
The Influence of Reading Fiction Upon Critical Thinking

Thesis submitted for the degree of Doctor of Philosophy in Information
Studies

Helena Hollis

Faculty of Arts and Humanities, Department of Information Studies

October 2021

I, Helena Hollis confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank Dr Oliver Duke-Williams, my supervisor, for all the guidance and support you provided during my research, particularly in helping me navigate my journey through learning statistical methods.

I would like to thank Dr Alison Hicks, my subsidiary supervisor, for your expert knowledge, and your insightful and thorough feedback which has helped refine my work every step of the way.

I would like to thank Dr Andreas Vlachidis, my subsidiary supervisor, for your suggestions and ideas that have provided inspiration for different directions.

I would like to thank Dr Charlie Inskip, who was my supervisor at the start of my PhD journey, for your belief in my ability, your encouragement, and your invaluable insight which helped shape the direction of this research project.

I am grateful to Dr Linda Elder and the Foundation for Critical Thinking for collaborating with me on validation of their critical thinking test. I would also like to extend my gratitude to The Sidney Perry Foundation for providing me a grant in support of my research.

Additionally, I would like to thank Geneviève Monastesse who introduced me to the art of trapeze during my time as a PhD student and thus provided me with an invaluable source of creativity and wellbeing.

Finally, I would like to thank each and every participant who took part in the studies conducted during this research process for their time, engagement, and support of my research.

Abstract

This thesis presents research into the relationship between reading fiction, as distinct from nonfiction, and critical thinking. Critical thinking is framed in the context of information literacy research. Prior research has shown increased fiction reading to be associated with social (Mumper & Gerrig, 2017), as well as cognitive and imaginative (Black et al., 2018; Oatley, 2011), capacities. These capacities are also associated with critical thinking (Byrne, 2016; R. H. Ennis, 2015; Thayer-Bacon, 2000). Thus, reading fiction may increase factors which in turn yield changes to critical thinking. To explore this potential relationship both normatively and subjectively, a sequential mixed methods approach was adopted. Four studies were conducted: one, an observational survey study assessing correlations between reading and factors associated with critical thinking; two, a reading log with pre- and post- critical thinking assessment, and experimental manipulation of assigned reading, testing a causal relationship; three, a reader interview study exploring experiences of reading and critical thought; four, a reading diary study exploring the day-to-day interplay of reading and critical thinking experiences. In conjunction, the four studies revealed: fiction reading was associated with experiential engagement in critical thought, while nonfiction was connected to the building of knowledge and procedure of critical thinking; fiction reading was predictive of disposition towards, change in, and improvement to critical thinking; nonfiction reading in long sessions was associated with improvement, but many short engagements were detrimental to critical thinking; assigning fiction reading to nonfiction readers was shown to be an efficacious critical thinking intervention. These findings suggest fiction is a utile resource for developing

critical thinking, and as such imply that the inclusion of fiction as part of information literacy and wider arts and humanities education, and across society through public library provision, is valuable.

Impact statement

This thesis offers findings, and methods, that can be utilised by researchers for future scholarship. As initial research into a previously un-explored area, the findings presented in this thesis validate a hypothesised relationship between reading fiction and critical thinking, and provide support for future detailed investigation. Findings on the different impacts of fiction and nonfiction reading contribute to research in psychology and literary studies where these differences have been explored in terms of driving social capacities, but without thus far addressing critical thinking as an outcome of reading. Furthermore, researchers focused on information literacy within information studies can deploy the conceptualisation of content evaluation and critical thinking presented in this thesis for the framing of future work. Information literacy practitioners can draw from these findings to design interventions and educational delivery including fiction reading materials. The findings pertaining to fiction and nonfiction reading impacts will also be of utility to researchers interested in public library provision outcomes. The effects of fiction, and nonfiction, reading upon critical thinking were small; they must also be interpreted within the timeframes of the studies, implying short term benefits. Longer term effects remain to be tested by future work. As such, reading, especially fiction, was shown to have small incremental impacts. This implies that ongoing engagement with fictional materials is recommended. As such, this research supports the case for future longitudinal work. This thesis also presents

methods that would be effective in further research. The design of the daily reading log form could be deployed in any research seeking to assess what participants read.

Likewise, the reading diary and approach to coding daily reading and critical thinking descriptions are both novel designs, and could be deployed in other research focusing on readers' experiences.

This research also entails benefits outside of academia, as critical thinking is widely considered socially valuable, and thus demonstrating ways in which reading fiction and nonfiction can influence critical thought provides a rationale for increasing access to reading materials. As such, this thesis presents evidence that could be utilised by public library advocates, and implies directions for public policy in enabling wide access to fiction as well as nonfiction reading materials. Furthermore, these findings contribute to the case for the value of the arts and humanities within the public and educational discourses. As critical thinking is likely to remain a highly valuable skill in future work, any means of increasing critical thinking capacities ought to be prioritised in educational programmes, and this research demonstrates that engagement with fiction is one means of doing so. Individuals seeking to challenge and bolster their critical thinking may be interested in the findings of this research, which implies reading fiction would be an impactful activity.

Table of Contents

1.	INTRODUCTION.....	1
1.1.	POSITIONALITY STATEMENT	8
2.	LITERATURE REVIEW	10
2.1.	CONTENT EVALUATION IN INFORMATION LITERACY AND CRITICAL THINKING.....	10
2.1.1.	<i>Content evaluation in conceptions of information literacy</i>	<i>12</i>
2.1.2.	<i>Critical thinking.....</i>	<i>14</i>
2.2.	READING FICTION AND NONFICTION	20
2.2.1.	<i>Fiction reading outcomes</i>	<i>20</i>
2.3.	EXISTING RESEARCH INTO FICTION READING AND CRITICAL THINKING	26
2.3.1.	<i>Narrative transportation and critical thinking.....</i>	<i>27</i>
2.3.2.	<i>Fiction reading and critical thinking</i>	<i>29</i>
2.4.	EMERGING RESEARCH DIRECTION.....	33
3.	METHODOLOGY	37
3.1.	GUIDING PRINCIPLES	37
3.1.1.	<i>Philosophical paradigm</i>	<i>37</i>
3.1.2.	<i>Mixed methods.....</i>	<i>40</i>
3.1.3.	<i>Validity.....</i>	<i>42</i>
3.2.	PRACTICAL APPROACHES	49
3.2.1.	<i>Sampling.....</i>	<i>49</i>
3.2.2.	<i>Piloting.....</i>	<i>50</i>
3.2.3.	<i>Methods for assessing reading.....</i>	<i>51</i>
3.2.4.	<i>3.2.4. Testing critical thinking.....</i>	<i>56</i>
3.2.5.	<i>Interviews</i>	<i>59</i>
3.3.	OVERVIEW OF THE FOUR STUDIES	60
3.3.1.	<i>Ethical issues.....</i>	<i>65</i>

4.	STUDY ONE: AN OBSERVATIONAL INVESTIGATION INTO THE RELATIONSHIP BETWEEN PRINT EXPOSURE AND FACTORS OF CRITICAL THINKING.....	66
4.1.	METHOD.....	70
4.1.1.	<i>Participants.....</i>	70
4.1.2.	<i>Measures.....</i>	70
4.1.3.	<i>Procedure.....</i>	72
4.2.	RESULTS.....	73
4.2.1.	<i>Descriptive statistics.....</i>	73
4.2.2.	<i>Modelling critical thinking disposition.....</i>	79
4.2.3.	<i>Modelling epistemological orientation.....</i>	84
4.3.	DISCUSSION.....	91
4.3.1.	<i>Limitations.....</i>	93
4.3.2.	<i>Relationship with the literature.....</i>	99
4.3.3.	<i>Implications for the next studies.....</i>	101
4.4.	CONCLUSIONS.....	103
5.	STUDY TWO: A READING LOG STUDY INVESTIGATING THE IMPACT OF FICTION AND NONFICTION READING ON CHANGE IN CRITICAL THINKING.....	105
5.1.	METHOD.....	111
5.1.1.	<i>Participants.....</i>	112
5.1.2.	<i>Measures.....</i>	117
5.1.3.	<i>Procedure.....</i>	122
5.2.	RESULTS.....	123
5.2.1.	<i>Descriptive statistics.....</i>	123
5.2.2.	<i>Inferential statistics.....</i>	140
5.2.3.	<i>Exploratory Analyses.....</i>	172
5.3.	DISCUSSION.....	189
5.3.1.	<i>Hypotheses testing summary.....</i>	189
5.3.2.	<i>Broader discussion of hypotheses, wider findings, and relation to literature.....</i>	192

5.3.3.	<i>Limitations</i>	204
5.3.4.	<i>Further directions</i>	212
5.4.	CONCLUSIONS	215
6.	STUDY THREE: AN EXPLORATION OF READER’S EXPERIENCES OF READING AND CRITICAL	
THINKING	217	
6.1.	METHOD	220
6.1.1.	<i>Participant recruitment</i>	220
6.1.2.	<i>Interviews</i>	220
6.1.3.	<i>Analysis</i>	221
6.2.	PARTICIPANTS.....	222
6.3.	RESULTS	225
6.3.1.	<i>Categories</i>	225
6.4.	DISCUSSION	258
6.4.1.	<i>Themes</i>	258
6.4.2.	<i>Relation to literature review</i>	265
6.4.3.	<i>Limitations and future directions</i>	273
6.5.	CONCLUSION	276
6.5.1.	<i>Answering the research questions</i>	276
6.5.2.	<i>Summary</i>	278
7.	STUDY FOUR: A DIARY STUDY OF DAILY READING AND CRITICAL THINKING EXPERIENCES ...	279
7.1.	METHOD	281
7.1.1.	<i>Participant recruitment</i>	281
7.1.2.	<i>Diaries</i>	282
7.1.3.	<i>Analysis</i>	283
7.2.	PARTICIPANTS.....	287
7.3.	RESULTS	287
7.4.	DISCUSSION	301
7.4.1.	<i>Relation to the literature</i>	301

7.4.2.	<i>Limitations</i>	304
7.5.	CONCLUSION	306
7.5.1.	<i>Answering the research questions</i>	306
7.5.2.	<i>Summary</i>	309
8.	DISCUSSION	311
8.1.	FICTION GIVES EXPERIENTIAL CRITICAL THINKING UNDERSTANDING	315
8.1.1.	<i>Time</i>	315
8.1.2.	<i>Counterfactual reasoning</i>	319
8.1.3.	<i>Imaginative capacities</i>	321
8.1.4.	<i>Openness, veracity, and bypassing defences</i>	323
8.1.5.	<i>Narrative transportation</i>	328
8.1.6.	<i>Fiction and critical thinking connection limitations and summary</i>	330
8.2.	NONFICTION BUILDS INFORMED CRITICAL THINKING PROCEDURE	331
8.3.	NUANCE IN READING MATERIALS	337
8.3.1.	<i>Diversity of reading</i>	338
8.3.2.	<i>Fiction and nonfiction</i>	341
9.	CONCLUSION	346
9.1.	CONTRIBUTIONS AND IMPLICATIONS.....	352
9.1.1.	<i>For theory</i>	352
9.1.2.	<i>For practice</i>	354
9.2.	FUTURE DIRECTIONS.....	358
9.3.	CHAPTER SUMMARY	361
	APPENDIX A: STUDY ONE	364
	MEASURES	364
	CTD MODEL SELECTION	368
	EO MODEL SELECTION	383
	ASSUMPTION CHECKING	397

MODEL ACCURACY TESTING	399
APPENDIX B: STUDY TWO	403
MEASURES	403
ASSUMPTION CHECKING	408
MODEL TESTING	422
APPENDIX C: VALIDATION OF ICTET-A WITH TEXT B	455
APPENDIX D: STUDY THREE.....	457
TOPIC SHEET FOR INTERVIEW SESSION 1.....	457
TOPIC SHEET FOR INTERVIEW SESSION 2.....	458
ELUCIDATING QUANTITATIVE FINDINGS.....	460
APPENDIX E: STUDY FOUR	466
APPENDIX F: COVID-19 PANDEMIC IMPACTS	470
REFERENCE LIST	472

Glossary of abbreviations

CT: critical thinking

“Active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends.” (Dewey, 1910, p. 6)

CTD: critical thinking disposition

“critical thinking dispositions have been defined as a constellation of attitudes, intellectual virtues or habits of mind, thus describing the way an individual reasons, argues and make decisions” (Sosu, 2013, p. 13)

EO: epistemological orientation

“In short, an individual’s epistemological orientation reflects his or her belief system about the nature of and acquisition of knowledge” (McGinnis, 2016, p. 279)

IL: information literacy

“Information literacy is the ability to think critically and make balanced judgements about any information we find and use. It empowers us as citizens to reach and express informed views and to engage fully with society” (CILIP, 2018)

NT: narrative transportation

“a state in which a reader becomes absorbed in the narrative world, leaving the real world, at least momentarily, behind.” (Green & Brock, 2013, p. 317)

ToM: theory of mind

“In brief, having a theory of mind is to be able to reflect on the contents of one's own and others' minds.” (Baron-Cohen, 2001, p. 169)

Statistical abbreviations are used in line with APA 7th edition style conventions (American Psychological Association, 2020).

List of Tables

Table 1: Sequence of research design	61
Table 2: Cluster centres for EO scores.....	77
Table 3: Results summary for cumulative link model for CTD	80
Table 4: Summary of results from generalised linear model of EO multiplist vs evaluativist	85
Table 5: Summary of results from generalised linear model of EO Absolutist vs Evaluativist.....	87
Table 6: Summary of results from generalised linear model of EO absolutist vs multiplist.....	89
Table 7: Assigned texts.....	110
Table 8: Participant blocks	116
Table 9: Description of total reading logged by experimental group	132
Table 10: Description of fiction reading logged by experimental group	132
Table 11: Description of nonfiction reading logged by experimental group	133
Table 12: Description of average NT by experimental group	137
Table 13: Comparison of regression models for CT t1 score and reading engagement	142
Table 14: Hierarchical regression results for t1 CT score with total engagement.....	143
Table 15: Hierarchical regression results for t1 CT score with reading engagement type	144
Table 16: Hierarchical regression results for CT t1 score with both fiction and nonfiction	146

Table 17: Hierarchical regression results for CT change with total read entries.....	149
Table 18: Hierarchical regression results for CT change with read entries by type ...	150
Table 19: Hierarchical regression results for CT change with total time	151
Table 20: Hierarchical regression results for CT change with reading time by type .	152
Table 21: Description of CT change across experimental groups	156
Table 22: Multiple regression results for CT change and experimental group with reading entries.....	158
Table 23: Multiple regression results for CT change and experimental group with reading time.....	159
Table 24: Comparison of models for CT improvement and total reading.....	164
Table 25: Results for logistic regression model of total entries and time on CT improvement	165
Table 26: Comparison of models for CT improvement and total reading.....	166
Table 27: Results for logistic regression model of fiction and nonfiction entries and time on CT improvement.....	166
Table 28: Results of logistic regression for improved or no improvement group with reading entries.....	169
Table 29: Results of logistic regression for improved or no improvement group with reading time.....	170
Table 30: Results of logistic regression with total reading time per entry upon CT improvement	173
Table 31: Results of logistic regression with fiction and nonfiction reading time per entry upon CT improvement.....	173
Table 32: Linear regression results for CT change and total time read per entry	175

Table 33: Linear regression results for CT change with NF t/e and F t/e.....	175
Table 34: Description of minimum and peak NT levels	177
Table 35: Description of reading entries at minimum and peak NT levels	177
Table 36: Description of reading time at minimum and peak NT levels.....	178
Table 37: Hierarchical regression results for CT change with peak NT reading entries	181
Table 38: Hierarchical regression results for CT change with peak NT reading time	181
Table 39: Description of minimum NT reading logged by CT improvement category	183
Table 40: Description of peak NT reading logged by CT improvement category....	183
Table 41: Comparison of models for CT improvement with peak and minimum NT reading.....	186
Table 42: Results for logistic regression model of fiction and nonfiction minimum and peak entries on CT improvement	187
Table 43: Results for logistic regression model of fiction and nonfiction minimum and peak time on CT improvement.....	188
Table 44: Table of Participants	224
Table 45: Summary of codes and categories	227
Table 46: Summary of code types	286
Table 47: Summary of categories and codes	288
Table 48: Comparison of single level models without interactions.....	376
Table 49: NT Interaction model comparison	378
Table 50: Random intercept model comparison.....	381
Table 51: Final model comparisons	383

