** One way of further defining how keystroke features may be related to specific writing processes is to examine bursts of text production. Bursts are defined as text production that is uninterrupted by a pause, insertion or revision. Bursts are identified based on how they are terminated. A burst that ends in a pause of two or more seconds is considered a P-burst, whereas a burst that is terminated by a revision or insertion is considered an R-burst (Chenoweth & Hayes, 2001). This distinction alone is relatively uninformative, as a burst that is terminated by a pause and starts with a pause would be treated in the same way as a burst that was terminated by a pause but started with a

revision. Clearly, though, the former may be more likely to represent processes associated with planning for the next unit of text. In contrast, the latter could relate to processes associated with the revision of content. Hence, Baaijen et al. (2012) went one step further and sub-categorised R-bursts and P-bursts into 13 different burst types, relating to the kind of pause, revision or insertion they were initiated and terminated by. Baaijen and colleagues' (2012) work on categorising bursts made it easier to accurately assess how bursts related to underlying writing processes, but there are some issues that remain. Notably, in their paper, the authors do not discuss whether bursts that are terminated by the combination of a pause and a revision should be treated as a P-burst, R-burst or combination of both. This also goes for identifying how the burst is initiated. Additionally, they do not discuss whether bursts that are terminated by a mouse movement or scroll should be considered P-bursts or R-bursts. Since there is no definite movement away from the leading edge, one could consider this a P-burst. However, because the burst is interrupted by a mouse input, rather than a pure pause, one could consider it an R-burst. In fact, when it comes to identifying the keystroke combinations that represent a revision, little is said about how they should be identified.

**Revisions.** Generally, revisions are thought to reflect semi-automatic adjustments within text production, such as correcting spelling errors. They can also reflect systematic attempts to change the semantic meaning of the content of a piece of writing (Baaijen et al., 2012). However, there is no standardised way of defining where a revision starts and ends in terms of keystrokes. For example, in cases where a single word may be deleted and replaced, it is sometimes difficult to understand whether researchers categorise just the deletion as the revision component or whether the typing of the replacement word is also considered part of the revision.

Additionally, many researchers fail to treat post-draft revision as a separate component to the revision that happens within normal text production. Often, when writers finish composing the first draft of a text, they will systematically go back through their draft and make revisions in a relatively linear fashion. Baaijen et al. (2012) argue that this type of revision should be considered separately from the revisions that happen within first draft text production because they could reflect different writing characteristics.

It is clear to see, then, that a lack of transparency in how researchers identify and analyse pauses, bursts and revisions, means that it is often difficult to replicate keystroke analysis. However, it is unlikely that researchers are actively trying to make their research irreproducible, and so it is essential to discuss why transparency and replicability issues may exist within the field.

### ***Why Do Reproducibility Issues in Writing Research Exist?***

The most apparent issue with the lack of transparency in how researchers identify and analyse pauses, bursts and revisions is that it is challenging to replicate analysis. This creates problems because replication allows researchers to test whether results are genuine (Munafò et al., 2017). Replication can take three forms. Most notably used in computational science, ‘exact replication’ refers to researchers taking data that was previously analysed and using the same procedures as in the original study to analyse the data in the hopes that they will obtain the same results (Peng, 2011). In terms of writing research, exact replication may be difficult to achieve even if the researcher shares their original data and the methods they used to analyse them. This is because keystroke datasets are so large that analysing them can be a complicated procedure. Hence, any variation from the original analytical techniques may lead to different results. Thus, researchers have to be highly transparent about the steps they used to obtain their results, and often, even if done accidentally, researchers may miss important analytical details.

The second type of replication is ‘direct replication’. That is, a study is reconducted by a different laboratory but using the same procedures as in the original study. This results in data being collected from a different (albeit similar) sample and then analysed using the same methods as in the original study (Simons, 2014). A strength of direct replication is that it allows researchers to see whether results can be reproduced with participants outside of the original study. Such findings would give evidence that the examined phenomenon is generalisable, rather than being a quirk of the original sample. However, when it comes to writing research, direct replication falls victim to the same hurdle as exact replication, in that any deviation from the original analytical methods (even if accidental) could result in different findings. Moreover, career incentives are not aligned with the notion of

conducting direct replications. Many writing researchers would not be able to ‘afford’ to try to replicate results from previous studies, as replication is not as valued as novel findings by many journals within the psychological sciences (Poldrack, 2019). As the replication crisis is well documented within psychology now, some funding has been made available for laboratories to conduct direct replications (Munafò et al., 2017). However, this funding does not seem to be very readily available within the field of writing research.

Additionally, the fact that many writing research journals have word limits for their articles means that even if researchers wanted to disclose their complete analytical procedures, they may not be able to do so in the number of words they are allowed. An increasing number of journals have tried to overcome this issue by allowing researchers to upload supplementary materials with their journal articles or encouraging them to upload additional information related to methods and analysis to an online repository such as the *Open Science Framework* (Springer, 2021). Nonetheless, the lack of emphasis on replication within the writing research field may discourage researchers from uploading supplementary materials to a journal or online repository. It is a time-consuming procedure that many researchers do not feel they *need* to commit to (nor have the time to) for their work to be published.

The third type of replication is ‘conceptual replication’, the idea of retesting a hypothesis, rather than a set of methods to reach a result (Simons, 2014). In the field of writing, this would mean that researchers could analyse keystrokes using varying approaches to test the same hypotheses about how pauses, bursts and revisions relate to writing processes. The issue with this, though, is that there is such wide variation in how pauses, bursts and revisions are identified and how their durations are calculated, that it is difficult to know whether writing researchers between studies are truly examining the same features. A side effect of this could be that whilst two separate studies appear to test the same relationship between, say, pauses and writing processes in university students’ essay writing, if the definitions used to identify those pauses are different between the two studies, we cannot be sure that they are measuring the same concepts. So, drawing comparisons between the two studies could be futile.

Given the reproducibility issues in writing research explained in this section, an apparent research problem exists. The lack of standardised, transparent approaches for analysing keystrokes

concerning underlying writing processes makes it hard to reproduce results. However, the difficulties associated with and lack of incentives encouraging researchers to create fully transparent and replicable keystroke analysis methods means that reproducibility issues could be standard within the field for the foreseeable future. Therefore, it is crucial to develop a transparent and reproducible method for defining, identifying and calculating the lengths of pauses, bursts and revisions through keystroke data.

### ***The Reliability of Mapping Keystroke Features onto Writing Processes***

Given that keystroke research is fraught with issues that can affect the reproducibility of findings, it is also important to discuss whether researchers can really use keystrokes to understand writing processes in an accurate way. As discussed already, pauses, bursts and revisions in a keystroke log can be representative of hidden elements of the writing process (e.g., planning and reflection), as well as explicit text production, so trying to map such features onto those processes poses particular challenges. However, some techniques have been developed to make that process both doable and more reliable.

**Mixture Modelling.** Mixture modelling has been used to examine how the complex structure of pause data may represent several underlying writing processes. Usually, pause data is very positively skewed, and whilst natural-log transformations help to reduce this skew, the data do not usually form a normal distribution (Baaijen et al., 2012; Conijn et al., 2019). Hence, writing researchers have attempted to use Gaussian mixture models (GMM) to test whether pause data are better represented by a mix of normal distributions rather than a single normal distribution. The theory is that using mixture modelling to reveal a multi-distribution latent structure in pause data can help narrow down how those latent components map onto processes. For example, Baaijen et al. (2012) examined linear pause data from an essay writing task and found that between-word pauses could be broken into a three-component GMM structure. They argued that the first distribution, which captured a high frequency of very short pauses, might represent word-retrieval processes. The middle distribution, which captured fewer mid-length pauses might describe phrase boundary processes, and

the right-hand distribution, which captured a tail of longer, miscellaneous pauses, likely represented higher-level processes such as planning or revision. Later work from Roeser et al. (2021) and Van Waes et al. (2021) found that inter-key intervals (the pauses between each individual key press) in copy task data could be modelled with two Gaussian distributions, which they theorised represented fluencies and disfluencies. The vital thing to note with all these studies is that a multi-component Gaussian mixture model better fit the pause data than a single Gaussian distribution model. Because it was more valid to model those pauses as a mixture of distributions, researchers believe that each distribution within a mixture model could represent a different group of writing processes.

It is also important to note that studies differ in terms of whether a two- or three-component model was best representative for their pause data. This was likely due to the different studies examining different pause boundaries. For example, Roeser et al. (2021) and Van Waes et al. (2021) examined every inter-key interval, whereas Baaijen et al. (2012) studied pauses within-words, between-words, between-sub sentences and between-sentences separately. So, differences in the number of components observed at varying pause intervals could be due to differences in the number of processes occurring at the different levels of text production.

Another factor likely to impact the number of components that can be fitted at pause boundaries include the types of writing task participants are assigned to. For example, a study that uses an argumentative writing task might elicit different processes in comparison to a study which uses a copy task. A copy task, for instance, does not require the writer to generate content, whereas an argumentative essay writing task would. Hence, it is very possible that fitting a GMM to pause data between-words in a copy task will produce fewer model components than fitting a GMM to between-word pause data in an essay writing task, because fewer cognitive processes are involved in copy task writing. This poses the question of whether manipulated differences in the types of tasks used in writing experiments can result in contrasts to the pause data structure observed with mixture models. If this could be done, it might help narrow down which processes pauses could be mapped onto because researchers could manipulate the writing task to elicit specific types of writing processes. Hence, judgements about the kinds of processes that writers engage in could be more accurate. To date though, this has not been tested.

**Global Linearity and Sentence Production Measures.** Pauses are not the only measure taken from keystroke logs that researchers have tried to map onto processes. Baaijen & Galbraith (2018) developed two-composite measures based on pauses, bursts and revisions (and some additional writing features) to assess global linearity and sentence production in computer-based writing. The authors found that the reliability of each measure was high (global linearity  $\alpha = .79$ ; sentence production  $\alpha = .80$ ). People who scored highly on the global linearity measure were observed to produce texts linearly - that is, they did not revise the structure of their text much throughout the process of writing. Hence, writers with low global linearity scores were thought to use revision processes more throughout writing than those with high global linearity. In terms of the sentence production measure, writers who had a high score were observed to pause for longer between sentences and made fewer sentence-level revisions. On the other hand, writers who had low scores on this measure were more likely to revise but pause only briefly throughout sentence production. Hence, writers with a high sentence production score were theorised to use more controlled and planned processes throughout writing than those with a low score, who were much more spontaneous in their sentence production.

The authors found that these two measures were related to two crucial writing outcomes: text quality and subjective understanding development. Chapter 1 (pp.63 - 65) discusses in detail exactly how global linearity and sentence production were related to these two outcomes. For the purpose of this study, however, the main point to consider is that the two composite measures were useful for aligning how keystroke features could map onto the two systems of the dual process model (Galbraith, 1999; Galbraith 2009; Galbraith & Baaijen, 2018), with high levels of revision to the text's global structure (low global linearity scores) implying engagement with the knowledge transforming system, and spontaneous sentence production (low sentence production scores) indicating engagement with the knowledge constituting system. To date, though, the composite measures used in Baaijen and Galbraith (2018) have only been examined within that study. Whilst the study itself implied that the two measures had good reliability, no one has tried to follow the methods used for compiling the measures in a separate study. Hence, we do not know whether they are easily replicated, and thus whether they can be used in further research.

### *Aims of the Current Study*

The discussion presented in this background section highlights three main issues with keystroke research. Firstly, there is no standardised, reproducible framework for the analysis of pauses, bursts and revisions from keystroke data. Secondly, whilst mixture modelling has recently emerged as a method for trying to map writing features onto processes, it is difficult to know how accurate that mapping is, because mixture modelling studies show that pauses at specific boundaries (e.g. between-sentences) are multi-modal, and hence representative of multiple underlying processes. Trying to narrow down what those processes might be is challenging. Finally, whilst two composite measures that incorporate pauses, bursts and revisions to analyse writing processes and their relationships with writing outcomes exist, they have only been employed in one study. As such, we do not yet understand the reproducibility of those measures, and whether they would provide similar results in other studies. Therefore, this study aims to address the following research questions:

1. How do we create a reproducible method for defining, identifying, and calculating pauses, bursts, and revisions from keystroke data?
2. Do Gaussian mixture models reveal differences in the durations and mixing proportions of pauses throughout text composition in the outline and synthetic drafting strategies?
3. Can the Baaijen and Galbraith (2018) two-component composite measures of global linearity and sentence production be reproduced on a new sample of university students' essay writing keystroke data?

## **Methods**

### *Participants and Data Collection Methods*

The data from this study were collected from a keystroke logging experiment conducted with UK university students who had English as a first language (N = 48). In the experiment, participants

completed an argumentative essay writing task in 30 minutes, during which time their keystrokes were logged. The essays were about one of two current-affairs topics: social media or veganism. Participants were randomly assigned to either outline planning with complete drafting, or synthetic planning with rough drafting writing conditions. Before essay writing, participants in the outline planning with complete drafting condition were given 5 minutes to create a handwritten essay outline which indicated their main ideas about their essay topic and the order the ideas were to go in for the essay. For the essay writing component, the participants were instructed to type a final version of an essay, which was well organised and had accurate spelling. In the synthetic planning with rough drafting condition, participants were given 5 minutes before writing to handwrite a single sentence that summed up their main ideas about their essay question. Notably, they were not allowed to write a full plan for their essays. Following this, the participants were asked to write a rough draft of an essay, in which they did not have to worry about spelling or how well organised the piece was. However, they were told to discuss the question with themselves and write down their thoughts as they occurred spontaneously. A detailed description of the methods for this study can be read in Chapter 2 (in the procedures section about Study 3, p.86).

### ***Experimental Manipulation to Elicit Different Writing Processes***

The design of this study replicated most of the procedures from Baaijen and Galbraith (2018) and Baaijen et al. (2014) with one clear difference. In the original studies, participants across both conditions were instructed to write a final version of an essay (i.e. a well-organised essay with accurate spelling). In this study, though, participants in the synthetic condition were told to write a rough draft of an essay. The reason for this change was to try to elicit different writing processes across the conditions. Specifically, it was hypothesised that writers in the outline condition would write in a more planned and controlled way due to being asked to create well-formed essays and also being allowed to create a structured plan before writing. In contrast, writers in the synthetic condition were hypothesised to write in a more spontaneous and unplanned way, due to being told to write

down their thoughts as they occurred, not having to worry about spelling or essay structure, and not being allowed to make a structured plan prior to writing.

No specific hypotheses were made about exactly how the experimental manipulations to elicit different writing processes would manifest in the mixture models of the pause data for the two groups. Rather, the aim was to explore whether the mixture models would show differences in the mixing proportions, means and standard deviations of the components depending on the kind of planning carried out in advance of writing the text.

### ***Creating Keystroke Logs***

The keystroke logs of the essays were collected using Inputlog Version 8 (Leijten & Van Waes, 2013). Using the Inputlog program, general analysis xml files were then made for each participant. Next, these files were converted into xlsx documents using the 2016 Windows version of Microsoft Excel. All initial preparation of the keystroke logs was conducted within Microsoft Excel.

### ***Creating a Reproducible Framework for Analysing Pauses, Bursts and Revisions***

After converting the data in the Excel general analysis files into a more reader-friendly format (using the first set of macros designed for this project, which were published in Hall et al. (2022) and can be accessed at: DOI 10.17605/OSF.IO/R53H2), all instances of pauses, bursts and revisions within the logs were identified. Since there are no standardised conceptual definitions for pauses, bursts and revisions, it was essential to develop these definitions first so that other researchers would be able to understand exactly how these features were defined within the study. These definitions are based on those used by Baaijen et al. (2012) but are further clarified here to make them completely explicit. Table 5 lists these definitions. The definitions of the linear pauses and non-linear events are also the definitions used in Hall, Baaijen and Galbraith (2022).

**Table 5***Conceptual Definitions of Keystroke Features*

Keystroke feature	Conceptual Definition
Linear pause	A clean, 'pure' transition between keystrokes. That is, there is only forward progression and no instances of revision, mouse movements, or insertion and edits away from the leading edge.
Linear within-word pause	The time between the letters pressed at the within-word level, when there are no instances of revision, mouse movements, or insertion and edits away from the leading edge.
Linear between-word pause	The time between the end of a word and the beginning of the next word when there are no instances of revision, mouse movements, or insertion and edits away from the leading edge.
Linear sub-sentence pause	The time between the end of a word that is followed by a comma, and the start of the next word that is preceded by the same comma, when there are no instances of revision, mouse movements, or insertion and edits away from the leading edge.

---

Linear between-sentence pause	The time between the end of a sentence and the beginning of the next sentence, when there are no instances of revision, mouse movements, or insertion and edits away from the leading edge.
R-burst	An overall category of uninterrupted text production that is terminated by a revision.
P-burst	An overall category of uninterrupted text production that is terminated by a pause lasting at least two seconds.
I burst	An overall category of uninterrupted text production which is inserted away from the leading edge.
PP burst	A subcategory of the P-burst that is initiated and terminated by pauses that last at least two seconds.
RP1 burst	A subcategory of the P-burst that is initiated by a revision and terminated by a pause of at least two seconds. This burst only includes the production of new text.
RP2 burst	A subcategory of the P-burst that is initiated by a revision and terminated by a pause of at least two seconds. This burst includes the replacement of previously written text and the production of new text.
RP3 burst	A subcategory of the P-burst that is initiated by a revision and terminated by a pause of at least two seconds. This burst only includes the replacement of previously written text.
PRL burst	A subcategory of the R-burst. It is initiated by a pause of at least two seconds, and is terminated by a revision at the leading edge.
PRI burst	A subcategory of the R-burst. It is initiated by a pause of at least two seconds and is terminated by a revision away from the leading edge.

---

---

PRLI burst	A subcategory of the R-burst. It is initiated by a pause of at least two seconds and is terminated by a revision at the leading edge and a revision away from the leading edge.
RRL burst	A subcategory of the R-burst. It is initiated by a revision and terminated by a revision at the leading edge.
RRI burst	A subcategory of the R-burst. It is initiated by a revision and terminated by a revision away from the leading edge.
RRLI burst	A subcategory of the R-burst. It is initiated by a revision and terminated by a revision at the leading edge and a revision away from the leading edge.
IG burst	A burst of text production that is initiated in the middle of sentence production but happens away from the leading edge. It is an insertion of text within the current sentence.
IR burst	A burst of text production that is initiated at the end of sentence completion. The insertion of text happens in the sentence that has just been finished.
IB burst	A burst of text production initiated in the middle of sentence production and happens away from the leading edge. The insertion occurs over the sentence boundary (outside of the sentence currently being produced). The insertion is no longer than a sentence.
Minor revision	The deletion of low-level spelling and grammar related edits (i.e. corrections to typos).
Major revision	Any deletions to the text that are not minor revisions. Minor revisions can also include the insertion of characters that are less than word length (e.g., a few letters or grammatical changes).

---

The conceptual definitions in Table 5 describe important writing features that can be extracted from keystrokes. However, they do not explain *how* to extract those features from a keystroke log. Thus, the next stage was to develop a framework for the identification of these features. Appendix K (pp.253-262) presents a full framework, demonstrating how the features were extracted from the keystroke logs on the basis of a set of defined rules. It also demonstrates how one would calculate the duration of each of the features.

In developing this framework, it became clear that whilst it was easy to create reproducible rules for the calculation of linear pauses, it was less easy to create these rules for the bursts and revisions, as so many different keystroke combinations could be used to create an example of these events, that it would be very difficult to dictate every single combination. Therefore, whilst the rules used to identify linear pauses were exact, the terms used to identify bursts and revisions were set by general guidelines rather than rigid instructions.

### ***Automating the Extraction of Pauses, Bursts and Revisions***

After creating the conceptual definitions of each pause, burst and revision type, along with rules for identifying them within the keystroke logs and the methods for calculating pause times, these features needed to be extracted from the keystroke logs. Part of this extraction was automated for several reasons. Firstly, using a set of code to run across all of the Excel keystroke files to extract the features makes the process fast. This is particularly useful in the analysis of keystroke logs because each file could have thousands of rows of data, so hand-coding the pauses, bursts, and revisions would have been highly time-consuming. Another reason is that if the same code is used to analyse each keystroke file, the procedure is standardised. This is useful because it ensures that each keystroke feature has been identified using the same rules. A final reason for automating the analysis is that the code used to do so can be made openly available so that other researchers can run the same analytical procedures on their data, or run a variation of the analysis by editing the original code. Making the code openly available means that the techniques used to analyse the keystrokes are transparent and reproducible.

However, in attempting to automate the analysis of pauses, bursts and revisions, several issues were encountered. Automating the analysis of linear pauses was relatively straightforward, as there were a limited set of keypresses that could be included within a linear pause. The code used to automate the analysis of pauses was written using a series of VBA macros in Excel, and the development of this code was an iterative process. After the initial set of code was created (based on the rules identified in the calculation framework), the macros were run on each participant's keystroke logs. The keystroke logs were then manually checked to ensure that the macros were correctly identifying pauses based on the rules defined in the code. The additional benefit of screening the macro outputs at this stage was that it enabled the identification of cases that could have been conceptually considered linear pauses, although they had not been identified in the calculation framework and subsequent macros. Two such instances were found.

The first example is demonstrated in Figure 5. The participant had come to the end of one sentence and continued linearly onto the next. Hence, the time between the start of the last character in the previous sentence and the start time of the first character in the new sentence should have been defined as a linear between-sentence pause. However, because the participant did not use the standard sequence of keypresses for a linear between-sentence pause (i.e., <FULL STOP>, <SPACE>, <LSHIFT>, <LETTER>), but rather swapped the <LSHIFT> and <SPACE>, the original macros did not pick up this pause. This happened with multiple participants. It is important to note that the Inputlog software also did not classify this sequence of key presses as a sentence-level transition but rather a between-word transition. Hence, if the automatic output from Inputlog or the initial set of macros had been used to identify the between-word transitions, several linear between-sentence pauses would have been omitted from further analysis. This could have affected the results from this study, and hence the additional keystroke combinations were incorporated into the macro scripts.

**Figure 5**

*Example of a Non-Regular Sentence Termination Keystroke Combination (Hall et al., 2022, p.14). Reproduced with Permission of the Rights Holder.*

output	startTime	endTime	actionTime	pauseTime	pauseLocationFull
i	569063	569123	60	136	BEFORE WORDS
t	569123	569212	89	60	WITHIN WORDS
.	569282	569362	80	159	AFTER WORDS
LSHIFT	569411	569560	149	129	COMBINATION KEY
SPACE	569451	569521	70	40	AFTER WORDS
T	569501	569610	109	50	BEFORE WORDS
h	569591	569670	79	90	WITHIN WORDS
i	569739	569809	70	148	WITHIN WORDS
s	569839	569890	51	100	WITHIN WORDS

The second example was less easy to amend. For one participant, 32 between-sentence transitions were observed. Still, only two of those transitions were coded as linear between-sentence pauses based on the original macros. Manual inspection of the keystroke log revealed that in the majority of cases where the participant was ending a sentence, they pressed <SPACE> followed by <BACKSPACE> and <FULL STOP>. However, the duration of each of these keypresses was very short. It quickly became apparent that this individual was habitually pressing the <SPACE> instead of the <FULL STOP> at the end of the sentence because the space bar was automatically and habitually pressed at the end of every word. Hence, it was decided that this keystroke sequence did not reflect a revision but rather an individual difference in automatic typing skill. When composing the work for Hall et al., (2022), our team discussed whether the conceptual definition of a linear-between sentence pause that we had identified needed to change to include these types of minor revisions. Other researchers had done this (Baaijen et al., 2012; Galbraith & Baaijen, 2019). However, to create as reproducible of a calculation framework as possible, in this instance, we decided not to include minor revisions in linear between-sentence pauses. This is because we would have had to identify the number of backspaces that could be included within a linear pause, and the duration of those

keypresses also. Additionally, whilst in this case, we were certain that the participant was habitually pressing the <SPACE> key instead of the <FULL STOP>, in many participants, such a revision may reflect a change in the writer's sentence structure (i.e., mid-sentence they decide to end the sentence, rather than continue it).

Hall, et al.'s (2022) strategy was also generally adopted for this study, but in cases where habitual <FULL STOP> pressing was evident (e.g. with the participant described here), these cases were indeed categorised as linear between-sentence transitions. More ambiguous cases were left coded as minor revisions.

After writing and running the macros for the linear pauses, burst and revisions were hand-coded. This was due to high levels of variation in how writers created these two features, essentially meaning that any macros written to identify them would have been too basic and not comprehensive enough to identify every instance within the keystroke log. The benefit of hand-coding these complex keystroke features was that ambiguous cases, which did not fall into the general guidelines outlined in Appendix K could be detected, and the guidelines could be edited to include those ambiguities. Every hand-coded burst and revision within the keystroke files was logged to make the analysis completely transparent for other researchers to view (available on the project's Open Science Framework repository, link on p.12).

### ***Separating First-Draft Text from Other Types of Text Composition***

Baaijen et al. (2012) highlighted the importance of separating first-draft text production from other types of text production when it came to analysing keystroke logs because each type of text production could reflect a very different set of processes. Hence, analysing a keystroke log as one undifferentiated whole risks aggregating measures of pauses, bursts and revisions across different types of text production. Therefore, in their study, the authors removed sections corresponding to the production of titles, explicit planning, first-draft text production and end revision/post-draft revision. To make these distinctions in the current study, the keystroke files were analysed alongside the screen

capture footage that Inputlog Version 8 (Leijten & Van Waes, 2013) records during a keystroke logging session. This made it easier to visualise the different types of text production.

In this experiment, participants tended to use the essay questions that they were assigned as their title, so identifying this type of text production was straightforward. Most participants wrote their titles at the beginning of the writing session. Still, a couple of participants inserted their titles towards the end of text production, so screening the keystroke files alongside observing the screen capture footage made these cases easier to identify.

No participants created explicit plans within their keystroke logs in this study. This is likely due to the fact that all participants took part in some form of planning activity prior to writing their essays, so they may have not felt the need to create explicit plans within their word documents. The guidelines from Baaijen et al. (2012) note that in instances where people do plan, this can usually be identified because participants bullet point their ideas, or put each new idea on a new line, and so examining the keystroke logs and screen capture footage combined can help identify such cases.

In terms of end-revision, Baaijen et al. (2012, p.255) identified end-revision using the following:

“Many writers...show evidence of writing an initial draft and then going back systematically through the text, editing and revising the initial draft. Although this may sometimes involve producing extended chunks of text, we did not include this in the analysis of text production, on the grounds that it may be different in character to text produced as part of the initial draft. We defined end revision as occurring when an individual made revisions outside the final paragraph while writing the final paragraph. In the majority of cases, this amounted to revisions made after the final sentence. But there were some cases where individuals broke off to make revisions and then returned to write a final summary sentence or two. In these cases, these sentences were also excluded from text production analysis.”

To add to this definition, it is essential to highlight that end-revision happens systematically. When the writer makes their end revisions, they generally make them sequentially, from the start of

their texts through to the end of their texts. However, some writers may start to make end revisions in the first paragraph of their text product, whereas others may start further into the text. So, some element of subjectivity is required in determining when end-revisions start. Also, contextual cues can be used to identify when the individual is planning to finish their texts and hence when they may be likely to start making end revisions. These clues are usually phrases, such as “finally...”, or “to conclude...”. However, not all writers use such expressions within their texts, and so one cannot rely on contextual clues only to determine when a writer is drawing to a close.

As can be inferred, identifying end-revision uses some element of personal judgement on the researcher’s behalf, leading to differences between researchers in how much of a keystroke log is defined as end-revision. Hence, to make this distinction as accurate and informed as possible, the keystroke logs in this experiment were examined by three separate researchers for cases of end-revision. We then discussed which participants we identified as having made end-revisions and which parts of the text we had categorised in that way. This meant that we could discuss any ambiguities that could have led to differences in how we identified end-revisions, and resolve them where necessary. In the dataset from this experiment, approximately 50% of participants across both conditions made end-revisions within their keystroke logs.

After preparing the keystroke logs using the methods outlined in this section, first-draft text production had been isolated from title production and end revisions. Also, pauses, bursts and revisions were identified using the conceptual and calculation frameworks.

### ***Extracting the Baaijen & Galbraith (2018) Global Linearity and Sentence Production***

#### ***Measures***

The next task was to extract the relevant measures to form two composite scales of global linearity and sentence production from the Baaijen & Galbraith (2018) study. Table 5 gives each of the definitions for the keystroke measures as set out by Baaijen & Galbraith (2018), and the method in the current study used to extract them from the keystroke logs.

**Table 6***Sentence Production and Global Linearity Measures*

Sentence production	Definition	Calculation method in the current study
Mean pause duration between sentences	Only conducted on the linear between-sentence pauses. This is the mean of all the pause durations between sentences.	Calculated with a macro. Calculated as the mean of all identified linear between sentence pauses in a participant's keystroke log.
Percentage of pauses over two seconds between words	Only conducted on linear between-word pauses. This measure is the number of pauses that lasted two seconds or longer occurring between words. It is calculated as the proportion of the total number of linear transitions between words.	Calculated with a macro. Calculated 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.
Percentage of R-bursts	The number of RRL and PRL bursts expressed as a percentage of the total number of bursts.	Calculated with a macro. Calculated as the number of RRL and PRL bursts, divided by the total number of RRL, RRI, RRLI, PRL, PRI, PRLI, RP1, RP2, RP3, PP, IR, IG and IB bursts.

Percentage of words produced in P-bursts	The total number of words produced in P-bursts, expressed as a percentage of the total number of words in the keystroke log.	Calculated with a macro. Calculated as the number of words in a PP burst, divided by the total number of words within the keystroke log (which are counted by totaling the number of “before word” records in the “PauseLocationFull” column, which occur during PP bursts, divided by the total number of “before word” record in the “PauseLocationFull” column).
Text modification	The total number of characters produced in the keystroke log, divided by the total number of characters in the final text.	<p>Total number of characters in the keystroke log: calculated with a macro that counted the number of characters, spaces and returns within the keystroke log.</p> <p>Total number of characters in the final text product: calculated by taking the number of characters, spaces and returns from the final text product Word document.</p> <p>Text modification: Calculated with a macro as the number of characters in the final Word document, divided by the number of characters in the keystroke log.</p>
<hr/>		
Global linearity	The number of linear transitions between-sentences as a proportion of the total number of sentence transitions.	Calculated with a macro. Calculated as the number of identified linear between sentence pauses, divided by the

Percentage of linear transitions between sentences		number of full stops in the output column in the keystroke log.
Percentage of linear transitions between words	The number of linear transitions between-words as a proportion of the total number of word transitions.	Calculated with a macro. Calculated as the number of identified linear between-word pauses, divided by the total number of “BEFORE WORDS” records in the “PauseLocationFull column”
Number of production cycles	A sequence of language bursts produced without interruption. Each break from the leading edge is defined as the start of a new cycle.	Calculated with a macro. Calculated as the count of every “PAUSE” or “REVISION” record in the “Burst” column, that is directly followed by a “PRI, PRLI, RRI or RRLI” record in the “Burst” column.
Percentage of time spent on events	The percentage of total time spent writing that is devoted to operations other than producing text or pausing.	Calculated with a macro. Calculated as total time for all “REVISION” cells in the “Burst” column (which is the adjacent pause time and action time) over the total time in first draft-text production (which is the sum of all the pause times and action times for each blank cell in the “TitlePD” column.