Table 52: Comparison of EO Multiplist vs Evaluativist GLM models without interactions	392
Table 53: Comparison of EO Multiplist vs Evaluativist interaction models	394
Table 54: ICTET-A questions	403
Table 55: Comparison of original and bootstrapped regression	437
Table 56: Comparison of original and bootstrapped regression	440
Table 57: Comparison of original and bootstrapped regression	445

List of Figures

Figure 1: Histogram of ART-G total scores	75
Figure 2: Histogram of ART-G fiction scores	75
Figure 3: Histogram of AR-G nonfiction scores.....	76
Figure 4: Cluster plot by principle components.....	79
Figure 5: Histogram of CT scores at t2.....	125
Figure 6: Histogram of CT scores at t1.....	125
Figure 7: Histogram of CT score change from t1 to t2.....	127
Figure 8: Box plots of fiction and nonfiction items logged by different experimental groups.....	136
Figure 9: Box plots of fiction and nonfiction time logged by different experimental groups.....	136
Figure 10: Box plot of CT score change in different experimental groups	157
Figure 11: Bar chart of CTD levels by educational level.....	369
Figure 12: Bar chart of CTD levels by gender	370
Figure 13: Bar chart of CTD levels by NT level.....	371

Figure 14: Box plot of ART-G total score by CTD levels.....	372
Figure 15: Box plot of ART-G fiction score by CTD levels	373
Figure 16: Box plot of ART-G nonfiction score by CTD levels	374
Figure 17: Box plot of Age by EO category	385
Figure 18: Bar chart of EO categories by gender	386
Figure 19:Box plot of ART-G total score by EO categories.....	387
Figure 20: Box plot of fiction score by EO categories	388
Figure 21: Box plot of ART-G nonfiction score by EO categories.....	389
Figure 22: Plots for cumulative link model described in Table 3.....	400
Figure 23: Plots for regression model described in Table 14	423
Figure 24: Plots for regression model described in Table 16	426
Figure 25: Plots for regression model described in Table 17	428
Figure 26: Plots for regression model described in Table 18	430
Figure 27: Plots for regression model described in Table 19	432
Figure 28: Plots for regression model described in Table 20	434
Figure 29: Plots for regression model described in Table 22	436
Figure 30: Plots for regression model described in Table 23	439
Figure 31: Plots for regression model described in Table 32	448
Figure 32: Plots for regression model described in Table 33	450
Figure 33: Plots for regression model described in Table 37	452

1. Introduction

This thesis presents an investigation into ways that reading fiction may influence critical thinking (CT), as a core facet of information literacy (IL). The aim of this research project was firstly to identify whether a relationship between fiction reading (as distinct from nonfiction) and CT could be found, and if so whether a causal direction from fiction reading to CT could be traced. Furthermore, the research aimed to explore connections within this stipulated relationship to seek associations between fiction reading and distinct factors of CT, provide possible explanation for any causal pathways, as well as integrating readers' experiences of reading and CT. As such, this research project aimed to establish a basis for future research into the influence of fiction upon CT. It also aimed to provide initial insights that can inform decisionmaking on the provision of fiction reading materials.

Prior research has proposed that fiction reading enables the reader to simulate the minds of other people, and counterfactual scenarios, and that this simulation trains our ability to understand other's perspectives and to imagine different realities (Oatley, 2011). By reading fiction, we thereby expand our imaginative repertoire, and through it our capacity to engage with different people and their ideas. The underpinning argument that commenced this research project is that these gains feed our aptitude to think critically, in ways that also fit under IL. I argue that when we are presented with an issue, opinion, or idea, we must imaginatively model the mind of the person proposing it, put ourselves into their position, and imagine their perspective so as to be able to fully engage with the information they are seeking to put across. Furthermore, in order to critically evaluate that information we must be able to imagine its

interconnected facets and implications, mentally modelling what the world would be like if it were true or not. This thesis therefore proposes that reading fiction uniquely enhances our ability to perform these simulations, and therefore increases our CT capacity. In short, it is hypothesised that by reading fiction we can improve our ability to think critically about information in everyday life.

In order to explore the effects of fiction reading, nonfiction reading is used as comparator. Although film, television, computer games and live plays may be highly analogous to reading texts in some responses they evoke (Oatley, 2011), differences between the responses to and effects of these media have also been found (e.g. Richardson et al., 2018; Stanovich & Cunningham, 1993). For the purposes of this research, the terms fiction and nonfiction will be used to describe print material such as books, articles, short stories, playscripts etc., inclusive of their performance in audiobook or similar formats. Additionally, in order to be able to investigate the effects of fiction reading in comparison with nonfiction reading, it is important to first have a means of classifying what is fiction and what is nonfiction. However, as Friend (2008) points out, a clear-cut distinction that permits a definitive sorting into these two categories is unobtainable as there will always be fringe cases and counterexamples, and differences in the ways people think about these terms. Rather than strict definitions, this thesis therefore uses general guidelines to classify items under these headings. For the purposes of this research, Gregory Currie's intentional approach for defining fiction was adopted. This places the classification of fiction and nonfiction in the intentions of the author, and the reader's recognition of those intentions. He encapsulates this in the formula:

“The author of fiction intends that the reader make-believe P, where P is the sentence or string of sentences he utters. And he intends that the reader shall come to make-believe P partly as a result of his recognition that the author intends him to do this.” (Currie, 1985, p. 387)

The term fiction will therefore refer to works presented in such a way that conveys the imaginative intention for the reader to make-believe their contents. However, it would be lackadaisical to merely call nonfiction “everything else” (Root, 2004); Root asserts that nonfiction is rooted in reality, both in its “outward show” and “inner core” (Root, 2004, p. 290). Such outward show can be interpreted as pertaining to all of the ways in which intentions can be presented by authors, publishers, etc. The inner core and its connection to reality can be characterised as the closeness of the contents to our actual world (Lewis, 1978). Therefore, the term nonfiction will refer to works presented with the intention of capturing reality and representing it as closely as possible, so that the reader is invited to believe the work’s contents. Though impossible to delineate perfectly, I take make-belief and belief to be the poles around which fiction and nonfiction cluster.

Reading fiction is often treated as an entirely different activity to nonfiction, with separate underlying processes and entirely distinct purposes and outcomes. By drawing such a bright line between fiction and nonfiction, we can imply that there are no commonalities between them and treat them as mutually exclusive categories. Many theorists take this position, for instance Bruner (1986) views fictional stories and nonfictional arguments as fundamentally different modes of thought that are irreducible into one another. He terms these: the “paradigmatic mode of thinking”, which is scientific, based upon hypothesis testing and empirical truth; the “narrative mode of

thinking”, which is concerned with human experiences, intentions, and actions. He argues stories are not aimed at uncovering truths in the same way as arguments are. This view is not uncommon. Such a division has also been present in library and information science (LIS), where there has arguably been “a long tradition involving a rationalistic utilitarian mode of reading, in contrast to an experiential, emotional mode of reading” (Hampson Lundh et al., 2018, p. 1048) dividing how nonfiction and fiction are treated. This echoes a wider emphasis on “efferent” reading, with a neglect of “aesthetic” reading, in broader educational contexts (Rosenblatt, 1982), and perhaps an undervaluing of arts and humanities (Dumitru, 2019). However, many challenge Bruner’s delineation, and see the role of creative, imaginative, fictional thinking as interwoven through all of our reasoning (e.g. Csikszentmihalyi, 1997; Damasio, 1998; Holmes, 1981; McLeish, 2019). Theoretical physicist Tom McLeish has argued that:

“Neither art nor science can exist in a solipsistic vacuum of their authors. Both must be listened to, observed, received, responded to.”
(McLeish, 2019, p. 13)

The terms “art” and “science” could here be replaced with “fiction” and “nonfiction”. To this end, I stipulate that reading fiction and nonfiction are distinct, yet closely related and interconnected processes. From this it can be hypothesised that there is enough interconnection between what Bruner terms the paradigmatic and narrative modes of reasoning that gains in narrative reasoning can impact paradigmatic reasoning. As McLeish says, both involve openness, observation, receiving others, and response. Yet, I also hypothesise that paradigmatic and narrative reasoning are not identical, and that fiction reading engages some processes that nonfiction reading does not, such that

fiction reading can provide unique gains to our observational, receptive, and responsive abilities that can then be manifested across reasoning modes.

The findings of this thesis contribute to the study of IL, which entails a concern both with reading and CT. Again, these terms require definition, and more attention will be given to explain them and their interlinkages in the literature review. I do not adopt a single definition of CT for this thesis, but rather develop a conception of CT drawn from multiple sources in the literature review. For the purposes of this thesis, the CILIP (2018) definition of IL is adopted as it is recent, and not tied solely to an academic context:

“Information literacy is the ability to think critically and make balanced judgements about any information we find and use. It empowers us as citizens to reach and express informed views and to engage fully with society” (CILIP, 2018)

This definition is also characteristic of a move towards a greater focus on the evaluation of the content of information, and thus CT, in recent conceptions of IL. This can be seen strongly in critiques of “source evaluation” focused IL as perpetuating models of authority rather than acknowledging potential value in, and need to think critically about, the contents of diverse sources (Whitworth, 2014). I have elsewhere argued that strong-sense IL and CT conceptions strongly overlap (Hollis, 2019), and as IL research and practice has been on a trajectory moving away from narrow skills focus towards more holistic conceptions (Hartel, 2019), this suggests that CT ought to be a focus within IL going forwards. This is in keeping both with the disruption of authority-based IL approaches (Whitworth, 2014), and with the social justice framing of IL that implies each person ought to be able to think critically about any information (hooks,

1994). The “fake news” phenomenon both brings the importance of IL to the fore (Batchelor, 2017), and I posit demonstrates the need for each person to think critically about the content of each piece of information they are presented with, regardless of its source. Indeed, IL authors have argued that “Information Literacy research would have much to gain from reintroducing an empirical focus on reading activities” (Hampson Lundh et al., 2018, p. 1043); this research aims to provide such a gain to the field.

This research can be framed within the context of studies on the impacts of reading fiction, with a comparative focus on how fiction and nonfiction differ in terms of impact upon social skill (e.g. Mar et al., 2006). However, these studies have not addressed CT, and thus have left a research gap. Furthermore, at the time of writing, I have not uncovered any wider prior research that has tested the effects of fiction reading on CT. In the literature review, studies that come closest to this will be reviewed, highlighting a gap in the ways fiction and CT are tested and explored. Studies that do connect fiction reading and IL or CT do not investigate how the reading experience can be impactful, but do give practical examples of how the content of fiction can be used to inform attitude change. Therefore, further investigation is needed.

This thesis presents four studies conducted as part of a sequential mixed methods approach, underpinned by a scientific realist philosophy. Each study offers some important findings in its own right, and they are brought together in the discussion chapter to draw out how the interplay of these findings yields deeper insights. The first study demonstrates a correlation between exposure to fiction and factors associated with CT, based upon an observational questionnaire. Study two shows evidence for a

causal relationship between reading fiction, which is distinctive to that of reading nonfiction, and change in CT test score using a pre- and post- testing method around a two week reading diary. Study two furthermore includes experimental manipulation, and thus demonstrates the efficacy of assigning fiction texts to habitual nonfiction readers in improving CT. Study three presents the experiences of readers in reading and thinking critically, with distinctive fiction and nonfiction connections to CT, and finds that being transported into a text integrates CT into the reading experience. Study four presents ways in which reading and CT were connected in participants' reading diaries on a day to day basis, finding overarching commonalities in how reading and CT were experienced. When combined, these findings in conjunction reveal ways that fiction reading experiences provide profound lived understanding, developed over an unfolding timespan through the narrative. Nonfiction reading experiences were found to offer an actively built-up base of knowledge and procedure for thinking. Both could inform CT, but in different, sometimes subtly intertwined, ways.

This thesis will first present a review of the literature, initially reviewing conceptions of content evaluation within IL and of CT in wider fields, then reviewing the literature on the impacts of reading fiction, and finally reviewing prior research relating fiction reading to CT. The literature review concludes with broad research questions derived from the literature. Next, the methodology chapter articulates the guiding principles of the research design, with respect to the scientific realist philosophy, the use of mixed methods, and approaches to validity. The methodology then discusses concrete practical considerations, such as sampling and selection of instruments. The sequential rationale for the four studies is then outlined. Following

the methodology chapter, each of the four studies is described in full. Each study chapter details the methods used, the results, and a discussion of the results. As such, each study chapter presents a complete description of the separate studies taken individually. The discussion chapter then combines the findings from across the four studies into overarching themes and integrates them into a broader contextualisation of wider issues raised. Finally, the conclusion summarises the key findings, articulates the contributions made, and suggests future directions and implications from this research.

The studies conducted as part of this research project have shown fiction reading to be an avenue for improving CT, with a distinctive role to nonfiction reading. These findings therefore provide evidence for the value of fiction reading across society. As such, this research contributes to arguments in favour of public library provision, and in favour of the inclusion of fiction reading in IL instruction. This research is a starting point in understanding the impact fiction reading may have on CT, and the findings validate a rationale for further research elaborating this relationship.

1.1. Positionality statement

The research direction of this project is in many ways a personal culmination of interests I have previously pursued, and therefore I believe it is important to articulate my own personal frame of reference. As a lifelong voracious fiction reader, I conform to the stereotype of the bookworm. I have a bachelor's degree in philosophy, giving me a specific grounding in CT. Coming from a working class background, I have depended on libraries for access to books, both for educational and personal use. Furthermore, I have qualified and worked as a librarian. As such I have a high level of

personal affiliation with libraries. As my research may yield evidence that can be used to argue for the value of libraries, and thus for the need to continue to fund them, I have a personal investment in the outcome of my research. All of these factors have influenced my relationship with both reading and CT, and therefore my approach towards to my research. I am not neutral. However, I am committed to the scientific method as the best means we have of obtaining understanding, and thus I am bound to seek the disproof of my hypotheses, not their acceptance, through my research design. I take my frame of reference to be a variable that is a part of my research, which cannot be set aside but which can be taken into account and bracketed. I aim to allow my findings to shift my position, and not to allow my position to skew my findings. To this end, I aim to deploy a bracketing strategy by parenthesising my pre-existing notions of reading and CT (Ahern, 1999; Gearing, 2004). Specifically, I will use a descriptive bracketing approach (Gearing, 2004), aiming at the suspension of presuppositions so as to be able to give an undistorted description of the phenomenon. I take Husserl's (1969) position that bracketing in this manner can aid me in deemphasising the specifics of my own experiences, and permit me to seek the underlying universals at play.

2. Literature Review

This literature review first presents a conceptualisation of CT that draws from the disciplines of LIS, philosophy, psychology, and education. This conceptualisation will be used to frame CT as the outcome variable of interest throughout this research project. Next, the literature on the outcomes associated with reading, particularly fiction, is reviewed. In addition to the disciplines already mentioned, the review of fiction reading research draws from literary studies and the wider humanities. The reviews of these areas shape the rationale underpinning the proposed causal relationship between fiction reading and CT. Finally, studies closest to this proposed relationship are reviewed, mapping the state of research into reading and its impact on CT at the time of writing, and demonstrating the space in this landscape that this research project seeks to address. The research questions derived from the literature are presented at the end of this chapter.

The literature was reviewed with a hermeneutic method, in an iterative process (Boell & Cecez-Kecmanovic, 2010). I took this approach due to the highly interdisciplinary nature of my research topics, the breadth of which would make the necessary limitations for a systematic literature review inappropriate. By undertaking the search for literature in an unbounded, exploratory manner I was able to retain openness in the process. This permitted the inclusion of highly diverse research, and ensured a greater level of scrutiny for the research gap identified.

2.1. Content evaluation in information literacy and critical thinking

Content evaluation is the key component of information literacy (IL) focused on throughout this research. This is commonly taken to be an individual's ability to

critically evaluate the content of a piece of nonfiction text, considering the authors' position and motivations, as well as the strength of the argument and any evidence presented, in order to judge the validity of the information, and revise one's beliefs and knowledge accordingly. Content evaluation therefore forms the cornerstone of modern and critical IL conceptions that have moved away from a focus on judging sources based on standards of authority (CILIP, 2018; Whitworth, 2014). Content evaluation is a factor of interest not only within LIS where it is conceptualised under the umbrella of IL. It is also a factor of interest in psychology, philosophy, education and other areas; authors from these fields typically term what I have introduced here as "content evaluation" as CT. Both IL and CT are argued to be essential for navigating modern information landscapes (Cardoso & Silva, 2019; Machete & Turpin, 2020). Furthermore, performance on IL and CT tests has been found to correlate (McMullin, 2018). While there has been some acknowledgement of the inter-relatedness of IL and CT (Albitz, 2007; Weiner, 2011), there has not been a thorough analysis of this relationship from either side. Conducting such an analysis in a separate paper, I have argued that views on content evaluation from authors discussing IL, and conceptions of CT, intersect and strongly overlap (Hollis, 2019). This literature review will discuss conceptions of content evaluation from IL theorists and practitioners, and identify conceptions of CT that align well with these views. Based upon this intersection in the literature, I will identify a formulation of CT that integrates content evaluation, to be utilised as the outcome variable in my research.

2.1.1. Content evaluation in conceptions of information literacy

IL has been defined in various ways since its conception as a term, but there is some general consensus of what is meant by it (Owusu-Ansah, 2003; Owusu-Ansah, 2005). Taking the definitions from the major bodies involved in the discussion of IL, it can be overall summarised as a combination of skills and knowledge pertaining to identifying what information is needed; finding information; managing information; using information effectively; creating new information; communicating information; understanding ethical and legal issues surrounding information; critically evaluating information (ACRL, 2016; ALA, 2000; CILIP, 2018; IFLA, 2012; SCOUNL, 2015; UNESCO, 2003). It is the final component of the concept of IL, critically evaluating information, that will be considered here. This aspect of IL has been given increasing prominence, as seen in the most recent CILIP definition of IL (CILIP, 2018) which takes focus away from searching for information (Goldstein, 2018), and towards CT or content evaluation as a primary concern. However, the evaluative element of IL is the least bounded, as it necessarily applies in other aspects. For example, in order to identify what information is needed, one must engage in some level of critical evaluation. Therefore, critical evaluation as a facet of IL underpins the concept as a whole, and is arguably the most integral. Furthermore, this move in reconceptualising IL is in line with newly emerging critical/radical approaches, as exemplified by Whitworth who posits that a resource-focus in IL is depersonalising (2014, p. 33), and therefore critical IL ought to extend beyond source evaluation into content.

Content evaluation predominantly refers to critique of message and argument (Grafstein, 2017; Metzger, 2007). However, this entails a wider range of

considerations. For instance, Deitering and Jameson (2008) argue that content evaluation requires openness and willingness to engage with different ideas, and an understanding of how knowledge can be built. Other authors go further, for instance Reed and Stavreva posit that content evaluation requires “intellectual curiosity, sustained attention, open-mindedness, creative courage, and multidimensional cognitive abilities” (2006, p. 439). Others have also emphasised the need for openness and curiosity in IL (e.g. Weiler, 2005). Taken together, these conceptions build an understanding of content evaluation that integrates both applied logical reasoning facets with characteristics such as openness and creative engagement. Whitworth (2014) takes the conception of content evaluation yet further, including the usage and production of text as a necessary component of the judgement of information, in addition to reflection. This integration of inwardly reflective and outwardly productive components of content evaluation is echoed by other IL authors, and is well summarised by Ward’s assertion that “If we fail to develop our inner information capacities, we fail to become fully information literate” (Ward, 2006, p. 22). Thus, the increasing emphasis on content evaluation is arguably widening and deepening the construct of IL.

Arguments have also been made for shifting away from only intentionally acquired information, to consider incidental information as part of the content we evaluate (Kohnen & Saul, 2018). Fiction can be thought of as a source of incidental information, as it is not typically sought for specific information it may contain. Additionally, reading has been argued to be the defining activity that has informed the development of thought in IL by Hampson Lundh and colleagues, who suggest that

“many information activities should be seen as expressions of reading activities” (2018, p. 1049). They specifically address the “informational reading” approach, where reading is done to extract facts and specific answers, as the root of contemporary source evaluation IL approaches. In their paper, Hampson Lundh et al. touch on “experiential reading” associated with fiction, and it seems that this could be a root for content evaluation, encompassing the whole process of engagement with the content of a text as necessary for truly evaluating it. These approaches mirror distinctions made in reader response theory between “efferent” goal-oriented reading, and the “aesthetic” experience of reading (Rosenblatt, 1982). Therefore it is appropriate to consider the relationship between content evaluation in IL and reading experiences and outcomes.

2.1.2. Critical thinking

CT became a prominent field of interest in the 1970s, expanding rapidly in the 1980s (R. H. Ennis, 2015). It is interesting that this is similar in timescale with the emergence and development of IL as a concept (Whitworth, 2014). Much as in the debate on defining IL (Owusu-Ansah, 2005), many definitions of CT exist, though consensus can also be found (R. H. Ennis, 2016). One expression of this consensus has been in a shift over time from narrow and logic-focused conceptions of CT to broader and more nuanced conceptions, including reflective and emotional facets as well as underlying traits and dispositions. The shifts and progression in thinking about CT seem very closely aligned to those that took place in the IL literature across a very similar timeframe (Hartel, 2019). CT is furthermore commonly broken down into cognitive skills and affective dispositions (Facione, 1990a), in a way that is analogous to

IL (ACRL, 2016; Fitzgerald, 2004); these distinctions can be thought of trait/state manifestations of the construct (Chaplin et al., 1988).

Creativity is an aspect of CT that is of particular interest with relation to reading, as both may involve creative and imaginative processes. Richard Paul has perhaps most systematically argued for the inclusion of creativity in conceptions of CT. For Paul and Elder (2009), creativity is interwoven into all aspects of CT through the primacy of the imagination, not only in the cognitive process of generating ideas, but also in engaging with the thinking of others. Paul (1987) argues that imagination is essential in empathetic engagement and reciprocity, which allow us to understand another's point of view. He further argues that an understanding of how all information is presented with a point of view, even seemingly unbiased information such as the news, is a critical feature of CT and is dependent on this type of imagining of other perspectives. This not only mirrors IL perspectives on information and its provenance, it also is remarkably prescient of the current debate surrounding "fake news" (Batchelor, 2017), and it is noteworthy that Paul's remedy of creative, imaginative CT has been absent from this recent discourse. Paul (1987) calls for a richer, deeper, multi-perspectival imagining as part of CT that goes beyond simply keeping an open mind. In addition to imagination deployed in engaging with the thinking of others, it is also needed to imagine various possibilities in counterfactual thinking (Byrne, 2016). This future imagining can be deployed in making judgements both about desirability, and likelihood, of different outcomes, which are key considerations in evaluating arguments (Hoeken et al., 2020). Therefore imagination can be taken as a key element of CT.

Additionally, interpersonal facets of CT are also of particular interest, as reading fiction is associated with social skill development (discussed later in this review). Thayer-Bacon (1992) uses feminist theory to conceptualise CT in an interpersonal manner. She takes Nel Noddings' concept of a universal experience of caring, and argues that caring is a vital component of CT. This inclusion of caring as "receiving others, feeling with others" (Thayer-Bacon, 1992, p. 10) could also be termed as "empathy" (Richmond, 2004). Thayer-Bacon also stresses the need to include the self in CT, rather than attempting to strip away one's own views and attempt impartiality which is in fact unobtainable. She emphasises the development of one's personal voice in a reflective manner through CT. This is also a valuable insight, echoing the calls for inclusion of reflection in conceptions of CT (e.g. Facione, 1990; Kuhn, 1991), and perhaps integrating them more fully with the latent self-development and self-fulfilment goals of the CT movement (Dewey, 1910). This also fits very closely with recent calls for this form of rich reflection to be included in conceptions of IL (Corrall, 2017; Critten, 2015), as well as for the inclusion of compassion in IL (Gorichanaz & Latham, 2019).

Further components of critical thinking

In addition to the discussed overall ways of conceptualising CT, it can also be broken down into component parts, or factors that contribute to critical thought. This offers a means of studying CT in smaller studies where the whole concept may not be addressed, but its factors can be. For the purposes of this research project, critical thinking disposition (CTD), and epistemological orientation (EO), will be treated in this manner.

CTD is an essential component of CT (R. H. Ennis, 2016; Facione, 1990a). It is defined as “a constellation of attitudes, intellectual virtues or habits of mind” (Sosu, 2013, p. 108) capturing one’s willingness to engage in CT, which implies at least some ability to do so. As CT is effortful (Byrnes & Dunbar, 2014), the disposition towards it has been argued as a necessary prerequisite to engaging in CT practice (Weinstock et al., 2017). CTD has been shown to correlate positively with performance on problem based learning tasks (Pu et al., 2019), and with critical thinking skill (Colucciello, 1997). Furthermore CTD has been correlated with social emotional learning (Arslan & Demirtas, 2016), which entails empathy and broader social skills as part of CT. These findings demonstrate the relevance of CTD as a variable for initial study. Furthermore, engagement with a digital storytelling activity has been shown to increase CTD, which the author argues to be due to the dialogic nature of the activity and its promotion of self-reflection (Chan, 2019); this gives an indication that engagement with narrative may be a relevant driver for CTD. These findings place CTD as potential factor connecting reading with CT.

Epistemological beliefs are those pertaining to what knowledge is and how it can be gained, and these have also been identified as an important aspect of CT (e.g. Kuhn, 1991). These beliefs can be grouped into orientations, typically divided into three categories: “absolutist”, entailing a belief that statements are either true or false; “multiplist”, entailing a belief that statements reflect an entirely subjective reality; “evaluativist”, entailing a belief that statements can be judged based on some evidence and criteria (Hofer & Pintrich, 1997; Kuhn et al., 2000). Unlike CTD which is framed as a spectrum from low to high, EO is framed as a developmental trajectory shifting from

one orientation to the next in increasing sophistication from absolutist, through multiplist, into evaluativist orientations (Hofer & Pintrich, 1997). Experimental findings support the supposition that EO is an important contributing factor for CT (Hyytinen et al., 2014; King & Kitchener, 2004; Kuhn, 1991; Liu et al., 2011; Tsui, 2000). Evaluativist EO is considered most conducive of CT (Kuhn et al., 2000), as one must believe that statements can be effectively judged in order to attempt to do so. This evaluative stance has been cast as “epistemic vigilance”, as it necessitates being on guard for both absolutist and multiplist assumptions that could curtail CT (Mercier & Sperber, 2011). This establishes EO as a relevant factor of CT worth addressing. In addition, EO has been associated with different reading approaches and comprehension (Strømsø et al., 2008). The possible influence of reading upon EO can be inferred from studies showing fiction reading association with more open perspectives (Fong et al., 2015), however this is an area for further investigation. Therefore including EO as a variable in this research project offers an exploratory avenue.

Critical thinking as the outcome variable of this research

From here on in, CT will be used as the term for the outcome variable of interest. It is not taken as a binary outcome (i.e. one is not either able, or not able, to think critically); a range of CT approaches and abilities is possible. CT is taken to encapsulate both content evaluation as discussed by LIS researchers and practitioners under the umbrella of IL, and the key aspects of the CT conceptions presented above. Though I will cease specifically referring to IL and content evaluation, these are entailed in my use of CT as a term. As such, CT as discussed throughout this document entails both trait and state manifestations of how one evaluates information,

incorporating logical reasoning about its consistency and possible bias, with an openness and sensitivity to points of view, an ability to imaginatively simulate other viewpoints empathetically, as well as an internal reflective sensitivity to one's own views and reasoning processes and epistemological beliefs, and broader creative and imaginative capacity to engage with and model alternative possibilities, actively and effortfully aimed towards a creative outcome that could be the generation of new information, behaviour change, or shift in belief. This conception of CT is presented in simplified form in Box 1 for easy reference. CT is a construct taken to underlie a variety of manifestations in different ways, measurable only indirectly (Messick, 1974). It is of vital importance to us as individuals and on a societal level; bell hooks (1994, p. 202) has argued that it is CT that is the primary enabling force for change, as all people of all genders, ethnicities, and backgrounds can use their capacity to think critically and thus change and shape society. Therefore, it is of interest to consider how CT could be encouraged, and which factors may lead to increased CT.

Box 1: Simplified CT conception

CT entails:

- Logic and bias detection
- Openness to different points of view
- Empathy
- Self-reflection
- Counterfactual imagination
- Making judgements informing action and/or belief

2.2. Reading fiction and nonfiction

As was described in the introduction of this thesis, fiction and nonfiction reading are often treated as entirely separate processes with different purposes and outcomes. This distinction can be seen in Bruner's (1986) narrative versus paradigmatic thinking modes; in the efferent and aesthetic reading distinction of reader response theory (Rosenblatt, 1982); informational and experiential reading in LIS (Hampson Lundh et al., 2018). CT, for Bruner, falls into the paradigmatic category, as it aims at uncovering truths. In LIS, content evaluation is also concerned with informational reading (Hampson Lundh et al., 2018). However, as the conceptualisation of CT derived from the reviewed literature includes many creative, imaginative, empathetic, and reflective elements that more closely align to narrative rather than paradigmatic modes, these dichotomies are challenged. The following literature review will outline some of the gains associated with fiction reading that could cross this boundary and influence CT on informational subject matter.

2.2.1. Fiction reading outcomes

Davies (2010, p. 57) neatly summarises what fiction can be thought to provide as: factual information or direct knowledge (e.g. learning that cyanide has the flavour of bitter almonds from Roald Dahl's *The Landlady*); knowledge and understanding of general and overarching principles (e.g. identifying some political principles that govern group behaviour in Ursula K. Le Guin's *The Dispossessed*); categorical classifications (e.g. classifying a bureaucracy as a Kafkaesque structure); experiential or affective insight (e.g. gaining an experience of what it might be like to be a stifled teenage girl from Jeffrey Eugenides' *The Virgin Suicides*). Each of these is supported by empirical findings, though the final point perhaps most expansively. These findings

will be reviewed here, as they form a basis for the hypothesised further outcome of increased CT in those with higher levels of fiction reading. As experimentally manipulating what and how much people read over their lifetimes is impossible both practically and ethically, most of the research into the effects of reading is observational, and subsequently only correlations can be concluded, not causation. However, some experimental studies have also been conducted, giving indications of causal direction. In this review, correlational studies will be presented first, followed by experimental studies. Following this, the suggested reasons for, and mechanisms of, the impacts of fiction will be discussed.

Observational studies

Raymond Mar and Keith Oatley's research group embarked on a series of studies investigating the effects of fiction reading in comparison with nonfiction. They hypothesised that reading fiction enables readers to simulate social processing, inferring the mental states of characters and engaging vicariously in their situations, and as such ought to provide practice that improves social ability in real life. This hypothesis has been supported across various studies, which have taken measures of empathy and theory of mind (ToM)¹ to test social ability (Mar et al., 2006). An early study showed a double-dissociation where fiction reading was positively correlated with empathy and ToM, and nonfiction reading had a negative correlation (Mar et al., 2006). However, the negative correlation with nonfiction has repeatedly failed to replicate (Fong et al., 2013; Mar et al., 2009; Tamir et al., 2016); it cannot therefore be concluded that nonfiction exposure is associated with lower empathy levels. The relationship between

¹ ToM is the ability to infer the mental states of others (Baron-Cohen, 2001).

higher fiction reading and ToM scores has also been found in other studies, notably in the successful validation of a ToM measure (Dodell-Feder et al., 2013). The correlation between empathy and fiction reading has been replicated (Mar et al., 2009; Stansfield & Bunce, 2014). Furthermore, a metaanalysis by Mumper and Gerrig (2017) focused on the relationship between reading and social cognition supported the findings of a stronger correlation between fiction reading and social skill than nonfiction. They argue that future research on the effects of reading should focus on causal direction, and on a more fine-grained understanding of the simulation processes thought to drive these effects (Mumper & Gerrig, 2017). Additionally to social skill, fiction reading has also been shown to correlate with increased counterfactual reasoning, both about what is permissible as well as what is possible (Black et al., 2018); this finding is of particular interest, being in line with CT research showing people make argument evaluations of both desirability and likelihood (Hoeken et al., 2020). Black et al. argue “the act of engaging with fiction may in some way serve as practice for thinking about what *could* be, rather than what *is*.” (2018, p. 1). Indeed, when it comes to envisioning possible futures, fiction has been treated as a research method in AI research (Avin, 2019). Overall, these studies suggest an association between increased fiction reading and increased ability to engage with the thoughts and feelings of others, as well as imagining wider possibilities.