Percentage of I-bursts	The number of I-bursts, expressed as a percentage of the total number of bursts.	Calculated as the number of IG, IB and IR bursts, divided by the number of RRL, RRI, RRLI, PRL, PRI, PRLI, RP1, RP2, RP3, PP, IR, IG and IB bursts.
Sentence linearity index	The proportion of the sentences in the final product that were produced in the same order as in the keystroke log.	<p>Calculated by hand, whilst observing the Inputlog screen capture footage for each keystroke log.</p> <p>Number of sentences out of order: calculated by counting every time an individual moved away from the leading edge to insert a new complete sentence. If a person inserted a new sentence away from the leading edge and another new sentence directly followed, another one is added to this score. The number of sentences out of order does not increase if the person just moves away from the leading edge to add text to or to edit an already existing sentence.</p> <p>Sentence linearity index: calculated as the number of sentences produced in order (which is the number of sentences in the final text product, minus the number of sentences made out of order) away from the total number of sentences in the final text product.</p>

---

Extracting these measures resulted in 11 items that theoretically represented either a global linearity component or a sentence production component. Previous research has indicated that the items load onto these two components with high reliability (sentence production  $\alpha = 0.79$ ; global linearity  $\alpha = .80$ ; Baaijen & Galbraith, 2018).

In what follows, an attempt to replicate the Baaijen and Galbraith (2018) measures is conducted. Reliability of the measures is also examined to see if similar alpha values can be achieved as in the original study. Before that, though, the study examines the linear pauses that were extracted from the keystroke logs in the outline/final drafting and synthetic/rough drafting writing conditions using Gaussian mixture models, to see whether experimental manipulation of the planning and drafting process resulted in differences in pause structures between the two conditions.

## Results

This results section is presented in two parts. First, the mixture models of the linear pause data from the experiment are presented to explore whether the writing condition manipulations have elicited different writing processes between the two groups. Then, the results from the reproduction of the Baaijen and Galbraith (2018) global linearity and sentence production measures are described.

### *Exploring whether Experimental Manipulation of Drafting Process Result in Variation in Linear Pause Structure with Gaussian Mixture Models*

Gaussian mixture models were used to examine the structure of the pauses in the outline and synthetic condition keystroke logs. The analysis was conducted in R Version 4.0.3 and with the package Mclust version 5.4.7. These models use the expectation-maximization algorithm (McLachlan & Peel, 2000), a technique for providing maximum likelihood estimation for data with an underlying latent variable structure (Do & Batzoglou, 2008). To do this, the EM algorithm first estimates a latent variable for each pause value within a participant's dataset. Following this, the algorithm optimises the parameters for the latent variables in the form of Gaussian distributions through an iterative

process until a suitable set of latent values that fits the data is found. Reproducible scripts for this analysis can be found in the project's Open Science Framework repository (link on p.12).

An initial screening of the pause data revealed that there were some unnaturally short pauses at the linear within-and between-word level (less than 50 milliseconds). Following the recommendations set out in previous pause modelling studies (Baaijen et al., 2012; Van Waes et al., 2021), the decision was made to eliminate these short pauses from the dataset because they were likely typos (indeed, inspection of the individual pauses revealed that they were typing errors, e.g. accidental <SPACE> presses). Histograms of the pause data also revealed that the data at each interval were positively skewed, particularly for the within- and between-word levels. Hence, natural log transformations were conducted to help with interpretation of the pause data structure. However, even after these transformations, the data remained positively skewed, especially for the within and between word pauses. Moreover, the histograms indicated that the pause data at each interval could be multi-modal. Previous research demonstrates that multi-component GMMs provide a better fit to pause data than single Gaussian models, and so that approach was taken for this study.

**Establishing a Common Model for Pauses at each Linear Interval.** To determine the maximum number of components to fit at each pause-interval, first a review of the literature on pause modelling with GMMs was conducted. This indicated that for linear within word pauses, or inter-key intervals, a two-component model was unanimously deemed the most representative for this type of data (Almond et al., 2012; Guo et al., 2018; Roeser et al., 2021; Van Waes et al., 2021). For linear between-word pauses, evidence points towards a three-component model being most representative (Baaijen et al., 2012). For linear sub-sentence and between-sentence pauses, evidence implies that a single Gaussian model may be most appropriate, but the evidence for this is less clear-cut (Baaijen et al., 2012).

Consequently, a preliminary analysis (published in Hall et al., 2022) was carried out on a subset of the log-transformed pause data from this study, using the 32 participants who took part in the study prior to the Covid-19 outbreak. Essentially, this analysis assessed the fit of one, two and three component models for each participant at each pause interval (i.e. within-word, between-word, sub-sentence and between-sentence). This was done to determine whether adding mixture components

provided a better fit than single Gaussian distribution models, and to test whether a common model for each pause interval could be established. Model fit was assessed using the Bayesian Information Criterion (BIC). The results are in Table 7.

**Table 7**

*Number of Components for the Best Fitting Models and Number/Percentage of Participants (Hall et al., 2022, p. 20). Reproduced with Permission of the Rights Holder.*

Linear pause boundary	Synthetic condition Best fitting number of components counts and percentages (BIC)	Outline condition Best fitting number of components counts and percentages (BIC)
Within-word	1 – 0 (0%) 2 – 12 (75%) 3 – 4 (25%)  (1845.504 to 5426.658)	1 – 0 (0%) 2 – 12 (75%) 3 – 4 (25%)  (2043.966 to 4806.208)
Between-word	1 – 0 (0%) 2 – 4 (25%) 3 – 12 (75%)  (593.8311 to 1899.715)	1 – 0 (0%) 2 – 5 (31%) 3 – 11 (69%)  (447.4839 to 1925.832)
Sub-sentence	1 – 2 (13%) 2 – 10 (63%) 3 – 1 (6%) NA – 3 (19%)  (0.992 to 64.975)	1 – 4 (25%) 2 – 7 (44%) 3 – 4 (25%) NA – 1 (6%)  (4.974 to 100.131)
Between-sentence	1 – 4 (25%) 2 – 7 (44%) 3 – 4 (25%) NA – 1 (6%)  (6.198 to 70.619)	1 – 4 (25%) 2 – 8 (50%) 3 – 2 (12.5%) NA – 2 (12.5%)  (2.949 to 85.096)

*Note.* NA\* indicates where model fit could not be assessed accurately for certain participants because there were too few pause observations.

The results in Table 7 demonstrate that a single distribution fitted the data best in only a few instances. In fact, at the within-word boundary, no participants had their pause data best represented by a single distribution. Moreover, at the between-sentence boundary, where it could be expected that linear transitions might uniformly reflect only higher order, complex processes (e.g. reflective thought or content planning), nearly 70% of the participants across conditions showed evidence of more than a single component. These results strongly suggest that the processes taking place at various pause boundaries are heterogeneous and therefore best represented by a mixture of components, rather than a single Gaussian distribution.

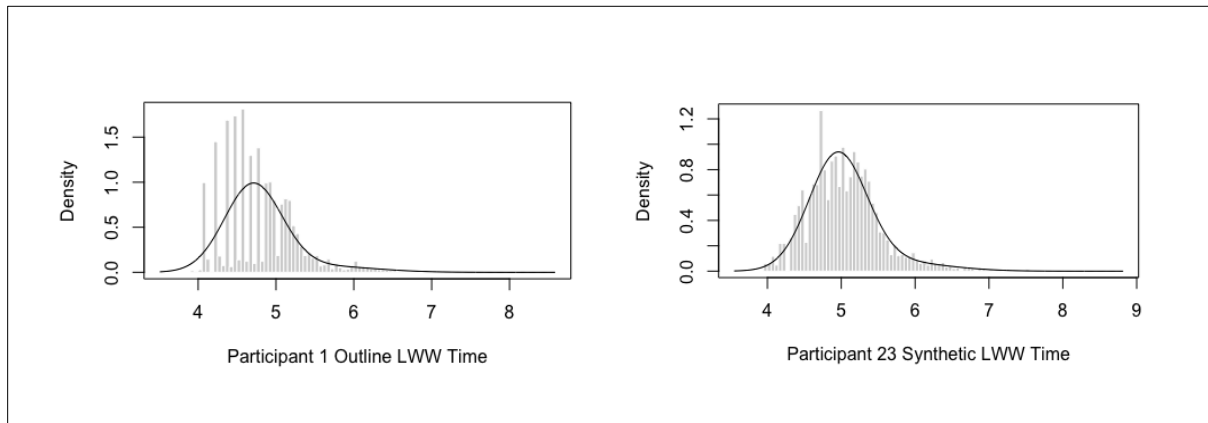
For within-word transitions, the majority of participants (75% across conditions) had a two-component model fit their data best, in line with recent findings (Roeser et al., 2021). For the between-word transitions, a three-component model was identified as the best fitting across conditions in the majority of cases (for 75% of participants in the synthetic condition and 69% in the outline condition). This is also in line with previous findings (Baaijen et al., 2012). For the sub-sentence and between-sentence pauses, a two-component model was identified to be best fitting in the highest number of participants (although, these findings were less well-defined than for the other two pause boundaries). Specifically, the two-component model for sub-sentence pauses was observed in 63% of participants in the synthetic condition and 44% of participants in the outline condition. For between-sentences, the two-component model was observed to be the best fitting in 44% and 50% in the synthetic and outline conditions respectively.

Accordingly, the models identified as most commonly best-fitting in the subset of data collected pre-pandemic were then fitted to the full set of data. The benefit of imposing the common models (three-component distribution for the linear transitions between words and two-component distributions for the transitions at other locations) was that it meant that the follow up analysis could be focused on comparing the properties of the distributions, to see whether the experimental manipulation to drafting strategy led to observable differences in the structure of those models between conditions. Three features were examined: proportion of pauses falling within each of the distributions, the mean of each component and the standard deviation of each component.

**Within-Words.** We focus first on the log-transformed within-word pauses, for which two components were fitted.

**Figure 6**

*Example GMMS for the Log Within-Word Pause Times for Outline and Synthetic Conditions*



**Table 8**

*Summary of Mixture Model Components Within Words for Outline and Synthetic Conditions*

Condition	Component	Mixing proportion (with SD)	Log mean (in milliseconds)	Log SD
Outline	1	0.83 (0.25)	4.88 (131.01)	0.35
	2	0.17 (0.25)	5.87 (355.36)	0.55
Synthetic	1	0.87 (0.20)	4.89 (132.65)	0.36
	2	0.13 (0.20)	5.86 (350.50)	0.59

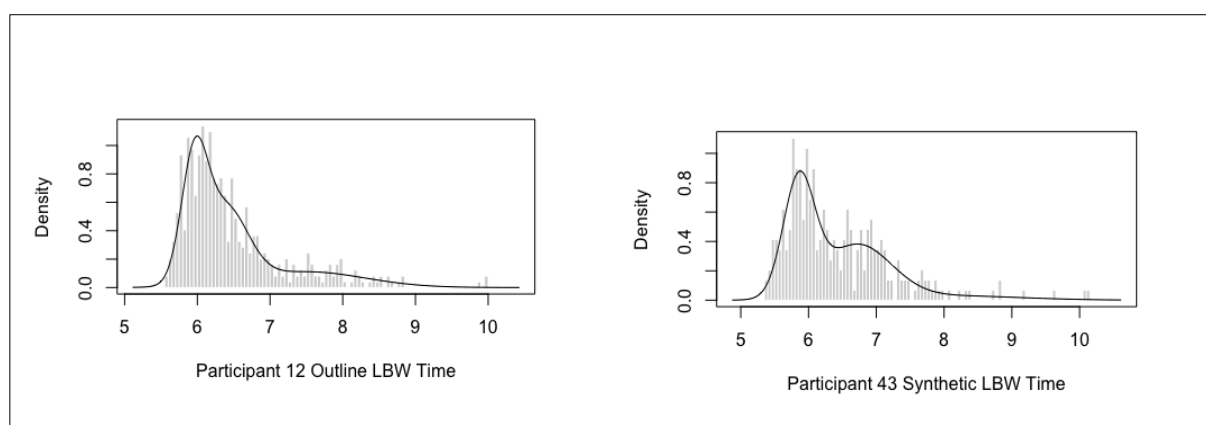
Table 8 outlines the means, standard deviations and mixing proportions for the two components. Figure 6 demonstrates examples of the fitted models in the outline and synthetic conditions. In terms of mixing proportions, the majority of within-word pauses were assigned to the left-hand distribution, regardless of condition (approximately 85%). The mean pause time for component 1 across conditions was around a tenth of a second, indicating that these pauses were representative of non-cognitively demanding, automated processes, most likely related to fluent typing, as theory would imply. For component 2, for which the mixing proportion was much smaller

across conditions (approximately 15%) average pause time was about a third of a second. These pauses are likely representative of disfluencies within typing, as previous evidence has suggested (Roeser et al., 2021; van Waes et al., 2021). Specifically, these pauses are still extremely short in nature, and so probably relate to disfluencies associated with the production of non-frequent bigrams, or bigrams for which letters are far apart on the keyboard. As can be seen, then, the majority of linear within-word pauses represented with the two component GMMs reflect relatively mechanical writing processes. However, those pauses extending towards the right-hand side of component 2 (reaching up to just under half a second in both conditions) could be representative of more cognitively-demanding writing processes, such as those related to word spelling. Indeed follow up T-tests comparing the mean mixing proportions, log means, and standard deviations within the synthetic outline and conditions revealed that there were no significant differences between these characteristics in the two conditions, and effect sizes were small (Cohen's  $d < 0.2$ ). This indicates that typing fluency was extremely similar regardless of the type of advance planning the participants carried out.

**Between-words.** We next move onto linear between-word pauses, for which three components were fitted to the data.

### Figure 7

*Example Mixture Models for the Log Between-Word Pause Times Across the Outline and Synthetic Conditions*



**Table 9**

*Summary of Mixture Model Components Between Words for Outline and Synthetic Conditions*

Condition	Component	Mixing proportion (with SD)	Log mean (in milliseconds)	Log SD
Outline	1	0.42 (0.19)	5.62 (275.31)	0.18
	2	0.41 (0.29)	6.09 (440.62)	0.33
	3	0.17 (0.36)	7.09 (1204.56)	0.80
Synthetic	1	0.46 (0.19)	5.59 (267.86)	0.18
	2	0.38 (0.29)	6.02 (410.63)	0.33
	3	0.17 (0.54)	7.11 (1225.13)	0.81

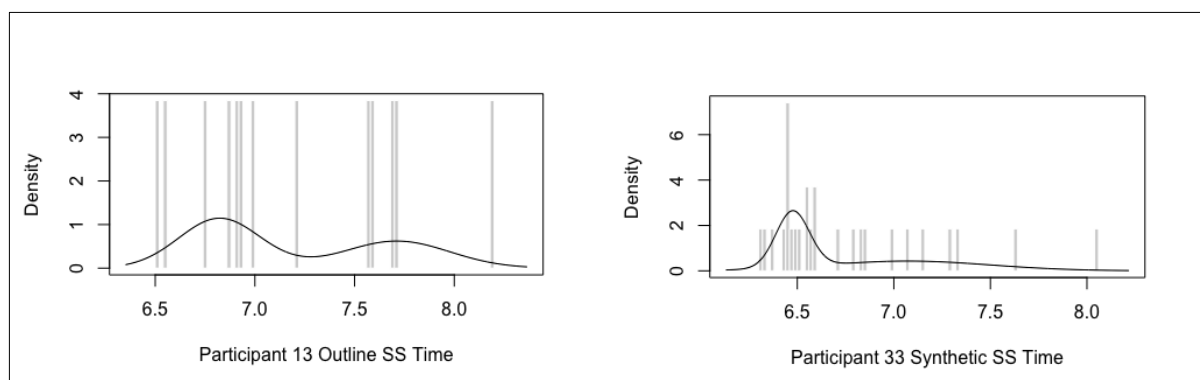
Importantly, the structure of the pauses identified between words here is similar to the structure of the pauses between-words identified in another argumentative essay writing task from Baaijen et al. (2012). They found that the majority of pauses were assigned to the left-hand distribution, followed by the middle distribution and finally the right-hand distribution. The same was observed in this study. As demonstrated in Table 9 and Figure 7, the majority of pauses in this study were assigned to component 1 (42% and 46% for the outline and synthetic conditions respectively, with the mean of this component falling just under a third of a second for both conditions). Hence, these pauses could indeed represent basic word retrieval processes, as suggested by Baaijen et al. (2012), and this is further evidenced by the mean of component 1 reaching just under a third of a second across both conditions. For component 2, approximately 40% of pauses were assigned across both conditions. The mean of this component was just under half a second in both conditions. Given that this is still a relatively short amount of time, the likelihood of these pauses being related to very cognitively demanding writing processes, (e.g. what message the writer wants to convey in the next large section

of a text), is low. However, these pauses could be related to phrase boundary processes. An example of this can be seen with the phrase “The brown dog chased the grey cat”. In composing this phrase, there might be an elevated pause before “chased” because this is the start of a new verb phrase. Hence, the writer needs to plan what to say about the “brown dog” and not just retrieve the next word. Finally, for component 3, 17% of pauses were assigned to this distribution for both conditions. The mean pause duration for component 3 was about 1.2 seconds, and so these pauses could be related to more conscious decisions about word choice (i.e., the writer trying to figure out what word to use to express their point clearly). Pauses that fall towards the end of this distribution, which can reach over 20 seconds in length, are likely to represent higher-level content planning - this was argued by Baaijen et al. (2012), who also found that between-word pauses modelled in the third component could extend beyond 20 seconds in duration. T-tests of the mixing proportions, overall log means, and standard deviations showed that there were no significant differences between how the components were fitted across conditions, and again effect sizes were small. This indicates that experimental manipulation of planning and drafting strategy did not lead to differences in linear pause durations between words for the two groups of writers.

**Sub-Sentences.** We next focus on the linear sub-sentence pauses, to which two distributions were fitted, demonstrated in Table 10 and Figure 8.

### Figure 8

*Example Mixture Models for the Log Sub-sentence Pause Times Across the Outline and Synthetic Conditions*



**Table 10**

*Summary of Mixture Model Components for Sub-Sentences in the Outline and Synthetic Conditions*

Condition	Component	Mixing proportion (with SD)	Log mean (in milliseconds)	Log SD
Outline	1	0.56 (0.17)	6.64 (764.81)	0.19
	2	0.44 (0.17)	7.83 (2510.14)	0.45
Synthetic	1	0.53 (0.21)	6.40 (599.04)	0.18
	2	0.47 (0.21)	7.46 (1730.80)	0.40

The majority of pauses were assigned to the left-hand distributions for both the outline and synthetic conditions (55% in the outline condition and 53% in the synthetic). However, component 2 also contained nearly half of the pauses (44% in the outline condition and 47% in the synthetic). For means, component 1 for the outline condition was 765 milliseconds and for the synthetic condition it was 600 milliseconds. Given that sub-sentence boundaries are indicated by punctuation markers, it is likely that these pauses are related to the planning of higher-level grammatical units. This is further supported by the fact that the mean transition time for component 1 at the sub-sentence level was longer than the mean pause time for component 1 between words, which is consistent with the idea that sub-sentence boundaries would be associated with the planning of larger units than single words (e.g. the content of the succeeding phrase). Component 2 at the sub-sentence level, which has higher mean pause time than component 1, is therefore likely to be representative of more reflective planning, where processing is more problematic and much more variable in duration. A T-test of the means for Component 1 across the two conditions revealed that they were significantly different from one another with a medium effect size ( $t(39) = 2.12$ ,  $p < 0.05$ ,  $d = 0.7$ ), with the synthetic condition having a lower mean than in the outline condition. This is in marked contrast to the boundaries within and between words, where there was virtually no evidence for differences in component means between the conditions. As a result, this finding implies that manipulations to drafting strategy affects higher level content generation processes but not processes relating to the formulation of individual

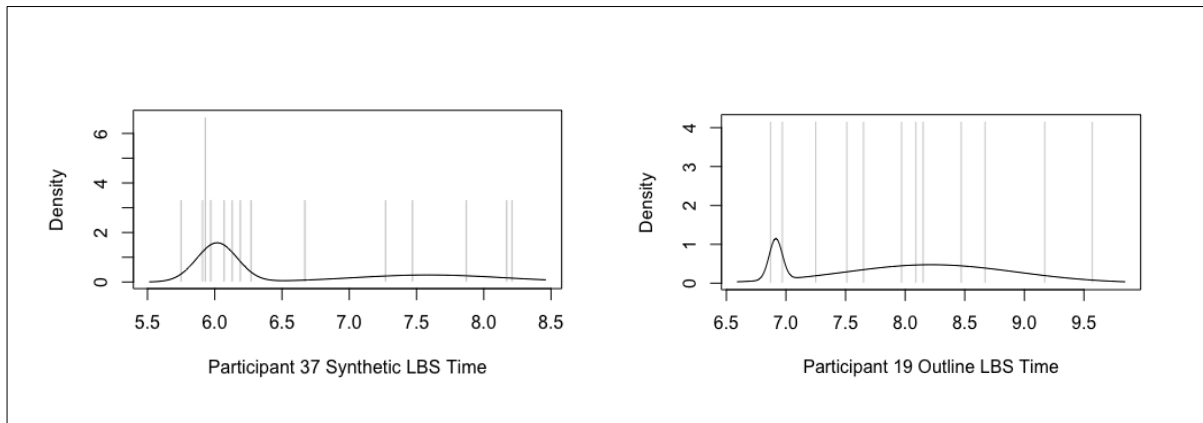
words. These findings also indicate that, in terms of grammatically related pausing, the rough drafting condition were pausing for shorter intervals in comparison to the outline drafting condition. This could be because in the rough drafting condition, participants were forming their essays without worrying about global organisation, whereas in the outline condition they were trying to make sure they produced a well-formed text. As a result, we would expect to see shorter pauses at the sub-sentence level in the rough drafting condition, because they did not need to be concerned about how well presented their final text products were. There were no significant differences between the means for component 2 across conditions.

It is also worth noting that for six participants in the synthetic condition, GMMs could not be fitted to the data due to too few observations of sub-sentence pauses, again suggesting that some of the writers in this condition engaged less with content-generation processes than others (only one participant in the outline condition could not have a two-component model fitted to their data due to too few observations). These results also imply that there might be significantly fewer linear sub-sentence transitions in the synthetic condition than in the outline condition. In order to check this, the percentage of sub-sentence transitions that were linear in the synthetic condition ( $M = 62\%$ ) was compared with the percentage of sub-sentence transitions that were linear in the outline condition ( $M = 65\%$ ) with a T-test. However, no significant differences were found ( $t(46) = 0.30$ ,  $p = 0.76$ ,  $d = 0.09$ ).

**Between-Sentences.** Finally, we focus on linear between-sentence pauses, for which two components were fitted. The results are demonstrated in Table 11 and Figure 9.

**Figure 9**

*Example Mixture Models for the Log Between-Sentence Pause Times Across the Outline and Synthetic Conditions*

**Table 11**

*Summary of Mixture Model Components for Between-Sentences in the Outline and Synthetic Conditions*

Condition	Component	Mixing proportion (with SD)	Log mean (in milliseconds)	Log SD
Outline	1	0.50 (0.20)	7.06 (1162.29)	0.30
	2	0.50 (0.20)	8.45 (4656.74)	0.48
Synthetic	1	0.56 (0.19)	6.92 (1009.08)	0.29
	2	0.44 (0.19)	8.26 (3871.56)	0.44

In terms of mixing proportions, for the outline condition, pauses were roughly equally distributed to components 1 and 2. However, in the synthetic condition, slightly more pauses were assigned to component 1 (56%) than component 2 (43%). In terms of the duration of between-sentence pauses in the outline and synthetic conditions, similar means were observed for component 1 (around a second). The fact that these pauses were longer than those observed in component 1 at all previous intervals implies that these transitions are reflective of larger units of content being planned, rather than processes related to grammatical boundaries and word retrieval. In terms of what the

pauses in the components 1 and 2 may represent, it is likely that component 1 reflects normal linear between-sentence transitions where the writer proceeds fairly rapidly to the next sentence. Component 2, however, likely represents more reflective, problem solving transitions, where the writer spent an extended amount of time working out the next piece of content.

T-tests of the mixing proportions, means and standard deviations for the components showed no significant differences in structure of the models between conditions, but it is important to note that in the synthetic condition, seven participants could not have models fitted to their data due to too few between-sentence pause observations, again implying a difference in the proportion of linear transitions between the two conditions. Again, in order to check this, the percentage of sub-sentence transitions that were linear in the synthetic condition ( $M = 41\%$ ) was compared with the percentage of sub-sentence transitions that were linear in the outline condition ( $M = 49\%$ ) with a T-test. However, no significant differences were found ( $t(46) = 1.38$ ,  $p = 0.17$ ,  $d = 0.39$ ).

To summarise then, the findings from the mixture modelling proportion of this study indicate three important things. Firstly, linear pauses are better model with a mixture of components than single Gaussian distributions. Secondly, as demonstrated in previous research (Baaijen et al., 2012), durations of pauses increase as the pause interval goes up (e.g. pauses between sentences are longer than pauses between words). This indicates that the higher the pause interval, the more complex the underlying writing processes. Finally, the modelling showed that pause structure was much the same for participants across the conditions at the within-word and between-word level, but structure began to vary at the sub-sentence and sentence level pause interval, implying that experimental manipulation to drafting strategy might impact higher-level content generation processes, rather than basic word-related processes.

### ***Reproduction of The Baaijen and Galbraith (2018) Global Linearity and Sentence***

#### ***Production Measures***

In the original Baaijen and Galbraith (2018) study, the authors used a principal-components analysis (PCA) with varimax rotation to define their two global linearity and sentence production

measures. Hence, the same methods were used in the present study. The analysis was conducted in R Version 4.0.3 with the Psych Version 2.1.9 package. Reproducible scripts for this analysis can be found in the project's Open Science Framework repository (link on p.12).

First, sampling adequacy of the 11 variables across the full dataset was tested (both outline and synthetic conditions combined) and produced a Kaiser-Meyer-Olkin test of .69, indicating that the sample size was adequate for PCA. Next, the sphericity of the data was checked with Bartlett's Test for Sphericity. This was highly significant ( $\chi^2(55) = 293.24$ ,  $p < .001$ ), indicating collinearity between the variables, and hence PCA was an appropriate method to use.

Table 12 shows the results of the PCA analysis and, for comparison, the Baaijen and Galbraith (2018) study results. Loadings of 0.3 or higher are in bold as they are interpreted as a meaningful correlation between the variable and the component (Tabachnick & Fidell, 2012).

**Table 12**

*Summary of Principal Component Analysis with Varimax Rotation for 2 Factor Solution in the Current Study and Baaijen and Galbraith (2018) Study*

Variables	Factor Loadings for principal components analysis (current study)		Factor Loadings for principal components analysis (Baaijen & Galbraith, 2018)	
	Component 1 Global Revision	Component 2 Sentence Production	Component 1 Global Linearity	Component 2 Sentence Production
1 Sentence linearity index	<b>-0.36</b>	-0.24	<b>0.849</b>	0.2
2 Percentage of I-bursts	<b>0.72</b>	0.2	<b>-0.83</b>	0.087
3 Percentage of time spent on events	<b>0.87</b>	-0.04	<b>-0.824</b>	-0.009
4 Percentage of linear transitions between sentences	<b>-0.71</b>	<b>0.45</b>	<b>0.773</b>	0.075

5 Number of production cycles	<b>0.58</b>	0.11	<b>-0.709</b>	-0.27
6 Percentage of linear transitions between words	<b>-0.74</b>	<b>0.43</b>	<b>0.601</b>	<b>0.363</b>
7 Percentage of bursts terminated by revision at the leading edge	-0.13	<b>-0.94</b>	0.113	<b>-0.901</b>
8 Percentage words produced in P-bursts	-0.19	<b>0.85</b>	<b>0.362</b>	<b>0.809</b>
9 Percentage of >2 second pauses between words	0.21	<b>0.79</b>	-0.072	<b>0.741</b>
10 Text modification index	<b>0.47</b>	<b>-0.51</b>	<b>-0.399</b>	<b>-0.675</b>
11 Mean pause duration between sentences	0	<b>0.66</b>	0.181	<b>0.636</b>
Eigenvalues	3.42	3.11	4.7	2.4
% Variance	31	28	42	21
$\alpha$	0.77	0.8	0.8	0.79

---

The first thing to note is the direction of the component loadings. Due to a trivial variation in how global linearity in the Baaijen and Galbraith (2018) study was scored in relation to the equivalent component in the current study, these scales need to be interpreted inversely. That is, in the original study, participants who scored highly on the global linearity measure had more linear texts. In contrast, participants who scored highly on the equivalent component in the current study have less linear texts. Thus, the global linearity component has been renamed ‘global revision’ in this study to make this difference more easily interpretable. When considering this, the directions of the loadings indicate that 11 items fall onto the two components in the same way across both studies.

Next, examining the strengths of the loadings, we can see the loadings from the original study and those from the current study are mostly similar. In particular, sentence linearity, percentage of I-bursts, percentage of time spent on events, percentage of linear transitions between sentences, number of production cycles and percentage of linear transitions between words all load more firmly onto the global revision component than the sentence production component. In contrast, percentage of bursts terminated with a revision at the leading edge, percentage of words produced in P-bursts, percentage of pauses between words that last longer than two seconds, text modification and mean pause duration between sentences all load onto the sentence production component more strongly than the global revision component.

Notably, though, sentence linearity produces a weak loading on the global revision component in this investigation whilst producing a very strong loading in the original study. After discussing why this may be with the authors of the original research, we found that we had used different methods to calculate the sentence linearity index. Specifically, there was a variation in how the two studies calculated the number of sentences produced out of order. In the current study, a sentence was deemed to be produced out of order when an individual moved away from the leading edge to insert a new complete sentence. The number of sentences out of order did not increase if the person moved away from the leading edge to add text or edit an already existing sentence. On the other hand, Baaijen and Galbraith (2018) *did* increase their ‘sentences out of order’ count when the participant moved away from the leading edge to add text to or edit an already existing sentence. Hence, this difference in calculation methods likely led to the discrepancies within the component loadings of the two models. The implications of this will be described in the discussion.

Collectively, the two components in the current study explained 59% of the variance of the 11 variables, whereas in the original study, 63% of variance was explained. This indicates that both models explained similar amounts of variance across the variables. Combining variables 1 to 6 from Table 11 produced a reliable scale ( $\alpha = 0.77$ ; note also that in dropping the sentence-linearity index variable, the reliability of this scale increases to  $\alpha = 0.80$ ). High scores on this scale represent non-linearly produced text with high levels of global revision, and low scores represent linearly produced text with low levels of global revision. Combining variables 7 to 11 in Table 12 produced a highly

reliable scale ( $\alpha = 0.82$ ). On this scale, writers who score highly pause for longer durations between sentences and use less within-sentence revision - their sentence production is more controlled, whereas writers who score low on this scale pause for shorter durations between sentences and use more within sentence revision - their sentence production is more spontaneous.

The results from the PCA analysis, therefore, indicate that the two composite measures of global linearity/global revision and sentence production from the Baaijen and Galbraith (2018) study can be reproduced, albeit with ten items, rather than eleven, due to different methods of calculation between the studies for the sentence linearity index.