Experimental studies

Experimental studies focusing in on empathy and ToM resulting from fiction reading have typically taken the approach of assigning participants some reading, with a pre- and post- measure of the outcome variable of interest. Johnson (2012) found an

increase in empathy immediately after reading a fictional story, and an increased likelihood of performing a prosocial behaviour, and both of these were mediated by the participants' narrative transportation (NT)². Bal and Veltkamp's (2013) research was the first to experimentally manipulate reading and compare empathy outcomes over a one-week timeframe, allowing for sleeper effects to develop. They argue that the effects of fiction unfold gradually over time, which is supported by other research (Kuiken et al., 2004). They found that in fiction reading high transportation levels predicted empathy increases while low transportation predicted empathy decreases, with nonfiction showing a converse effect. This offers an indication of the importance of transportation for both fiction and nonfiction effects. In addition to empathy, Kidd and Castano (2013) conducted a series of experiments that showed an immediate improvement in ToM following the reading of a literary fiction text. They also found supporting evidence for fiction reading overall being correlated with higher ToM. However, their findings of an immediate boost to ToM after reading have failed to replicate (Panero et al., 2016, 2017). Nonetheless, the replication study did support finding that lifetime fiction exposure is predictive of ToM, suggesting a slow cumulative effect rather than an immediate gain. Interestingly, Panero et al. also did not find any significant relationship between NT and ToM, contrary to other research (Mar et al., 2009). Overall, they suggest that "any immediate effect of reading on theory-of-mind abilities is likely to be fragile and depend not only on the individual reader and the text, but also the relationship between the two" (Panero et al., 2016, p. e53), thus suggesting a direction for future research.

² NT is the immersive experience of being absorbed into a text, such that awareness of the external world is diminished and what is being read dominates one's conscious experience (Nell, 2013).

Other experimental research has addressed the process of reading, and addressed reflection and evaluation in addition to ToM. Özyürek and Trabasso (1997) conducted a study into the evaluations readers make whilst reading. They stipulate that just as we evaluate and reflect upon events we experience, we also evaluate events in narratives, including fictional texts. The evaluations their participants made while reading fiction served different purposes ranging from monitoring the concerns of the character and reacting as an observer, to constructing coherent interpretations of events. The authors believe “these inferences form causal chains and elaborate and fill in the situation model being constructed by the reader” (Özyürek & Trabasso, 1997, p. 325). This type of agent tracking and situation modelling is stipulated to be the mediating factor in correlation studies between fiction reading and other variables such as ToM and empathy (Mar et al., 2006). Therefore it seems that evaluation and inference making are integral in fiction reading, and it can be argued that if ToM and empathy are increased through practice by reading fiction that other inference and evaluation based abilities ought to also be improved.

Why reading fiction has these effects

It is important to consider how and why reading, especially fiction reading, could yield the previously discussed outcomes. Stanovich and Cunningham (1992) compare the time spent on reading with time spent exercising muscles, implying that reading gives us practice that builds up particular abilities. Zunshine (2006) similarly argues that a central function of reading fiction is the exercise of ToM. However, fiction is fundamentally different from real interactions, and thus these systems are argued to act upon it in a distinctive “offline” manner; therefore, fiction reading is

widely discussed as a simulation activity (Oatley, 2011). Mar, Djikic and Oatley (2008) posit that we are adept at understanding single, step-by-step processes (such as clouds forming and leading to rain), but poor at understanding complex interactions between multiple processes (such as the interactions of different weather systems). For the latter, we require simulations. They argue that fiction provides a simulation for complex social interactions and that this simulation constitutes a form of practicing skills, much like a flight simulator allows pilots to practice flight without leaving the ground. This explanation is supported by neuroimaging evidence showing that networks associated with simulation are involved with fiction reading (Tamir et al., 2016). The simulation explanation of the power of fiction fits with findings that NT drives its outcomes (D. R. Johnson, 2012; Mar et al., 2009), as entering into a simulation is an act of being transported:

“The traveller goes some distance from his or her world of origin, which makes some aspects of the world of origin inaccessible.”
(Gerrig, 1993, pp. 10–11)

Indeed, Busselle and Bilandzic (2008) argue NT takes place in the construction of mental models of the narrative, which unifies conceptions of NT with those of fiction reading as simulative mental modelling. This suggests fiction reading simulates reality, and by being transported into the simulation and out of reality allows the safe “offline” training of abilities. These abilities can then be applied outside of the simulation, having returned from the transported state.

The “offline” nature of fictional simulation may have a distinct effect to nonfiction in bypassing our defences. Djikic and colleagues (2009a) found that avoidant individuals experienced greater emotion change when reading literary fiction. They

argue that literary fiction is both safe, as the book can be put down, and yet also contains elements that cannot be defended against as “How does one defend against a juxtaposition of images or thoughts?” (Djikic et al., 2009a, p. 15). Bal and Veltkamp (2013) make a similar argument when they suggest that fiction reading can increase empathy in a way that nonfiction does not, because fiction does not come with any sense of obligation that we may be resistant to. By bypassing our defences, fiction reading may be a more powerful agent of change than nonfiction. Mar, Djikic and Oatley (2008) also note that readers commonly self-report changes to themselves as a result of books they have read. Furthermore, they have demonstrated experimental evidence of change to participants’ personality trait profiles following the reading of a short story, mediated by emotion (Djikic et al., 2009b). Importantly, this is not a specific change as a result of persuasion, rather reading opens one up to “fluctuation” that can take an individual direction (Djikic & Oatley, 2014). This implies a greater openness and flexibility in the directions fiction can take us in. Therefore transportation into simulations prompted by fiction reading may be uniquely able to drive change in our social and self schemata.

2.3. Existing research into fiction reading and critical thinking

Based upon the previously discussed research, how much fiction reading we engage in has been established as an individual difference worth investigating in relation to various outcome variables (Stanovich, 1993). As this research project aims to investigate the link between fiction reading and CT, I will here review the literature that comes closest to this. Firstly, the relationship between NT and CT will be reviewed, given that NT has been argued as central to the mechanism through which fiction reading may confer gains. Secondly, existing research addressing fiction reading

and CT will be discussed, with an aim to carve out a space for new directions and contributions.

2.3.1. Narrative transportation and critical thinking

NT is argued to be an active, effortful process (Green & Brock, 2013). In addition, empathy has also been argued to be a requirement for NT (van Laer et al., 2013). Retrospective reflection has also been found to be a mediator of NT, as the ability to link aspects of the narrative to one's past experiences facilitates the model building process (Hamby et al., 2017). Thus, NT has been associated with some aspects of CT. A positive relationship between NT and CT could therefore be hypothesised; however the existing research into such a relationship is contradictory.

Research has shown that being highly transported by an advertising narrative reduces CT while increasing affect, leading to more emotional and less critical brand evaluations (Escalas, 2004). Furthermore, there is less of a tendency to think of counterarguments or to be resistant towards to what is being presented when highly transported in a narrative (Dal Cin et al., 2004; Green & Brock, 2013). These findings have led many to hypothesise that NT reduces CT, a position described as:

“narratives do not invite people to systematically consider the issue, but rather the opposite: Narratives prohibit people in generating thoughts relevant to the issue, and their persuasive impact may even depend on this capacity to suppress such thoughts.” (Hoeken & Fijkers, 2014, p. 85)

Busselle and Bilandzic (2008) hold to this view and propose a model under which critical evaluations are only performed when the reader is bored, or when the reader is not able to be transported. This would suggest that NT in fact reduces CT, and by this

token spending more time transported into fiction ought to be associated with lower CT. However, Hoeken and Fikkers (2014) dispute this hypothesis and argue that the relationship between NT and CT is in fact far more complex. They found a high level of issue-relevant evaluative thinking when participants reported thoughts during reading a narrative, and these thoughts in conjunction with the level of protagonist identification predicted the participants attitudes (Hoeken & Fikkers, 2014). This is in keeping with findings of evaluative thinking taking place throughout the reading of narratives (Özyürek & Trabasso, 1997). These studies therefore present evidence that complicates the relationship between NT and CT. This suggests there is room to further explore the nuances of this relationship. Furthermore, it should be noted that these studies, both suggesting CT does and does not take place when highly transported, all investigate the effects of state NT by a given narrative upon attitudes towards the topic of that narrative, e.g. the effect of being transported into a narrative about homosexual characters upon attitudes towards homosexuality (Mazzocco et al., 2010). They do not investigate a relationship between trait NT and trait CT dispositions and abilities. This is an area that requires further study.

Finally, it is notable that while NT is most commonly discussed in association with fictional narratives, it is not exclusive to fiction. Indeed, studies finding NT in some way inhibiting CT depend upon there being some factual content to the material in question that they argue participants ought to think critically about (e.g. Dal Cin et al., 2004; Escalas, 2004). Furthermore, NT is often argued to be a driver of engagement with nonfiction that aims to educate or shift perspectives on a given topic (Borum Chattoo & Feldman, 2017). If using the definition of fiction and nonfiction

adopted for this research project (Currie, 1985), then nonfiction is that which invites readers to believe its contents; Green et al. (2012) found no difference in response to texts that had been labelled fiction or nonfiction, implying the same level of NT could arise when participants were invited to believe or make-believe the text. Therefore NT is taken into consideration for both fiction and nonfiction as part of this research project.

2.3.2. Fiction reading and critical thinking

Previous research has linked generally increased reading habit to CT, without distinguishing fiction and nonfiction (Bulgurcuoglu, 2016; Genç, 2017; Hawkins, 2012; Kilic et al., 2017; Terenzini et al., 1995; Ulu, 2019). These studies show a correlation between reading and CT in the higher education (HE) context, but offer no further insight into the role of fiction reading. When it comes to linking fiction reading specifically to CT, research is sparse. In discussing the power of fiction, authors across many contexts do posit its power to shape our thinking. This is often subsumed into the wider argument for the value of the arts and humanities as expanding CT capacities (e.g. Dumitru, 2019). However, when focused in, the relationship between CT and reading is often framed in terms of reading comprehension as a requirement for critical thought (Aloqaili, 2012; Farley & Elmore, 1992). In line with this framing, research into CT in reading centres around “critical reading” or “close reading”: reading with the aim of analysing the text for various features (Bialostosky, 2006). This is typically associated with the study of texts in different educational settings, most commonly in the study of literature (Bialostosky, 2006; Tabačková, 2015). However, this is a specific and narrow application of CT to a set text, and does not match my aim

of seeking ways in which fiction reading may influence our wider CT beyond the text at hand. Therefore the research on CT deployed in critical or close reading within texts will not be reviewed further here.

There has been some research conducted in the field of English as a Second Language (ESL), and in law and accountancy teaching, connecting literature reading with CT. Some ESL researchers present arguments for the use of literary texts in improving students CT, although without empirical backing (Erkaya, 2005; Pardede, 2019; Shihab, 2011; Shukri & Mukundan, 2015; Tabačková, 2015; Tang, 2016). Additionally, in both fields it is suggested that the enjoyment in reading literary texts may make them more motivating for CT (Crumbley & Smith, 2000; Dorocak & Purvis, 2004; Shukri & Mukundan, 2015). However, all of these authors utilise literature instructionally, in a classroom setting, where fiction is treated as “a teaching aid” (Dorocak & Purvis, 2004, p. 66) more than as a means for development. This is reflected in the empirical ESL studies that have found positive associations between literature reading and CT levels, as in all cases classroom instruction formed a core aspect of the study (Armstrong, 2015; Khatib & Alizadeh, Iman, 2012; Khatib & Mehrgan, 2012; Tung & Chang, 2009). At the time of writing, only one study in the field of ESL was found that focused on fiction and did not use structured critical reading programme, but gives insight into the role of literature reading independently of instruction (Prinsloo, 2018a, 2018b). Prinsloo assigned students short stories and had them complete open questions responding to them (2018a, 2018b). Short stories were found to fulfil a “thinking function” for students of varied disciplines, including creative, interpretive, and critical thinking (2018a, p. 148). Students described seeking

out hidden meanings, engaging with different perspectives, and reflecting of social themes, all of which fed into their personal growth (2018b). Furthermore, they found reading short stories helped them to concentrate, through being a pleasurable activity (2018b). Thus, Prinsloo's findings suggest that students' experiences corroborate the benefits theoretically stipulated for the use of literature in garnering CT within the ESL context. Overall, current research offers arguments for the value of reading literature to improve CT, but while some studies have supported these, most empirical research in this field is bound to critical reading instruction. Nonetheless, this does offer some indication that the link between reading fiction and CT is worth pursuing in broader research contexts.

Within the field of LIS, some authors have made connections between fiction reading and IL. In a blog post, Via Rivera describes how fiction reading has been used to improve IL with respect to the specific topic of ecology; i.e. reading fiction about ecological issues as a means of developing an "ecological sensibility" (Via Rivera, 2019); a rare approach in a field where the contributions of science and literature are more commonly strictly delineated (Aberbach, 2018). Via Rivera argues that fiction is specifically valuable as it can "help us imagine alternative scenarios, challenge our ideological biases, and allow us to grasp large interconnected networks of ideas" (2019). However, Via Rivera's suggested outcomes of fiction reading are entirely coupled to the subject matter read. Similarly, fiction reading has been used in IL instruction around specific subjects such as history (Paterson & Gamtso, 2019); feminism, and considering future AI developments (Dunham-LaGree et al., 2017). However, these approaches again anchor the gains from fiction to the specific subject matter of the

reading material. IL approaches also align closely with those in ESL that treat fiction as an effective classroom material, rather than considering its efficacy without instruction (Dunham-LaGree et al., 2017; Paterson & Gamtso, 2019). These limitations are emphasised by Harlan (2019) who presents perhaps the most robust case for the use of fiction to promote IL. Harlan argues that fiction ought to be treated as an information source in IL instruction, as fictional texts contain information that can be incorporated into our knowledge schema. Harlan also considers the emotional and exploratory nature of fiction as important, with specific reference to Rosenblatt's (1982) reader response theory. Harlan's overall argument is that "an aesthetic reading experience can be considered an information experience" (2019, p. 2). However, when discussing the ways students responded in questionnaires and interviews about their learning from reading specific fiction novels, Harlan takes a view that learning from the novel must be about the external world. She is critical of students who describe personal, reflective learning from the novel as a "mirror", rather than the explicit social information she takes to be a "window into learning about the world" (Harlan, 2019, p. 6). This forms a rationale for Harlan's argument that instruction is necessary for fiction reading to be informative:

"To engage in critical literacy readers need context, they need to be questioned, and they need to engage in dialogue. As evidenced in how students read for identity management to engage in the complexity of a story, a way for reading fiction to learn about other rather than the self" (Harlan, 2019, p. 7)

Thus, Harlan crystallises the views that fiction is valuable due to its specific content and subject matter, and that it is valuable as a tool for instruction. She actively negates the value fiction reading may have beyond this. As such, Harlan is not considering the self-

reflective aspects of IL (Corrall, 2017), nor is she considering the value of personal change through fiction reading (Djikic & Oatley, 2014). She treats fiction purely as a source of intentionally acquired information, without receptiveness to the possibilities of unintentionally acquired information (Kohnen & Saul, 2018). Thus, while Harlan does acknowledge the aesthetic experience of reading, she is firmly focused on “efferent” or “informational reading” and misses the potential of “experiential reading” (Hampson Lundh et al., 2018; Rosenblatt, 1982). Overall, these instances of fiction used to promote IL suggest that information presented through fiction may offer a unique means of developing sensitivities and understandings within specific subject areas. However, the role of fiction in developing IL/CT more widely is not considered or is actively dismissed. Thus, there is a space for further research into whether fiction reading without a set topic focus, or instruction, may serve to develop IL/CT.

2.4. Emerging research direction

Based upon the findings and theories presented in this literature review, it can be hypothesised that reading fiction can lead to increases in CT, encompassing content evaluation facets of IL. I posit that fiction can increase CT by increasing the elements listed in Box 1 that comprise it. While nonfiction reading can also lead to CT gains through knowledge building, the gains from fiction reading may be more diverse and more broadly reflective of CT in its wider conceptions.

It is crucial to this proposed relationship that fiction and nonfiction reading processes, or Bruner’s (1986) narrative and paradigmatic reasoning modes, are not mutually exclusive and compartmentalised. This is reflected in the conception of CT that is utilised throughout this research project, which incorporates interpersonal,

imaginative, creative and reflective aspects, not merely syllogistic and logical reasoning. Conceptualised in this manner, given the way that many have argued that fiction reading trains various capacities like exercise trains muscles, I argue that fiction reading trains CT. However, it is also essential to my argument that reading fiction and nonfiction are tapping into different processes, and are not identical; fiction has something unique to offer, and it trains and improves our ability to think critically about information in a way that reading nonfiction alone does not. I am suggesting that “aesthetic” or “experiential” reading of fiction may develop our CT abilities such that we become better at “efferent” or “informational” reading (Hampson Lundh et al., 2018; Rosenblatt, 1982).

If we return to Richard Paul’s characterisation of the process of thinking critically, which he describes as:

“We must first of all imagine ourselves in a given frame of reference. Then we must imaginatively construct some reasons to support it. Next we must step outside the framework of those acts and imagine ourselves responding to those reasons from an opposing point of view to create a response to the objection we just created. Next we must change roles again and create a further response. And so on.” (Paul, 1987, p. 143)

then the similarity between CT and reading fiction is made clear. This mental perspective switching is the fundamental characteristic of reading fiction, in which we routinely imagine the world through frames of reference alien to our own (e.g. Hakemulder, 2008). It is essential to my argument that fiction readers train this capability to step outside of themselves, into often very different points of view, and maintain the ability to evaluate those differing views as they do so (Özyürek &

Trabasso, 1997). This activity utilises ToM, which has been robustly shown to be improved through fiction reading in the afore discussed research. It also utilises counterfactual reasoning, and broader imaginative capacity, also linked to fiction reading. Fiction reading can thus be associated with gains in different factors that each contribute to increased CT.

Another means by which fiction may offer unique gains not available from nonfiction lies in its enjoyment value, and offline and safe processing. The pleasure of reading may enable more engagement with it (Crumbley & Smith, 2000; Prinsloo, 2018b), than with other possible means of training CT. The offline safety may allow fiction to bypass our defences (Bal & Veltkamp, 2013; Djikic et al., 2009a). The themes and characters we willingly engage with in fiction reading seem more diverse (Hakemulder, 2008), and more risky (Zunshine, 2006), than those we typically encounter in reality. For example, many readers seem open and willing to engage with the point of view of a paedophile in Vladimir Nabokov's *Lolita*, or a psychopath in Bret Easton Ellis' *American Psycho*, in a way that seems very different than if they were presented with a nonfiction article written by Humbert Humbert or Patrick Bateman. The fictional invitation to make-believe seems to broaden our openness to a greater variety of points of view. Given that openness to different points of view is a crucial component of CT, this seems to be a strong candidate for an avenue by which fiction reading can expand our CT capacity. After all, engaging with the point of view of someone as morally repugnant as Nabokov and Easton Ellis' creations is essential for thinking critically about what they have to say. Indeed, avoiding views we disagree with or find unpleasant is part of confirmation bias (Nickerson, 1998), a major barrier

to CT. Perhaps if we read more fiction, inevitably being exposed to unpleasant characters as well as those we like, in a manner that bypasses our defences and is in fact pleasurable, we can practice our capacity to engage despite discomfort, and reduce our avoidance of uncomfortable topics and voices outside of fiction also.

Overall research questions can be articulated stemming from this argument for a causal relationship between fiction reading and increased CT. Firstly, this research falls within a much wider area of interest that seeks to answer:

- What form of relationship does reading have with CT?

This broad question has been approached in many ways, in the studies included in this literature review and far wider. The space within this question that this particular research project will address can be delineated through the following specific research questions:

- Are there differences in fiction and nonfiction readers in their CT approaches?
- Can reading, particularly fiction, cause increases in CT?

3. Methodology

This research project was comprised of four discrete studies, each of which utilised a different methodological approach; therefore a detailed discussion on concrete methods will be presented in each study chapter. In this chapter, the underlying research approaches that guided all four studies are discussed, as these formed the basis for the research as a whole. The philosophical paradigm guiding the research design is described; the approach to, and aims of, utilising mixed methods are discussed; the approach to ensuring validity is detailed. Following these overarching principles, more applied issues with respect to the studies overall are covered: sampling strategies; approaches to assessing reading; ways of assessing CT. Finally, the four studies conducted are described in sequence with a focus on how they relate and combine into the overall design of this research project. It should be noted that some methodological choices I initially made had to be revised due to the restrictions arising from the COVID-19 pandemic; changes made as a result of this are detailed in Appendix F.

3.1. Guiding principles

3.1.1. Philosophical paradigm

This research project utilised a scientific realist philosophy (Maxwell & Mittapalli, 2010). The philosophical or scientific realism adopted is characterised by Lakoff as a position that maintains both that there is a singular external reality and that “there can be more than one scientifically correct way of understanding reality in terms of conceptual schemes with different objects and categories of objects” (1987, p. 265). As such, I adopt a realist position on causality, while holding causal relationships to be context-dependent (Shadish et al., 2002). In the quantitative studies, this means that

the results of the models are interpreted within the context of the sampling; for example in study one, the sample was mainly female and highly educated, thus the correlations found can only be generalised within this context. Similarly, I have taken a realist stance on mental states, entailing multiple perspectives on the world but rejecting relativist “multiple realities” (Maxwell & Mittapalli, 2010). By taking this position, I acknowledge the importance of perspective, meaning, and context. This research project was concerned with individuals’ actual states and their own point of view on those states. This fitted the subject matter of the research, as reading is both quantifiable (Stanovich & Cunningham, 1992), and also deeply personal and individual (Djikic & Oatley, 2014). This is also in keeping with conceptions of CT that entail both creativity and flexibility in perspectives, alongside normativity (Siegel, 1997). Finally, as I take diversity to be a real phenomenon (Maxwell & Mittapalli, 2010), individual differences have been included as an exploratory area of interest, thus aligning with calls for a finer-grained approach to research on the effects of reading (Mumper & Gerrig, 2017).

Adopting a scientific realist paradigm positions this research within a particular approach to IL in LIS, and to CT in other disciplines particularly education and pedagogy. This framing can be articulated in relation to research approaches in IL, which range from highly instruction-focused normative paradigms (e.g. Leichner et al., 2014), through to descriptive constructivist approaches (e.g. Lloyd, 2012). The former is an approach to IL research that is framed within HE contexts, and tends towards narrow skills-based conceptions of IL. The latter form of IL research tends towards a relativist stance towards epistemology, rejecting validity judgements in favour of

treating all understandings as of equal validity (e.g. Elmborg, 2006; Hall, 2010).

However, for content evaluation in IL/CT to be coherent as a concept, different views must be taken to have differing levels of validity, even if that is contextual, or else there is little to evaluate and think critically about. Equally, the conception of CT used in this research is broader than a simple set of directly teachable skills. Thus, the scientific realist paradigm allows for both normativity and subjectivity, and perhaps fills a gap in these commonly delineated approaches. In order to utilise normative criteria that are in keeping with more nuanced and contextual understandings of IL and CT, I draw from Whitworth's use of Mikhail Bakhtin's concepts: "polyphony", "dialogue" and "openness to transformation" (Whitworth, 2014, p. 121). Polyphony is a requirement for multiple perspectives to be presented. Dialogue describes how information is unfinalized; views are continuously refined, changed and discussed. Openness to transformation preserves these criteria by preventing congealment to a monologic, static presentation of information in hegemony. CT underpins these criteria, as being able to critically evaluate the content of what is said enables other voices to present themselves in a dialogue, and possibility of continuous evaluation and critique maintains the possibility of transformation. I therefore adopt these criteria for my research, and have sought to apply them in designing, conducting, and analysing each study undertaken; the specifics of their application will be noted in describing the individual studies.

3.1.2. Mixed methods

In keeping with the scientific realist approach, this research project has both tested hypotheses, and undertaken exploration, of the impacts of reading on CT. In this way, I align with reading researcher Gerald Cupchik when he writes:

“Qualitative methods offer an in-depth account of underlying processes and can help frame hypotheses that test specific functional relationships, while empirical findings related to processes can suggest areas which might benefit from detailed descriptive examination.”

(Cupchik, 2001, p. 1)

It has thus been my aim to make use of what both qualitative and quantitative methods have to offer in a complementary manner. Concretely, the research followed a sequential and explanatory paradigm (Ivankova et al., 2006; Subedi, 2016).

Quantitative data was collected and analysed first, followed by qualitative data in order to enrich and explain the quantitative findings, and to elucidate a causal explanation for quantitative casual descriptions (Shadish et al., 2002). As such, priority was given to the quantitative components of the research design, but with flexibility to re-weigh the balance of priority later dependent upon findings (Ivankova et al., 2006); this re-balancing will become evident in the overall discussion. This research was therefore designed with a results point of interface, with the quantitative results forming the framework for integrating the qualitative results, which in turn embellish the quantitative findings and add important detail (Maxwell & Mittapalli, 2010).

Triangulation was a key aim in undertaking four distinct studies, two quantitative and two qualitative, each with a unique research method. Specifically, conducting multiple different studies has allowed between-methods triangulation,

compensating for some of the limitations of each method in each study through the deployment of others (Denzin, 1970). For example, a quantitative reading log permits readers to log all of their daily reading yielding a picture of their complete reading habits, but without exploring how they felt and thought about what they read as this would become too taxing; a daily reading diary permits readers to detail their experiences of examples of their reading in full, but without capturing everything they read. Using both methods allows for different emphases without demanding excessive time commitments on participants by asking for both types of information. However, this approach to triangulation does not attempt “quasi-correlation” (Flick, 2018, p. 16); I do not take my quantitative and qualitative studies to be capturing necessarily the same phenomenon, such that it can be more accurately portrayed. Rather, triangulation was pursued as an opportunity for enrichment through deepening (gaining detail) and widening (increasing avenues for exploration) the scope of my research. As Flick puts it:

“This aim is pursued via more adequacy and comprehensiveness in grasping the issue under study and not by a unilateral or mutual validation of the single-method results.” (Flick, 2018, p. 16)

As such, I did not seek congruence in my results, but rather a convergence towards a way of answering my overall research questions, elaborating upon my hypotheses, and identifying new potential areas of exploration. In Flick’s terms, this research project aimed for “a strong programme of triangulation” in which including multiple methods offered “extra knowledge” and acted as an extension of the research (2018, p. 19). Returning to reading log and diary example, the qualitative diary provided extra information that extended the findings of the quantitative log into further areas.

Though Flick has criticised Denzin's (1970) view that different methods can compensate for each other's limitations, I take the stance that their positions can be reconciled; I do not take the balancing of limitations to be a gap-filling exercise in completing a picture of a single phenomenon, rather I treat the knowledge extension granted through additional methods as the means of addressing limitations in the bounded nature of any single method alone.

3.1.3. Validity

The validity of the research design was continually assessed at every stage. Validity was taken to be necessarily context specific, adopting the position that there is no way to make general claims about the validity of a method, it must always be assessed for its validity within the context of the specific study (Shadish et al., 2002). In order to ensure validity within the context of each study conducted, I sought alternative explanations of my findings at every stage of my research process, and sought to weigh these alternatives up against my interpretations, such as to provide robust evidential and theoretical support for my choices. Such alternative explanations are noted in the study discussions. Another way to frame this is as vigilance to "validity threat" (Shadish et al., 2002); identifying possible ways my conclusions may be wrong, and seeking to eliminate these threats through further data analysis or through supplementary data collection. One example of this is in the interpretation of the observational study (study one), where findings based on educational level could be interpreted as significant, but also be interpreted as an artefact of the sample which skewed towards higher educational levels. Identifying the threat to the validity of any conclusion on the role of education given the sampling issues, the investigation into

the role played by education was taken forward into the reading log study in which a targeted sampling strategy was used to avoid the same validity threat arising, and through which this association could be further tested. Another example is in the approach taken in categorising codes in study four, in which I did not differentiate fiction and nonfiction, but rather categorised them together; similarly in study three one of the themes derived combined fiction and nonfiction. This permits disconfirmatory explorations into how fiction and nonfiction may be similar, not only different, and thus provides a more solid grounding for articulating differences against this background of commonality.

As a further part of seeking validity in my research, I reflected upon and bracketed my pre-existing biases in planning my studies and in analysis, as well as in a post-analysis reflection (Ahern, 1999). In order to effect this bracketing strategy, I utilised a reflective diary keeping approach, in which I detailed my experiences, feelings, and thoughts on a routine basis (Moon, 1999). This reflection and bracketing took place throughout the research process; for instance, I noted my bias in my dislike of certain authors who participants I interviewed in study three made reference to, and actively ensured I asked as many follow-up questions as when exploring authors I personally liked; in coding the data, I re-checked my coding pertaining to my less preferred author examples. By having noted my preferences in my reflective diary, I was able to systematically check my coding against these. Bracketing also took place in my broader coding and interviewing approaches, by noting topics and themes which I found salient in the data, but which were not relevant to the study at hand. This was particularly the case in study three, where interviewees made reference to current

affairs issues such as Black Lives Matter, the climate crisis, and others that I felt personally engaged in. I sought to balance giving participants the space to express their thoughts, but without seeking elucidation into these topics but rather asking follow-up questions targeting their thinking *about* the topics. Similarly, I did not code for any topics such as these in the qualitative studies, but focused on coding for concepts relating to my research questions.

Finally, Whitworth's (2014) criteria of polyphony, dialogue and openness to transformation were deployed in the research design. Polyphony was sought in sampling strategies; dialogue in the iterative interplay between quantitative and qualitative studies; openness to transformation in changes made to research plans arising from new situations and insights.

All studies were pre-registered, to ensure transparency as part of my validity concerns. For the quantitative studies, this enabled a clear delineation between hypothesis testing and exploratory analyses, and accountability in analysis choices (Toth et al., 2020). For the qualitative studies, pre-registration formed part of an "audit trail" (Shenton, 2004) of my decision making, in delineating the starting point from which the research progressed, and making my subjective positioning explicit (Haven & Van Grootel, 2019). As such, the pre-registered plans for both types of study were not set in stone, and I was flexible in order to retain "openness to transformation" (Whitworth, 2014). Deviations were made from my pre-registered plans, and these were tracked and are discussed in each study chapter to preserve the transparency aims of pre-registration (Haven & Van Grootel, 2019; Toth et al., 2020).

In order to ensure both validity and sensitivity in the quantitative studies, thus avoiding both type I and type II errors, sample sizes were determined primarily by statistical power estimates (Lipsey & Hurley, 2009). A medium effect size (0.3) and power of 80%, with the standard .05 alpha error probability, were used as target values. These standard values were used due to a lack of prior research for more fine-grained calibration, and likewise two-tailed tests were favoured (Lipsey & Hurley, 2009). The G*Power statistical software was used to make sample size calculations (Faul et al., 2009). Where power estimation could not be used to determine a minimum sample size, for example in study one, a ratio of participants to scale items was deployed (Anthoine et al., 2014).

In addition to sample size, the choice of instruments used to measure the variables of interest was also considered in terms of validity and reliability, and thus only previously validated measures, with respect to construct, face, criterion validity as well as internal reliability, were used (Coolican, 2014); specific measures are described in the study chapters. Each of these measures was identified through the previous literature, to methodologically align my research with the previous body of evidence I sought to build upon (for example using the ART-G (Mar & Rain, 2015), in keeping with the studies on the social impacts of reading fiction). No un-validated measures were used, however. For example, in study two the ICTET (The Foundation for Critical Thinking, 2019) was used as a measure of CT, however only one sample text had been validated with the test (Hollis et al., in press). Two texts were needed to enable a pre- and post- test paradigm, and therefore a second text was tested for consistency against the previously validated version, essentially validating a secondary

version of the test, as part of this study. This maintained the aim of relying upon validated measures only.

Validity was also the key concern in selecting statistical models to use in analysing the quantitative data. Models were selected to be best fit the data. In the case of categorical outcome variables, cumulative link models were used for ordinal outcomes, and binary generalised linear models for binary outcome variables, as these classes of model offered both power and flexibility while preserving the data categories (Christensen, 2015; Field et al., 2012). Where outcome variables were numerical, linear (multiple) regression models were used, as these models have been extensively used in the literature upon which this research project builds (e.g. Black et al., 2018; Mar et al., 2006). Additionally, using multiple regression models with a hierarchical logic allows for a conservative approach in partialling out other variables (i.e. control variables, nonfiction) before testing the effect of fiction (Stanovich et al., 1995; Stanovich & Cunningham, 1992, 1993, 2004); this is in keeping with the overarching vigilance to validity threat (Shadish et al., 2002) adopted throughout the research process. Finally, to ensure validity in all statistical models used to test hypotheses, assumption checking was conducted and model accuracy was tested (i.e. seeking outliers and influential cases, verifying all necessary assumptions were met) to evaluate if the models sufficiently met the data and could be generalised from (Field et al., 2012); these analyses are presented the appendices. In the case of meeting the assumptions of the ANOVA, equality of variances was considered as more critical than normal distribution in study one which had a large sample size, but a more

conservative approach was taken in study two where the sample was smaller (Blanca et al., 2017).