## **Discussion**

There were three general aims for this chapter. The first was to establish a reproducible framework for the analysis of pauses, bursts and revisions from keystroke data. The second was to test whether experimental manipulation of drafting strategy would result in differences in the pause structures observed in keystroke logs in order to improve the accuracy of relating pauses to underlying processes. The third was to examine the reproducibility of the two-component composite measures of global revision and sentence production first identified by Baaijen and Galbraith (2018). This discussion section explains the implications of the methods applied and findings within the study.

### ***Evaluating the Reproducible Framework for Analysing Pauses, Bursts and Revisions***

In regards to establishing a reproducible framework for the analysis of pauses, bursts and revisions, the first task was to create a set of definitions for the various features that needed to be extracted from the logs. Identification and calculation definitions were made, which were designed to demonstrate exactly how one would extract a pause, burst, or revision from a keystroke log. Whilst it was possible to automate the analysis of the pauses (accompanied with thorough manual checking), bursts and revisions were less easy to automate because of the high levels of variation in how they could be produced. As such, it was determined that it was better to be completely transparent about the processes used to extract bursts and revisions, which were based general guidelines rather than

rigid identification rules. Full transparency was achieved by sharing the final coded data (which can be seen in the project's Open Science Framework repository, link on p.12).

This is important because it highlights that whilst automated analysis is beneficial for making analysis reproducible (due to its standardised nature), it is not *necessary* in order to make the analysis reproducible. Simply being transparent about how data is analysed is adequate for other researchers to be able to observe, critique and, if necessary, replicate the findings. This also applies to the process of identifying different types of text production (e.g. titles and post-draft revisions) because high variation between participants in how these elements are produced makes it challenging to automate their extraction from a keystroke log. Hence, transparent reporting of the manual categorisations of different parts of the text was necessary. In short, clear, transparent criteria and instructions to ensure reliable subjective judgments over how to categorise the features of a keystroke log are, in many ways, more necessary than complete *standardisation* of the extracting methods.

### ***Using GMMs to Map Pauses onto Processes in Experimentally Manipulated Writing Tasks***

This analysis aimed to see whether experimental manipulation to the type of drafting strategy that students used in essay writing task would narrow down the types of processes that pauses could be mapped onto. The writing conditions used within this study were designed to elicit different levels of engagement with certain writing processes. It was expected that the writers within the outline condition would create more planned and controlled texts, whereas the writers in the synthetic condition would be less planned and more spontaneous in their text production and that this might lead to differences in the structure of the linear pause components observed in the mixture models. The main finding was that there appeared to be no significant difference in the structure of the GMMs at the within- and between-word level across the outline planning/final drafting strategy and the synthetic planning/rough drafting strategy, implying that the experimental manipulation did not impact relatively automated word-level processes. By contrast, there was some evidence to suggest that the structure of pauses at the sub-sentence and between-sentence level were affected by drafting strategy manipulation. This section discusses those findings in more detail.

Focusing first on linear within-word pauses, the 2-component GMMs showed that across both conditions, the left-hand distributions picked up the majority of observations, which were extremely short in nature, and the right-hand distributions picked up fewer, longer pauses. This reiterates the notion that within-word pauses are likely to reflect very automated, mechanical writing processes, with the left-hand distribution representing fluent typing and the right-hand distribution representing typing disfluencies (e.g. for spelling related disruptions, or the production of non-frequent bigrams). The fact that there were no significant differences in the mixing proportions, component means and component standard deviations between conditions in the within-word level models is perhaps not that surprising given that previous research has indicated that within-word pauses can be modelled with two components across *various* types of writing tasks, including copy tasks and essays (Baaijen et al., 2012; Roeser et al., 2021). Hence, it is plausible that the processes captured within-words are so low-level that experimental manipulation to planning and drafting strategy does not impact pause durations of this nature.

For between-word pauses, again the characteristics of the 3-component models fell in line with how this type of pause data had been previously modelled. There were no significant differences in how the 3-component models were fitted across conditions, suggesting again that experimental manipulation to planning and drafting strategy did not lead to differences in the ways in which outline and synthetic condition writers engaged in between-word writing processes. To reiterate, this is not that surprising considering that between-word pauses are predominantly theorised to represent processes related to word-retrieval, rather than higher-level content planning, and so engagement in word-level processes would likely be similar across the groups.

Where potential differences in pause structure did begin to arise was at the sub-sentence and between-sentence transitions. The first thing to note for sub-sentence pauses was that there was a significant difference in means for component 1 across conditions, indicating that writers in the outline condition were being more controlled and deliberate in the structural planning of their sentences.

Interestingly, there were six participants in the synthetic condition whose data could not be represented with a two-component GMM due to too few observations of linear sub-sentence

transitions. This finding provides additional evidence for outline condition writers being more engaged in sub-sentence writing processes than synthetic condition writers. One potential reason for this finding could be that text structure was regarded as an important writing outcome of the outline/final drafting condition due to the nature of the instructions they were given (i.e. they had to create an *organised* outline plan and produce a *final* version of an essay). Hence, it is reasonable to assume that these writers would spend more time on polishing the structure of their sentences, which may have resulted in longer pauses at the sub-sentence level. In contrast, because writers in the synthetic condition were explicitly told not to be concerned about text structure, it makes sense that several participants in this condition could not have GMMs fitted to their data due to too few linear sub-sentence observations. Specifically, it could be that because the writers in the synthetic condition were prompted to write down their thoughts as they occurred, and to not worry about the presentation of their texts, they were less concerned about sentence-level presentation (e.g. using the correct type of clause) than the writers in the outline condition.

These findings were also somewhat echoed at the between-sentence pause level, although to a lesser extent. Specifically, seven participants in the synthetic condition did not have GMMs fitted to their data due to too few linear between-sentence pause observations. This was compared to only two participants in the outline condition. Again, the reason for this could be due to the fact that the synthetic condition was told to write their thoughts down as they occurred, which could have resulted in some of the synthetic participants writing their texts as if they were spontaneous streams of consciousness, rather than well-formed essays made of complete sentences.

At this point in the discussion, though, it is salient to emphasise the exploratory nature of this GMM analysis. Whilst there is some evidence to suggest that writers in the outline and synthetic conditions engaged to different extents with sub-sentence and sentence-level processes, these findings should be considered only preliminary for several reasons. Firstly, the sample size for this study was small ( $N=48$ ), and so any variation between the groups might not be upheld if the participant pool was larger. Secondly, the writing tasks were only 30 minutes in length, which meant that even for the participants who had enough linear sub-sentence and sentence pauses for 2-component GMMs to be fitted to their data, there were still only a few observations of these pauses per participant. This means

that the accuracy of these models is likely to be low. Therefore, mixture modelling may still be a fruitful technique for analysing sub-sentence and sentence level processes, particularly in combination with experimentally manipulated drafting strategies, but it would be more accurate if there were a greater number of these pause observations.

### ***Evaluating the Reproduction of the Baaijen and Galbraith (2018) Global Linearity and Sentence Production Measures***

After creating the methods for identifying pauses, bursts and revisions, the extracted measures were used to recreate the global revision and sentence production composite measures first identified by Baaijen and Galbraith (2018). Using the same PCA methods as in the original study, the replication was largely successful, with the relevant items loading onto the global revision and sentence production measures in the replication as in the original analysis. Because this study used a different dataset from the original study, the results imply that the global revision and sentence production measures are reproducible with other writers. However, it is important to consider that the writing tasks used, and participant demographics, were similar across both studies. Consequently, it would be worthwhile trying to reproduce the measures again within a different group of writers (e.g. sixth-form students) and another type of writing task (e.g. subject-based essays). This would help investigate the generalisability of the measures to other writing contexts.

The one item that did not reproduce well was the sentence linearity index. In this study, it loaded weakly onto the global revision component. In the original, it loaded strongly onto it. When discussing with the authors of the original paper why this discrepancy may have occurred, we quickly determined that different methods had been used to create the item in each study. This highlights the importance of being completely transparent in describing the methods used in creating writing measures. Ambiguity in the description of the sentence linearity index within the original study (i.e. the omission of information about including partially produced sentences and fully completed sentences) meant that including partially produced sentences when creating the measure was not factored into the analysis in the current study.

In general, though, the global revision and sentence production measures were reproduced, and the reliability of the components in the reproduced scales was high. In future, research should continue to look at how these measures might relate to various writing outcomes, such as text quality and subjective knowledge development.

To conclude, the aims of this chapter were straightforward: to create a reproducible framework for the analysis of pauses, bursts and revisions; to test whether experimental manipulation of writing might help in mapping pause data onto specific writing processes, and to replicate the Baaijen and Galbraith (2018) global revision and sentence production measures. However, the outcomes of the study were far from straightforward. In particular, it is essential to emphasise that it might be impossible to develop completely standardised methods for identifying pauses, bursts and revisions. So instead, it is necessary to focus on reporting procedures transparently, to enable other researchers to observe and reproduce exactly how analyses are conducted.

Additionally, the GMM findings from this study and previous studies strongly demonstrate that pause data should be modelled as a mixture of distributions, rather than a single, aggregated measure. In terms of aligning pauses with processes, the results from this study preliminarily suggest that using experimentally manipulated drafting strategies in combination with mixture modelling could be a valuable technique for better defining how writing processes relate to pause data. However, for essay writing tasks such as the one in this study, differences in pause structures may be more likely observed at the sub-sentence and sentence level, rather than within- and between-words. As such, 30 minutes might not be a long enough time in order to collect enough pause data for these models to be accurate, so future studies should take this into account.

Finally, the Baaijen and Galbraith (2018) composite measures were reliably reproduced in a new sample of participants, albeit with ten rather than eleven items, due to some ambiguity of reporting around the methods used to create the sentence linearity index in the original study. These results imply that the composite measures are suitable for usage in future studies looking at writing processes through keystrokes.

## 5 Study 3: Exploring the Impact of Writing Beliefs, Writing Processes and Drafting Strategy on Subjective Understanding Development

### Introduction

This chapter starts with a discussion about how subjective understanding development is regarded as an integral outcome of writing under the dual process model (Galbraith, 1999; Galbraith, 2009; Galbraith & Baaijen, 2018). It briefly recaps some of the most important research presented in Chapter 1, the Baaijen et al. (2014) and Baaijen and Galbraith (2018) studies, which are the only two pieces of research that have explicitly examined how drafting strategy, writing beliefs, and writing processes may collectively be implicated in the development of subjective understanding. Primarily, the background section of this chapter focuses on the limitations of these two studies, namely that they use a somewhat limited tool to examine writing beliefs, and that the experimental manipulation to drafting strategy only differentiated the type of planning that the writers used.

In response to those issues, Study 2 is presented, which sought to add to the original investigations by Baaijen et al. (2014) and Baaijen and Galbraith (2018) by also examining how different drafting strategies, writing processes and writing beliefs influenced the amount of subjective understanding that university students developed over the course of essay writing. There were a few important differences between this study and the originals. Firstly, in the current study, the WBI (Galbraith and Baaijen, in preparation; examined in Study 1) was used to investigate the impact of writing beliefs on subjective understanding. Secondly, the experimental manipulation to writing task was extended so that not only did the two conditions use different planning strategies, but participants were also required to *draft* their texts using different techniques. In combination, these changes resulted in two different conditions: synthetic planning with rough drafting and outline planning with final drafting. Finally, this study was novel because it used a multi-item, rather than single-item scale to examine two components of subjective change in understanding: *organisation* and *how much you feel you know*. This study also used the Baaijen and Galbraith (2018) composite measures of global

revision and sentence production (reproduced in Study 2) to examine the relationships between writing processes and change in subjective understanding.

The chapter concludes with an examination of the study's results, which indicate that sentence production processes are related to both components of the subjective understanding scale, suggesting writers who engage more spontaneously with sentence production experience greater increases in subjective understanding during writing. It discusses how these findings support the assumptions of the dual process model, but highlights that further evidence is needed about the role of drafting strategies and writing beliefs on the development of subjective understanding.

## **Background**

### ***Important Outcomes of Writing: Text Quality and Subjective Understanding Development***

Traditionally, research has focused predominantly on text quality as an outcome of writing. This is perhaps because one of the most heavily cited theories of writing within the field, the Flower & Hayes (1981) cognitive model of writing, essentially states that writing is a problem-solving process. The writer's aims are to meet the rhetorical goals of the task at hand which, if writing for an audience, usually involves expressing their thoughts in a clear communicative way.

Perhaps also though, much writing research has focused on quality as an outcome because in many real-world scenarios, a high standard of writing quality is imperative. Take academic writing, for example. In order to obtain good marks on an essay, or to get a research paper accepted within a journal, the writer needs to be able to clearly communicate their ideas in a convincing way to the reader.

Moreover, research has indicated that certain writers are better at producing good quality texts than other groups. For example, university students with dyslexia have been observed to consistently produce lower quality texts than students without dyslexia in writing experiments (Connelly et al., 2006; Galbraith et al., 2012; Sumner & Connelly, 2020; Wengelin, 2007). In other research, writers who display certain preconceived beliefs about writing are seen to produce better quality texts than writers who do not. For example, writers with high transmissional beliefs tend to produce poorer

quality pieces of writing than writers with low transactional and high transactional beliefs (Baaijen et al., 2014; Sanders-Reio et al., 2014; White & Bruning, 2005). Studies have also demonstrated that the way in which writers engage with the writing process can be implicated in the quality of texts they produce (Baaijen & Galbraith, 2018). Hence, systematic differences in groups of writers have been seen to lead to discrepancies in writing quality. As a result, researchers may gravitate towards investigating how to reduce such disparities so that, optimistically, writing can become more of an equal playing field for all.

Whilst text quality is an important outcome of writing, so is the development of new ideas, or ‘discovery’, and subsequently a change in the writer’s subjective understanding. This is apparent in the most influential models of writing. The problem-solving accounts of writing (Bereiter & Scardamalia, 1987; Flower & Hayes, 1981; Hayes, 2012), for example, deem subjective understanding a by-product of the writing process. That is, in learning to write more expertly, through controlled problem-solving processes, individuals incidentally also develop their understanding of the topics they are writing about. The dual-process view of writing (Galbraith, 1999; Galbraith, 2009; Galbraith and Baaijen, 2018) implies that the development of subjective understanding involves the careful balancing of spontaneous text production (knowledge constituting) and re-organisation of content to meet rhetorical goals (knowledge-transforming). Yet, subjective understanding as a whole is less well-researched than text quality.

### ***Summary of the Baaijen et al. (2014) and Baaijen and Galbraith (2018) Results***

Chapter 1 discussed much of the research that has examined how subjective understanding develops throughout the course of writing, and how varying approaches to drafting strategy, individual differences in writing beliefs and writing processes may have an impact. As such, it is assumed that the reader is now familiar with this evidence. Consequently, this section focuses on recapping the results from Baaijen et al. (2014) and Baaijen and Galbraith (2018), because these are the only two published pieces of research that have *collectively* looked at the role of writing beliefs, writing processes and drafting strategies on subjective understanding development.

One component of writing that Baaijen & Galbraith (2018) examined, which had not previously been investigated in studies looking at the development of subjective understanding, was the role that writing processes play. The results from their investigation indicated that writing processes were indeed related to subjective understanding. Specifically, global revision was significantly associated with increases in understanding. That is, writers who revised the global structure of their texts had higher increases in subjective understanding than writers who did not revise their texts. This finding is in line with expectations of the knowledge-transforming system within the dual process model, which links subjective understanding development to a reorganisation of content to meet rhetorical goals, rather than a simple translation of thoughts into text.

There was also a significant two-way interaction between planning condition and sentence production processes. That is, the relationship between sentence production processes and subjective understanding development varied based on the type of planning the writer engaged with. Writers in the synthetic condition developed their understanding more when they used spontaneous, rather than controlled sentence production processes. These findings support the assumptions of the dual process model because they suggest when synthetic planning is used, and little has been pre-planned in terms of how a text should meet rhetorical goals, sentence production is instead shaped by the implicit organisation of thoughts in semantic memory (a knowledge-constituting process). More spontaneous sentence production processes (i.e. more engagement with the knowledge constituting system) leads to greater increases in subjective understanding. By contrast, for writers in the outline condition, increases in understanding were generally low, and spontaneous text production was, at the extreme, associated with decreases in understanding.

The most important point to take away from Baaijen and Galbraith's (2018) findings was that global revision and sentence production processes made independent contributions towards subjective understanding development. This implies that the knowledge transforming and knowledge constituting components of the dual process model have independent effects on the development of subjective understanding. To date, however, this is still the only study that has explicitly examined the effect of writing processes involved in the dual process model on subjective understanding.

Unlike previous research on the development of subjective understanding, Baaijen and Galbraith (2014) examined how participants' beliefs about writing were implicated in the relationship between different planning conditions and text quality and the development of understanding. They found that participants' writing beliefs moderated the relationship between planning strategy and understanding development. For example, writers with high-transactional beliefs (i.e. writers who cognitively and emotionally engaged in the writing process and believed that the main purpose of writing was to develop their understanding about the topic) developed their subjective knowledge about the writing topic more than those with low transactional beliefs, regardless of the type of planning condition they were assigned to. On the other hand, writers who had high transactional *and* high transmissional beliefs were better at developing their subjective understanding of the writing topic when they were in the synthetic condition (to such an extent that they had bigger increases in subjective knowledge than any other group of writers). Contrastingly, when these writers were in the outline planning condition, they experienced the smallest increase in subjective understanding development compared to other writers.

These findings support the assumptions of the dual process model. The fact that writers with both high transactional and transmissional beliefs experienced the most development in subjective understanding when they synthetically planned their texts, but the least amount when they outline planned, implies that the process of creating an outline plan stops the knowledge-constituting process from happening. Specifically, because these writers had created a plan in advance of writing which aligned with rhetorical goals (an important consideration for high transmissional writers) they were unable to engage with the spontaneous, unplanned text production that forms the knowledge constitution process. On the other hand, when these writers synthetically planned, they were not defining rhetorical goals in an advance plan, and so their text production could be more spontaneous, allowing the writers to develop their understanding considerably (an important consideration for high transactional writers).

### ***Limitations of the Evidence about Subjective Understanding Development over the Course of Writing***

Whilst the findings from Galbraith and colleagues provide evidence about how writing processes, writing beliefs and drafting strategies are involved in the development of subjective understanding under a dual process perspective, there are three main limitations within the studies.

First, although the synthetic and outline conditions promoted different types of pre-planning for participants, all participants were required to write high-quality final products, so a rhetorical goal of the writing task would have been effective communication, regardless of condition. Although this is useful for investigating how writers may balance the knowledge transforming and constituting processes, this specific manipulation may have limited the full extent to which participants could have engaged with knowledge constituting processes. The outline condition might have been expected to promote knowledge-transforming but was likely to reduce knowledge-constituting, due to the instructions for the condition being very much centred around organising content to meet rhetorical goals, rather than producing text spontaneously.

By contrast, the synthetic planning strategy in the other condition should have promoted knowledge constitution. However, because these writers were also required to produce a well-formed final product that met rhetorical goals, text production could not have been completely spontaneous, so maximal engagement in the knowledge constitution system would not have been achieved.

If, instead, a synthetic planning strategy was used in combination with a rough drafting strategy, in which participants did not need to worry about how well-formed their final text product was, this should have maximised the development of subjective understanding, but not knowledge transforming, because there would be no requirement to meet rhetorical goals. This was actually tested in Galbraith et al. (2006; see chapter 1, p.49). The synthetic condition was required to produce a rough draft of an essay, in which they had to write down their thoughts as they occurred, but not worry about how well organised or expressed the final texts were. The writers in the outline condition were told to write a well-structured essay, in which they communicated their ideas as clearly as they could to the reader, but without worrying too much about the mechanical features of text composition

(note, though, that this study did not examine how writing beliefs and writing processes may have also been involved in the relationship with subjective understanding).

Second, the Baaijen et al. (2014) and Baaijen and Galbraith (2018) studies relied on single-item measures to examine development of subjective understanding, which may not have been sensitive enough to pick up the complexities of understanding. The dual process model implies that the development of subjective understanding can involve both knowledge transforming and knowledge constituting processes. To the extent that knowledge transforming involves reorganising thought, one might expect it to lead to changes of subjective organisation. However, the item used in these two studies to measure subjective understanding (*How much do you feel you know about the topic?* A 7-point scale, where 1 = very little and 7 = a great deal) did not examine subjective organisation, only subjective knowledge.

This demonstrates the need for an understanding scale potentially capable of differentiating between changes in subjective organisation and knowledge. In fact, Galbraith et al. (2023) created a multi-item scale designed to capture these specific components – *organisation* and *how much you feel you know*. Under the assumptions of the dual process model, the *organisation* component maps onto the knowledge transforming system, as it captures how the writer's organisation of ideas changes over the course of writing. The *how much you feel you know* component maps onto the knowledge constituting system because it captures how the writer has developed their ideas and understanding around a topic over the course of writing.

Third, Baaijen and Galbraith's (2014) research, which investigated the role of writing beliefs in subjective understanding development, utilised the White and Bruning (2005) measure of writing beliefs. This questionnaire only captures transmissional and transactional beliefs (and recursive process beliefs under the assumption that these are represented by transactional items). However, as demonstrated by the work from Sanders-Reio et al. (2014) and the evidence presented in Study 1, it is also important to consider how audience and planning beliefs are implicated in the relationship between writing beliefs and subjective understanding. Audience beliefs may be important because, if an individual prioritises the needs of their audience within their writing, this rhetorical goal may limit the extent to which they engage in the knowledge-constituting system. Additionally, beliefs about

planning might have an impact on how drafting strategy affects subjective understanding (e.g. do high planning beliefs eliminate the effect that synthetic planning with rough drafting could have on maximising subjective understanding change, because these writers use a high level of planning throughout their writing anyway, limiting the extent to which they can spontaneously develop their ideas?). Yet, the White and Bruning (2005) measure does not examine planning or audience beliefs. As such, future research should consider using a tool that examines a more comprehensive breadth of writing beliefs.

Quite apart from the important methodological limitations described above, the Baaijen and Galbraith (2018) study is the only study that has examined the role of writing processes in the development of subjective understanding. Hence, there is an important need to determine whether those findings can be reproduced.

### ***Aims of the Current Study***

On the basis of the issues presented above, the current study used a design similar to Baaijen and Galbraith (2018) and Baaijen et al. (2014), but with systematic modifications to address the limitations of those studies. The overall aim with this research was to provide further evidence for the role of writing beliefs, writing processes and drafting strategies on subjective understanding change over the course of writing, and whether this evidence provides further support for the dual process model.

First, the manipulation of planning strategy was extended into a fuller drafting strategy. Synthetic planning was paired with rough drafting, a technique that relaxes the general constraints associated with normal writing and should result in maximum engagement with knowledge constitution. In the outline planning condition (where participants are required to write an organised and hierarchical plan before essay writing), participants were instructed to write a final version of an essay. Hence, under the assumptions of the dual process model, these writers should experience the smallest engagement with knowledge constitution, because they are constrained by more rhetorical task goals.

Second, to capture how writers' engagement in the knowledge-constituting and knowledge-transforming systems is linked to understanding, the current study also uses the newly developed Subjective Understanding Scale (Galbraith et al., 2023) splitting subjective understanding into two components – *organisation* (mapping onto the knowledge transforming system) and *how much you feel you know* (mapping onto the knowledge constituting system).

Third, the Galbraith and Baaijen (in preparation) Writing Beliefs Inventory, which was validated in Study 1, was used in this study. The questionnaire examines planning beliefs as well as transactional, transmissional, revision and audience beliefs and has been chosen for this study because it should give a broader insight into how writing beliefs affect subjective understanding.

Based on the aims outlined above, the research question for this study was:

Do drafting conditions, writing beliefs and writing processes affect the development of subjective understanding during essay writing?

## Methods

The data for this study were collected from an experiment involving UK university students ( $N = 61$ ). The participants were assigned to either an outline or synthetic drafting condition. They were told to write an essay about one of the following questions, which were counterbalanced across the conditions:

1. Does social media do more harm than good?
2. Should we all become vegans?

As in Baaijen et al. (2014) and Baaijen and Galbraith (2018), participants in the outline condition were given five minutes before writing to create an organised outline to inform their essays. They were told to work out an outline that indicated their opinion, their main ideas, and the order they would go in. After completing the outline, they were required to write a 'complete draft' of an essay, consisting of a well-organised text expressed in continuous prose with accurate spelling. The critical writing goal for these participants was a good quality, clearly communicative text product.

In the synthetic condition, the participants were given five minutes to work out what they thought about the question and to write a single sentence that summed up their opinion and main ideas. In contrast to the original studies, these participants were then instructed to write a rough draft of an essay. They were told to discuss the essay question with themselves, writing down their thoughts as they occurred and then forming a conclusion. Furthermore, they were told not to concentrate on the organisation of the essay but just to focus on directly expressing their thoughts as they unfolded, expressed in continuous prose but without worrying about accurate spelling. Participants composed their texts on laptops with the Inputlog Version 8 software (Leijten & Van Waes, 2012) recording keystrokes throughout the duration of the writing task. The participants were given 30 minutes to write their essays.

Before essay writing, all participants completed the Baaijen and Galbraith (in preparation) WBI. They were also asked to write a list of ideas about their topic and then order the ideas based on how important they thought each one was. This task aimed to create an explicit set of ideas that the writers were likely to refer to during their essays. Participants also completed the 12-item version of the Galbraith et al. (2023) Subjective Understanding Scale to assess their subjective understanding of the topic they were writing about. After finishing the essays, participants immediately filled in the subjective understanding scale again, referring to how much they understood about the writing topic *after* writing their essays. The difference in pre- and post-understanding scale scores was used to assess subjective understanding development over the course of writing.

A detailed breakdown of the tools and procedures in this study is in Chapter 2 (pp.81-84).

### ***Background of the Galbraith et al. (2023) Subjective Understanding Scale***

The Subjective Understanding Scale (Galbraith et al., 2023) originally consisted of 12 items asking questions about how much participants felt they knew about a writing topic, and how organised their thoughts about the writing topic were. Galbraith et al. (2023) conducted preliminary testing of this tool in a survey study, in which 165 university students were asked to imagine that they had been instructed to write an essay about one of six current affairs topics, which they were

randomly allocated to. The participants then completed the 12 items of the scale in reference to their current subjective understanding about their assigned topic.

A principal components analysis was then applied to examine the structure of the scale. Galbraith et al.'s (2023) initial analyses demonstrated that a one-component solution (with all items) accounted for most variance within participants' ratings (eigenvalue = 8.04). This was closely followed by a second possible component with an eigenvalue of 0.95, but additional components added progressively less variance. Moreover, observation of the PCA's scree plot indicated that one component was enough to model participants' ratings, although the plot also suggested that a two-component model may be appropriate. Importantly, though, the model-fit indices for the PCA (shown in the top section of table 13) showed that none of these models fitted satisfactorily.

**Table 13**

*Summary of Goodness of Fit Statistics for 1-3 Component Solutions, Before and After Item Removal from Galbraith et al. (2023). Reproduced with Permission of the Rights Holder.*

Model	$\chi^2/\text{df}$	NoFParm	CFI	TLI	RMSEA	AIC	BIC	CorrBIC	SRMR
Initial 1-factor model	340.54/54	36	.843	.808	.179	5412.04	5523.86	5409.89	.059
Initial 2-factor model	214.76/43	47	.906	.855	.156	5308.25	5454.23	5305.43	.040
Initial 3-factor model	107.77/33	57	.959	.918	.117	5221.26	5398.30	5217.84	.024
Final 1-factor model	156.67/20	24	.870	.817	.204	3746.04	3820.58	3746.04	.065
Final 2-factor model	17.36/13	31	.996	.991	.045	3620.73	3618.87	3618.87	.014
Final 3-factor model	7.06/7	37	1.00	1.00	.007	3622.43	3737.35	3620.21	.010

*Note.* NoFParm = number of free parameters; CFI = comparative fit index; TLI = Tucker-Lewis

Index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square

residual; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; CorrBIC = sample-size adjusted BIC.

Modification indices were then used to progressively remove poorly fitting items from the components. This resulted in a two-component solution (also in table 13) which provided good fit. As can be seen, all fit indices for the final two-component model were either excellent (CFI and TLI) or at least good (RMSEA). This model also provided significantly better fit to the data than the one-component reduced-item model ( $\chi^2(7)=139.30$ ,  $p<.00005$ ) and no worse than the three-component reduced-item model ( $\chi^2(6)=10.30$ ,  $p=.11$ ).

**Table 14**

*Loadings for 2-Factor Solution using Maximum Likelihood Estimation with Geomin Oblique Rotation from Galbraith et al. 2023). Reproduced with Permission of the Rights Holder.*

Items	Understanding	Organisation
1. How much you feel you know about the topic.	<b>0.841*</b>	0.020
2. How well you understand the topic.	<b>0.880*</b>	0.062
3. How organised your thoughts about the topic are.	0.225*	<b>0.678*</b>
4. How well you could explain the topic.	<b>0.450*</b>	<b>0.462*</b>
5. How coherent your thoughts about the topic are.	0.002	<b>0.871*</b>
6. How structured your thoughts about the topic are.	-0.155	<b>1.040*</b>
7. How much you can make sense of the topic's issues.	<b>0.418*</b>	<b>0.400*</b>
8. How well-ordered your thoughts about the topic are.	0.082	<b>0.790*</b>
Cronbach's alpha	.88	.93

*Note.* Loadings  $\geq .4$  appear in bold. \* $p<.05$ . Correlation between factors = .66.

The authors argued that the two components represented *understanding* (i.e. how much participants felt they knew about the topic) and *organisation* (i.e. how ordered and coherent participants' thoughts about the topic were). The factors were relatively highly correlated with one

another ( $r = .66$ ), suggesting that the concepts of *organisation* and *understanding* share common ground but are indeed separate components.

In the current study, only the eight items that loaded onto the *understanding* and *organisation* components in the Galbraith et al. (2023) investigation were analysed. However, the full 12-item version was administered to participants. This was because this study ran simultaneously with the assessment of the questionnaire's underlying structure, so the reduced-item version was not available at the time of the present study. In the current investigation, the *understanding* component has also been renamed to *knowledge*, to make it completely clear that the component is capturing how much respondents think they know about the topic in question. It is also salient to note that because Galbraith et al.'s analysis showed that items '*How well you could explain the topic*' and '*how much you can make sense of the topic's issues*' loaded onto both components, in the analysis for the current study, these items are used in the calculation of both the overall *organisation* and *knowledge* scores for each participant.

## Results

To obtain a global revision and sentence production score, the items that made up each scale (see Chapter 4, pp.141-144) were z-transformed. An average score taken from each of these standardised groupings, giving an overall global revision and sentence production score for each participant. To get a participant-level score for writing beliefs, the mean of the items from each latent factor within the WBI (transmissional, transactional, revision, planning and audience) was calculated. To obtain overall measures of understanding change over the course of writing for each participant, the difference between pre-test and post-test scores for each subjective understanding scale item was calculated. The mean was taken from the differences of the items that loaded onto the *knowledge* component, and the same was done *organisation*, resulting in two separate scores, one for each component, per participant.