In order to ensure validity and reliability within the qualitative studies, Guba's (1981) criteria as operationalised by Shenton (2004) were deployed: credibility; transferability; dependability; confirmability. In order to ensure credibility, Shenton's recommendations were followed: "random sampling" in avoiding purposive participant selection beyond the criteria that they consider themselves to be readers; "iterative questioning" in having participants complete multiple diary entries and conducting two interviews with each participant; "triangulation" as these studies fit into the wider research project, and as such findings will be comparatively integrated in the discussion chapter. With concern for transferability and dependability, I consider these studies to be a small-scale starting point for how the topics of CT and reading could be approached with broader participant types by further researchers; these studies taken in isolation are not intended to produce generalisable data. Finally, considering confirmability, I aimed to minimise my own researcher bias to as great an extent as is possible. To this end, I opted for open coding approaches instead of the application of any pre-formed coding schemes, to enhance openness in my interpretation of the data. Furthermore, I engaged in reflexivity, both introspective and methodological (Patnaik, 2013); the former by reviewing both my personal experiences and thought processes throughout the studies with a view to bracket my own biases as much as possible (Ahern, 1999). In terms of the latter, I tracked my research procedures across the course of the studies and maintained a stable approach to conducting each interview, to

reading the diary content, and to each round of coding so as to achieve parity across participants.

Limitations

While careful consideration was given to validity so as to ensure that each study was as robustly designed as possible and that the four studies could complement one another and offer an integrated combination of results, this research design nonetheless has limitations. Each individual study design entailed limitations that will be discussed within the chapters describing each study. Taken in conjunction, the overall research design spanning the four studies also has limitations.

While the aim of the mixed methods approach adopted for this research project was to combine and integrate the findings from the four studies, they were nonetheless conducted separately, and are presented in separate chapters of this thesis. This presents a limitation to the extent to which they can be taken in combination as one research contribution, as there is an inherent separation between the different approaches (Halcomb, 2019). In the discussion chapter, overall themes combining all four studies are presented, but there is limitation to this integration. While “seamless” transfer of knowledge derived from different methods into a singular overall mix of findings may be the aim of mixed methods (Castro et al., 2010), it is important to acknowledge the limits to which that is truly achievable.

A limitation present across all four studies is the fact of my undertaking this research project as a sole investigator. This resulted in decisionmaking taking place from a single perspective, and without the potential benefit of multiple researchers’ views and approaches. This is a particularly concerning limitation when it comes to the

coding of the qualitative data in studies three and four, as the codes, categories, and themes derived from the data in these studies came from my own individual readings of the data, and reflect my own particular interpretation. As the sole coder of the data, inter-rater reliability could not be assessed. However, by holding regular meetings with my supervisors during the data analysis phases, I was able to discuss examples of my decisionmaking, and talk through the approaches I was taking as I was in the midst of the analyses. This permitted a level of sense-checking, and broadening of perspective, thus introducing some polyphony into the coding process. However, the results from these studies must be interpreted with this limitation in mind.

There are also limitations in the extent to which this research project can address the research questions derived from the literature review. Rather than seeking to provide definitive answers and fully define a relationship between fiction reading and CT, this research is rather a first step in scoping out the potential interplay between these factors. The ways in which this research is limited in the extent to which it can shed light on this relationship is further discussed in the discussion chapter.

3.2. Practical approaches

3.2.1. Sampling

As each study entailed a specific sampling strategy, these will be discussed separately in the study chapters. However, some overall sampling approaches were adopted. This research was concerned with reading and CT in adults only, and as such only participants who were 18 years old or above were included. Mirroring the move in reading research from the wider impacts of reading towards more specific outcomes associated with different kinds of reading habits (Stanovich & Cunningham, 1993),

sampling started with a wide inclusion of readers and non-readers, and then moved into a focus on readers of different types. Concretely, study one included any participant willing to take part, with no criteria for reading; following this recruitment narrowed down to participants who identified themselves as readers, and subsequent studies focused in on the differences that may be present between those who primarily read fiction or nonfiction. In seeking polyphony, participants from different contexts were sought through varied advertising approaches (e.g. both online posts and physical poster displays).

In questionnaire studies, paying participants has been shown to consistently increase response rate (Ripley, 2006). Participant payments must be balanced between fairness and incentive, and avoiding undue influence; however, recent arguments have been made in favour of higher payments, as undue influence is argued to only occur in a minority of cases (Largent & Lynch, 2017). Therefore in questionnaire based studies, voucher rewards were offered for participation, and participants were compensated as well as possible given available funds. For study one which was very brief, a prize draw was used. For the diary studies which required substantial time commitments, guaranteed vouchers were given to each participant upon completion. Rewards mirrored the scope of sampling; in study one, a voucher for a wide-ranging retailer was offered; in the subsequent studies book vouchers were used.

3.2.2. Piloting

Due to time and resource constraints, full pilot studies were not conducted as part of this research project. However, all four studies were informally tested and refined prior to data collection. For the questionnaire studies, this informal testing

involved sending the online questionnaire forms to my peers and asking for feedback on ease of use, clarity of instructions, general clarity of language, as well as length of time taken to complete each questionnaire. Based upon peer feedback, I then amended the questionnaire forms to maximise clarity and ease of use. This also permitted for more accurate estimates of the time commitment needed to be provided to participants. For study three, the interviews were similarly piloted with volunteers from my peer group. This allowed for a refining of the language used in the interview questions, where any ambiguity or misunderstanding was apparent in the pilot interviews. This also permitted me to ensure my planned questions could be asked within the time limit set for the interviews. Overall, this informal piloting process ensured that all four studies were designed to maximise ease for participants and pre-empted possible issues with clarity or timing.

3.2.3. Methods for assessing reading

In the quantitative studies, participants' reading of fiction and nonfiction were treated as predictor variables; therefore some means of quantifying reading was required. Asking participants to self-report how much they read may be the obvious approach, however as reading is a socially desirable and aspirational activity, participants tend to report higher levels of reading than they have in fact performed. For instance, in P. H. Ennis' (1965) research participants who read as little as 1-4 books per year felt themselves to be "moderate" or "heavy" readers, making self-report of this kind highly dubious, and therefore demonstrating a need for a testing measure. One means of measuring reading without self-report is to use print exposure as an indicator. Stanovich and West (1989) designed a recognition measure in which participants are

asked to check names they recognise from a list of authors that includes foils (i.e. false names); the Author Recognition Test (ART). This aims to eliminate social desirability effects, as guessing on the ART, or trying to check more items so as to present a higher level of reading, would be a poor strategy due to the presence of the false items (Stanovich & Cunningham, 1992). It is important to note that the test does not measure how many of the authors participants have themselves read, merely how many they recognise. However, to recognise the names without reading them one would need to enter into environments where they are present (i.e. bookshops and libraries) or engage with conversation or media where they are featured, and these would be indirect indicators in engagement with reading (Stanovich et al., 1995). Following on from Stanovich et al.'s work, others have diversified the ART to include distinctive fiction and nonfiction subscales (Mar & Rain, 2015), and to add further genre scales (Black et al., 2018), or to make it more regionally specific (Masterson & Hayes, 2007). All of these further validate the efficacy of the ART's design. A variant of the ART was therefore utilised as one means of measuring reading in study one of this research project. However, given the limitations of this measure design, it was only used for an initial observational study that piloted the hypotheses of the research. I aimed to gain an indication of a possible relationship between reading and CT in this way, and then move on to using more fine-grained diary measures to test a causal relationship.

Diary methods

Two of the four studies involved participants keeping a reading diary for a two week period: the first a log of the items they read each day (referred to as the "reading log"); the second a diary describing their experience of the main item they read that

day (referred to as the “reading diary”). Here, the overarching aims in utilising both forms of diary method will be described, with methodological detail given in the separate study chapters. One key aim in using diary methods was to maximise ecological validity; many people routinely keep reading diaries, either logging items or more reflectively writing about their reading, for themselves with no research motivation (Thelwall & Kousha, 2017). In addition to this realism, a daily diary may overcome issues of recall that would arise if asking participants to report on the phenomena of interest after a much longer interval (Alaszewski, 2006; Tourangeau, 2009), although some level of retrospection is always inherent in making a diary entry. The specificity of daily diary reporting ought also minimise desirability effects (e.g. P. H. Ennis, 1965), as reporting concrete daily reading is not open to such estimating as reporting one’s general reading levels.

However, diary paradigms also come with limitations, particularly pertaining to participant motivation. Using a diary paradigm to make comparisons between non-readers and readers would not be suitable due to motivation, as non-readers would have little to report and likely have lower engagement in the subject of the diary (M. C. Smith, 2000). Therefore diary methods were only deployed in studies focusing solely on readers. The issue of motivation also generates limitations based on cost; rewards for diary keeping are suggested to ensure sufficient motivation (Alaszewski, 2006; Nezlek, 2012; M. C. Smith & Stahl, 1993). Cost therefore formed an important factor in the number of participants recruited for the diary studies. Motivation has also been found to be improved if a diary is kept over a period of time when higher rates of change are likely (Bolger et al., 2003), and so reading diaries with varied reading taking

place are more likely to hold participants' interest. In designing the diary studies, I therefore sought to maximise the likelihood of variety by allowing a longer window of time, permitting participants to start their diary at a time of their choosing, and in study two by assigning readings. A two-week period was set for both diary studies, as this has been previously found to permit for sleeper effects to develop (Bal & Veltkamp, 2013), and is a duration considered representative as a sample of reading habits (M. C. Smith & Stahl, 1993).

A further issue with diary methods is that they introduce the diary itself as variable that may alter the participants' normal behaviour (Nezlek, 2012), as the diary may motivate participants to read and therefore log/write more. Participants were therefore able to select an option for no reading completed each day, to remove pressure to read something for the purpose of recording it. Nonetheless, the diaries do constitute a divergence from participants' normal habits. All of these limitations will be taken into account in the conclusions drawn from the diary studies.

Differentiating fiction and nonfiction

In order to compare fiction and nonfiction reading in the reading log (study two) and reading diary (study four), participants' inputs needed to be differentiated. However, as previously discussed, there is no bright line between fiction and nonfiction and many texts fall into a liminal space between them. To make the categorisation process as consensus-driven and independent of the researcher as possible, I opted for using the Goodreads platform (Thelwall & Kousha, 2017). Readers "shelve" books to genres on the platform, and thus classification can be left to popular opinion of users by selecting the most frequent category. However, this

approach is only applicable to books. In the case of other reading material such as blog, news, and magazine articles, no single stable method was possible, and judgements had to be made individually. To ensure as much consistency as possible, general principles were applied, such as considering all biographical work to be nonfiction. While controversial (Freeman, 2003), this decision was felt to be in keeping with the overarching definition adopted for this research project (Currie, 1985).

Engaging in this form of categorisation poses a broader issue, as these categories are not fixed, and material such as biography (among others) is open to interpretation. Thus in study one and two, where fiction and nonfiction are clearly delineated into separate predictor variables, the clarity of the conclusions drawn is compromised. The testing of a hypothesis such as “fiction will have a greater effect than nonfiction” must be viewed with this in mind. It is more accurate to think of this as “material that tends more towards inviting make-believe will have a greater effect than material seeming more to invite belief”, but for the sake of brevity the labels of “fiction” and “nonfiction” are used. This problem of dichotomising fiction and nonfiction for quantitative analysis is at least in part addressed through the qualitative studies. In the interview study (study three), participants were asked about their experiences of fiction and nonfiction reading, without imposing or requiring any definition. This permitted participants to describe their experiences of different materials in the way that they had received them, and to articulate both differences and commonalities in their reading of different texts. In the reading diary study (study four), participants described their reading experiences of any material of their own choice, and these experiences were coded in combination with their descriptions of CT, with fiction and nonfiction only

classified at the end of this process to see if either was more present under any of the codes. This foregrounded the reading experience over the read material. In this way, the mixed methods approach was used to both differentiate fiction and nonfiction in order to make comparisons, and to balance out the artificiality of such a separation by including the broad plethora of reading experience.

3.2.4. 3.2.4. Testing critical thinking

In order to avoid desirability effects and the inaccuracies inherent in our assessments of our own abilities (Kruger & Dunning, 1999), I utilised a test of CT in study two, so as to be able to test for a causal relationship between reading and CT. The ICTET-A (The Foundation for Critical Thinking, 2019) test was selected following a review of available tests that is briefly summarised here.

To select a measure that matches the conception of CT utilised in this research project as a construct that incorporates IL content evaluation approaches, both IL and CT tests were evaluated. To be suitable, the measure needed to be a good match for the construct of CT as described in Box 1; must not be dependent upon specific contexts for use (e.g. limited to HE settings); must be validated; must be freely available for research use. Previous research has identified a lack of validated IL tests that are freely available, and found that all such tests are designed for education (almost exclusively HE) contexts, and contain further specificities limiting their use (Hollis, 2018). Tests of CT were reviewed, and found to entail the same issues as IL tests, and to be very narrowly construed and ill-matched to broader CT conceptions³. Of the

³ e.g.: the Watson-Glaser Critical Thinking Appraisal (WGCTA) (Watson & Glaser, 1980), the Cornell Critical Thinking Test (CCTT) (R. H. Ennis & Millman, 1985); the California Critical Thinking Skills Test (CCTST) (Facione, 1990b), and California Critical Thinking Disposition Inventory (CCTDI) (Facione & Facione, 1992).

reviewed multiple choice CT assessments, only the Halpern Critical Thinking Assessment (HCTA) has a little more scope, as it combines multiple choice with free response options (Halpern, 2007), and thus the free response options for the test items could provide room for the inclusion of the kinds CT factors that are omitted in the other tests, but it is unclear how these would be scored. The test is overall still designed for argument analysis, not for the wider conception of CT, and requires payment for use.

This review of the available tests of CT/IL led me to seek a test that was open-ended rather than multiple choice, and allowed flexibility in terms of the subject matter of the test. I aimed to use a test that permitted for Whitworth's (2014) "polyphony" by permitting various different answers, rather than seeking a single correct formulation. Thus, essay tests or rubric-based assessments were preferred, as a means of providing space for many different ways of responding while still maintaining a fair and validated grading scheme. Within the field of IL, most rubric assessments are highly context and subject specific (e.g. Fagerheim & Shrode, 2010; Makani-Lim et al., 2014; Rosman et al., 2015); thus, these rubrics have the same limitations as IL tests that are designed for HE course assessments. In terms of CT essay tests, those based around a single assigned text carry the same issues of specificity that makes them ill applicable to wider contexts (e.g. R. H. Ennis & Weir, 1985). Ultimately, the International Critical Thinking Essay Test (ICTET) (The Foundation for Critical Thinking, 2019), was selected. This test was designed in line with Paul's (1981) conception of CT, which informed the formulation of the construct utilised in this research project. The test uses a sample text, with a series of questions covering a broad range of CT aspects. As the sample

text can be changed, this permits use across varied contexts. The test has 2 components; Form A of the ICTET has been tested for construct validity and reliability (Hollis et al., in press), and found to be a reliable measure of CT. Therefore the ICTET-A was used in study two.

While the ICTET-A was the most suitable test of CT found at the time of this research, it is not by any means a comprehensive test of the whole of the construct of CT as it is used in this research project (Box 1). The ICTET-A may be used to assess a state instance of evaluating, logically reasoning about, thinking about the points of view of, considering alternative possibilities with relation to, and to some extent reflecting on a piece of information (the set text). However, it cannot be used to assess all of this construct. Firstly, it does not evaluate trait aspects of CT. To compensate for this omission, study one used scales to assess participants trait disposition towards thinking critically, and also their epistemological beliefs, thus covering these elements of CT which the ICTET-A in study two could not. Secondly, the ICTET-A cannot be used to capture the more personal, imaginative, and deeply reflective aspects of this construal of CT. Nor can it assess behaviour or belief change. The way in which it assesses those elements that it does capture is furthermore inherently limited to the fairly brief written answers it elicits to set questions. Therefore the qualitative studies of the adopted mixed methods paradigm were used to address the facets of CT missing from the ICTET-A, and to give space for more in-depth and individual expressions of those elements it does cover. In this way, the four different studies form a mosaic approach to building a picture of CT as a whole.

3.2.5. Interviews

In order to gain insights into readers' experiences of both reading and thinking critically, interviews were conducted in addition to the qualitative reading diary. Using two different approaches to gather reader experiences offered the capture of different ways of framing and describing these experiences. While both interviews and diary entries are necessarily capturing remembered experiences, not the direct experiences themselves, they give a different time scale to the recalled description. A daily reading diary entry is likely to be more representative of reading and CT on a daily basis. For instance, prior research has shown people are more likely to select "highbrow" media for the longer term, while being likely to engage with "lowbrow" media in the short term (Read et al., 1999). Therefore daily descriptions of reading are more likely to capture overall habits. By contrast, interviews may be capturing the broader, more longer term ways in which reading is perceived by participants. This permits for participants to make judgements on the personal significance of different reading and CT experiences and discuss those they select as important and meaningful to them. While the reading and CT taking place on any one given day may not be very important and may be quickly forgotten, when asked to reflect overall participants can recall examples that have been memorable and hold personal meaning. In this way, the interview and diary approach complement each other by encompassing both day-to-day and more holistic experiences.

Semi structured interviews were used as a means of enabling a targeted questioning guided by theory, with an openness to participants' lived experiences (Galletta, 2013). This reflected the design of the overall study, as an argument derived

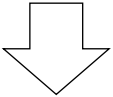
from the literature review presented a possible connection between reading and CT, but this connection required further exploration and interrogation. In practical terms, a set of interview questions were designed based upon the literature review and the findings of the quantitative studies, but participants were also told the interview was more of a conversation and they were encouraged to discuss ideas beyond the questions directly asked. Allowing this divergence also permitted for “openness to transformation” (Whitworth, 2014).

3.3. Overview of the four studies

Four studies were conducted as part of this research project, two quantitative followed by two qualitative. Each study is summarised in a box in the introduction of each study chapter: Box 2: Study one summary, Box 4: Study two summary, Box 8: Study three summary, Box 9: Study four summary. As described above, the research design followed Subedi’s holistic sequential explanatory process. Table 1 shows an adapted version of Subedi’s table (2016, p. 574), with specific detail pertaining to the studies conducted, in order to display the progression of the research from the top of the table to the bottom.

Table 1: Sequence of research design

Sequence	Phase	Procedure	Product
Study one	Observational data collection	Cross sectional survey	Numeric and categorical data
	Observational data analysis	Descriptive and inferential statistics	Models of associations between variables with correlational inferences
Study two	Reading log data collection	Daily log of reading with pre and post CT test	Numeric and categorical data
	Reading log data analysis	Descriptive and inferential statistics	Models of associations between variables with causal inferences
Qualitative design	Connecting quantitative and qualitative phase	Development of research questions, interview questions and diary topics	Interview protocol and diary design
Study three	Reader interviews data collection	Two phase in-depth interview	Textual data
	Reader interviews data analysis	Coding, categorisation of codes, and thematic analysis	Themes, and elucidation of quantitative findings

Study four	Reading diary data collection	Daily descriptive journal of reading and CT experiences	Textual data
	Reading diary data analysis	Iterative coding and categorisation of codes	Categories
	Overall integration	Integration of the quantitative and qualitative results	Interpretation and explanation of all results in relation to one another, and the project's research questions
			Discussion, implications, conclusions, suggestions for future research

Study one was designed so as to establish whether the direction of the overall research project was sensible, and in order to decide on which variables would be worth pursuing further. To this end, it was designed as an observational study, which aimed to establish whether there is a correlation between reading fiction and CT that would be worth pursuing further. It also aimed to assess whether NT ought to be included as a relevant variable in the further studies, and which demographic variables may be of relevance. The results of the observational study affirmed the direction of the research project overall. It furthermore informed the design of study two, and generated avenues for exploration in the qualitative studies.

Study two was designed to test a causal relationship between fiction reading and CT. The aim of testing of a causal relationship informed the choice to use pre- and post- CT testing and the inclusion of experimental manipulation (Shadish et al., 2002). The reading log was chosen as a means of capturing individual differences in reading choices alongside manipulated reading through an assigned text, spanning the two dominant ways of testing the outcomes of reading identified in the literature. The findings from this study form a means of addressing the hypotheses derived from the literature review, and offer new insight into the influence of fiction and nonfiction reading upon CT, offering evidence of a relationship that may be worth pursuing in further research. The results of this study were used to inform the overarching research questions for the qualitative studies, as well as the design of the reading diary and interview protocol.

The two qualitative studies were designed simultaneously, and the reading diary data was collected prior to the reader interviews but analysed after the interviews. Therefore these two studies fall together in the sequence of the research. As the interview data was analysed first, this analysis then informed the approach to the reading diary data. Therefore the interview study is presented prior to the diary study in the chapter sequence of this thesis.

Study three aimed to capture broad ways that readers experienced their reading and CT, and their views on any relationship between them. Reader interviews chosen to explore readers' experiences within the framing of these research aims, with space for participants' to voice wider ideas and experiences. Findings from the interviews both corroborated the initial links proposed as a result of the literature, and also

expanded them. Furthermore, the findings provided some explanation for the associations found in the quantitative studies. The themes derived from this study offer ways of understanding reading and CT that could be valuable in future research formulations. The findings from this study informed some of the decision making in the data analysis of the reading diaries.

Finally, study four aimed to capture day-to-day reading and CT experiences to give a different time frame from study three and complement the bigger-picture interview discussion with daily detail. To this end, the diary design was chosen to collect daily experiences of reading and CT, such that these could be combined to seek commonalities between them. By taking reading and CT together in this way, ways of engaging in both emerge in the resulting categories. These suggest shared reading and CT approaches that may offer a conceptualisation of the interplay of these activities that suggests avenues for further investigation. The findings from this study completed the data collection and analysis cycles of this research project.

In order to connect these four studies together and integrate their findings, a thematic approach was taken (Castro et al., 2010). Themes were identified that could span the findings from all four studies, and these themes were used to “recontextualise” (Castro et al., 2010, p. 354) findings by grouping the statistical trends and participant experiences together. By relating the quantitative and qualitative findings under theme headings in this way, the boundaries of individual studies were broken and an overall understanding emerged.

Taken together, these studies offer some possible answers to these questions: does reading fiction influence CT? Is such an influence distinct from that of reading

nonfiction? How do people experience these influences? How do these influences manifest in day to day reading and CT experiences? Furthermore, these studies give indications of how NT fits into the relationship between reading and CT, along with other exploratory avenues.

3.3.1. Ethical issues

Ethical approval for all four studies was granted by the UCL Research Ethics Committee: study one: 15397/001; study two: 15397/002; studies three and four: 15397/003.

As there was scope for sensitive or personal topics to be raised across the studies, especially given the nature of fictional material and how it can prompt emotional and deeply personal responses (e.g. Djikic et al., 2009b), in each of the studies care was taken in study design to ensure participants could easily quit at any time, or could have as much flexibility as needed to accommodate individual responses. This is detailed in each study chapter.

No participants raised any concerns or issues with the researcher, the supervisory team, or the ethics committee resulting from any of the studies conducted. Only one issue arose during a study: in study two, one participant noted disapproval in a free text entry box at having been asked to read Arthur C. Clarke as an assigned text, as they believed Clarke to have been guilty of paedophilia. However, based on press reporting of historical accusations against Clarke being disproven (Associated Press, 2008), and given the participant did not make use of any available channel to reach out to the research team or ethics committee about the issue, and the fact the participant was anonymous, no action was taken as a result.

4. Study one: An observational investigation into the relationship between print exposure and factors of critical thinking

This study aimed to investigate how different levels of exposure to fiction and nonfiction impact individuals' critical thinking disposition (CTD) and their epistemological orientation (EO). These outcome variables were selected as key facets of critical thinking (CT), as discussed in the literature review. The study also investigated the role of narrative transportation (NT) as a mediator between reading and its outcomes. The variables of this study are listed in Box 3 for reference. It aimed to obtain observational, correlational data to give an initial indication of whether there is a relationship between the variables of interest worth pursuing, as a first step in this research project. It also aimed to elucidate which variables of interest are most relevant for further investigation in the following studies, thus informing the design of the further research. Box 2 summarises the study.

Box 2: Study one summary

This study aimed to test correlations between reading exposure, and EO and CTD as factors of CT. Self-selecting participants ($N = 335$) completed an online survey including an author recognition test and self-report scales. Total print exposure did not improve model fit for either outcome. Fiction scores were significantly associated with higher critical thinking disposition. Fiction readers were more likely to be evaluativist than absolutist, and nonfiction vice-versa. Nonfiction readers were more likely to be absolutist than multiplist, and evaluativist than multiplist. Narrative transportation mediated the relationship upon critical thinking disposition, but not on epistemological orientation.

Hypotheses:

H₀: print exposure will not have any significant correlation with either levels of critical thinking disposition, or with sophistication of epistemological orientation.

H₁: print exposure will correlate with stronger critical thinking disposition.

H₂: print exposure will correlate with more sophisticated epistemological orientation.

H₃: fiction reading will correlate more strongly than nonfiction reading with stronger critical thinking disposition.

H₄: fiction reading will correlate more strongly than nonfiction reading with more sophisticated epistemological orientation.

H₅: narrative transportation will mediate the relationships predicted in H₁ through H₄.

This study was preregistered: <https://osf.io/97pmj>

A paper based upon this study has been published:

Hollis, H. (2021). An investigation into the relationship between fiction and nonfiction reading exposure, and factors of critical thinking. *Scientific Study of Literature*, 11(1), 108–141. <https://doi.org/10.1075/ssol.20014.hol>

The data from this study has been published:

Hollis, H. (2021). *Fiction and Nonfiction print exposure, Narrative Transportation, Critical Thinking Disposition, and Epistemological Orientation, questionnaire responses* [Data set]. University College London. <https://doi.org/10.5522/04/14743347.v1>

Box 3: Study one variables

Outcome variables:

- CTD: critical thinking disposition
- EO: epistemological orientation:
 - Absolutist
 - Multiplist
 - Evaluativist

Predictor variables:

- Print exposure:
 - Fiction score
 - Nonfiction score
 - Total score

Mediator:

- NT: narrative transportation

Control variables:

- Age
- Gender
- Educational level

4.1. Method

4.1.1. Participants

Adult (18 years of age or above) participants were recruited through varied avenues. Online calls for participation were posted on social media (twitter and reddit), as well as online study promotion websites (callforparticipants.com, Survey Circle, Survey Swap). The study was also promoted in London public places such as coffee shops, supermarkets, community centres, public libraries (in short, any venue hosting a public notice board) using posters and bookmarks, to reach a wider sample including those who may not engage with the utilised online platforms. Participants self-identified as being fully proficient in English, and no further restrictions were placed on participation. A £20 amazon.co.uk voucher was offered a prize for 10 randomly selected participants. A ratio of participants to scale items was used determine a target sample size (Anthoine et al., 2014). Dividing the ART into separate genres (i.e. each genre being treated as a separate variable), resulted in a maximum of 15 predictor variables in total, and thus applying a ratio of 10:1 participants to scale items would call for 150 participants as a minimum. 200 was selected as the target participant number to be in excess of this minimum value. No cap was placed on the maximum number of participants. A total of 353 responses were obtained. From these, 18 were removed for incompleteness, yielding 335 responses.

4.1.2. Measures

Demographic information (*age, gender, educational level*) was collected using open questions, and responses for *gender* and *educational level* were then coded into categories. All of the following measures are available in Appendix A under Measures.

The ART-G was used as a measure of print exposure (Mar & Rain, 2015), 2015). This is checklist of 200 author names, including 40 foil items. Participants were asked to check those names in the list which they recognised to be authors. The scale was scored as a numerical variable, yielding the following scores: *total score*, i.e. the number of correctly identified authors minus foils checked (160 items in total); *fiction score*, i.e. the number of fiction authors checked (110 items in total); *nonfiction score*, i.e. the number of nonfiction authors checked (50 items in total).

The Fantasy subscale of the Interpersonal Reactivity Index (IRI-F) was used as a measure of *NT* (Davis, 1983). The scale was selected in line with other similar research on reading and its outcomes, and in keeping with prior studies one item of the scale was omitted which does not pertain to immersion in a narrative, and the removal of which has been found to increase its internal reliability (Mar et al., 2009). This resulted in a 6 item Likert scale, with two items reverse coded, with 5 options of agreement rating from strongly disagree to strongly agree. No instruction for scoring and division into categories is present in the manual (Davis, 1983), and previous studies have treated scores from the IRI-F as a numerical variable (Mar et al., 2009). However, for the purposes of this study the IRI-F was scored as a categorical variable, in keeping with the view that Likert scales do not represent interval levels of measurement (Jamieson, 2004). Responses were categorised using the SD and mean as cut-offs to create 4 categories of scores (*low, moderately low, moderately high, high*).

The Critical Thinking Disposition Scale (CTDS) was used as a measure of *CTD* (Sosu, 2013). This is a Likert scale with 1-5 agreement rating, containing 11 items.

Scoring was performed by total score with subdivision into three categories (*low, moderate, high*) based upon predetermined scoring boundaries in the test manual.

The Justifying Conclusions Inventory (JCI) was used as a measure of *EO* (McGinnis, 2016). This is a Likert scale with 1-6 agreement rating, containing 23 items. The items represent three categories of *EO*: *absolutism, multiplism, evaluativism*. Scoring was performed using mean responses for each category of *EO*. Cluster analysis was then used to group participants into response categories.

4.1.3. Procedure

Responses were collected between 25/06/2019 and 17/07/2019. Participants followed a link to the online survey, where they were presented with the information sheet and consent form, including questions checking they were 18 years of age or above, and considered themselves fully proficient in English. Upon consenting and meeting the inclusion criteria, participants were asked to complete demographic information. They were then given the scales in the following order: IRI-F; ART-G; CTDS; JCI. Finally, participants were offered the option of signing up for the prize draw, and debriefed. All responses were fully anonymous.

Participants were excluded based on how many foil items in the ART-G they checked, so as to remove responses with excessive guessing. The cut-off for foil checking was set at 1 *SD* above the mean; the mean number of foils checked was 0.4, with a *SD* of 1.22, therefore 2 items was the maximum permitted number of foils checked. Based on this, 11 participants were excluded. A further 3 participants were removed as they did not select any items in the ART-G. This yielded 321 responses.

Furthermore, 10 participants did not complete the JCI, which was presented last in the order of items in the survey, though they did complete all other parts of the survey. These responses were kept for analysis with *CTD* as the outcome variable, but excluded for analysis of *EO* as the outcome variable.

Finally, participants who did not disclose their *gender* ($n = 9$) or identified as neither male nor female ($n = 1$) were excluded in models with *gender* as a variable. This is due to the low sample size these *gender* categories would contain if included in the analysis.

Analysis was conducted using R 4.0.4. in R Studio version 1.4.1103.

4.2. Results

4.2.1. Descriptive statistics

Demographics

As some participants did not complete the JCI, a smaller subset ($n = 311$) was used in analyses with *EO* as the outcome variable. Participants who completed both scales ($n = 321$) were used in analyses with *CTD* as the outcome variable. All following descriptive statistics (unless otherwise stated) will be of the larger sample of participants ($n = 321$), including those who did not complete the JCI.

Participants had a mean *age* of 35 ($SD = 12.64$), median 30, with the youngest being 18 and oldest 77.

Participants included 216 female; 86 male; neither 1; undisclosed 8.

Although initially the ONS coding of *educational level* was used (Office for National Statistics, 2011), this did not fit the distribution of education in the sample

well as it yielded a very large sample in the category of “bachelor’s and above”, with very small samples in the lower levels. This indicates that participants in this study represent a highly educated sample, more so than the wider population. *Educational level* was therefore coded into the following categories: *GCSE or Equivalent and Vocational*: 8% ($n = 26$); *A level*: 7% ($n = 23$); *Bachelor’s degree*: 34% ($n = 111$); *Master’s degree*: 42% ($n = 132$); *Doctoral*: 9% ($n = 29$).

ART-G

The ART-G *total score* is the total number of correct items selected, minus the number of foils selected. The mean *total score* was 30.62 ($SD 22.78$), with a median of 26. The minimum total score was 1, and the maximum 116. As the maximum possible *total score* was 160, the mean response represents a 19% correct recognition rate.

The mean *fiction score* checked was 23.57 ($SD 18.57$), with a median of 19, and responses ranging from 0 to 92. As the maximum possible *fiction score* was 110, the mean response represents a 21% correct fiction recognition rate. The mean *nonfiction score* was 7.26 ($SD 7.29$), with a median of 5, and responses ranging from 0 to 45. As the maximum possible *nonfiction score* was 50, the mean response represents a 15% correct nonfiction recognition rate.