**Table 15**

*Means, Standard Deviations and Bivariate Correlations of All Variables in the Understanding Change Analysis*

Variables	Mean	SD	Correlations											
			1	2	3	4	5	6	7	8	9	10	11	12
1. Transmissional	2.92	0.67												
2. Transactional	2.21	0.57	-0.23											
3. Revision	2.24	0.53	-0.24	<b>-0.62</b>										
4. Audience	2.87	0.56	-0.18	<b>-0.64</b>	<b>-0.59</b>									
5. Planning	1.9	0.57	<b>-0.29</b>	<b>-0.29</b>	<b>-0.30</b>	<b>-0.31</b>								
6. Condition	1.5	0.51	0.21	-0.22	<b>-0.43</b>	<b>-0.34</b>	-0.09							
7. Pre-test Knowledge	4.98	1.01	-0.10	-0.16	-0.22	-0.24	-0.18	0.14						
8. Knowledge Difference	0.17	0.66	-0.05	-0.26	-0.25	-0.26	0.03	0.13	<b>0.74</b>					
9. Pre-test Organisation	4.61	0.94	-0.05	0.19	0.21	0.19	0.26	-0.15	<b>-0.40</b>	-0.16				
10. Organisation Difference	0.28	0.78	0.04	<b>0.35</b>	0.22	0.22	0.19	-0.10	-0.07	-0.25	<b>0.67</b>			
11. Sentence Production	0.01	0.4	0.03	0.14	<b>0.32</b>	0.16	0.11	-0.20	-0.08	-0.03	-0.22	-0.27		
12. Global Revision	0	0.42	-0.19	0.14	-0.01	-0.15	0.01	0.02	0.21	0.09	-0.19	-0.03	-0.02	
13. Format	1.67	0.47	0.06	0.09	0.25	0.22	0.20	0.05	0.09	0.03	0.15	0.12	-0.03	-0.15

*Note. N = 46; Significant correlations ( $p < 0.05$ , two-tailed) are in bold; To aid with interpretation, correlations involving writing beliefs have been reversed so that positive correlations relate to stronger, rather than weaker, writing beliefs.*

<sup>a</sup> Dummy coded, data collected online = 0, data collected in person = 1; <sup>b</sup> Dummy coded, outline = 0, synthetic = 1

Table 15 shows the descriptive statistics and correlations for the variables included in the analyses. Focusing on the relationships between the writing beliefs first, revision, audience and planning beliefs were all moderately negatively related to transactional beliefs. This is interesting because although in the writing beliefs investigation presented in Study 1, the planning, revision and audience factors were also seen to be negatively related to transactional beliefs, those inter-factor correlations were very small and non-meaningful (all below 0.25, see Appendix J for breakdown of inter-factor correlations). In this investigation, revision and audience beliefs are much more negatively related to transactional beliefs. Similarly, in this investigation the negative relationships between revision beliefs with audience and planning beliefs are much stronger than they were in Study 1. Note, that this does not invalidate the findings from Study 1, but instead implies that there was something different about how the participants filled in the questionnaire within this study. It also implies that relationships between the writing beliefs with subjective understanding might not be observed in the same way as we would expect them to, based on theory and previous evidence. The results from the writing beliefs analysis will indicate whether that is indeed the case, and any implications of that will be explored in the discussion of this chapter.

There were also small to moderate negative relationships between drafting condition with revision and audience beliefs. This indicated that participants in the synthetic condition had slightly stronger revision and audience beliefs than participants in the outline condition. However, the correlations between condition and revision/audience beliefs was not strong enough for these variables to be considered multicollinear.

A significant small/moderate positive correlation was also observed between transactional beliefs and difference in organisation scores, indicating that writers with high transactional beliefs rated their thoughts as more organised after writing than in writers with lower transactional beliefs. This theoretically makes sense, as writers with high transactional beliefs are deemed to be more cognitively engaged with writing than those with low transactional beliefs. Hence, they are more likely to experience knowledge transforming processes rather than basic knowledge telling processes, which would likely help those writers to better organise their ideas. However, the correlational analysis

implied that no other writing beliefs appeared to be strongly related to difference in organisation and knowledge.

Pre-test knowledge was significantly and positively related with difference in knowledge, indicating that participants who rated their subject knowledge as high pre-essay writing, also rated it as high post-essay writing. This was also observed with pre-test organisation and difference in organisation, suggesting it would be necessary to statistically control for pre-test scores on the two components of the subjective understanding scale in later analyses.

The correlations seemed to indicate that drafting condition was not strongly related to differences in understanding and organisation, but of course, these correlations do not control for the effects that pre-test knowledge and organisation scores might have on the relationship.

Finally, and importantly, because of the Covid-19 outbreak, this study was conducted in two different formats. Before the outbreak, the experiment was conducted in person. However, afterwards it was conducted online (more information about how the experiment was adapted can be read in Chapter 2). Hence, it was necessary to check whether the format of the experiment was having an unexpected impact on the other examined variables. However, no strong or significant relationships were observed, so this was not included as a control variable in any of the subsequent models.

### ***Effects of Drafting Strategy on Change in Subjective Understanding***

Examination revealed that the mean change in knowledge scores pre- and post-writing was smaller in the synthetic condition ( $M = 0.08$ ,  $SD = 0.68$ ) than it was in the outline condition ( $M = 0.27$ ,  $SD = 0.63$ ). Difference in organisation pre- and post-writing scores was also smaller in the synthetic condition ( $M = 0.20$ ,  $SD = 0.87$ ) in comparison to the outline condition ( $M = 0.36$ ,  $SD = 0.68$ ). Importantly, though, the differences between the conditions were very small, indicating that drafting strategy may not have had an impact on the development of subjective understanding.

To formally test whether drafting strategy was indeed related to subjective understanding development over the course of writing, a multivariate analysis of (co)variance (MANCOVA) was conducted on the complete dataset ( $N = 61$ ). In this model, condition, pre-test knowledge and pre-test

organisation were the predictor variables. Difference in knowledge and difference in organisation were the outcome variables<sup>1</sup>.

The results showed that, when controlling for pre-test knowledge and pre-test organisation scores, drafting strategy did not have a statistically significant effect on change in organisation ( $F(1,57) = 0.67, p = 0.42, \eta^2 = 0.01$ ) and knowledge ( $F(1,57) = 0.20, p = 0.14, \eta^2 = 0.003$ ). Effect sizes were also small.

### ***Effects of Writing Processes on Subjective Understanding Development***

Whilst drafting conditions did not appear to be associated with subjective understanding development, this was not unexpected (Baaijen and Galbraith (2018) did not find such effects either). Drafting conditions would only be expected to be associated with the development of understanding to the extent that they led writers to implement relevant processes: Baaijen & Galbraith (2018) found, for example, that synthetic planning was only associated with increased understanding when text production was carried out spontaneously. The dual process model would also theorise that high levels of global revision would be related to an increase in the writer's organisation of thoughts, which could be related to an increase in their subjective understanding.

To examine whether sentence production and global revision processes were related to the development of subjective understanding, two regression analyses were conducted. The first regressed change in knowledge on sentence production and global revision (controlling for pre-test knowledge). In order to replicate Baaijen and Galbraith's (2018) examination, initially the full model they used was fitted. This included: the main effects of condition, sentence production, global revision and the control variable pre-knowledge; all two-way interactions between condition, sentence production and global revision; and the 3-way interaction between these three predictors. However, the sample size for this analysis was considerably smaller than it was in the Baaijen and Galbraith (2018) study (current study  $N = 47$ , original study  $N = 84$ ), so it was clear that in replicating the analysis, sample

---

<sup>1</sup> MANCOVA model assumptions were achieved; Q-Q plots revealed that these data had multivariate normality, Box's M Test showed that covariance matrices were homogenous across conditions ( $\chi^2 = 3.86, p = 0.28$ ), and Shapiro Wilk's test showed that the dependent variables were normally distributed ( $W = 0.97, p = 0.1$ ).

size in the current study was likely too small to pick up any interaction effects. Hence, non-significant interaction terms were progressively removed, starting with the highest order interaction, to simplify the model. The same was done for regressing organisation onto sentence production and global revision (controlling for pre-test organisation). The results are in Tables 16 and 17<sup>2</sup>.

**Table 16**

*Simplified Regression of knowledge on Sentence Production and Global Revision*

Variable	$\beta$	Std. Error	t	p
(Intercept)	0.69	0.59	1.18	0.24
Condition	-0.23	0.23	-1.02	0.31
Sentence Production	-0.60	0.29	-2.07	0.04*
Global Revision	-0.03	0.28	-0.09	0.98
Pre-test KNOWLEDGE	-0.06	0.12	-0.50	0.62

Note. \* indicates significance at the 5% level.  $R^2=0.106$ ,  $F = (4,42) = 1.24$ ,  $p=0.31$ .

**Table 17**

*Simplified Regression of Organisation on Sentence Production and Global Revision*

Variable	$\beta$	Std. Error	t	p
(Intercept)	1.30	0.55	2.35	0.024*
Condition	-0.20	0.22	-0.88	0.38
Sentence Production	-0.59	0.29	-2.11	0.04*
Global Revision	-0.012	0.27	-0.05	0.96
Pre-test Organisation	-0.20	0.12	-1.69	0.10

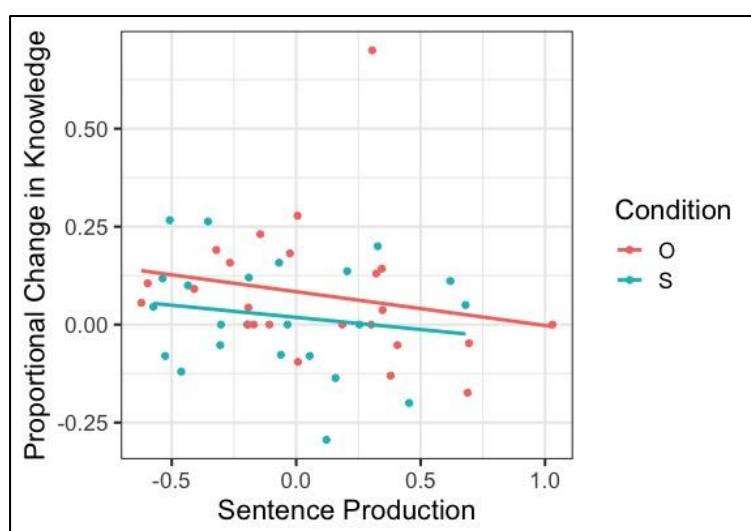
Note. \* indicates significance at the 5% level.  $R^2=0.16$ ,  $F = (4,42) = 1.96$ ,  $p=0.12$ .

<sup>2</sup> All models met the assumptions for multiple regression, including linear relationships between variables, normally distributed residuals and homoscedasticity. However, for the knowledge on sentence production/global revision model, there was one outlier in knowledge scores (more than 1.5x the interquartile range). Inspection of their data however revealed no reason for systematic removing them. Cook's distance also revealed no highly influential cases within either of the models.

The results indicate that there was a significant main effect of sentence production on knowledge ( $\beta = -0.60$ ,  $t = -2.07$ ,  $p = 0.04$ ), suggesting that more spontaneous sentence production was associated with greater increases in understanding. This relationship is in the same direction as the one discovered by Baaijen and Galbraith (2018). However, unlike Baaijen and Galbraith's findings, this effect was not moderated by type of planning. Figure 10, depicts how the effect is similar across both planning conditions. Furthermore, and again unlike Baaijen and Galbraith (2018), there was no significant relationship between global revision and the development of understanding.

**Figure 10**

*Proportion of Change in Understanding Over the Course of Writing by Condition*



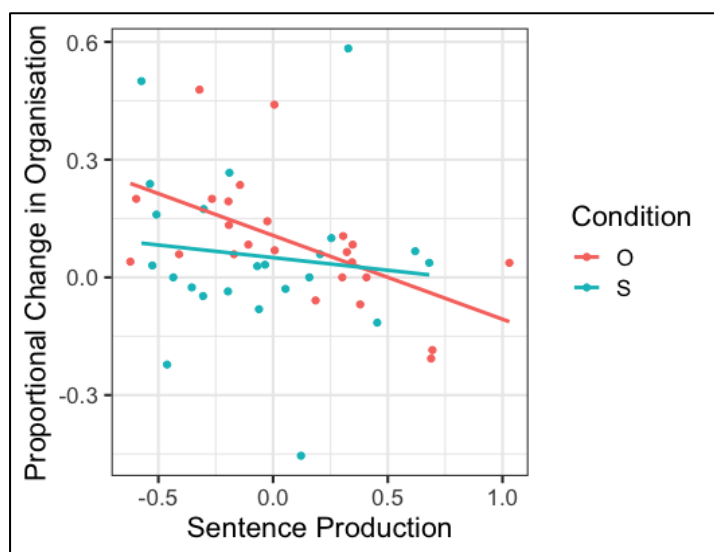
*Note. Proportion of knowledge calculated as difference in knowledge divided by pre-test knowledge.*

Baaijen and Galbraith (2018) did not explicitly measure organisation as a component of understanding, so the analysis presented in Table 15 is relatively novel. This analysis shows essentially the same pattern as found for the knowledge variable: a main effect of spontaneous sentence production on increased organisation ( $\beta = -0.59$ ,  $t = -2.11$ ,  $p = 0.04$ ). Surprisingly, though, there was no relationship between global revision and change in organisation.

One feature to note here is that the effect on organisation appears to be stronger for outline planning than it was for synthetic planning (see Figure 11). However, this had a small effect size and was not statistically significant ( $\beta = 0.63$ ,  $t = 0.58$ ,  $p = 0.28$ ,  $f^2 = 0.02$ ).

**Figure 11**

*Proportion of Change in Organisation Over the Course of Writing by Condition*



*Note. Proportion of organisation calculated as difference in organisation divided by pre-test organisation.*

### ***Effects of Writing Beliefs on Subjective Understanding Change***

The original intention with the next section of analysis was to try to replicate the findings from Baaijen et al. (2014) and then extend them by also incorporating planning beliefs and audience beliefs into their models. However, the sample size for this analysis was too small for the number of interactions that this type of analysis required. So, it was unachievable to conduct an exact replication of the original study, because statistical power would have been extremely low. Hence, the decision was taken to try to partially replicate some of the analysis from Baaijen et al. (2014), focusing explicitly on whether transactional and transmissional beliefs interacted with planning strategy to affect subjective understanding.

Because the original Baaijen et al. (2014) study used the White and Bruning (2005) scale to examine writing beliefs, it is necessary to highlight that their examination of transactional items also included the examination of revision items. To run the equivalent examination in the current study, each participant was given a combined transactional-revision (TAR) score (representative of the transactional factor in the original study). This was then interacted with transmissional (TM) beliefs

and drafting strategy. The benefit of doing this in comparison to a four-way interaction between transactional beliefs, transmissional beliefs, revision beliefs and drafting strategy was that there were fewer model parameters (a four-way interaction would have created far too many variables to legitimately be examined, given this study's sample size). The model results are outlined in Tables 18 and 19.

**Table 18**

*Equivalent Replication of Baaijen et al. (2014) Analysis of knowledge Regressed on Writing Beliefs*

Variable	$\beta$	Std. Error	t	p
(Intercept)	3.79	4.64	0.82	0.42
Pre-test Knowledge	-0.23	0.09	-2.28	0.09
TAR	-0.75	2.12	-0.35	0.73
TM	-0.79	1.54	-0.51	0.61
Condition	-3.05	5.27	-0.58	0.57
TAR*TM	0.23	0.67	0.33	0.75
TAR*Condition	0.58	2.43	0.21	0.83
TM*Condition	1.26	1.70	0.75	0.46
TAR*TM*Condition	-0.28	0.77	-0.37	0.71

*Note.*  $R^2 = 0.28$ ,  $F = (8, 37) = 1.80$ ,  $p = 0.12$ .

There were no significant effects of transactional-revision, transmissional beliefs and condition on knowledge (note that even reducing the model by removing non-significant interaction terms did not lead to an effect of writing beliefs on knowledge). This means that the effects of writing beliefs and drafting condition on knowledge examined in Baaijen et al. (2014) were not conceptually replicated within this study.

**Table 19**

*Equivalent Replication of Baaijen et al. (2014) analysis of knowledge regressed on writing Beliefs*

Variable	$\beta$	Std. Error	t	p
(Intercept)	3.93	5.68	0.69	0.49
Pre-test Organisation	-0.11	0.13	-0.86	0.40
Condition	-2.16	6.40	-0.34	0.74
TAR	-0.89	2.58	-0.35	0.73
TM	-1.19	1.87	-0.64	0.53
Condition*TAR	0.049	2.95	0.02	0.99
ConditionS*TM	1.57	2.05	0.77	0.45
TAR*TM	0.377	0.85	0.44	0.67
Condition*TAR*TM	-0.392	0.93	-0.42	0.68

*Note.*  $R^2=0.25$ ,  $F = (8,37) = 1.57$ ,  $p=0.17$ .

In terms of the impact of transactional-revision beliefs, transmissional beliefs and condition on subjective organisation, again no significant, notable effects were discovered, even after simplifying the model by removing non-significant interaction terms. This suggested that transactional-revision and transmissional writing beliefs did not impact subjective organisation within this study.

Finally, a key aim within this research was to see whether planning beliefs and audience beliefs contributed to subjective understanding development, because these relationships had not been examined in previous studies. As such, exploratory analysis was conducted by incorporating planning and audience beliefs into the pre-existing writing beliefs regression models. Interaction terms were completely removed from the models due to small sample size and their lack of significance in the previous analyses. However, neither planning nor audience beliefs were observed to have a statistically meaningful impact on knowledge or organisation outcomes and change in effect size.

To summarise, then, the results from the writing beliefs analysis was inconclusive. The equivalent replications of the Baaijen et al. (2014) study did not provide any further evidence of an effect of transactional and transmissional (and revision) beliefs on subjective understanding. Furthermore, the exploratory analysis showed no evidence for a simple effect of planning and

audience beliefs on the two components of subjective understanding. The discussion section of this chapter will explore potential reasons for the lack of observable relationships.

## **Discussion**

### ***The Impact of Drafting Strategy on Subjective Understanding Development***

This study aimed to investigate whether subjective understanding development over the course of writing differed between participants who used outline and synthetic drafting strategies. The results implied little to no difference between the drafting conditions. In terms of why this may be, a simple reason could be that the variations between conditions did not result in differences in the level to which writers developed their subjective understanding during writing. The mean levels of increase for knowledge and organisation were low across synthetic and outline condition writers ranging from  $M = 0.057$  to  $0.433$ . Hence, it appears that development of subjective understanding stayed similar from pre- to post-writing. There are several reasons why this might have happened. Firstly, the essay topics chosen may have had an impact. The two essay questions were “does social media do more harm than good?” and “should we all become vegans?”. These topics were chosen because they reflect current affairs that the participants were likely to have an opinion about. As such, it was assumed that they would be able to argue for or against the question based on those opinions. However, perhaps the students already had relatively high subjective understanding of these topics, so increases in their understanding may not have changed much throughout writing. In actual fact, the overall means for pre-test understanding and organisation scores ( $M = 4.98$  and  $4.55$ , respectively) were very close to the overall means for post-test understanding and organisation ( $M = 5.16$  and  $4.90$ , respectively) so, in general, participants’ subjective understanding of the writing topics stayed at a middling level throughout the writing task. This is likely because participants were only given 30 minutes to write, and five minutes to plan, and this could have been too little time to experience and substantial development in understanding.

Nevertheless, previous research found significant differences in students' understanding development scores in outline and synthetic conditions, so it is possible that an effect exists, but that the current study did not identify it. Moreover, sample size was an issue. An a priori power analysis indicated that to obtain a medium effect size with sufficient statistical power ( $1 - \beta = 0.8$ ), 128 participants would have been required for the study, so it is unsurprising that an effect of drafting strategy was not observed, and in fact implies that if the study had a larger sample size, there may have been evidence of a relationship.

### ***The Impact of Writing Processes on Subjective Understanding Development***

Another reason why the two conditions might not have led to differences in subjective understanding development is because the manipulations to writing may not have produced their intended effects on the writing process. Hence, it was also essential to examine the relationships between underlying writing processes and subjective understanding development. This analysis revealed that sentence production processes had an impact on subjective understanding development. Specifically, writers who produced sentences more spontaneously had larger increases in subjective understanding during writing than those who wrote their sentences in a more controlled way. These findings fall in line with those of Baaijen and Galbraith (2018), who theorised that spontaneous sentence production reflected spontaneous formulation of ideas during writing, forming part of the knowledge constituting process.

Conversely to Baaijen and Galbraith (2018), however, the current study showed no strong relationship between global revision and subjective understanding development. In the former study, high levels of global revision were linked to an increase in understanding, indicating that editing the structure of a text during text production helped the writer to organise their thoughts (knowledge transforming) and consequently increase their understanding of the topic. It is plausible that the lack of replication of these findings could be related to the way in which the current study extended the writing manipulation. That is, in the previous study, all writers across conditions were told to write a final version of an essay that was well formed and had accurate spelling. In this study, only the outline

condition was instructed to write polished essays. The writers in the synthetic condition were told to write rough drafts. To reiterate, they were told to write their thoughts down as they occurred, to not worry about the structure of their texts and to not be concerned about spelling. Hence, global revision was not likely to play a big role within the synthetic condition, as text structure was not highlighted as an important outcome. Similarly, in the outline condition, global revision may have been reduced because of the structured pre-planning task that writers completed. As writers in the outline condition were asked to make a hierarchical plan with the ideas they wanted to include in their texts, it is possible that much of the text structure for these writers was determined in the outline planning component of the task. Consequently, the likelihood of these writers editing the global structure of their texts would have been reduced.

As a result, it is possible that the impact that global revision can have on subjective understanding development, as indicated in Baaijen and Galbraith (2018), was ‘levelled out’ in this study. The requirements of the synthetic condition potentially eliminated the need for global revision because writers were just asked to write down their thoughts as they occurred. In the outline condition, the hierarchical planning element likely reduced the need for writers to revise the structure of their texts throughout essay writing, because the order of their text was determined in their outline plan. It would be interesting to conduct replications of the Baaijen and Galbraith (2018) study and the current study. If the findings of both investigations were upheld, it would provide strong evidence that the effects of global revision on subjective understanding are eliminated under rough drafting scenarios.

### ***The Impact of Writing Beliefs on Subjective Understanding***

No strong relationships were observed between writing beliefs and both components of subjective understanding in this study. In terms of why this happened, the most obvious potential reason is that the sample size too small to detect any meaningful effects.

It is also worth noting, though, that the bivariate correlations presented at the beginning of the chapter may have played a contributing role. These indicated moderately strong negative relationships

between planning, revision and audience scores, which were not observed in the validation of the WBI (Galbraith & Baaijen, in preparation) in Study 1. This implies that these beliefs may have been related to each other differently in this study than in the validation study, which ultimately could have had an impact on the relationship with understanding. This is further supported by the fact that there were moderate correlations between revision beliefs with audience and planning beliefs, which were also not observed in Study 1.

It is important to highlight that these findings do not invalidate the results from Study 1, because not only did that study indicate that the five-factor ESEM provided good fit to the data, but also it demonstrated that the five-factor structure could also be observed in a group of dyslexic students.

Instead, these results stress the need for additional testing of the WBI in different writing scenarios. In this study, participants knew that they were going to write an argumentative essay prior to completing the WBI. Hence, when they filled it in, they may have been reporting their beliefs specifically in relation to argumentative essay writing. Conversely, in Study 1, participants were told to think about their beliefs in regards to general academic writing – a much broader and ambiguous categorisation of writing than an argumentative essay. Ultimately this could have meant that the writing beliefs captured by the WBI in this experiment did not act in the same way as in the validation study, because the writers were thinking about their beliefs in reference to different writing tasks. This subsequently might have led to the writing beliefs not being related to subjective understanding development in the way that was expected.

As such, future research should consider not only validating the WBI in different populations of writers, but also in reference to different writing tasks, in case this has an impact on how writers group their beliefs, and the relationships those beliefs might form with understanding.

### ***Evaluating the Utility of a Multi-Item Scale of Subjective Understanding Development***

In terms of evaluating the utility of using a multi-component scale for examining understanding change, the findings from this study indicated that the link between sentence production with

knowledge and organisation appeared to follow the same directional relationship. That is, increases in organisation and how much writers felt they knew were linked to more spontaneous sentence production processes, rather than more controlled processes. As the direction of these relationships were the same, one could argue that it is not necessary to split the Subjective Understanding Scale into separate organisation and knowledge sub-components.

However, there are two things to note here. Firstly, the strength of those relationships *were* different, indicating that the extent to which spontaneous sentence production impacts subjective understanding development was stronger for organisation than knowledge. Secondly, the fact that the relationship observed between sentence production processes and the two components of the understanding scale followed the same direction supports the assumptions of the dual process model. This is because the model assumes that organisation of ideas (knowledge transforming) and production of ideas (knowledge constituting) work together to increase the writer's understanding. Hence, one would expect that increases in scores on the organisation component would also relate to increases in scores on the knowledge component. Using a two-component measure of subjective understanding meant that this theoretical directional relationship could be observed.

To conclude, then, this study suggests that sentence production processes have an impact on subjective understanding development over the course of writing, with writers who produce their sentences more spontaneously experiencing the biggest overall increases in subjective organisation and knowledge. These findings support those of Baaijen and Galbraith (2018), and they support the assumptions of the Dual Process Model of Writing (Galbraith & Baaijen, 2018). The findings also imply that the manipulation to the type of writing that the participants were required to produce (synthetic with rough drafting and outline with complete drafting) eliminated the impact of global revision on the development of subjective understanding throughout writing, rather than extending it, as was originally hypothesised.

Nevertheless, the study produced inconclusive results regarding the impact of drafting strategy and writing beliefs on the development of subjective understanding, and this was most likely due to small sample size and low statistical power, but might have also been because of issues surrounding how the writer interpreted the WBI items in reference to argumentative essay writing. It is absolutely

necessary that future studies investigating the effects of drafting strategies, writing processes and writing beliefs obtain large samples of participants for the results to be reliable.

## **6 General Discussion**

### **Introduction to General Discussion**

This final chapter will bring together the findings of all three studies within the thesis to discuss whether the evidence provided supports the assumptions of the dual process model in relation to development of subjective understanding. It will discuss the strengths and weaknesses of the research conducted, and as such will make recommendations for future research within this area. Finally, this chapter delves into the real-world implications of the project's results.

### **A Brief Summary of the Results from the Three Studies**

This project consisted of three studies designed to estimate the effects of drafting strategies, writing beliefs and writing processes on subjective understanding development in university students' essay writing. The first study focused on determining whether the WBI (Galbraith & Baaijen, in preparation) was an appropriate tool for examining the writing beliefs of university students with and without dyslexia. The second study highlighted transparency and reproducibility issues in keystroke analysis methods, and presented a reproducible framework for analysing keystroke data in response to those issues. It also explored whether Gaussian mixture modelling could be a valuable tool for examining writing processes when used in combination with experimental manipulation to the planning and drafting strategy of a writing task, with the aim of making the procedure for mapping pauses onto processes easier. Finally, Study 2 examined the reproducibility of two writing process measures - global revision and sentence production (Baaijen & Galbraith, 2018). The final study brought together the WBI (Galbraith & Baaijen, in preparation), examined in Study 1, and the sentence production and global revision measures (Baaijen & Galbraith, 2018), analysed in Study 2, to investigate how writing beliefs and writing processes, along with drafting strategy, could impact subjective understanding development throughout writing.

The most pertinent findings from Study 1 were that the WBI appeared to have a five-factor underlying ESEM structure, indicating that it captures five different types of writing belief (transmissional, transactional, planning, revision and audience). Additionally, the WBI had a generalised five-factor structure across two different groups of writers: university students with and without dyslexia. Also, examination of latent factor means across the two groups showed that students with dyslexia had higher transmissional and planning beliefs, and lower transactional beliefs than students without dyslexia.

For Study 2, the most important findings were as follows. Firstly, the process of creating a completely reproducible framework for the analysis of keystroke features was difficult, because high levels of variation within and between writers meant that pauses, bursts and revisions presented themselves in many ways. This highlights that transparency in the reporting of methods used to extract and analyse keystroke features are perhaps more important than trying to develop a standardised, systematic framework to do the same. Secondly, the results from Study 2 implied that mixture modelling may provide a useful tool for mapping keystrokes onto writing processes when experimental manipulation is used to elicit particular processes, but further evidence is needed to confirm the extent of that utility. Finally, the Baaijen and Galbraith (2018) measures of global revision and sentence production did appear to be reliable and reproducible (apart from one sub-measure, the sentence linearity index, which was described ambiguously in the original research).

For Study 3, the most salient findings were that sentence production processes were observed to impact subjective understanding development over the course of writing, with writers who produced sentences more spontaneously experiencing the biggest increases in subjective understanding. Moreover, the extension to the experimental manipulation to planning strategy in the original Baaijen and Galbraith (2018) and Baaijen et al. (2014) studies to also include drafting strategy could have resulted in a 'levelling-out' of the effects that global revision had on subjective understanding across the two writing conditions. This would have been because the two groups were required to create different end products (synthetic planning with rough drafting, where global revision is not required, and outline planning with final drafting, where global organisation is established in the planning task). Additionally, the study provides support for the usage of a multi-

item scale of subjective understanding development that examines how much writers feel they know about the topic and how they organise their thoughts around a topic as two separate components.

## **Discussion of Findings in Relation to the Dual Process View of Subjective**

### **Understanding Development**

This project provided evidence supporting the dual process model's view of how subjective understanding develops during writing. Firstly, in Study 1, dyslexic students were observed to have a lower latent mean for transactional beliefs than the students without dyslexia. Typically, high transactional beliefs have been related to higher development of subjective understanding (Baaijen et al., 2014) and better text quality (White and Bruning, 2005). This is perhaps because writers with high transactional beliefs are more cognitively and emotionally involved in the writing process than writers with low transactional beliefs, meaning that they are more likely to engage in knowledge transforming and knowledge constitution processes, than just simple knowledge telling processes. The fact that the dyslexic participants showed lower transactional beliefs than the students without dyslexia indicates that they are less cognitively and emotionally involved in writing than students without dyslexia (perhaps because the word-level decoding and encoding issues they experience take up too much cognitive resource during writing), and so they may view academic writing as a knowledge telling process, resulting in little subjective understanding development during the course of writing. Given that students with dyslexia have reported that organisation and expression of ideas are aspects of academic writing that they struggle with (Mortimore and Crozier, 2006), this could be because of their low transactional beliefs, and a tendency to view writing as a knowledge telling rather than knowledge constituting/knowledge transformation process.

This is further exemplified by the fact that the dyslexic students had a higher latent mean for the transmissional factor than the students without dyslexia, indicating these writers prioritise conveying accurate information to the audience. As such it would make sense that dyslexic writers would engage more in straightforward knowledge telling processes than knowledge transforming/knowledge constitution processes because they would want to express the knowledge

they already have, rather than manipulate it or even develop new ideas, on the basis that doing so might jeopardise the extent to which the information they were reporting was indeed accurate.

Finally, the dyslexic writers had a higher latent mean score for planning beliefs than in the group without dyslexia. This indicates that the writers with dyslexia thought that planning was an important aspect to consider prior to and throughout writing, and more so than the students without dyslexia. However, as indicated in previous research (Baaijen and Galbraith, 2014), and also based on the assumptions of the dual process model, planned and controlled text production that fulfils rhetorical constraints may limit the extent to which writers can develop their understanding of a writing topic (via the knowledge constituting system) because thoughts are not being produced spontaneously.

To this end, the fact that the dyslexic participants had a combination of lower transactional beliefs but higher transmissional and planning beliefs provides evidence that writers with dyslexia may gravitate more towards basic knowledge telling processes during writing than more cognitively engaged knowledge transforming and knowledge constituting processes. Hence, it makes sense that these writers would struggle with the expression of ideas and organising their thoughts when writing, as indicated by Mortimore and Crozier (2006).

Yet, it is important to consider that the results from Study 1 only provide indirect support for the two systems in the dual process model. They imply that students with dyslexia, who have commonly been defined as a group that struggles with academic writing, may not engage with processes related to subjective understanding development. However, because Study 1 did not also examine participants' academic writing, the implications of the WBI analysis are only theoretical. Future research should use experimental techniques to examine whether the writing beliefs that students with dyslexia report are related to their writing outcomes (e.g. text quality and subjective understanding). It would be useful to also compare these results to those of writers without dyslexia, to see whether any discrepancies between the two groups might go some way as to explaining why students with dyslexia find academic writing difficult (beyond the word-level issues such as reading and spelling which are commonly associated with dyslexia).

Nevertheless, further preliminary support for the knowledge-constitution system of the dual process model comes from Study 2's mixture modelling results. Specifically, the study revealed that writers who were assigned to different essay planning/drafting strategies, designed to elicit different writing processes, showed evidence for varying processes at the sub-sentence and sentence level of text production. Firstly, there was a significant difference in Component 1 sub-sentence pause means, demonstrating that writers in the outline condition were typically pausing for longer. It was reasoned that the pauses in this component were likely representative of higher-level content planning, rather than word-level planning, because the length of these pauses was longer than those for component 1 at the within- or between-word boundaries. The fact that the sub-sentence pausing was longer in the outline condition implies that these writers were using more planned and controlled text production processes in order to meet the rhetorical goals of the task (i.e. creating a well-drafted final text product). Controlled sentence production is linked to lower engagement with the knowledge constituting system of the dual process model (Baaijen & Galbraith, 2018), which the model would predict to be the case in an outline planning/final drafting text condition, as the strictness around the rhetorical constraints of the task would limit the extent to which the writers could develop their personal understanding of the topic through spontaneous sentence production processes.

There were also six participants in the synthetic planning/rough drafting condition whose data could not be represented with a two-component GMM due to too few observations of sub-sentence pause types. Under the assumptions of the dual process model, this makes sense because these writers were explicitly told not to be concerned about text structure and communicative goals. Because the writers in the synthetic condition were prompted to write down their thoughts as they occurred, these writers were more likely to be writing down their thoughts as they unfolded, subsequently engaging more in knowledge constitution processes, rather than focusing on applying conventional syntactical and grammatical practices associated with high-quality text production, which would inhibit knowledge constitution.

These findings were also somewhat echoed at the between-sentence pause level, albeit to a lesser extent. Seven participants in the synthetic condition did not have GMMs fitted to their data due to too few linear between-sentence pause observations. This was compared to only two participants in

the outline condition. Again, because the participants in the synthetic condition were told to write their thoughts down as they occurred, this could have resulted in some of the synthetic participants writing their texts as if they were spontaneous streams of consciousness (knowledge constitution), rather than well-formed essays, and hence not writing many full sentences.

It is important to highlight, though, that the results from Study 3 indicated that there was no definitive impact of drafting strategy on subjective understanding development. So, although the mixture modelling results from Study 2 implied that writers in the synthetic condition were engaging in more spontaneous text production processes, the extent to which this then improved their subjective understanding around the writing topic is unclear.

Nevertheless, Study 3 also provided evidence in support of the dual process view of subjective understanding development, again, in relation to the writing processes that participants were engaging in. In terms of support for the knowledge constituting system, writers who had higher scores on the sentence production measure (i.e. they produced their texts more spontaneously), experienced larger increases in subjective understanding than writers with lower sentence production scores. This finding directly supports the assumptions of the dual process model, because it indicates that increased subjective understanding is related to spontaneously produced text.

On the other hand, global revision was observed to have no relationship with subjective understanding. This finding directly opposes those of Baaijen and Galbraith (2018). In the former study, high levels of global revision were linked to an increase in understanding developed over the course of writing, indicating that editing the structure of a text during text production helped the writer to organise their thoughts (knowledge transforming) and consequently increase their understanding of the topic. It is plausible, however, that the lack of replication of Baaijen and Galbraith's (2018) finding in this project could be related to the way in which it extended the writing manipulation. That is, in the previous study, all writers across conditions were told to write a final version of an essay that was well formed and had accurate spelling. In this project, only the outline condition was told to do this. The writers assigned to the synthetic strategy were told to write rough drafts. Hence, global revision was unlikely to play a large role within the synthetic condition, as text structure was not highlighted as an important outcome. Similarly, in the outline condition, global

revision may have been reduced because of the structured pre-planning task that writers completed. As writers in the outline condition were asked to make a hierarchical plan with the ideas they wanted to include in their texts, it is possible that much of the text structure for these writers was determined in the outline planning component of the task. Hence, the likelihood of them needing to edit the global structure of their texts during writing would have been reduced.

As a result, it is plausible that the impact that global revision can have on subjective understanding development, as indicated in Baaijen and Galbraith (2018), was levelled out in this study. The requirements of the synthetic condition potentially eliminated the need for global revision during writing. In the outline condition, the hierarchical planning element likely reduced the need for writers to revise the structure of their texts during writing. Both would have reduced the need for knowledge transforming processes.

Consequently, future research should aim to replicate the conditions of this study (to see whether findings can be upheld), but also of Baaijen and Galbraith's (2018) writing conditions, in order to determine whether it really was changes to the final text product in the synthetic condition that led to a lack of relationship between global revision and change in subjective understanding. If the results from both sets of studies could be upheld, it would imply that global revision, and hence knowledge transforming, is related to an increase in subjective understanding.

It may also be useful to assess change in understanding at multiple points throughout the writing task – once before planning, once directly after planning and once after the writing task has been completed, to see how much the planning and writing tasks *separately* relate to an increase in subjective understanding. Moreover, the rough-drafting strategy could be extended further, so that after writing their initial rough drafts, these writers then had to write a final draft. One would expect that during the final-drafting stage, these writers would experience greater global revision processes, which subsequently could lead to further understanding development. Hence, it would also be useful to examine understanding at an additional time point for the synthetic strategy writers - after their final drafts had been written.

Finally, the evidence produced in regards to the multi-component Subjective Understanding Scale (Galbraith et al., 2023) provides support for the dual process model. Specifically, the scale's

two components of *organisation* (which maps onto the knowledge transforming system) and *knowledge* (which maps onto the knowledge constitution system), were both observed to show a positive relationship with sentence production processes. The dual process model assumes that organisation of ideas (knowledge transforming) and production of ideas (knowledge constituting) work together to increase the writer's understanding, and so one would expect that increases in scores on the organisation component would also relate to increases in scores on the knowledge component, which was indeed the case in this study.

Needless to say, there are some questions that remain around the dual process model's assumptions about how subjective understanding develops throughout writing, not answered within this PhD project. As briefly touched upon already, evidence for the knowledge transforming system's role in subjective understanding development within this project is limited. Theoretical support comes from Study 1. The dyslexic students (who in previous research have been identified as a group who struggle to organise their ideas throughout writing) had lower latent means on the transactional factor, which is associated with engagement in the knowledge transforming system, but higher scores on the transmissional factor, which may be associated with lack of engagement with knowledge transforming. However, because subjective understanding as an outcome of writing was not formally assessed using the participants in this research, these findings do not supply *empirical* support for the role of the knowledge transforming system in subjective understanding development.

Moreover, the lack of effect of global revision processes on subjective understanding in Study 3 means that the only empirical evidence that Study 3 provides for the dual process model comes from the relationship between sentence production processes and development of subjective understanding. However, these results only provide support for the knowledge constituting system. Although the lack of impact of global revision processes on text quality may be down to how the writing conditions were designed, rather than global revision not playing a role in subjective understanding development at all, without further research on global revision processes, we will not know whether this is the case.

Another area that this PhD project did not provide empirical evidence for was the role that writing beliefs play in subjective understanding development. Study 3 did not reproduce the results that Baaijen et al. (2014) found in terms of the impact that transactional, transmissional and revision beliefs had on subjective understanding development. However, this was likely due to the study not having enough statistical power (the implications of which will be elaborated on in the next section of this discussion). Subsequently, Study 3 also produced null results in relation to the impact that planning and audience beliefs could have on subjective understanding development. As such, future research should focus on further investigating the impact of writing beliefs on subjective understanding development. It would also be beneficial for future studies to look at how writing beliefs might affect how writers engage with sentence production and global revision processes. This project did not consider the moderating role that writing beliefs may have had on writing processes because of issues with sample size making it impractical to run such an analysis. However, it is important to consider how writer's preconceived ideas about writing are actually implicated in the types of writing processes they engage in, because this may ultimately have an effect on subjective understanding development.

### **Strengths and Limitations of the Project**

As already touched upon, the most significant limitation of this project was that the sample size in each study was relatively small. In Study 1, although the sample of non-dyslexic students was large ( $N = 493$ ), the sample size for dyslexic students was very small in comparison ( $N = 69$ ). Hence, the statistical power of the invariance testing proportion of the study was likely to be low. As such, the results presented in Study 1 should only be considered preliminary and should be backed up with studies that utilise larger samples of dyslexic students to determine not only whether the WBI is genuinely suitable for usage amongst students with and without dyslexia, but also whether the results in relation to different latent means between the two groups can be upheld. However, the pragmatic issue in attempting this is that there are significantly fewer students with diagnosed dyslexia than students without in the general population. In Study 1, the number of students with dyslexia was

roughly proportionally representative of the number of people in the UK with dyslexia. Additionally, there was limited interest from students with dyslexia to participate in the study, perhaps because the research was about writing. This may have been unappealing to dyslexic students due to the nature of their struggles with academic writing. Hence, future research should consider ways of increasing the participation of students with dyslexia as part of the research design for the project. At the very least, this could be by emphasising how the research could be beneficial for dyslexic students.

In study 2, the sample size was also small ( $N=48$ ). However, because this study focused on demonstrating methodology, the number of participants was generally less of an issue. Yet, it would have been more beneficial to have a larger pool of participants for the GMM analysis. As a result, it was hard to determine whether there was a real trend in pause structure at different intervals, or whether the findings within the study were only representative of the participants within the group.

One reason why sample size was small in Study 2 is because data collection had to be transferred from face-to-face to online during the COVID-19 pandemic. This created multiple pragmatic challenges. For example, participants had to use their personal laptops to complete the experiment. In some instances, the Inputlog Version 8 keystroke logging software did not work on participants' computers, which resulted in fourteen cases where keystroke data could not be obtained. The primary reason for this issue was that Inputlog was not compatible with the operating system and version that some participants had on their laptops. As such, future research should consider using keystroke logging programs that are compatible with different operating systems, or programs that can be used within a web browser. The accessibility of the keystroke logging program in this study was poor. This is essential to consider given that much research is now taking place remotely, so participants need to be able to access keystroke logging software from their own personal devices.

In study 3, the sample sizes were small for the types of statistical analyses conducted. However, throughout the chapter, it was made clear that the study's findings were exploratory and intended to produce an initial insight into the impact that writing beliefs, drafting strategies and writing processes could have on subjective understanding change. The small sample size was again an outcome of the challenges that the COVID-19 pandemic created. So, whilst future studies should indeed strive to achieve a sample size that is big enough for high statistical power, it is crucial to

consider that in today's research climate, that may be very difficult. Hence emphasis should also be put on novel ways of ensuring that studies reach a large enough sample size. One way of doing this would be to run projects collaboratively across multiple institutions so that separate participant groups from each institution can be pooled into one large sample. This is also more likely to make the research representative of the general student population, as participants would not be primarily from one university.

One final important limitation with Study 3 was that it focused solely on subjective understanding as a writing outcome. As previous research has indicated (Baaijen et al. 2014; Baaijen & Galbraith 2014), drafting strategy, writing processes and writing beliefs all impact text quality as well as subjective understanding development. The decision was made *only* to examine subjective understanding as a writing outcome throughout this PhD project because research on academic writing outcomes is already heavily skewed towards text quality. As such, one aim of this project was to bring research on subjective understanding development over the course of writing more into focus. However, in retrospect, it would have been interesting to also examine text quality as a writing outcome because text quality and subjective understanding are likely interlinked, as Baaijen and colleagues' original research implies. Therefore, future research should consider examining text quality and subjective understanding development in unison.

Additionally, the study did not examine how writing beliefs might have moderated sentence production and global revision processes to impact understanding because small sample size would have resulted in incredibly low statistical power. Yet, the existing literature has not investigated the effects of writing beliefs on global revision/sentence production processes, and so this should be a priority area of focus in future research.

One major strength of this project was its use of open methods. Whilst this includes the transparent reporting of research methods and analytical decisions, advocated throughout the thesis, the project also used open methods not discussed in great detail. For example, the anonymous data and scripts for data analysis throughout this project have been made openly available via this project's Open Science Framework repository (link on p.12), which means that they can be fully scrutinised by other researchers, or used in other research projects. Also, for most of the analysis conducted within

this project, non-proprietary software has been used so that others can replicate the analysis accessibly. For instances where proprietary software was employed (e.g. MPlus and Excel), the scripts for the analyses have still been made publicly available, increasing the accessibility of the study's analytical methods.

Finally, the most important strength of this research project was the fact that it built upon the existing evidence base around the role of subjective understanding development during the course of writing. Specifically, the thesis provided preliminary evidence for areas not deeply explored within the existing literature. These include the role that beliefs about planning may have on writing, using reproducible methods to analyse keystrokes in relation to writing processes, utilising experimental manipulation of drafting strategy to aid with the mapping of pauses onto writing processes and, finally, using a multi-item scale to assess subjective understanding development throughout the course of writing. Although the thesis only provides preliminary evidence within these areas, the findings indicate that all these aspects are important to continue considering in future research.

### **Real-World Implications**

As reiterated throughout the thesis, the majority of the results in this project provide preliminary evidence for phenomena that should be further pursued in later research. Nevertheless, some of these results have important real-world implications.

Firstly, the findings have implications for how future writing research should be conducted. The results from Study 1 imply that ESEM is a better technique for investigating the structure of questionnaire tools than traditional techniques such as CFA and EFA. Specifically, cross-loadings within the ESEM framework allow you to understand how items may relate to more than one underlying construct. Although EFA also allows for cross-loadings, the benefit of ESEM over EFA is that it can be used alongside follow-up invariance testing, so one can investigate whether a tool's underlying structure is consistent across multiple groups of users. This is a particularly important thing to consider in writing beliefs research, because the most commonly used inventories (White and Bruning, 2005; Sanders-Reio et al., 2014) have had their structures exclusively examined in university

student populations without any form of writing difficulty. Yet, ultimately, writing difficulties may impact the way that writing beliefs manifest themselves, so this should be considered in the development of these types of questionnaires.

Study 2 demonstrated the necessity of transparent methodological and analytical reporting in keystroke research. Until now, methods for extracting and analysing pauses, bursts and revisions from keystroke logs have typically not been well and accurately explained. Study 2 has provided a framework for making transparent reporting in keystroke studies easier, and if the techniques demonstrated are used by other researchers, they should lead to keystroke analysis results being more reproducible.

Study 2 also demonstrated how mixture modelling as a technique for inferring processes from pause data might be better enhanced when it is conducted on experimentally manipulated writing, rather than data from non-experimental contexts such as copy tasks (frequently used in mixture modelling studies). The experimental manipulation enables researchers to direct participants to write in a particular way, which results in different underlying writing processes being used (e.g. high or low levels of spontaneous text production). This can be useful for then determining whether those manipulations have produced their intended effects on the pause data, which can help to narrow down the types of processes those pauses can be paired with. This technique is not perfect by any means, because ultimately without looking at pauses on a case-by-case basis, and without talking to the participants producing those texts, it is difficult to understand exactly what was happening in their minds during text production. However, it has the potential to improve the accuracy of mapping pauses in comparison to studies that rely on non-experimentally manipulated writing data. On a pragmatic basis, this is particularly useful in contexts where it is difficult to examine every individual pause on a case-by-case basis, such as in an extended writing task, because such tasks can produce extremely large datasets.

Study 2 also saw the reproduction of the Baaijen and Galbraith (2018) measures of global linearity and sentence production. This has implications for future research. Not only does the reproduction indicate that the two measures are relatively straightforward to compose (albeit without the sentence linearity index), but the PCA analysis indicated that the two measures were highly

reliable. This suggests that they are robust and trustworthy tools for examining writing processes with keystroke data. These findings also demonstrate the benefit of analysing pauses, bursts and revisions collectively, in the form of composite measures, rather than in isolation, and this should be considered in future research.

Finally, in terms of the real-world methodological implications from Study 3, the findings indicated that subjective understanding development may be better examined when a multi-item rather than single item scale is used. The evidence presented in the study revealed that a multi-item scale was better for detecting nuances in how writers felt they developed the organisation of their ideas, and how much they developed their knowledge around a topic during writing. These two components of subjective understanding link to the knowledge transforming and knowledge constituting components of the dual-process model respectively. As such, other researchers may not only consider using a multi-item scale because it is useful for examining the multi-faceted nature of subjective understanding, but also because it can be used to examine subjective understanding development in direct comparison with the dual process view of writing.

As well as having an impact on future research, the findings in this project have the potential to lead to change within academic writing practice (if followed up with more confirmatory research). For example, the fact that the results of Study 1 indicated that university students with and without dyslexia may have varying levels of writing beliefs suggests that this is likely to be implicated in how they perceive academic writing tasks, and subsequently their writing outcomes. The beliefs that the dyslexic students displayed (high transmissional and low transactional beliefs) have typically been associated with poorer writing outcomes (Baaijen et al., 2014; Sanders-Reio et al., 2014; White & Bruning, 2005). As such, it may be important for academic writing teaching practice to focus somewhat on writing beliefs. If students were trained to understand more about how their pre-conceived ideas about writing may be implicated in the way they engage with writing, then this might encourage those writers to alter their beliefs about writing, instead motivating them to think less in terms of transmissional beliefs, and more in terms of transactional beliefs.

Additionally (although this project did not provide empirical evidence for this theory), the thesis emphasises that the way in which drafting strategies are used in an academic writing context might

have implications on important writing outcomes. Whilst notable research from Kellogg (1988;1990) suggests that hierarchical outline planning can improve text quality, other research has indicated that such outline planning, which is typically advocated for by course lecturers and writing tutors, can have a negative impact on other important academic writing outcomes such as understanding development (Baaijen et al., 2014; Baaijen & Galbraith, 2018; Galbraith & Torrance, 2006). As such, it may be important for those teaching academic writing skills to be wary of the fact that a strategy aimed at improving one academic writing outcome (outline planning on text quality) might have a negative impact on another (understanding development). As such, it is important for teachers of academic writing to consider how different drafting techniques may be used to target different outcomes.

## **Conclusion**

The overall goal of this project was to investigate the role of writing beliefs, writing processes and drafting strategies in the development of subjective understanding during writing, under the perspective of the dual process model of writing (Galbraith, 1999; Galbraith, 2009; Galbraith & Baaijen, 2018). The project achieved this goal in several ways. Study 1 provided evidence to suggest that a group of writers that commonly finds academic writing challenging (students with dyslexia) had writing beliefs that were consistent with a lack of engagement with the knowledge constituting and knowledge transforming components of the dual process model, and more consistent with basic knowledge telling processes (Bereiter and Scardamalia, 1987). The implications of these results are that, theoretically, students with dyslexia will likely find developing their subjective understanding throughout the process of writing difficult. Study 2 indicated that some writers who were instructed to write synthetically-planned rough drafts with few rhetorical constraints showed evidence of producing their texts more spontaneously than participants who wrote under outline-planned, final drafting conditions (because of shorter linear pauses between sub-sentences, and a general lack of linear pauses in some writers at the between sub-sentence and sentence level). The results imply that these writers may have been engaging in knowledge constituting processes, rather than basic knowledge

telling processes. Study 3's results showed that participants who had higher scores on the sentence production measure from Baaijen and Galbraith (2018), and thus produced their sentences more spontaneously, had larger increases in subjective understanding than writers who produced more controlled sentences. This finding directly supports the assumptions of the knowledge constituting system within the dual process model, and demonstrates that subjective understanding development is likely to be at least partially due to spontaneous, rather than controlled, text production.

Methodologically, this PhD also produced important findings. Study 1 provided evidence for using ESEM as a technique to examine the factor structure of a questionnaire with theoretically related constructs over more traditional techniques such as CFA and EFA. Study 2 highlighted the importance of being methodologically transparent when reporting analytical techniques within a study so that those results can be reproduced, and the same techniques can be used by other researchers in other studies. Finally, Study 3 demonstrated that in order to properly examine subjective understanding development under the dual process account of writing, it may be important to use a multi-item scale which captures items relating to knowledge transformation and knowledge constitution separately.

This work is relevant because it challenges the traditional, and most frequently referred to models of writing in the field, the problem-solving accounts of writing. Contradictory to the problem-solving models' claims, the results from this project demonstrate that subjective understanding development appears not to be a by-product of learning to write more effectively, through *controlled* problem-solving processes, but instead involves active knowledge constitution, which is elicited by *spontaneous* text production. In doing so, this thesis strongly advocates that researchers underpin future work on subjective understanding development with the dual process explanation.

Yet, it is necessary to stress that much of the work presented in this thesis is novel, whether that is because it looks at novel concepts (e.g. planning beliefs) or because it uses advanced, relatively new methods for the field of writing research (e.g. ESEM, mixture modelling with experimental manipulation to drafting strategy, and multi-item scales of subjective understanding). As such, many of the findings within this project should only be considered an initial insight into the role that writing

beliefs, writing processes and drafting strategies have on subjective understanding development throughout writing.

Future research needs to continue exploring the impact that these three components have on subjective understanding, with a particular focus on academic writing. This is because, ultimately, academic writing is not just about creating a clear text product that informs another reader about a particular argument, theory or piece of research. It is also about giving the writer the opportunity to develop and organise their ideas in relation to a specific topic so that, throughout the process of creating that text, they can develop understanding.

### References

- Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., ... Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, 17(5), e3000246. <https://doi.org/10.1371/journal.pbio.3000246>
- Almond, R., Deane, P., Quinlan, T., Wagner, M., & Sydorenko, T. (2012). *A preliminary analysis of keystroke log data from a timed writing task* (Research Report No. RR-12-23). Educational Testing Service.
- 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>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397-438. <https://doi.org/10.1080/10705510903008204>
- Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances. *Journal of Management*, 41(6), 1561–1577. <https://doi.org/10.1177/0149206315591075>
- 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>
- 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](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 32(2), 139-161. [https://doi.org/10.1016/0167-8116\(95\)00038-0](https://doi.org/10.1016/0167-8116(95)00038-0)

- Beck, R. N. (1979). *Handbook in social philosophy*. Macmillan.
- Berchtold, A. (2010). Sequence Analysis and Transition Models. In M. D. Breed & J. Moore (Eds.), *Encyclopedia of Animal Behavior* (pp. 139–145). Academic Press. <https://doi.org/10.1016/B978-0-08-045337-8.00233-3>
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Routledge.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation. *Psychometrika*, 50, 229–242. <https://doi.org/10.1007/BF02294248>
- Breetvelt, I., van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between Writing Processes and Text Quality: When and How? *Cognition and Instruction*, 12(2), 103–123. <http://www.jstor.org/stable/3233677>
- Britton, J. (1982). Shaping at the point of utterance. In G. M. Pradl (Ed.), *Prospect and retrospect: Selected essays of James Britton* (pp.139–145). Montclair, NJ: Boynton/Cook.
- Britton, J., Burgess, T. , Martin, N. , McLeod, A. , & Rosen, H. (1975). *The development of writing abilities 11-18*. London: Macmillan.
- Brookings, J. B., & Bolton, B. (1988). Confirmatory factor analysis of the interpersonal support evaluation list. *American Journal of Community Psychology*, 16(1), 137-147. <https://doi.org/10.1007/BF00906076>
- Byrne, B. M. (2013). *Structural Equation Modeling with Mplus*. Routledge. <https://doi.org/10.4324/9780203807644>
- Chen, F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chenu, F., Pellegrino, F., Jisa, H., & Fayol, M. (2014). Interword and intraword pause threshold in writing. *Frontiers in Psychology*, 5, 182. <https://doi.org/10.3389/fpsyg.2014.00182>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*. Routledge. <https://doi.org/10.4324/9781315456539>

- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, 8(2), 240–247. [https://doi.org/10.1016/S0022-5371\(69\)80069-1](https://doi.org/10.1016/S0022-5371(69)80069-1)
- Conijn, R., Roeser, J., & van Zaanen, M. (2019). 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>
- Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2006). Contribution of Lower Order Skills to the Written Composition of College Students With and Without Dyslexia. *Developmental Neuropsychology*, 29(1), 175–196. [https://doi.org/10.1207/s15326942dn2901\\_9](https://doi.org/10.1207/s15326942dn2901_9)
- Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*, 26(8), 897–899. <https://doi.org/10.1038/nbt1406>
- Fecher, B., Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In: Bartling, S., Friesike, S. (eds) *Opening Science*. Springer, Cham. [https://doi.org/10.1007/978-3-319-00026-8\\_2](https://doi.org/10.1007/978-3-319-00026-8_2)
- Firestone, W. A. (1993). Alternative Arguments for Generalizing From Data as Applied to Qualitative Research. *Educational Researcher*, 22(4), 16–23. <https://doi.org/10.3102/0013189X022004016>
- Flower, L. (1994). *The construction of negotiated meaning: A social cognitive theory of writing*. Carbondale, IL: Southern Illinois University Press.
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365. <https://doi.org/10.2307/356600>
- Flower, L., & Hayes, J. R. (1984). Images, Plans, and Prose: The Representation of Meaning in Writing. *Written Communication*, 1(1), 120–160. <https://doi.org/10.1177/0741088384001001006>
- Foster, M. (1983). Writing and the Writer. *Journal of Education*, 165(2), 221–224. <https://doi.org/10.1177/002205748316500209>
- Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, 21, 45–71. <https://doi.org/10.1007/BF00119655>
- Galbraith, D. (1999). Writing as a knowledge-constituting process. In *Knowing what to write: Conceptual processes in text production* (pp. 139–160). Amsterdam University Press.
- Galbraith, D. (2009). Writing as discovery. *British Journal of Educational Psychology*, 2(6), 5–26. <https://doi.org/10.1348/978185409X421129>

- Galbraith, D. (2015). Conditions for writing to learn. *Journal of Writing Research*, 7(1), 215–226.  
<https://doi.org/10.17239/jowr-2015.07.01.09>
- 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., Smith-Spark, J., & Torrance, M. (2012). The effects of dyslexia on the writing processes of students in Higher Education. In L. Torrance, M. Alamargot, D. Castello, M. Ganier, F. Kruse, O. Mangen, A. Tolchinsky, L. van Waes (Ed.), *Learning to Write Effectively: Current Trends in European Research* (pp. 195–198). Brill Publications.
- 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(2), 113–145. <https://doi.org/10.1007/s10674-005-0119-2>
- Galbraith, D., Hallam, J., Le Bigot, N., & Olive, T. (2009). The role of different components of working memory in writing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31(31).  
<https://escholarship.org/uc/item/9z87v98t>
- Galbraith, D., Peters, A., Hall, S., & Baaijen, V.M. (2023) Measuring discovery through writing. In R. Horowitz (2<sup>nd</sup> Ed.), *The Routledge Handbook of International Research on Writing*.
- Galbraith, D., & Torrance, M. (2004). Revision in the Context of Different Drafting Strategies. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision Cognitive and Instructional Processes* (pp. 63–85). Springer Netherlands. [https://doi.org/10.1007/978-94-007-1048-1\\_5](https://doi.org/10.1007/978-94-007-1048-1_5)
- Galbraith, D., Torrance, M., & Hallam, J. (2006). Effects of writing on conceptual coherence. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 28, 1340-1345.
- Garcia, A. S., Morrison, K., Tsoi, A. C., & He, J. (2014). *Managing complex change in school: Engaging pedagogy, technology, learning and leadership*. Routledge. <https://doi.org/10.4324/9781315814186>
- Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, 56(2), 79–89.  
<https://doi.org/10.1080/00461520.2021.1898961>
- Given, L., (2008). *The SAGE Encyclopedia of Qualitative Research Methods*. SAGE Publications.  
<https://doi.org/10.4135/9781412963909.n467>

- Goertzen, M. J. (2017). Introduction to Quantitative Research and Data. *Library Technology Reports*, 53(4), 12–18.
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341), 341ps12. <https://doi.org/10.1126/scitranslmed.aaf5027>
- Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2008). How do secondary school students write poetry? How creative writing processes relate to final products. *L1-Educational Studies in Language and Literature*, 8(3), 57–80. <https://doi.org/10.17239/L1ESLL-2008.08.03.01>
- Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M., & Bennett, R. E. (2018). Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement*, 55(2), 194-216. <https://doi.org/10.1111/jedm.12172>
- Hall, S., Baaijen, V. M., & Galbraith, D. (2022). Constructing theoretically informed measures of pause duration in experimentally manipulated writing. *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10284-4>
- Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *The British Journal of Educational Psychology*, 72(1), 119–133.
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, 29(3), 369–388. <https://doi.org/10.1177/0741088312451260>
- Hayes, J. R., & Chenoweth, N. A. (2006). Is Working Memory Involved in the Transcribing and Editing of Texts? *Written Communication*, 23(2), 135–149. <https://doi.org/10.1177/0741088306286283>
- Hayes, J. R., Schriver, K., Hill, C., & Hatch, J. (1990). *Seeing problems with text: How students' engagement makes a difference*. Center for the Study of Writing.
- Higher Education Statistics Authority. (2020). Who's studying in HE? <https://www.hesa.ac.uk/Data-and-Analysis/Students/Whos-in-He>.
- Janssen, T., & Overmaat, M. (1990). *Tekstopbouw en stelvaardigheid. Een onderzoek naar de effecten van twee experimentele methoden voor tekstopbouw [Text structure and writing competence. A study on the effects of two experimental methods for text structure]*. Amsterdam: Swets & Zeitlinger.
- Keil, F. C. (2006). Explanation and Understanding. *Annual Review of Psychology*, 57, 227–254. <https://doi.org/10.1146/annurev.psych.57.102904.190100>

- 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. <https://doi.org/10.1037/0278-7393.14.2.355>
- Kellogg, R. T. (1990). Effectiveness of Prewriting Strategies as a Function of Task Demands. *The American Journal of Psychology*, 103(3), 327. <https://doi.org/10.2307/1423213>
- Kellogg, R. T. (1994). *The psychology of writing*. Oxford University Press.
- Kieft, M., Rijlaarsdam, G., Galbraith, D., & van den Bergh, H. (2011). The effects of adapting a writing course to students' writing strategies. *The British Journal of Educational Psychology*, 77, 565–578. <https://doi.org/10.1348/096317906X120231>
- Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2008). An aptitude-treatment interaction approach to writing-to-learn. *Learning and Instruction*, 18, 379–390. <https://doi.org/10.1016/j.learninstruc.2007.07.004>
- Kirk, J., Miller, M. L., & Miller, M. L. (1986). *Reliability and Validity in Qualitative Research*. SAGE.
- Klein, P. D., Haug, K. N., & Arcon, N. (2017). The effects of rhetorical and content subgoals on writing and learning. *Journal of Experimental Education*, 85(2), 291–308. <https://doi.org/10.1080/00220973.2016.1143795>
- Lavelle, E. (1993). Development and validation of an inventory to assess processes in college composition. *British Journal of Educational Psychology*, 63(3), 489–499. <https://doi.org/10.1111/j.2044-8279.1993.tb01073.x>
- Leggette, H., 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. (2013). Keystroke Logging in Writing Research. *Written Communication*, 30(3), 358–392. <https://doi.org/10.1177/0741088313491692>
- Marsh, H. W., & Hau, K. T. (1999). Confirmatory Factor Analysis: Strategies for Small Sample Sizes. *Statistical Strategies for Small Sample Research*, 1, 251–284.

- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471–491. <https://doi.org/10.1037/a0019227>
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. *Annual Review of Clinical Psychology*, 10(1), 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439–476. <https://doi.org/10.1080/10705510903008220>
- McClelland, J. L., & Rogers, T. T. (2003). The Parallel Distributed Processing Approach to Semantic Cognition. *Nature Reviews Neuroscience*, 4, 310–322. <https://doi.org/10.1038/nrn1076>
- McLachlan, G. J., & Peel, D. (2000). *Finite Mixture Models*. Wiley
- McNulty, M. A. (2003). Dyslexia and the Life Course. *Journal of Learning Disabilities*, 36(4), 363–381. <https://doi.org/10.1177/00222194030360040701>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- 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), 1–32. <https://doi.org/10.1016/J.JECP.2003.09.003>
- Mohammadzadeh, A., Ahour, T., & Saeidi, M. (2020). A Sociocultural Perspective on Second Language Writing: The Effect of Symmetrical versus Asymmetrical Scaffolding on Intermediate EFL Learners' Writing Accuracy, Fluency, and Complexity and Their Attitudes. *Education Research International*, <https://doi.org/10.1155/2020/5292356>
- Morin, A. J. S., & Maïano, C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport and Exercise*, 12(5), 540–554. <https://doi.org/10.1016/j.psychsport.2011.04.003>

- Mortimore, T., & Crozier, W. R. (2006). Dyslexia and difficulties with study skills in higher education. *Studies in Higher Education*, 31(2), 235–251. <https://doi.org/10.1080/03075070600572173>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1–9. <https://doi.org/10.1038/s41562-016-0021>
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171–189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, B., & Muthén, L. K. (2015). *Mplus User's Guide*. Muthén & Muthén.
- National Health Service. (2020). Dyslexia—NHS. <https://www.nhs.uk/conditions/dyslexia/>
- Nystrand, M. (2006). The Social and Historical Context for Writing Research. In *Handbook of writing research* (pp. 11–27). The Guilford Press.
- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>
- Peñuelas, A. B. C. (2012). The writing strategies of American university students: focusing on memory, compensation, social and affective strategies. *ELIA: Estudios de Lingüística Inglesa Aplicada*, 12(12)
- Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard Power as a Protection Against Imprecise Power Estimates. *Perspectives on Psychological Science*, 9, 319–332. <https://doi.org/10.1177/1745691614528519>
- Poldrack, R. A. (2019). The Costs of Reproducibility. *Neuron*, 101(1), 11–14. <https://doi.org/10.1016/j.neuron.2018.11.030>
- Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, 47(11), 1451–1458. <https://doi.org/10.1016/j.ijnurstu.2010.06.004>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 3, 354–373. <https://doi.org/10.1037/a0029315>

- Rijlaarsdam, G., van Weijen, D., & van den Bergh, H. (1994). Relations Between Writing Processes and Text Quality: When and How? *Cognition and Instruction*, 12(2), 103-123.  
[https://doi.org/10.1207/s1532690xc1202\\_2](https://doi.org/10.1207/s1532690xc1202_2)
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (2013). Schemata and Sequential Thought Processes in PDP Models. In *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*. Morgan Kaufmann. <https://doi.org/10.1016/B978-1-4832-1446-7.50020-0>
- Roeser, J., De Maeyer, S., Leijten, M., & Van Waes, L. (2021). Modelling typing disfluencies as finite mixture process. *Reading and Writing*. <https://doi.org/10.1007/s11145-021-10203-z>
- Rose, M. (1985). The Language of Exclusion: Writing Instruction at the University. *College English*, 47(4), 341–359. <https://doi.org/10.2307/376957>
- Rose, S. J. (2009). Identifying and Teaching Children and Young People with Dyslexia and Literacy Difficulties An independent report. Nottingham: DCFS Publications.
- Rozenblit, L., & Keil, F. (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](https://doi.org/10.1207/s15516709cog2605_1)
- 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>
- Schilperoord, J. (2001). On the cognitive status of pauses in discourse production. In T. Olive & M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp. 61–90). Kluwer.
- Schraw, G., & Bruning, R. (1996). Readers' Implicit Models of Reading. *Reading Research Quarterly*, 31(3), 290-305. <https://doi.org/10.1598/rrq.31.3.4>
- Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In *Handbook of research methods for studying daily life* (pp. 302–320). The Guilford Press.
- Silva, T., & Nicholls, J. G. (1993). College students as writing theorists: Goals and beliefs about the causes of success. *Contemporary Educational Psychology*, 18(3), 281–293. <https://doi.org/10.1006/ceps.1993.1021>
- Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science*, 9(1), 76–80. <https://doi.org/10.1177/1745691613514755>

- Sumner, E., & Connelly, V. (2020). Writing and Revision Strategies of Students With and Without Dyslexia. *Journal of Learning Disabilities*, 53(3), 189-198 002221941989909.  
<https://doi.org/10.1177/0022219419899090>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). New York: Harper and Row.  
<https://doi.org/10.1037/022267>
- Torrance, M., Thomas, G., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education*, 27, 379–392. <https://doi.org/10.1007/BF03179901>
- Tóth-Király, I., Bőthe, B., Rigó, A., & Orosz, G. (2017). An Illustration of the Exploratory Structural Equation Modeling (ESEM) Framework on the Passion Scale. *Frontiers in Psychology*, 8, 1968.  
<https://doi.org/10.3389/fpsyg.2017.01968>
- Trochim, W. M. K. (2020). Research Methods Knowledge Base: Positivism & Post-Positivism.  
<https://conjointly.com/kb/positivism-and-post-positivism/>
- Van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2021). Measuring and assessing typing skills in writing research. *Journal of Writing Research*, 13(1), 107–153.  
<https://doi.org/10.17239/jowr-2021.13.01.04>
- Vygotsky, L. S., & Cole, M. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.
- Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*. Wiley.  
<https://doi.org/10.1002/9781118356258>
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Wengelin, Å. (2007). The Word-Level Focus in Text Production in Adults with Reading and Writing Difficulties. In Mark. Torrance, L. van (Luuk) Waes, & D. Galbraith (Eds.), *Writing and Cognition* (pp. 67–82). Elsevier.

White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality.

*Contemporary Educational Psychology*, 30(2), 166–189.

<https://doi.org/10.1016/j.cedpsych.2004.07.002>

## Appendix A Writing Beliefs Inventory

### A.1 Writing Beliefs Online Questionnaire

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton

## Academic Writing Beliefs

**Researcher:** Sophie Hall

**ERGO number:** 47141

To help you decide whether you would like to take part in this survey study, it is important that you understand why the research is being done and what it will involve. Please read the information below carefully before deciding to take part. If you are happy to participate, you will need to indicate this by ticking the consent box at the end of this form.

### What is the research about?

This study forms part of a wider PhD project looking at how students with and without dyslexia write for academic purposes. I am a University of Southampton PhD student and I am conducting this survey to gain a better understanding of how university students' beliefs about writing affect the way they approach academic writing tasks. Combined with the findings of another study I am running, this project will give a deeper insight into the writing processes used by students with and without dyslexia. This will lead to greater knowledge about the similarities and differences between dyslexic and non-dyslexic writers. The results from this project, along with further research, will hopefully influence the types of writing support on offer for students at university.

This work is funded by the South Coast Doctoral Training Partnership through the Economic and Social Research Council (ESRC). ESRC funded projects aim to produce real world impact that can make a positive difference to people's lives. More information about the ESRC can be found at <https://esrc.ukri.org/>.

### Why have I been asked to participate?

I am interested in how writing beliefs concerning academic tasks vary across disciplines and years of study. Therefore, I am contacting current undergraduate, masters and PhD students from a range of disciplines.

### What will happen to me if I take part?

This questionnaire consists of two sections. The first section will ask you about your background, including whether you have a dyslexia diagnosis or not, which university you attend, the course you are studying and the grades that you have achieved in your studies so far. The second section will consist of a set of statements that students have made about academic writing, and will ask you to rate the extent to which you agree/disagree with them. None of the statements are necessarily "correct" beliefs - we all differ in how we approach writing, so please rate them as honestly as possible in terms of how well they correspond with your personal beliefs.

The questionnaire should take approximately 15 minutes to complete. No further involvement is required.

### Are there any benefits in my taking part?

By taking part in this study, you will be adding to the current academic knowledge base we have about students' writing beliefs. In the future, the evidence obtained from this questionnaire, along with other writing studies, could lead to developments in the academic provision for students in higher education in general, as well as to the development of writing support for students with dyslexia.

At the end of this survey you will be asked if you are interested in taking part in my second study which focuses on the writing processes used by dyslexic and non-dyslexic students. This separate study will offer you the opportunity to be entered into a prize draw, giving you the chance to win one of five £100 Amazon Vouchers.

### What data will be collected?

All of your questionnaire responses will be anonymous and stored in password protected, encrypted files on the University of Southampton's secure server.

If you wish to be contacted about the second study mentioned in the above section, your name and email address will be recorded, but these will be stored separately from your questionnaire responses on a secure database only accessible to the main researcher. You must be a University of Southampton student to take part in the second study.

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton

**Are there any risks involved?**

There are no special risks involved in participating in this survey.

**Will my participation be confidential?**

This study complies with GDPR, the Data Protection Act and the University of Southampton ethics policy, and your participation will be completely confidential. Your responses will be stored anonymously in an encrypted, password protected file on the University of Southampton's secure server, separate from your e-mail address (which you may supply if you want to be involved in the second study).

On completion of my PhD, the anonymous questionnaire data from this study will be made available to other researchers in line with the University of Southampton and the ESRC's research data management policies. Your participation will still remain anonymous and confidential.

**Do I have to take part?**

No, it is entirely up to you to decide whether or not to take part. If you decide you want to take part, you will need to confirm this by ticking the consent box at the end of this form.

**What happens if I change my mind?**

You have the right to change your mind and withdraw at any time whilst filling in the questionnaire without giving a reason and without your participant rights being affected. However, please be aware that once you have submitted the questionnaire withdrawal will no longer be possible. This is because your submitted data will be anonymised and therefore difficult to trace back to you.

**What will happen to the results of the research?**

The results from this survey will be written up as part of my PhD thesis. The research findings made available in any reports or publications will not include information that can identify you.

**Where can I get more information?**

If you would like more information about the study, you can contact Sophie Hall ([s.m.hall@soton.ac.uk](mailto:s.m.hall@soton.ac.uk)) at the University of Southampton.

**What happens if there is a problem?**

If you have a concern about any aspect of this study, you should speak to the researcher, Sophie Hall ([s.m.hall@soton.ac.uk](mailto:s.m.hall@soton.ac.uk)). She will do her best to answer your questions.

If you remain unhappy or have a complaint about any aspect of this study, please contact the University of Southampton Research Integrity and Governance Manager (023 8059 5058, [rgoinfo@soton.ac.uk](mailto:rgoinfo@soton.ac.uk)).

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton



## Section 1. Section 1. Background Information

### Question 1.1

Please enter your date of birth.

1 / January / 2019

### Question 1.2

Please select your gender.

- ☐ Female
- ☐ Male
- ☐ Prefer not to say

### Question 1.3

Do you have a dyslexia diagnosis?

- ☐ Yes
- ☐ No

### Question 1.4

Please enter the title of the degree you are taking (e.g. BSc Education and Psychology; MSc Statistics, PhD Physics...).

### Question 1.5

Please select the year of your degree that you are in.

- ☐ First year
- ☐ Second year

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton

- ☐ Third year  
☐ Fourth year  
☐ Fifth or more year

## Question 1.6

Would you say that English is your first language?

- ☐ Yes  
☐ No

## Question 1.6b

Please state your first language.

## Question 1.6c

Please give the results of any English proficiency test you have taken.

## Question 1.6d

Using the scale below, please make a judgement on your written English proficiency.

Very bad 1                      2                      3                      4                      5                      Very good

☐                      ☐                      ☐                      ☐                      ☐

## Question 1.7

Please estimate what proportion of your **COURSEWORK** involves writing essays or "essay-like" text. This includes anything involving full text production (e.g. reports, essays, reviews etc.) but not assessments like multiple choice tests, oral presentations or statistical tests.

None    20%    30%    40%    50%    60%    70%    80%    90%    All

☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐

## Question 1.8

Please estimate what proportion of your **EXAMS** involve writing essays or "essay-like" text. This includes anything involving full text production (e.g. reports, essays, reviews etc.) but not assessments like multiple choice tests, oral presentations or statistical tests.

None    20%    30%    40%    50%    60%    70%    80%    90%    All

☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐

## Question 1.9

Please select the mark below that is closest to your average **COURSEWORK** mark in your current year of study.

0    30    38    42    45    48    52    55    58    62    65    68    72    78    87    100    N/A

☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐

## Question 1.10

Please select the mark below that is closest to your average **EXAM** mark in your current year of study.

0    30    38    42    45    48    52    55    58    62    65    68    72    78    87    100    N/A

☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐    ☐

## Question 1.11

Please select your nationality.

- |   |  |
|---|--|
| a) White  | d) Black/African/Caribbean/Black British |
| <input type="radio"/> Welsh/English/Scottish/Northern Irish/British | <input type="radio"/> African            |
|   | <input type="radio"/> Caribbean          |

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton

- ☐ Irish  
☐ Gypsy or Irish Traveller  
☐ Any other White background  
**b) Mixed/Multiple ethnic groups**  
☐ White and Black Caribbean  
☐ White and Black African  
☐ White and Asian  
☐ Any other Mixed/Multiple ethnic background  
**c) Asian/Asian British**  
☐ Indian  
☐ Pakistani  
☐ Bangladeshi  
☐ Chinese  
☐ Any other Asian background  
☐ Any other Black/African/Caribbean background  
**e) Other ethnic group**  
☐ Arab  
☐ Any other ethnic group, please describe  
☐ Do not state

Question 1.12

What university do you attend?

## Section 2. Section 2. Writing Beliefs

This questionnaire has 43 statements about academic writing. For each statement, please check whether you Strongly Agree, Agree, Don't Know, Disagree or Strongly Disagree with the statement. There is no right way of writing. It depends on what you believe about academic writing.

Remember your answers are **CONFIDENTIAL**; try to answer as honestly as possible.

Question 2.1

Writing Belief Statements

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Good writers include a lot of quotes from authorities in their writing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When writing, it's best to use proven formats and templates, and then fill in the important information.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The key to successful writing is to stick to one's plan.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good writers are reader-friendly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A primary goal of writing should be to have to make as few changes as possible.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good writing requires making a detailed outline before writing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important to develop a distinctive writing style.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good writing involves developing ideas over a series of drafts.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good writers anticipate and answer their audience's questions.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A good writer makes sure they know what they think before they start to write.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The key to good writing is revising.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important for writers to get their ideas straight before they start to write.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The most important reason to write is to report	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton

what authorities think about a subject.

Writing should involve a personal interpretation of content.

☐ ☐ ☐ ☐ ☐

Writing requires going back over it to improve what has been written.

☐ ☐ ☐ ☐ ☐

It is important to thoroughly revise initial drafts of text.

☐ ☐ ☐ ☐ ☐

Thorough planning is the most important aspect of writing.

☐ ☐ ☐ ☐ ☐

My thoughts and ideas become clearer to me as I write and rewrite.

☐ ☐ ☐ ☐ ☐

Writing is often an emotional experience.

☐ ☐ ☐ ☐ ☐

Writing is a process involving a lot of emotion.

☐ ☐ ☐ ☐ ☐

Good writing involves expressing a distinctive point of view.

☐ ☐ ☐ ☐ ☐

Writing's main purpose is to share the information in sources accurately.

☐ ☐ ☐ ☐ ☐

It's important to select the words that suit your purpose, audience, and occasion.

☐ ☐ ☐ ☐ ☐

The key to successful writing is making a well-organised plan.

☐ ☐ ☐ ☐ ☐

Writing should focus on the information in books and articles.

☐ ☐ ☐ ☐ ☐

Writers need to immerse themselves in their writing.

☐ ☐ ☐ ☐ ☐

Writing helps me understand better what I'm thinking about.

☐ ☐ ☐ ☐ ☐

Writing helps me see the complexity of ideas.

☐ ☐ ☐ ☐ ☐

Writing's main purpose is to give other people information.

☐ ☐ ☐ ☐ ☐

Good writing involves getting each sentence right before moving on to the next.

☐ ☐ ☐ ☐ ☐

Revising involves reorganising the structure of the text.

☐ ☐ ☐ ☐ ☐

The key to successful writing is accurately reporting what authorities think.

☐ ☐ ☐ ☐ ☐

Good writers adapt their message to their readers.

☐ ☐ ☐ ☐ ☐

Good writers try to be objective.

☐ ☐ ☐ ☐ ☐

Revision is a multi-stage process.

☐ ☐ ☐ ☐ ☐

Writing is primarily about transmitting information.

☐ ☐ ☐ ☐ ☐

Good writers keep their audience in mind.

☐ ☐ ☐ ☐ ☐

The key to good writing is conveying information clearly.

☐ ☐ ☐ ☐ ☐

It's important to keep your overall purpose in mind while writing.

☐ ☐ ☐ ☐ ☐

Writing is a process of reviewing, revising, and rethinking.

☐ ☐ ☐ ☐ ☐

Writing helps new ideas emerge.

☐ ☐ ☐ ☐ ☐

Good writing involves editing many times.

☐ ☐ ☐ ☐ ☐

Good writers thoroughly explain their opinions and findings.

☐ ☐ ☐ ☐ ☐

24/10/2022, 11:51

iSurvey - Online Questionnaire Generation from the University of Southampton

**Section 3. Section 3. Optional contact details to take part in second writing study**

If you would like to be contacted about the opportunity to take part in another writing study in which you will get the chance to win one of five £100 Amazon gift cards, please enter your name and e-mail address below. These will be stored separately from your responses to the questionnaire, and will only be used to contact you regarding the second study. Your responses to the questionnaire will be completely confidential. If you do not want to give your contact details, please fill the boxes in with "NA" .

Question 3.1

Name (Forename and surname)

Question 3.2

E-mail address

Thank you for completing this questionnaire. Your data will help improve our understanding of the beliefs that dyslexic and non-dyslexic students hold about academic writing. Write ups of this study will not include your name or any other identifying characteristics. The research did not use deception.

If you have any further questions please contact me, Sophie Hall, at S.M.Hall@soton.ac.uk.

## A.2 Writing Beliefs Items and Respective Hypothesised Factors

Question numbering in analysis	Question numbering in online survey	Question 1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, 5 = neutral	Hypothesised factor
1	18	Good writing involves expressing a distinctive point of view.	TA
2	32	A good writer makes sure they know what they think before they start to write.	P
3	29	The key to successful writing is to stick to one's plan.	P
4	41	Good writers anticipate and answer their audience's questions.	A
5	24	The key to good writing is revising.	R
6	40	It's important to keep your overall purpose in mind while writing.	A
7	12	Writing is a process involving a lot of emotion.	TA
8	5	Writing is primarily about transmitting information.	TM
9	10	Writers need to immerse themselves in their writing.	TA
10	11	Writing helps me understand better what I'm thinking about.	TA
11	31	Good writing requires making a detailed outline before writing.	P
12	14	Writing is often an emotional experience.	TA
13	4	The key to successful writing is accurately reporting what authorities think.	TM
14	38	Good writers thoroughly explain their opinions and findings.	A
15	33	Good writing involves getting each sentence right before moving on to the next.	P

16	7	Writing's main purpose is to share the information in sources accurately.	TM
17	13	Writing helps me see the complexity of ideas.	TA
18	22	Revising involves reorganising the structure of the text.	R
19	9	My thoughts and ideas become clearer to me as I write and rewrite.	TA
20	27	A primary goal of writing should be to have to make as few changes as possible.	R
		(REVERSE CODED)	
21	20	Good writing involves editing many times.	R
22	21	Writing is a process of reviewing, revising, and rethinking.	R
23	23	Revision is a multi-stage process.	R
24	28	It is important for writers to get their ideas straight before they start to write.	P
25	2	The most important reason to write is to report what authorities think about a subject.	TM
26	8	Good writers try to be objective.	TM
27	6	Writing's main purpose it to give other people information.	TM
28	35	When writing, it's best to use proven formats and templates, and then fill in the important information.	P
29	25	Good writing involves developing ideas over a series of drafts.	R
30	16	It is important to develop a distinctive writing style.	TA
31	30	The key to successful writing is making a well-organised plan.	P
32	1	Good writers include a lot of quotes from authorities in their writing.	TM

33	19	Writing requires going back over it to improve what has been written.	R
34	37	The key to good writing is conveying information clearly.	A
35	26	It is important to thoroughly revise initial drafts of text.	R
36	17	Writing should involve a personal interpretation of content.	TA
37	3	Writing should focus on the information in books and articles.	TM
38	15	Writing helps new ideas emerge.	TA
39	43	Good writers keep their audience in mind.	A
40	34	Thorough planning is the most important aspect of writing.	P
41	36	Good writers adapt their message to their readers.	A
42	39	Good writers are reader-friendly.	A
43	42	It's important to select the words that suit your purpose, audience, and occasion.	A

*Note. TA = transactional; TM = transmissional; P = planning; R = revision; A = audience*

## Appendix B Understanding Scale

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

Knowledge about the topic.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How well you understand the topic.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How organised your thoughts about the topic are.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How well you could explain the topic.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How clear your thoughts about the topic are.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How clear your interpretation of the topic is.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How coherent your thoughts about the topic are.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How structured your thoughts about the topic are.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How much you can make sense of the topic's issues.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How clear the relationships between your ideas about the topic are.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How well you comprehend the topic's issues.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

How well-ordered your thoughts about the topic are.

Very little 1	2	3	4	5	6	A great deal 7
------------------	---	---	---	---	---	----------------------

## Appendix C – First and Second Ideas Lists

## POINTS 1

Instructions.

I want you to write down a list of all the ideas you can think of relevant to the question: should we all be vegan?/ Does social media do more harm than good?

Don't worry about how well expressed these are, or whether anyone else could understand what you mean, just jot down a sentence, word or phrase representing the idea for yourself. A point can be as short as a single word but no longer than sentence. If a point has several different aspects which you can't write as a single sentence then put these down as separate points.

You will have ten minutes to do this in. The aim is to write down as many relevant ideas as you can think of in the time available.

List of points.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.
- 11.

## POINTS 2

Instructions.

I want you to write down a list of all the ideas you can think of **NOW** relevant to the question of whether social media does more harm than good/should we all become vegans?

Don't worry about how well expressed these are, or whether anyone else could understand what you mean, just jot down a sentence, word or phrase representing the idea for yourself. A point can be as short as a single word but no longer than sentence. If a point has several different aspects which you can't write as a single sentence then put these down as separate points.

You will have ten minutes to do this in. The aim is to write down as many relevant ideas as you can think of in the time available.

List of points.

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.
- 11.

## Appendix D – First and Second Ratings of Importance

## RATINGS OF IMPORTANCE 1 AND 2

Read through the list of points you have just made and rate how important each idea is on the form below. Use the following scale (**note: 5 = major ideas, 1 = minor ideas**)

1\_\_\_\_\_2\_\_\_\_\_3\_\_\_\_\_4\_\_\_\_\_5  
**Minor** **Major**  
 points points

**Point number.**

1.	11.	21.
2.	12.	22.
3.	13.	23.
4.	14.	24.
5.	15.	25.
6.	16.	26.
7.	17.	27.
8.	18.	28.
9.	19.	29.
10.	20.	30.

## Appendix E – Degree of Correspondence

**DEGREE OF CORRESPONDENCE****Instructions.**

I want you to compare the ideas in the lists you produced before and after writing.

The numbers of the points in list 2 are written in the left hand column.

I want you to take each of the points in this list in turn, and read through list 1 to see if there are any corresponding point(s). If there are, write the number(s) of the corresponding point (s) in the second column.

If there are any corresponding points(s) please rate how similar they are to the point in list 2, using the scale below. Put this rating in the third column headed "degree of correspondence".

If there is more than one point in list 1 corresponding to a point in list 2, then rate the similarity of the group as a whole to that point (i.e. give it a single rating). If there are no corresponding points in list 1 then leave the second column blank and put a rating of 6 in the third column.

\_\_\_\_1\_\_\_\_  
An identical  
point

\_\_\_\_2\_\_\_\_

\_\_\_\_3\_\_\_\_

\_\_\_\_4\_\_\_\_

\_\_\_\_5\_\_\_\_

\_\_\_\_6\_\_\_\_  
No corresponding  
point

Number of point in list 2	Corresponding point(s) in list 1	Degree of correspondence
1		
2		
3		
4		
5		
6		

7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		

23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		

## Appendix F – Writing Task Instructions

**Outline condition**

**"Should we all become vegans?" / "Does social media do more harm than good?"**

In a moment I will want you to write a complete draft of an essay discussing this question. I want you to consider arguments for and against veganism/social media and to come to a conclusion about the question.

By 'complete draft', I mean that your aim is to produce a well-organised essay. This should be expressed in continuous prose (i.e. in complete, and connected sentences, rather than in note-form) and spelling should be accurate.

Before doing this, I am going to give you five minutes in which I want you to work out an outline for the essay, indicating your opinion, what the main ideas are, and the order they are to go in. So, by the end of the 5 minutes I want you to have made an organised outline of the essay.

**When the 5 minutes is up, allow them to finish writing whatever they are currently doing. They can keep the outlines.**

Now I want you to write the complete draft. This should be expressed in continuous prose (i.e. in complete, and connected sentences, rather than in note-form) and spelling should be accurate.

You have half an hour in which to do this; should you finish before the half hour is up, please write down the time at which you finish.

**Ask if they understand what they have to do, then tell them they can start, and note the time at which they do so.**

**Rough draft**

**"Should we all become vegans?" / "Does social media do more harm than good?"**

In a moment I will want you to write a rough draft of an essay discussing this question. I want you to consider arguments for and against veganism/social media and to come to a conclusion about the question.

By 'rough draft', I mean that your aim is to discuss the question with yourself, writing down your thoughts as they occur to you, and forming a conclusion. Don't worry about how well organised the text is, just concentrate on directly expressing your thoughts as they unfold. Your thoughts should be expressed in continuous prose (i.e. in complete, and connected, sentences rather than in note-form) but don't worry about spelling.

Before doing this, I am going to give you five minutes to work out what you think about the question. By the end of the 5 minutes I want you to have written down a single sentence summing up your opinion about your main ideas.

**When the 5 minutes is up, allow them to finish writing whatever they are currently doing. Allow them to keep the 'sentences' in front of them.**

Now I want you to write the rough draft of the essay. Remember that your aim is to discuss the question with yourself, writing down your thoughts as they occur to you, and forming a conclusion. Don't worry about how well organised the text is, just concentrate on directly expressing your thoughts as they unfold. This should be expressed in continuous prose (i.e. in complete, and connected sentences, rather than in note-form) but don't worry about spelling.

You have half an hour in which to do this; should you finish before the half hour is up, please write down the time at which you finish.

**Ask if they understand what they have to do, then tell them they can start, and note the time at which they do so.**

## Appendix G – Study 1 Ethical Approval

Approved by Faculty Ethics Committee - ERGO II 47141

UNIVERSITY OF  
Southampton

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

Submission ID: 47141

Submission Title: The academic writing beliefs of students with and without dyslexia

Submitter Name: Sophie Hall

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.

Comments:

- 
- The amendments requested following the original ethics submission have been completed satisfactorily in the revised version. Good luck with the research.

[Click here to view the submission](#)

## Appendix H – Study 2/3 Ethical Approval

Approved by Faculty Ethics Committee - ERGO II 46126.A1

UNIVERSITY OF  
Southampton

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

Submission ID: 46126.A1

Submission Title: The effects of dyslexia on writing (Amendment 1)

Submitter Name: Sophie Hall

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.

Comments:

- Dear Sophie

Thanks for a detailed and comprehensive application. I am happy to approve this, but just a few pointers. You are correct to point out that traditional poster advertisement is not going to happen, so you will need to think about recruitment strategies via social media which enabling services may be able to help with. This may be important if the pool of students from the psychology bank is not sufficient. It would also be a good idea reiterating participants' right to withdraw on the PIS.

[Click here to view the submission](#)

---

## Appendix I – Data Management Plan

<b>About your Research</b>			
PhD title:	The academic writing processes of students with and without dyslexia		
Student name:	Sophie Hall		
Supervisor(s):	David Galbraith James Hall		
Ethics No. (if appropriate)	Study 1: 46126 Study 2: 47141		
<b>About this plan</b>			
Date of plan:	29/04/2019	Frequency of reviews	3m
Date of next review:	29/06/2019 (and three monthly from thereon)		
Agreed actions to help you implement the plan	<i>Unique ID database for anonymization of data.</i>		
Agreed equipment and/or resources required:	<i>Locked drawers in office for storing paper-based data.</i>  <i>Electronic data saved in password protected files on Southampton University's secure server. Accessible on a password protected University of Southampton computer.</i>		
Further information (as appropriate):			
<b>Version Table</b>			
Version	Changes made	Date	
1		29/04/2019	
2	No changes made, but because all data collection is now taking place online, there are no paper-based documents. Therefore, all data collection and storage procedures now follow the electronic data guidelines	06/10/2020	

**1. Project Description:**

This is a two-study project investigating the writing processes of students with and without dyslexia. Questionnaire and experimental techniques will be used to obtain data about student's writing beliefs, literacy and cognitive skills, and writing composition to find out about how dyslexic and non-dyslexic student groups differ in the ways they write for academic tasks. This information will hopefully be used to inform dyslexia education policy and practice in order to better equip dyslexic students with the skills for writing well at university.

**2. What policies will apply to your research?**

University of Southampton's privacy policy - anonymization and confidentiality  
 University of Southampton's data management policies  
 University of Southampton's ethics policy  
 GDPR

**3. What data/research material will you collect or create?**

**Study 1 (writing beliefs questionnaire):** Online questionnaire data collected on the University of Southampton's iSurvey website. This will include information about students' demographics, average grades and writing beliefs. At the end of the questionnaire, students also have the opportunity to share contact details if they want to be contacted about getting involved in my second study. These contact details will be stored separately from the rest of the questionnaire data for anonymization purposes.

**Study 2 (writing experiment):** Two-session experimental study involving the collection of computer and paper based data. Computer based data includes:

- Working memory experiment scores
- Baseline typing speed (copy task)
- Final essays
- Keystroke logs taken during essay writing task

Paper based data includes:

- DAST measures (spellings and task scores)
- Writing belief questionnaire
- Idea lists
- Understanding scale
- Outline/synthetic plans
- Idea comparisons

#### **4. How will your data/research material be documented and described?**

All paper based documentation will be filed and stored in a locked cabinet in my PGR office. Some information (i.e. DAST scores, number of generated ideas and understanding scale rankings) will be transcribed into an excel database which can be converted to an SPSS file, csv file etc. for further analysis.

The excel database file will also contain demographic information about participants, scores from the working memory tasks, participants' baseline typing speed and essay quality scores. This excel file is password protected and saved in a research data folder on the University's secure server. Backups are also saved on the secure server and on my university Office 365 OneDrive account. Other electronic data includes the idfx keystroke logging files from the essay writing tasks and the MS Word documents containing the finished essays. These are also saved and backed up in password protected files on the university's secure server and on my OneDrive account.

#### **5. How will you deal with any ethical and copyright issues?**

Participant confidentiality and anonymity are the main ethical issues posed by my project. For the questionnaire data, all responses are anonymous and therefore cannot be traced back to the respondent. The only case in which this is possible is if participants leave their contact details at the end of the questionnaire in order to be contacted about taking part in my second study. This information will be initially saved alongside the questionnaire responses on the iSurvey database, but will only be accessible to me. To ensure participant anonymity beyond me, I will manually create two new data files - one for storing contact details and one for storing questionnaire responses. When each participant's original questionnaire responses are transcribed into the two new files, I will delete their iSurvey questionnaire. The two new files will not be able to be linked and therefore the questionnaire responses will have been anonymised.

In terms of confidentiality, the contact details obtained from the questionnaires will only be viewed by me, the researcher. I will not share these details with anybody else and therefore all participants' involvement in the study will remain confidential.

For the experimental study, no study documentation will be named. Instead, participants will be given a unique identifying number at the beginning of the study which will be used to link all of their data. I have created a database which matches participants with their identifying number. Once all study documentation is, their name will be removed from the unique ID database so that only their ID number remains. This way, the data collected in the experiment will not be traceable back to the participant.

In terms of confidentiality, only I will know who has taken part in the study. No documentation or write ups of the study will contain any information that could identify participants. Therefore, all participant's involvement will remain confidential.

#### **6. How will your data/research materials be stored, and backed up?**

**Electronic data** - Password protected files in a password protected folder call 'DATA' which is stored and backed up on the University of Southampton's secure server. A second back up of all data is stored on my passwords protected university OneDrive account. Secure passwords have been used to protect the data. These procedures are in line with the University of Southampton's ethics and privacy policies.

**Paper-based data** - These data are filed and stored in a locked filing cabinet in my research office. Only I have access to this filing cabinet. I ensure that it is locked at all times other than when I need access to the data. Electronic copies of all paper-based data will be made and stored in the same way as the rest of the electronic data. Once the electronic copies have been created, the original paper versions will be destroyed.

## **7. What are your plans for the long-term preservation of data/research materials supporting your research?**

The data from this study will be kept by the University of Southampton for a minimum of 10 years. To meet ESRC guidelines, the excel files containing scores and averages for each participant will also be made available to other researchers to increase the replicability and transparency of the project. My supervisor and I will also keep copies of the data from this project, again for any potential future analysis. Metadata will be created for all excel files, most importantly the keystroke logging files which may be hard to interpret without this additional information.

## **8. What are your plans for sharing the data/research materials after the submission of your thesis?**

To meet ESRC expectations, the anonymised excel files containing scores and averages for each participant on each measure will be made openly available for other researchers to conduct analysis with. My supervisor and I will also keep copies of all data, including the essays, keystroke logging files and metadata. The essays and keystroke files will be made accessible to other researchers on request.

The University of Southampton Library has developed this Doctoral Research Data Management Plan and guidance notes based on material adapted from the Australian National Data Service, Sheffield Hallam University, the Open

University and the universities of Bath and Newcastle.



## Appendix J – Additional Results Study 1

**Table 20**

*Factor Mean Differences using the Scalar Model when the No Dyslexia Group is Constrained to Zero*

Dyslexia				
Factor	Estimate	S.E	Est/S.E.	two-tailed p-value
TM	-0.477	0.17	-2.808	0.005
TA	0.779	0.221	3.529	0
R	0.122	0.152	0.8	0.424
P	-0.321	0.15	-2.145	0.032
A	0.198	0.19	1.041	0.298

*Note. 1 is 'strongly agree', 5 is 'strongly disagree'*

**Table 21***39-item WBI 5-factor ESEM Loadings for Students Without Dyslexia*

Item numbers	Estimate	S.E.	Est./S.E.	Two-tailed P-value
Transmissional by				
Q8	0.273	0.307	0.892	0.373
Q13	0.608	0.262	2.319	0.02
Q16	0.292	0.316	0.926	0.354
Q26	0.478	0.239	1.996	0.046
Q32	0.081	0.229	0.354	0.723
Q37	0.251	0.138	1.823	0.068
Q1	-0.164	0.335	-0.49	0.624
Q2	0.236	0.177	1.331	0.183
Q3	0.459	0.209	2.195	0.028
Q4	0.069	0.207	0.332	0.74
Q5	0.305	0.164	1.864	0.062
Q6	0.033	0.202	0.161	0.872
Q7	-0.148	0.221	-0.671	0.502
Q9	-0.055	0.352	-0.156	0.876
Q10	-0.082	0.198	-0.416	0.677
Q11	0.547	0.234	2.343	0.019
Q14	0.146	0.266	0.549	0.583
Q15	0.394	0.279	1.412	0.158
Q17	0.036	0.178	0.201	0.841
Q18	-0.038	0.248	-0.152	0.879
Q19	0.06	0.252	0.24	0.81
Q20	-0.041	0.285	-0.145	0.885
Q21	0.029	0.161	0.182	0.856
Q22	-0.012	0.135	-0.087	0.931
Q23	0.004	0.291	0.014	0.988
Q28	0.58	0.296	1.959	0.05
Q29	-0.146	0.324	-0.45	0.653
Q30	0.145	0.196	0.74	0.459
Q31	0.151	0.138	1.091	0.275
Q33	0.156	0.174	0.897	0.369
Q34	0.052	0.118	0.439	0.661
Q35	0.522	0.411	1.27	0.204
Q36	0.029	0.27	0.108	0.914
Q38	-0.054	0.114	-0.474	0.635
Q39	-0.057	0.181	-0.316	0.752
Q40	0.302	0.157	1.919	0.055
Q41	-0.104	0.179	-0.577	0.564
Q42	0.108	0.21	0.515	0.606
Q43	-0.099	0.235	-0.422	0.673
Transactional by				
Q1	0.211	0.177	1.194	0.233

Q7	0.394	0.264	1.496	0.135
Q9	0.382	0.229	1.669	0.095
Q10	0.715	0.141	5.075	0
Q17	0.765	0.139	5.483	0
Q19	0.531	0.283	1.878	0.06
Q30	0.298	0.224	1.33	0.184
Q36	0.054	0.113	0.481	0.63
Q38	0.638	0.131	4.879	0
Q2	0.053	0.108	0.49	0.624
Q3	0.083	0.143	0.579	0.563
Q4	0.172	0.13	1.324	0.185
Q5	0.235	0.13	1.805	0.071
Q6	-0.071	0.077	-0.93	0.352
Q8	-0.18	0.09	-1.999	0.046
Q11	0.247	0.13	1.895	0.058
Q13	0.1	0.154	0.649	0.516
Q14	-0.049	0.136	-0.359	0.719
Q15	-0.083	0.153	-0.543	0.587
Q16	-0.144	0.108	-1.334	0.182
Q18	0.073	0.128	0.567	0.57
Q20	-0.2	0.152	-1.315	0.189
Q21	0.244	0.141	1.727	0.084
Q22	0.19	0.137	1.387	0.165
Q23	0.086	0.109	0.791	0.429
Q26	0.004	0.128	0.031	0.975
Q28	-0.228	0.147	-1.553	0.121
Q29	0.371	0.176	2.109	0.035
Q31	-0.084	0.086	-0.975	0.33
Q32	0.349	0.156	2.247	0.025
Q33	0.189	0.098	1.931	0.053
Q34	-0.006	0.062	-0.09	0.928
Q35	0.15	0.207	0.727	0.467
Q37	0.133	0.122	1.09	0.276
Q39	0.149	0.095	1.573	0.116
Q40	-0.011	0.107	-0.099	0.921
Q41	0.009	0.103	0.089	0.929
Q42	0.037	0.115	0.317	0.751
Q43	0.078	0.066	1.178	0.239
Revision by				
Q5	-0.005	0.177	-0.026	0.979
Q18	-0.056	0.307	-0.182	0.856
Q20	0.58	0.284	2.043	0.041
Q21	0.211	0.21	1.006	0.315
Q22	0.446	0.235	1.897	0.058
Q23	0.173	0.332	0.52	0.603
Q29	0.281	0.27	1.039	0.299
Q33	0.308	0.097	3.169	0.002
Q35	0.085	0.179	0.472	0.637

Q1	-0.017	0.264	-0.065	0.948
Q2	-0.063	0.129	-0.491	0.623
Q3	-0.013	0.265	-0.05	0.96
Q4	-0.013	0.147	-0.092	0.927
Q6	0.234	0.133	1.762	0.078
Q7	-0.235	0.192	-1.223	0.221
Q8	0.069	0.124	0.557	0.578
Q9	-0.082	0.19	-0.434	0.664
Q10	0.149	0.165	0.906	0.365
Q11	0.096	0.224	0.429	0.668
Q13	-0.022	0.194	-0.112	0.911
Q14	-0.043	0.121	-0.354	0.723
Q15	-0.517	0.267	-1.935	0.053
Q16	0.048	0.104	0.458	0.647
Q17	0.044	0.144	0.304	0.761
Q19	0.477	0.291	1.64	0.101
Q26	-0.133	0.154	-0.861	0.389
Q28	-0.218	0.201	-1.082	0.279
Q30	-0.52	0.143	-3.624	0
Q31	0.181	0.203	0.888	0.374
Q32	-0.431	0.179	-2.409	0.016
Q34	0.224	0.155	1.443	0.149
Q36	-0.082	0.2	-0.412	0.68
Q37	-0.042	0.155	-0.274	0.784
Q38	0.146	0.1	1.471	0.141
Q39	0.068	0.1	0.675	0.5
Q40	0.098	0.189	0.518	0.604
Q41	-0.109	0.094	-1.157	0.247
Q42	-0.014	0.158	-0.088	0.93
Q43	0.034	0.147	0.228	0.82
Planning by				
Q2	0.019	0.156	0.122	0.903
Q3	0.458	0.215	2.127	0.033
Q11	0.271	0.269	1.005	0.315
Q15	0.307	0.253	1.214	0.225
Q28	0.159	0.223	0.713	0.476
Q31	0.394	0.155	2.534	0.011
Q40	0.483	0.208	2.324	0.02
Q1	0.297	0.363	0.819	0.413
Q4	-0.023	0.171	-0.132	0.895
Q5	-0.009	0.192	-0.045	0.964
Q6	0.209	0.186	1.124	0.261
Q7	0.304	0.245	1.24	0.215
Q8	0.164	0.231	0.712	0.476
Q9	0.203	0.336	0.604	0.546
Q10	-0.205	0.173	-1.183	0.237
Q13	0.144	0.215	0.672	0.502
Q14	-0.158	0.285	-0.553	0.58

Q16	0.129	0.257	0.5	0.617
Q17	-0.332	0.12	-2.764	0.006
Q18	0.319	0.246	1.299	0.194
Q19	-0.001	0.361	-0.003	0.998
Q20	-0.222	0.293	-0.757	0.449
Q21	0.131	0.227	0.576	0.565
Q22	0.192	0.186	1.033	0.302
Q23	0.428	0.357	1.199	0.231
Q26	-0.006	0.171	-0.033	0.973
Q29	0.16	0.209	0.763	0.445
Q30	0.101	0.318	0.316	0.752
Q32	0.209	0.24	0.871	0.384
Q33	0.13	0.186	0.698	0.485
Q34	0.209	0.138	1.518	0.129
Q35	-0.32	0.387	-0.826	0.409
Q36	0.212	0.243	0.874	0.382
Q37	0.098	0.135	0.724	0.469
Q38	-0.202	0.139	-1.458	0.145
Q39	-0.059	0.13	-0.458	0.647
Q41	0.031	0.198	0.157	0.875
Q42	-0.136	0.181	-0.753	0.451
Q43	0.159	0.166	0.958	0.338
Audience by				
Q4	0.277	0.217	1.276	0.202
Q6	0.534	0.164	3.249	0.001
Q14	0.445	0.19	2.345	0.019
Q34	0.517	0.14	3.682	0
Q39	0.492	0.203	2.42	0.016
Q41	0.522	0.12	4.36	0
Q42	0.554	0.187	2.956	0.003
Q43	0.459	0.113	4.045	0
Q1	-0.121	0.158	-0.767	0.443
Q2	0.486	0.154	3.154	0.002
Q3	-0.15	0.242	-0.618	0.537
Q5	0.124	0.13	0.954	0.34
Q7	-0.138	0.269	-0.511	0.609
Q8	0.173	0.108	1.607	0.108
Q9	0.086	0.153	0.56	0.576
Q10	0.01	0.209	0.048	0.961
Q11	-0.032	0.163	-0.198	0.843
Q13	-0.35	0.191	-1.83	0.067
Q15	0.062	0.141	0.441	0.659
Q16	0.122	0.108	1.132	0.258
Q17	0.099	0.089	1.112	0.266
Q18	0.31	0.187	1.656	0.098
Q19	-0.14	0.166	-0.848	0.396
Q20	0.039	0.138	0.28	0.78
Q21	0.045	0.119	0.382	0.703

Q22	0.023	0.138	0.166	0.868
Q23	0.116	0.117	0.993	0.321
Q26	0.151	0.163	0.93	0.352
Q28	0.147	0.267	0.552	0.581
Q29	0.257	0.235	1.093	0.274
Q30	0.129	0.183	0.704	0.482
Q31	0.376	0.097	3.878	0
Q32	0.04	0.113	0.353	0.724
Q33	0.008	0.137	0.059	0.953
Q35	0.072	0.171	0.421	0.674
Q36	0.049	0.113	0.429	0.668
Q37	0.201	0.113	1.773	0.076
Q38	0.132	0.1	1.315	0.188
Q40	0.034	0.095	0.359	0.72
TR + TA	0.11			
TR + P	0.22			
TR + R	-0.05			
TR + A	0.25			
TA + P	0.15			
TA + R	0.16			
TA + A	0.21			
P + R	-0.07			
P + A	0.25			
R + A	0.14			

---

*Note.* Positive inter-factor correlations are negative relationships due to reverse coding of items.

## Appendix K – Calculation Framework for Pauses, Bursts and Revisions

Feature Type	Rule	Agreed way to calculate pause and notes
Linear within word pause	1) Letter/number 2) Letter/number  Or  (To pick up words at the beginning of a sentence or proper nouns)  1) Letter/number 2) Combination key 3) Letter/number	<b>NOTE: Character = letter/number/apostrophe/quote marks - based on QWERTY keyboard.</b>  Start time of current character – start time of character before.
Linear between word pause	1) Letter/number/apostrophe/quotes 2) Space 3) Letter/number/apostrophe  OR 1) Letter/number/apostrophe/quotes 2) Space 3) Combination key 4) Letter/number/quotes	Start time of character after SPACE – start time of character before SPACE  Combination key could be pressed before or after the space
Linear Between sub sentence pause	1) Letter/number/quotes/apostrophe 2) Comma 3) Space 4) Letter/number/apostrophe  OR 1) Letter/number/quotes/apostrophe 2) Comma 3) Space 4) Combination key 5) letter/number/quotes	Start time of the character after the comma - start time of character before comma  Combination key could be pressed before or after the comma.
Linear Between sentences pause	4 KEYSTROKE COMBINATIONS: 1) <LETTER> / <NUMBER> / <QUOTES> / <APOSTROPHE> 2) <FULL STOP> 3) <SPACE> 4) <LETTER> / <NUMBER> / <APOSTROPHE>  5 KEYSTROKE COMBINATIONS: 1) <LETTER> / <NUMBER> / <QUOTES> / <APOSTROPHE> 2) <FULL STOP> 3) <SPACE> 4) <COMBINATION KEY> 5) <LETTER> / <NUMBER> / <QUOTES> / <APOSTROPHE>  OR	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.  The combination key can fall anywhere between the full stop and first character of the next sentence.

	<ol style="list-style-type: none"> <li>1) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> <li>2) &lt;FULL STOP&gt;</li> <li>3) &lt;SPACE&gt;</li> <li>4) &lt;SPACE&gt;</li> <li>5) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt;</li> </ol> <p>6 KEYSTROKE COMBINATIONS</p> <ol style="list-style-type: none"> <li>1) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> <li>2) &lt;FULL STOP&gt;</li> <li>3) &lt;SPACE&gt;</li> <li>4) &lt;SPACE&gt;</li> <li>5) &lt;COMBINATION KEY&gt;</li> <li>6) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> </ol> <p>5 KEYSTROKE COMBINATION WHEN FIRST SENTENCE ENDS IN ? OR !</p> <ol style="list-style-type: none"> <li>1) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt;</li> <li>2) &lt;COMBINATION KEY&gt;</li> <li>3) &lt;?&gt; OR &lt;!&gt;</li> <li>4) &lt;SPACE&gt;</li> <li>5) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;APOSTROPHE&gt;</li> </ol> <p>6 KEYSTROKE COMBINATIONS</p> <ol style="list-style-type: none"> <li>1) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> <li>2) &lt;COMBINATION KEY&gt;</li> <li>3) &lt;?&gt; OR &lt;!&gt;</li> <li>4) &lt;SPACE&gt;</li> <li>5) &lt;COMBINATION KEY&gt;</li> <li>6) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> </ol> <p>OR</p> <ol style="list-style-type: none"> <li>1) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> <li>2) &lt;COMBINATION KEY&gt;</li> <li>3) &lt;?&gt; OR &lt;!&gt;</li> <li>4) &lt;SPACE&gt;</li> <li>5) &lt;SPACE&gt;</li> <li>6) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt;</li> </ol> <p>7 KEYSTROKE COMBINATIONS</p> <ol style="list-style-type: none"> <li>1) &lt;LETTER&gt; / &lt;NUMBER&gt; / &lt;QUOTES&gt; / &lt;APOSTROPHE&gt;</li> <li>2) &lt;COMBINATION KEY&gt;</li> <li>3) &lt;?&gt; OR &lt;!&gt;</li> <li>4) &lt;SPACE&gt;</li> <li>5) &lt;SPACE&gt;</li> <li>6) &lt;COMBINATION KEY&gt;</li> </ol>	
--	---	--

	7) <LETTER> / <NUMBER> / <QUOTES> / <APOSTROPHE>	
Linear Between paragraphs pause	<p>4 KEYSTROKES</p> <ol style="list-style-type: none"> <li>1) letter/number/apostrophe/ quotes</li> <li>2) full stop</li> <li>3) return</li> <li>4) letter/number apostrophe</li> </ol>	Start time of the first character of the new paragraph – start time of the character before the full stop/question mark/exclamation mark.
	<p>5 KEYSTROKES</p> <ol style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) full stop</li> <li>3) return</li> <li>4) combination key</li> <li>5) letter/number/quotes/ apostrophe</li> </ol> <p>OR</p> <ol style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) full stop</li> <li>3) space</li> <li>4) return</li> <li>5) letter/number/apostrophe</li> </ol> <p>OR</p> <ol style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) full stop</li> <li>3) return</li> <li>4) return</li> <li>5) letter/number/apostrophe</li> </ol> <p>6 KEYSTROKES</p> <ol style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) full stop</li> <li>3) return</li> <li>4) return</li> <li>5) return</li> <li>6) letter/number/apostrophe</li> </ol> <p>OR</p> <ol style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) full stop</li> <li>3) space</li> <li>4) return</li> <li>5) combination key</li> <li>6) letter/number/apostrophe/ quote marks</li> </ol> <p>7 KEYSTROKE COMBINATION</p> <ol style="list-style-type: none"> <li>1) letter/number/apostrophe/quote marks</li> <li>2) full stop</li> <li>3) return</li> <li>4) return</li> <li>5) return</li> <li>6) combination key</li> <li>7) letter/number/apostrophe/quote marks</li> </ol> <p>OR</p> <ol style="list-style-type: none"> <li>1) letter/number/apostrophe/quote marks</li> <li>2) full stop</li> </ol>	<p>The combination key can fall anywhere between the full stop and first character of the new paragraph.</p>

	<ul style="list-style-type: none"> <li>3) space</li> <li>4) return</li> <li>5) return</li> <li>6) combination key</li> <li>7) letter/number/apostrophe/quote marks</li> </ul> <p>8 KEYSTROKE COMBINATION</p> <ul style="list-style-type: none"> <li>1) letter/number/apostrophe/quote marks</li> <li>2) full stop</li> <li>3) space</li> <li>4) return</li> <li>5) return</li> <li>6) return</li> <li>7) combination key</li> <li>8) letter/number/apostrophe/quote marks</li> </ul> <p>THEN THE SAME AGAIN WITH ? OR !</p> <p>5 KEYSTROKES</p> <ul style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) Combination key</li> <li>3) ! or ?</li> <li>4) Return</li> <li>5) Letter/number/quotes/apostrophe</li> </ul> <p>6 KEYSTROKES</p> <ul style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) Combination key</li> <li>3) ! or ?</li> <li>4) Return</li> <li>5) combination key</li> <li>6) letter/number/quotes/ apostrophe</li> </ul> <p>OR</p> <ul style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) combination key</li> <li>3) ? or !</li> <li>4) Space</li> <li>5) Return</li> <li>6) letter/number/apostrophe</li> </ul> <p>OR</p> <ul style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) combination key</li> <li>3) ? or !</li> <li>4) Return</li> <li>5) return</li> <li>6) letter/number/apostrophe</li> </ul> <p>7 KEYSTROKES</p> <ul style="list-style-type: none"> <li>1) letter/number/quotes/apostrophe</li> <li>2) combination key</li> <li>3) ! or ?</li> <li>4) Return</li> <li>5) Return</li> <li>6) Return</li> </ul>	
--	--	--

	<p>7) letter/number/apostrophe</p> <p>OR</p> <p>1) letter/number/quotes/apostrophe 2) combination key 3) ? or ! 4) Space 5) Return 6) combination key 7) letter/number/apostrophe/ quote marks</p> <p>6 KEYSTROKE COMBINATION</p> <p>1) letter/number/apostrophe/quote marks 2) Combination key 3) ? or ! 4) return 5) return 6) return 7) combination key 8) letter/number/apostrophe/quote marks</p> <p>OR</p> <p>1) letter/number/apostrophe/quote marks 2) combination key 3) ? or ! 4) Space 5) return 6) return 7) combination key 8) letter/number/apostrophe/quote marks</p> <p>9 KEYSTROKE COMBINATION</p> <p>1) letter/number/apostrophe/quote marks 2) combination key 3) ? or ! 4) Space 5) Return 6) return 7) return 8) combination key 9) letter/number/apostrophe/quote marks</p>	
PP Burst	<p>&lt;PAUSE OF 2 SECONDS&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt;. &gt; &lt;, &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;PAUSE OF 2 SECONDS&gt;</p>	<p>A PP-burst must contain the production of language, not just non-letter/numerical characters.</p> <p>A pause can be an interval between two key presses which lasts for 2-seconds or longer, or it can be a within-word,</p>

		<p>between-word, between-subsentence, between-sentence or between-paragraph pause that lasts for two seconds or longer.</p> <p>PP bursts can contain minor revisions.</p>
RP burst	<p>&lt;MAJOR REVISION&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt;. &gt; &lt;, &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;PAUSE OF 2 SECONDS&gt;</p>	<p>Major revisions can be any substantial revisions either at the leading edge or away from the leading edge</p> <p>RP1 bursts must contain only new text production</p> <p>RP2 bursts must contain replacement of old text and new text production</p> <p>RP3 bursts must contain only the replacement of old text</p> <p>All RP-bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>A pause can be an interval between two key presses which lasts for 2-seconds or longer, or it can be a within-word, between-word, between-subsentence, between-sentence or between-paragraph pause that lasts for two seconds or longer.</p> <p>RP bursts can contain minor revisions.</p>
PRL burst	<p>&lt;PAUSE OF 2 SECONDS&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt;. &gt; &lt;, &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;MAJOR REVISION ONLY AT THE LEADING EDGE&gt;</p>	<p>Major revisions can be any substantial revisions either at the leading edge or away from the leading edge.</p> <p>All PRL-bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>A pause can be an interval between two key presses which lasts for 2-seconds or longer, or it can be a within-word, between-word, between-subsentence, between-sentence or between-paragraph pause</p>

		<p>that lasts for two seconds or longer.</p> <p>PRL bursts can contain minor revisions.</p>
PRI burst	<p>&lt;PAUSE OF 2 SECONDS&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt;. &gt; &lt;, &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;MAJOR REVISION ONLY AWAY FROM THE LEADING EDGE&gt;</p>	<p>Major revisions can be any substantial revisions either at the leading edge or away from the leading edge.</p> <p>All PRI-bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>A pause can be an interval between two key presses which lasts for 2-seconds or longer, or it can be a within-word, between-word, between-subsentence, between-sentence or between-paragraph pause that lasts for two seconds or longer.</p> <p>PRL bursts can contain minor revisions.</p>
PRLI burst	<p>&lt;PAUSE OF 2 SECONDS&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt;. &gt; &lt;, &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;MAJOR REVISION AT THE LEADING EDGE AND AWAY FROM THE LEADING EDGE&gt;</p>	<p>Major revisions can be any substantial revisions either at the leading edge or away from the leading edge.</p> <p>All PRLI-bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>A pause can be an interval between two key presses which lasts for 2-seconds or longer, or it can be a within-word, between-word, between-subsentence, between-sentence or between-paragraph pause that lasts for two seconds or longer.</p> <p>PRL bursts can contain minor revisions.</p> <p>Revision at the leading edge and away from the leading edge can happen in any order.</p>

RRL burst	<p>&lt;MAJOR REVISION AT THE LEADING EDGE OR AWAY FROM THE LEADING EDGE&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt; . &gt; &lt; , &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;MAJOR REVISION AT THE LEADING EDGE&gt;</p>	<p>Major revisions can be any substantial revisions either at the leading edge or away from the leading edge.</p> <p>All RRL-bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>RRL bursts can contain minor revisions.</p>
RRLI burst	<p>&lt;MAJOR REVISION AT THE LEADING EDGE OR AWAY FROM THE LEADING EDGE&gt;</p> <p>A sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt; . &gt; &lt; , &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;MAJOR REVISION AT THE LEADING EDGE AND AWAY FROM THE LEADING EDGE&gt;</p>	<p>Major revisions can be any substantial revisions either at the leading edge or away from the leading edge.</p> <p>All RRLI bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>RRLI bursts can contain minor revisions.</p> <p>Terminating revision at the leading edge and away from the leading edge can happen in any order.</p>
IG burst	<p>&lt;MOVEMENT AWAY FROM LEADING EDGE TO A POSITION WITHIN THE CURRENT SENTENCE&gt;</p> <p>Insertions of a sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt; . &gt; &lt; , &gt;, &lt;!&gt;, &lt;?&gt;</p> <p>&lt;MOVEMENT BACK TO THE LEADING EDGE&gt;</p>	<p>All IG bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>Text insertion should be substantial (i.e. not just changes to spelling) but should also be no longer than one or two sentences. Extended insertions (i.e. inserting whole paragraphs) would be deemed as text production at the leading edge</p> <p>IG bursts can contain minor revisions.</p>
IR burst	<p>&lt;MOVEMENT AWAY FROM LEADING EDGE AT THE END OF SENTENCE PRODUCTION TO A POSITION WITHIN THE SENTENCE JUST WRITTEN&gt;</p> <p>Insertions of a sequence of &lt;CHARACTERS&gt;,</p>	<p>All IR bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>Text insertion should be substantial (i.e. not just changes to spelling) but should also be no longer than one or two sentences. Extended insertions</p>

	<p>&lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt; . &gt; &lt; , &gt;, &lt; ! &gt;, &lt; ? &gt;</p> <p>&lt;MOVEMENT BACK TO THE LEADING EDGE&gt;</p>	<p>(i.e. inserting whole paragraphs) would be deemed as text production at the leading edge</p> <p>IR bursts can contain minor revisions.</p>
IB burst	<p>&lt;MOVEMENT AWAY FROM LEADING EDGE IN THE MIDDLE OF SENTENCE PRODUCTION TO A POSITION WITHIN ANOTHER SENTENCE&gt;</p> <p>Insertions of a sequence of &lt;CHARACTERS&gt;, &lt;COMBINATION KEYS&gt;, &lt;SPACES&gt;, &lt; . &gt; &lt; , &gt;, &lt; ! &gt;, &lt; ? &gt;</p> <p>&lt;MOVEMENT BACK TO THE LEADING EDGE&gt;</p>	<p>All IR bursts must contain the production of language, not just non-letter/numerical characters.</p> <p>Text insertion should be substantial (i.e. not just changes to spelling) but should also be no longer than one or two sentences. Extended insertions (i.e. inserting whole paragraphs) would be deemed as text production at the leading edge</p> <p>IR bursts can contain minor revisions.</p>
Major revision	<p>Revision keypresses are: &lt;BACK&gt;, &lt;DELETE&gt;, &lt;UP&gt;, &lt;DOWN&gt;, &lt;LEFT&gt;, &lt;RIGHT&gt;, &lt;[N:N]&gt;</p> <p>Revision mouse inputs are: &lt;LEFT CLICK&gt;, &lt;RIGHT CLICK&gt; &lt;SCROLL&gt; &lt;MOVEMENT&gt;</p>	<p>Single letter/number keypresses that are sandwiched between revision keypresses/mouse inputs should be included within the revision as they do not represent language production.</p> <p>Non letter/numerical characters that are sandwiched between revision keypresses should be included as part of the revision.</p> <p>&lt;SCROLL&gt; and &lt;MOVEMENT&gt; should not be considered part of a revisions unless they are used in combination with other revision keypresses/mouse clicks.</p> <p>If &lt;SCROLLS&gt; and &lt;MOVEMENTS&gt; happen not in combination with other revision keypresses/mouse movements and last longer than 2 seconds, they should be considered as pause indicators.</p> <p>If a &lt;PAUSE&gt; occurs before a major revision or during a major revision, it should be included as part of the major revision.</p>

Minor revision	Revision keypresses are: <BACK>, <DELETE>, <UP>, <DOWN>, <LEFT>, <RIGHT>, <[N:N]>  Revision mouse inputs are: <LEFT CLICK>, <RIGHT CLICK>, <SCROLL> <MOVEMENT>	Minor revisions are any small-scale revisions related to minor spelling or punctuation errors.
----------------	--	--

*Note. The keystroke combinations are based on a QWERTY keyboard on a Windows computer.*

<sup>a</sup> 'character' is a letter, number, quotation marks or an apostrophe.

<sup>b</sup> the combination key can fall anywhere between the full stop and first character of the next sentence.

*This applies to all given keystroke combinations.*

## Appendix L – Full Mixture Model Results Study 2

**Table 22***Descriptive Statistics for Each Linear Within-Word GMM*

Condition	Pro 1	Pro 2	Mean 1	Mean 2	Var 1	Var 2	SD 1	SD 2
Outline	0.8854833	0.1145167	4.704859	5.579214	0.1349363	0.440037	0.36733677	0.66335285
	0.97053985	0.02946015	4.687914	6.101757	0.1249467	0.3577791	0.353478	0.59814639
	0.923125	0.076875	4.89576	5.698302	0.1504304	0.3940072	0.38785358	0.62769993
	0.96180636	0.03819364	5.043543	6.571736	0.1976044	0.1424168	0.44452716	0.3773815
	0.97998861	0.02001139	5.033547	6.41883	0.1629268	0.1116666	0.40364192	0.33416553
	0.91741275	0.08258725	5.048465	6.024526	0.08986686	0.4787661	0.29977802	0.69192926
	0.95304125	0.04695875	5.00883	5.911988	0.1672537	0.6850537	0.40896662	0.82767971
	0.1381374	0.8618626	4.347781	5.0872	0.03480822	0.17058791	0.18656961	0.41302289
	0.95491735	0.04508265	4.956286	6.426781	0.1394031	0.1131792	0.37336725	0.33642116
	0.90799658	0.09200342	4.945418	5.63017	0.1558465	0.3463216	0.39477399	0.58849095
	0.2361084	0.7638916	4.382934	4.9398	0.0384215	0.1798733	0.19601403	0.42411473
	0.7872876	0.2127124	5.199702	5.788127	0.05452649	0.36335765	0.23350908	0.60279155
	0.95282036	0.04717964	4.870756	6.313762	0.1592389	0.1995821	0.39904749	0.44674612
	0.97989657	0.02010343	5.089054	6.541256	0.1586474	0.1199575	0.39830566	0.34634881
	0.932825	0.067175	4.817098	5.687255	0.1522448	0.4818341	0.3901856	0.69414271
	0.97202048	0.02797952	4.971333	6.337932	0.14635923	0.09674559	0.38256925	0.31103953
	0.95040318	0.04959682	4.909953	6.114996	0.1538822	0.1727393	0.39227822	0.41561918
	0.94027881	0.05972119	4.707725	5.72683	0.1374953	0.5437231	0.37080359	0.73737582
	0.8929892	0.1070108	4.877544	5.515892	0.09581647	0.40636841	0.30954236	0.63747032
	0.8212515	0.1787485	5.188188	5.887222	0.1105227	0.3978718	0.33244955	0.6307708

Synthetic	0.7456614	0.2543386	5.15552	5.631317	0.1047723	0.5989844	0.3236855	0.77394082
	0.1551515	0.8448485	4.287127	5.006115	0.03630886	0.13264479	0.19054884	0.36420432
	0.8764907	0.1235093	4.951676	5.716397	0.1507646	0.4742693	0.38828417	0.68867213
	0.8964931	0.1035069	4.884244	5.939169	0.1632249	0.4843783	0.40401101	0.69597292
	0.94245643	0.05754357	4.916525	6.231479	0.1334312	0.280342	0.36528236	0.52947332
	0.94591076	0.05408924	4.918745	6.259948	0.1697935	0.2823826	0.41206007	0.53139684
	0.3037422	0.6962578	4.328773	4.840041	0.03451424	0.17319913	0.18578009	0.416172
	0.94291325	0.05708675	4.732565	5.604813	0.1296311	0.3847129	0.36004319	0.62025229
	0.92207203	0.07792797	4.988959	5.823011	0.1504674	0.4697851	0.38790128	0.68540871
	0.96124234	0.03875766	4.92311	6.254569	0.1400087	0.170084	0.37417736	0.41241241
	0.95728127	0.04271873	4.789274	6.129571	0.127112	0.2352244	0.3565277	0.48499938
	0.1887582	0.8112418	4.406496	4.912023	0.03492491	0.14956394	0.18688208	0.38673497
	0.93186902	0.06813098	4.860911	5.997624	0.1064211	0.4608771	0.32622247	0.6788793
	0.7840641	0.2159359	4.662326	5.27529	0.1312219	0.3149512	0.36224563	0.56120513
	0.91607876	0.08392124	5.278673	6.003681	0.1231004	0.5179553	0.35085667	0.71969111
	0.94001814	0.05998186	4.926415	5.65636	0.1575203	0.3669622	0.39688827	0.60577405
	0.9435998	0.0564002	5.147759	6.068377	0.148993	0.6533491	0.38599611	0.80830013
	0.93951002	0.06048998	5.181477	6.009274	0.1270696	0.4892531	0.35646823	0.6994663
	0.8336659	0.1663341	5.016385	5.654191	0.1290367	0.3518876	0.35921679	0.59320115
	0.94671807	0.05328193	4.76321	5.550636	0.1499675	0.3212683	0.38725638	0.56680535
	0.98320969	0.01679031	5.006073	6.170537	0.21624	0.96261	0.46501613	0.9811269
	0.96665595	0.03334405	4.822361	6.153936	0.1539002	0.1450417	0.39230116	0.38084341
	0.8032111	0.1967889	5.191966	5.63369	0.1259426	0.5226134	0.35488392	0.72292005
	0.96810632	0.03189368	4.876017	6.234379	0.1367345	0.1839071	0.36977628	0.42884391
	0.97088122	0.02911878	4.76704	5.765308	0.1203511	0.492948	0.34691656	0.70210256
	0.96748969	0.03251031	4.817743	6.191515	0.1551919	0.1461212	0.39394403	0.38225803
	0.95163924	0.04836076	4.849094	6.153942	0.1639659	0.2958487	0.40492703	0.54391976
	0.97500741	0.02499259	5.162287	6.422625	0.1327407	0.3092228	0.36433597	0.55607805

*Note. Pro 1 = proportion of pauses assigned to component 1; Pro 2 = proportion of pauses assigned to component 2; Mean 1 = mean of component 1; Mean 2 = mean of component 2; Var 1 = variance of component 1; Var 2 = variance of component 2; SD 1 = standard deviation of component 1; SD 2 = standard deviation of component 2.*

**Table 23***Descriptive Statistics for Each Linear Between-Word GMM*

Condition	Pro 1	Pro 2	Pro 3	Mean 1	Mean 2	Mean 3	Var 1	Var 2	Var 3	SD 1	SD 2	SD 3
Outline	0.2339719	0.5753213	0.1907068	5.546796	6.153785	7.291341	0.02990994	0.13604	0.72952986	0.1729449	0.36883601	0.8541252
	0.2907122	0.4166155	0.2926723	5.195369	5.445534	6.75257	0.03289116	0.04294054	1.25267797	0.1813592	0.20722099	1.11923097
	0.52975547	0.3930584	0.07718613	5.692058	6.159907	7.19524	0.03848433	0.16165473	0.97179113	0.19617423	0.40206309	0.98579467
	0.2741323	0.5881245	0.1377432	5.62194	6.248203	7.395195	0.02527833	0.12171952	1.04485644	0.1589916	0.34888325	1.0221822
	0.4341266	0.4082546	0.1576188	5.60838	5.915875	6.947435	0.02961656	0.14043814	0.53076105	0.17209463	0.37475077	0.72853349
	0.2881194	0.2582941	0.4535865	5.662192	5.843588	6.853006	0.0190508	0.02401295	0.67333325	0.13802464	0.15496112	0.82056886
	0.5267633	0.346208	0.1270287	5.755324	6.362139	7.72473	0.03838136	0.16012228	0.5560399	0.19591161	0.40015282	0.74568083
	0.359235	0.4928916	0.1478735	5.485867	5.586575	6.307207	0.0214367	0.07688567	0.44663669	0.14641277	0.27728265	0.66830883
	0.3686188	0.434228	0.1971532	5.598461	6.20969	7.191659	0.03731377	0.13523511	0.86278713	0.19316773	0.36774327	0.92886335
	0.4196837	0.4153053	0.1650111	5.734816	6.008017	7.12993	0.02239979	0.05687523	0.61792668	0.14966559	0.23848528	0.78608313
	0.43851345	0.50086113	0.06062541	5.203343	5.623856	7.183818	0.05238026	0.20021094	0.18156342	0.22886734	0.44744937	0.42610259
	0.371258	0.4054532	0.2232888	5.958493	6.384711	7.501105	0.03189074	0.09433396	0.64438667	0.17857979	0.30713834	0.80273699
	0.5462895	0.3277123	0.1259982	5.532894	6.310121	6.916136	0.05444231	0.0687047	0.68292519	0.23332876	0.26211581	0.82639288
	0.3994053	0.3702524	0.2303424	5.774466	6.25903	7.070406	0.04765769	0.12372136	0.62092191	0.21830641	0.35174047	0.78798598
	0.3718724	0.4762454	0.1518821	5.60814	6.211236	6.905028	0.07514387	0.13522757	0.75075186	0.27412382	0.36773301	0.86645938
	0.52235862	0.40871896	0.06892241	5.596741	6.200716	7.066634	0.02909442	0.17867938	0.71542104	0.17057087	0.42270484	0.84582566
	0.55963057	0.38558961	0.05477982	5.670042	6.285201	7.736929	0.06298523	0.17517354	0.42317407	0.25096858	0.41853738	0.65051831
	0.3249486	0.4881792	0.1868722	5.336591	5.678961	6.478665	0.01594123	0.0289398	0.64566562	0.12625858	0.17011702	0.80353321
	0.5439068	0.36971091	0.08638229	5.56403	5.911416	7.026582	0.02869517	0.07159563	0.56298273	0.16939649	0.2675736	0.75032175
	0.370057	0.380351	0.249592	5.825849	6.328962	7.395298	0.01798594	0.09471315	0.50510647	0.13411167	0.30775502	0.71070843
	0.5203114	0.2662092	0.2134795	5.901378	6.585516	7.233894	0.04387842	0.06977834	0.41175754	0.20947176	0.2641559	0.64168336
	0.2418667	0.5803245	0.1778087	5.551245	5.77777	6.681955	0.0137826	0.04966557	0.5047363	0.11739932	0.22285773	0.71044796
	0.5694272	0.2873134	0.1432593	5.6753	6.105916	6.955744	0.04114002	0.15038471	0.51950434	0.20283003	0.38779468	0.72076649

Synthetic	0.4431894	0.3535115	0.2032991	5.532872	6.147315	6.781942	0.03242276	0.16397526	0.53811087	0.18006321	0.40493859	0.73356041
	0.6228447	0.2404415	0.1367137	5.814944	6.461093	7.624234	0.04512219	0.18270252	0.90645548	0.21241984	0.42743715	0.95207956
	0.5749043	0.2739333	0.1511624	5.516251	6.235566	7.408186	0.03728684	0.20847685	1.00810054	0.19309801	0.45659265	1.0040421
	0.3184582	0.4820158	0.1995261	5.312628	5.633124	6.487783	0.02079151	0.04315209	0.78065963	0.14419261	0.20773081	0.88354945
	0.2119928	0.5641613	0.2238459	5.24337	5.528342	6.759755	0.01714211	0.08209105	1.52587795	0.13092788	0.28651536	1.23526432
	0.52843911	0.44259122	0.02896967	5.72525	6.223264	7.959683	0.04837422	0.2009768	0.35550463	0.2199414	0.44830436	0.59624209
	0.4948943	0.2974682	0.2076375	5.545465	6.010038	6.816236	0.03297212	0.1810356	0.52600615	0.18158227	0.42548278	0.72526281
	0.6976546	0.2220582	0.0802872	5.498171	6.038078	7.097156	0.03358642	0.16265371	0.70233911	0.18326598	0.4033035	0.83805675
	0.3498086	0.4883579	0.1618335	5.407935	5.533311	6.353637	0.01400144	0.04907524	0.54687158	0.11832768	0.22152932	0.73950766
	0.2438124	0.5030477	0.2531398	5.4779	5.838992	6.602096	0.01577315	0.06302411	0.43747674	0.1255912	0.25104603	0.66142024
	0.4077328	0.4532884	0.1389788	5.530269	5.790681	7.300901	0.03717991	0.17600965	0.65045791	0.19282093	0.41953504	0.80650971
	0.3472404	0.3832394	0.2695202	5.913427	6.12436	7.010475	0.01961347	0.0420552	0.49631964	0.1400481	0.20507365	0.70449957
	0.3597578	0.4429127	0.1973295	5.49097	5.910151	6.792819	0.02182373	0.09285003	0.72197538	0.14772857	0.30471303	0.84969134
	0.3449886	0.4050968	0.2499146	5.739474	6.173063	7.295633	0.03025858	0.11703992	0.64002441	0.17394994	0.34211098	0.80001526
	0.3960659	0.401512	0.202422	5.86125	6.396115	7.70691	0.01826761	0.12000288	0.86362889	0.13515772	0.34641432	0.92931636
	0.559671	0.3245017	0.1158273	5.70695	6.149908	7.367478	0.03830594	0.10505575	0.66218722	0.19571903	0.32412305	0.81374887
	0.68515658	0.25226236	0.06258106	5.40855	5.810671	7.15772	0.05696423	0.222498	0.48451604	0.2386718	0.47169694	0.69607186
	0.3133703	0.4625514	0.2240784	5.730413	6.003903	7.095484	0.05221087	0.08854642	0.84430826	0.22849698	0.2975675	0.91886248
	0.5487998	0.38073577	0.07046443	5.526868	6.046877	7.188046	0.03089719	0.12802627	0.61490809	0.17577597	0.35780759	0.78416076
	0.4551988	0.46305994	0.08174126	5.857243	6.717816	8.02871	0.05318219	0.25286037	1.13737808	0.23061264	0.50285223	1.06647929
	0.69787266	0.21680081	0.08532653	5.626316	6.264975	8.346961	0.04599834	0.20075875	0.39235887	0.21447224	0.4480611	0.62638556
	0.4693633	0.3311381	0.1994986	5.48388	5.748988	6.863523	0.03266869	0.07873613	0.61165456	0.18074482	0.28059959	0.78208347
	0.6412189	0.1033213	0.2554598	5.518735	5.820814	6.44017	0.04151543	0.00344906	0.46788734	0.20375335	0.05872872	0.68402291
	0.3439621	0.4264033	0.2296346	5.55305	6.201235	6.447248	0.05214343	0.12365767	0.793243	0.22834936	0.35164993	0.8906419
	0.4767523	0.3640985	0.1591491	5.906591	6.206763	7.021925	0.02687314	0.06525613	0.50113588	0.16393029	0.25545279	0.70790951

*Note.* Pro 1 = proportion of pauses assigned to component 1; Pro 2 = proportion of pauses assigned to component 2; Pro 3 = proportion of pauses assigned to component 3; Mean 1 = mean of component 1; Mean 2 = mean of component 2; Mean 3 = mean of component 3 Var 1 = variance of component 1; Var 2 =

variance of component 2; Var 3 = variance of component 3; SD 1 = standard deviation of component 1, SD 2 = standard deviation of component 2; SD 3 = standard deviation of component 3.

**Table 24***Descriptive Statistics for Each Linear Sub-Sentence GMM*

Condition	Pro 1	Pro 2	Mean 1	Mean 2	Var 1	Var 2	SD 1	SD 2
	0.4945126	0.5054874	6.606694	7.729667	0.02113301	0.71849094	0.14537197	0.84763845
	0.3087238	0.6912762	5.721021	7.55597	0.00570244	1.47725663	0.07551452	
	NA	NA	NA	NA	NA	NA	NA	NA
	0.3049774	0.6950226	6.891658	7.993744	0.00623866	0.35822407	0.07898519	0.28104305
	0.4287796	0.5712204	6.546004	7.772197	0.02329151	0.61197086	0.15261556	0.39066042
	0.8243989	0.1756011	7.313604	9.255679	0.13941	1.009072	0.37337649	0.6110454
	0.4912729	0.5087271	7.140494	8.269784	0.00102007	0.36346967	0.03193852	0.17871351
	0.644909	0.355091	6.286044	6.904589	0.02148166	0.32221693	0.14656623	0.38283969
	0.349515	0.650485	6.646748	7.987356	0.0367855	0.1512715	0.19179546	0.43794459
	0.8998922	0.1001078	6.842135	9.008344	0.1715046	0.1653531	0.41413114	0.64353022
	0.7218508	0.2781492	5.811626	6.995726	0.03933792	0.30785518	0.19833789	0.44535143
	0.7933511	0.2066489	6.980304	8.150102	0.02425735	0.14274133	0.15574771	0.39464885
	0.585934	0.414066	6.822822	7.711139	0.04179952	0.07072377	0.20444931	0.45216071
	0.5193959	0.4806041	6.607074	7.56056	0.01852773	0.13288682	0.1361166	0.36893984
	0.7828234	0.2171766	6.220602	7.153241	0.04001924	0.0621178	0.20004809	0.44726736
	0.5	0.5	6.946918	7.883573	0.00876445	0.00125456	0.09361865	0.30597164
	0.4590081	0.5409919	6.509447	7.646575	0.04515705	0.20127451	0.21250188	0.46097926
	0.4989023	0.5010977	6.178731	6.636296	0.02461739	0.29742665	0.1568993	0.39610516
	0.4510704	0.5489296	6.448567	6.943015	0.00722766	0.40666513	0.08501563	0.29157441
	0.5760889	0.4239111	7.015935	8.670239	0.00901839	1.30157598	0.09496523	0.30816428
	0.5293837	0.4706163	7.017913	8.112606	0.07276664	0.17967687	0.26975292	0.51937744
	0.7279414	0.2720586	6.53367	7.717053	0.0553548	0.03533187	0.23527601	0.48505258
	0.4551273	0.5448727	6.475431	7.712343	0.09695561	0.35003751	0.31137696	0.55801161

Synthetic	0.3921321	0.6078679	6.349485	7.452711	0.0194916	0.6209432	0.13961232	0.37364732
	0.6666916	0.3333084	7.438284	9.05178	1.93E-01	6.73E-05	0.4398427	0.66320638
	0.587717	0.412283	6.078548	7.105455	0.01600226	1.37989696	0.12650004	0.35566844
	0.7814946	0.2185054	6.814539	8.641079	0.1907135	0.1108203	0.43670757	0.66083853
	0.5140626	0.4859374	6.196929	6.978934	0.04478871	0.6258093	0.21163343	0.46003634
	0.6640843	0.3359157	6.666209	7.874701	0.1230662	0.239932	0.35080792	0.59229041
	0.6398142	0.3601858	6.455621	8.139163	0.02739482	0.48819304	0.16551381	0.40683388
	0.3871262	0.6128738	6.791366	7.558289	0.04119742	0.14236671	0.20297148	0.45052356
	0.8573955	0.1426045	6.470888	8.039835	0.1849666	0.1839758	0.43007743	0.65580289
	0.5403784	0.4596216	6.476131	7.071096	0.00751106	0.18237713	0.08666637	0.29439152
	0.2441108	0.7558892	6.257132	6.568703	8.28E-05	1.49E-01	0.009099	0.09538866
	0.771528	0.228472	6.115339	7.123543	0.07084776	0.39836906	0.26617243	0.51591901
	0.1660713	0.8339287	6.625868	7.626324	0.00118202	0.4977623	0.03438052	0.18541984
	0.725573	0.274427	6.4937	7.749001	0.04757069	0.12688956	0.21810706	0.46701934
	0.6345121	0.3654879	6.50316	7.26426	0.03560796	0.09768038	0.18870072	0.43439696
	0.3199868	0.6800132	6.172048	7.803552	0.01003114	0.48957791	0.10015558	0.31647366
	0.4854764	0.5145236	6.120253	7.290232	0.01454975	0.28745289	0.12062234	0.34730728
	0.4270003	0.5729997	6.142924	6.396213	0.01489636	0.15976408	0.12205065	0.34935747
	0.19324	0.80676	6.340007	7.527353	0.00213905	0.57621936	0.04624981	0.21505769
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA

*Note.* Pro 1 = proportion of pauses assigned to component 1; Pro 2 = proportion of pauses assigned to component 2; Mean 1 = mean of component 1; Mean 2 = mean of component 2; Var 1 = variance of component 1; Var 2 = variance of component 2; SD 1 = standard deviation of component 1; SD 2 = standard deviation of component 2; NA = GMMs could not be fitted due to too few pause observation

**Table 25***Descriptive Statistics for Each Linear Between-Sentence GMM*

Condition	Pro 1	Pro 2	Mean 1	Mean 2	Var 1	Var 2	SD 1	SD 2
Outline	0.6087404	0.3912596	7.047136	8.554741	0.07314617	0.35509022	0.27045549	0.59589447
	0.3127706	0.6872294	6.184467	8.716824	0.08862658	0.46254349	0.29770217	0.6801055
	NA	NA	NA	NA	NA	NA	NA	NA
	0.7542792	0.2457208	8.001531	9.347028	0.27145782	0.03183794	0.52101614	0.17843189
	0.4302754	0.5697246	7.125433	8.674193	0.1615614	0.1849691	0.40194701	0.43008034
	0.8579815	0.1420185	8.415823	10.411932	0.39858556	0.00286934	0.63133633	0.05356623
	0.5300056	0.4699944	7.239688	9.03987	0.05034825	0.30137256	0.22438416	0.5489741
	0.6843622	0.3156378	6.664538	7.658973	0.04449671	0.0803094	0.21094243	0.28338913
	0.5429561	0.4570439	7.364762	8.762967	0.3384253	0.1366383	0.58174333	0.36964618
	0.252141	0.747859	6.588791	7.934642	0.00317964	0.44127833	0.05638826	0.66428783
	0.37803	0.62197	6.043688	7.225871	0.03037013	0.60771594	0.17427028	0.77956138
	0.7701992	0.2298008	8.050496	9.800948	0.2183634	0.02276272	0.4672937	0.15087319
	0.4586712	0.5413288	7.285362	8.134842	0.00462121	0.36133066	0.0679795	0.60110786
	0.3599106	0.6400894	7.264345	8.298355	0.09220868	0.34764726	0.30365882	0.5896162
	0.4223321	0.5776679	6.881367	8.671092	0.2022661	0.2285878	0.44974004	0.47810857
	0.4810571	0.5189429	7.113826	7.788498	0.00321007	0.09982495	0.05665744	0.31595086
	0.1315117	0.8684883	6.275186	7.072798	0.01144335	0.2476768	0.10697359	0.49767138
	0.3768426	0.6231574	7.48905	8.937325	0.0249454	0.3994474	0.15794113	0.63201851
	0.1456657	0.8543343	6.9146	8.216975	0.00302745	0.51420482	0.05502229	0.71708076
	0.5825815	0.4174185	6.692209	7.727005	0.01263769	0.2727633	0.11241748	0.52226746
	0.4497554	0.5502446	7.106918	8.159682	0.241123	0.3999269	0.49104277	0.63239774
	NA	NA	NA	NA	NA	NA	NA	NA
	0.4566127	0.5433873	7.236594	8.502275	0.2459724	0.5583634	0.49595605	0.74723718

Synthetic	0.8004317	0.1995683	7.201676	8.872317	0.29995971	0.00034774	0.54768578	0.01864789
	0.6892169	0.3107831	6.149815	7.750479	0.06224272	0.49027621	0.24948491	0.70019727
	0.8418684	0.1581316	7.114607	9.805116	0.5472694	0.1014555	0.73977659	0.3185208
	NA	NA	NA	NA	NA	NA	NA	NA
	0.4030499	0.5969501	6.543779	8.362788	0.1222816	1.2857425	0.34968786	1.13390586
	0.8404361	0.1595639	6.586765	7.491387	0.04878818	0.07342532	0.22088047	0.27097107
	0.5598936	0.4401064	6.915584	8.159146	0.03416446	0.11211193	0.18483631	0.3348312
	0.7004057	0.2995943	6.88334	7.848102	0.07273679	0.00248822	0.26969758	0.04988203
	NA	NA	NA	NA	NA	NA	NA	NA
	0.2741645	0.7258355	6.423174	7.726317	0.02356802	0.36015099	0.15351879	0.60012581
	0.2614069	0.7385931	7.360568	8.409703	0.0080634	0.17305968	0.08979642	0.41600442
	0.4401666	0.5598334	7.186203	8.279832	0.02694837	0.41381886	0.16415959	0.64328754
	0.6571129	0.3428871	7.165403	8.577703	0.1263682	0.2258278	0.35548305	0.47521343
	0.5935773	0.4064227	6.017591	7.589206	0.02254732	0.32536223	0.15015765	0.57040532
	0.368336	0.631664	7.589526	7.919015	0.06751152	0.42851582	0.25982979	0.6546112
	0.5749026	0.4250974	7.143649	8.707543	0.18620932	0.06009483	0.43151978	0.24514247
	NA	NA	NA	NA	NA	NA	NA	NA
	0.8240417	0.1759583	6.94164	9.135848	0.3791006	0.2486732	0.61571146	0.49867143
	0.5121256	0.4878744	6.758107	7.379272	0.09650536	0.00105301	0.31065311	0.03245013
	0.5622769	0.4377231	7.121906	8.530219	0.01637067	0.12666834	0.12794792	0.35590496
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA	NA	NA

*Note.* Pro 1 = proportion of pauses assigned to component 1; Pro 2 = proportion of pauses assigned to component 2; Mean 1 = mean of component 1; Mean 2 = mean of component 2; Var 1 = variance of component 1; Var 2 = variance of component 2; SD 1 = standard deviation of component 1; SD 2 = standard deviation of component 2; NA = GMMs could not be fitted due to too few pause observations

## Appendix M - Additional Study 3 Model Outputs

For the following model outputs:

- 1) Condition = drafting condition
- 2) PreUND = pre-test knowledge score
- 3) PreORG = pre-test organisation score
- 4) DiffUND = difference in pre- and post-test knowledge scores
- 5) DiffORG = different in pre- and post-test organisation scores
- 6) SentenceProduction = sentence production score
- 7) GlobalRevision = global revision score
- 8) TM = transactional beliefs score
- 9) TAR = transactional/revision beliefs score
- 10) R = revision beliefs score
- 11) A = audience beliefs score
- 12) P = planning beliefs score

# Full Model with Interactions output – Processes on KNOWLEDGE (UND) and Organisation (ORG)

Call:

```
lm(formula = DiffUND ~ Condition * SentenceProduction + Condition *
    GlobalRevision + Condition * SentenceProduction * GlobalRevision +
    Condition + SentenceProduction + GlobalRevision + PreUND,
    data = Combined2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.32035	-0.38750	-0.00304	0.31486	1.12341

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.50767	0.47193	3.195	0.00281 **
ConditionS	-0.23427	0.18130	-1.292	0.20412
SentenceProduction	-0.47358	0.30414	-1.557	0.12773
GlobalRevision	-0.07061	0.33504	-0.211	0.83422
PreUND	-0.24279	0.09405	-2.581	0.01382 *
ConditionS:SentenceProduction	-0.02352	0.46628	-0.050	0.96003
ConditionS:GlobalRevision	-0.10224	0.45275	-0.226	0.82256
SentenceProduction:GlobalRevision	-1.26774	0.73253	-1.731	0.09163 .
ConditionS:SentenceProduction:GlobalRevision	0.96838	0.96860	1.000	0.32374

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.595 on 38 degrees of freedom

Multiple R-squared: 0.3105, Adjusted R-squared: 0.1654

F-statistic: 2.139 on 8 and 38 DF, p-value: 0.0556

Call:

```
lm(formula = DiffORG ~ Condition * SentenceProduction + Condition *
    GlobalRevision + Condition * SentenceProduction * GlobalRevision +
    Condition + SentenceProduction + GlobalRevision + PreORG,
    data = Combined2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.98997	-0.30829	-0.03796	0.21263	2.12222

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.3687	0.5591	2.448	0.0191 *
ConditionS	-0.2655	0.2204	-1.204	0.2359
SentenceProduction	-0.9107	0.3697	-2.463	0.0184 *
GlobalRevision	0.2603	0.3999	0.651	0.5190
PreORG	-0.1971	0.1206	-1.634	0.1105
ConditionS:SentenceProduction	0.6187	0.5667	1.092	0.2818
ConditionS:GlobalRevision	-0.2615	0.5535	-0.472	0.6394
SentenceProduction:GlobalRevision	-1.9628	0.8877	-2.211	0.0331 *
ConditionS:SentenceProduction:GlobalRevision	1.6348	1.1705	1.397	0.1706

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7216 on 38 degrees of freedom

Multiple R-squared: 0.2789, Adjusted R-squared: 0.1271

F-statistic: 1.837 on 8 and 38 DF, p-value: 0.1001

## Writing Beliefs – Reduced Models (Equivalent to Baaijen et al., 2014)

Call:

```
lm(formula = DiffUND ~ PreUND + TAR + TM + Condition, data = Combined3)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.57209	-0.36273	-0.02159	0.39295	1.10657

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.522590	0.634883	2.398	0.0211 *
PreUND	-0.241243	0.093517	-2.580	0.0136 *
TAR	-0.241205	0.217150	-1.111	0.2731
TM	0.134169	0.151376	0.886	0.3806
ConditionS	-0.005951	0.208208	-0.029	0.9773

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6182 on 41 degrees of freedom

Multiple R-squared: 0.1957, Adjusted R-squared: 0.1173

F-statistic: 2.494 on 4 and 41 DF, p-value: 0.05766

Call:

```
lm(formula = DiffORG ~ PreORG + Condition + TAR + TM, data = Combined3)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.08503	-0.36613	-0.06904	0.34852	1.90083

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.76752	0.78189	2.261	0.0292 *
PreORG	-0.14585	0.12649	-1.153	0.2555
ConditionS	0.05651	0.25581	0.221	0.8263
TAR	-0.46506	0.27381	-1.698	0.0970 .
TM	0.06650	0.18619	0.357	0.7228

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7627 on 41 degrees of freedom

Multiple R-squared: 0.1307, Adjusted R-squared: 0.04586

F-statistic: 1.541 on 4 and 41 DF, p-value: 0.2084

## Planning and Audience Beliefs Models

```
Call:
lm(formula = DiffUND ~ Condition + PreUND + TM + TAR + P + A,
    data = Combined3)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.34300	-0.40720	0.01238	0.36959	1.16762

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.961591	0.705302	2.781	0.0083 **
ConditionS	0.004563	0.208889	0.022	0.9827
PreUND	-0.236997	0.093913	-2.524	0.0158 *
TM	0.186183	0.154852	1.202	0.2365
TAR	-0.185795	0.272878	-0.681	0.5000
P	-0.272181	0.180764	-1.506	0.1402
A	0.021717	0.185487	0.117	0.9074

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6162 on 39 degrees of freedom  
 Multiple R-squared: 0.24, Adjusted R-squared: 0.123  
 F-statistic: 2.052 on 6 and 39 DF, p-value: 0.08156

```
Call:
lm(formula = DiffORG ~ Condition + PreORG + TAR + TA + P + A,
    data = Combined3)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.7400	-0.3769	-0.0013	0.2238	1.9668

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.28951	0.87127	2.628	0.0122 *
ConditionS	-0.05551	0.25301	-0.219	0.8275
PreORG	-0.16286	0.12893	-1.263	0.2140
TAR	0.15502	0.62590	0.248	0.8057
TA	-0.51492	0.50148	-1.027	0.3108
P	-0.19504	0.22154	-0.880	0.3840
A	0.06602	0.23165	0.285	0.7771

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7661 on 39 degrees of freedom  
 Multiple R-squared: 0.1657, Adjusted R-squared: 0.03737  
 F-statistic: 1.291 on 6 and 39 DF, p-value: 0.284