Figures 1-3 show ART-G score distributions.

Figure 1: Histogram of ART-G total scores

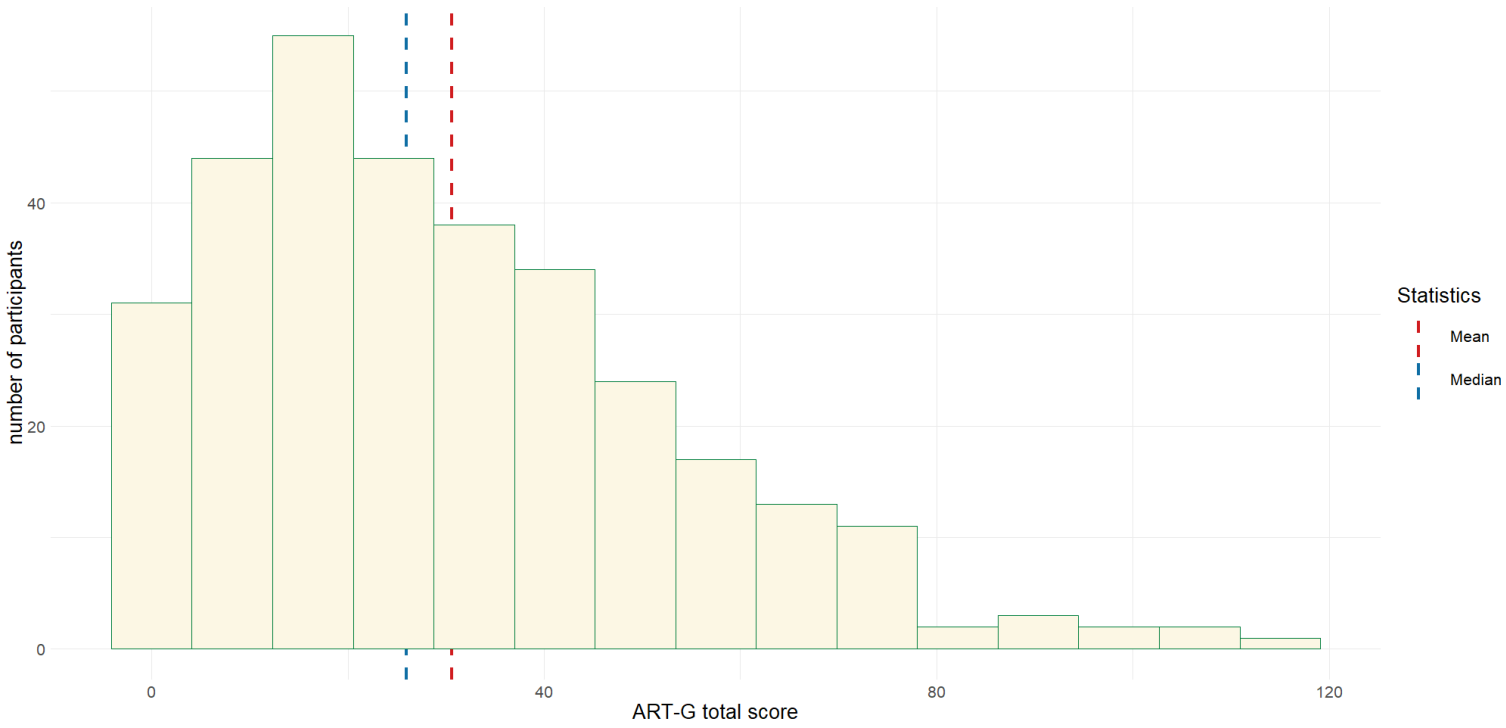


Figure 2: Histogram of ART-G fiction scores

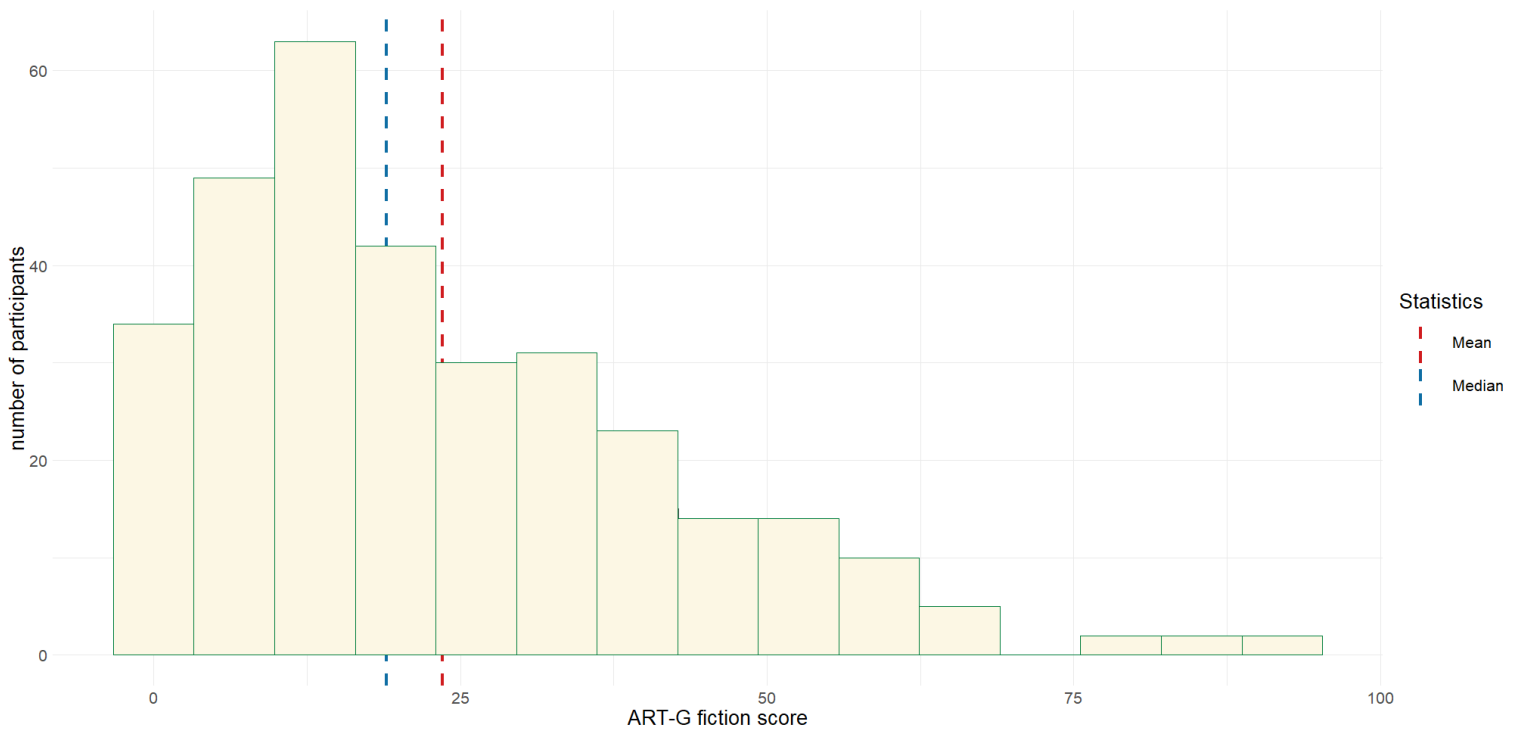
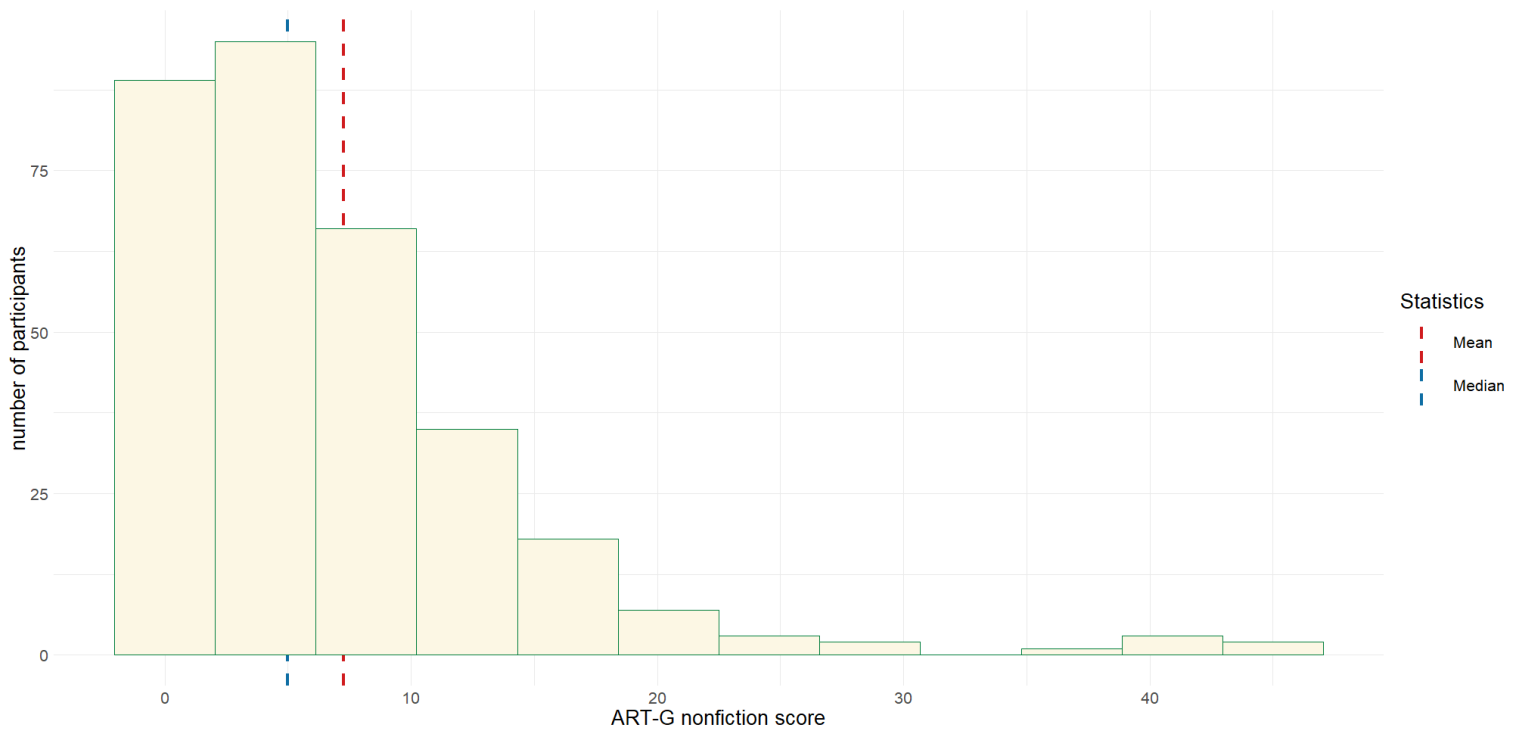


Figure 3: Histogram of AR-G nonfiction scores



IRI-F

The mean score on the IRI-F was 24.11, with a *SD* of 4.78. This was used to determine cut-off points for categorisation of responses, such that 4 *NT* categories were created with cut-offs at 1 *SD* rounded to the nearest whole number below the mean (19), the mean (24), and 1 *SD* above the mean (29). This allowed for division of participants into categories of *low* ($n = 57$), *moderately low* ($n = 96$), *moderately high* ($n = 134$), and *high* ($n = 34$). This means that of the total sample, 18% were *low*, 30% *moderately low*, 42% *moderately high*, and 11% *high* on *NT*. *Low* was used as the reference category.

CTDS

Based on predetermined score cut-off points, participants were grouped into *low* ($n = 17$), *moderate* ($n = 118$) and *high* ($n = 186$) *CTD* levels; 5% were *low*, 37% were *moderate* and 58% were *high* in their *CTD*. *Low* was used as the reference category.

JCI

The subset ($n = 311$) of participants who completed the JCI was used for these analyses. In order to derive *EO* categories, mean responses to the three item types (*absolutism*, *multiplism*, *evaluativism*) were grouped using cluster analysis. Using the NbClust function which provides 30 indices determining the best number of clusters for a given data set (Charrad et al., 2014), 3 clusters were found to be the best fit in this case. *K*-means cluster analysis was therefore used to cluster data into 3 groups. The cluster centres for each grouping are shown in Table 2.

Table 2: Cluster centres for *EO* scores

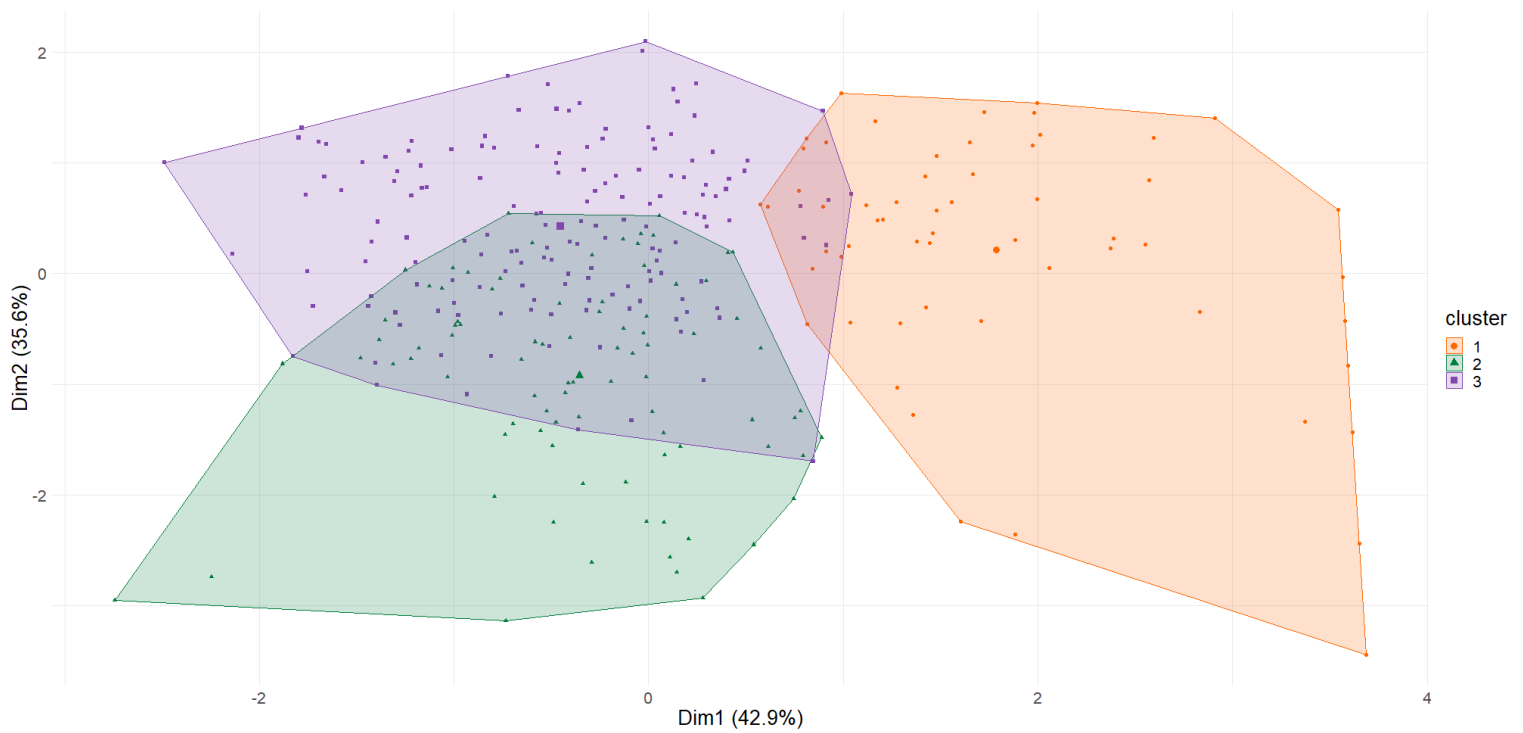
	Absolutism	Multiplism	Evaluativism
1: Absolutist	4.31	2.04	5.28
2: Multiplist	2.97	3.95	4.68
3: Evaluativist	1.82	3.05	5.25

Based on this analysis, evaluativism has the highest mean score across all clusters. However, cluster 1 can be characterised as *absolutist* with low multiplism and high evaluativism; cluster 2 can be characterised as *multiplist* with relatively lower evaluativism and moderate absolutism; cluster 3 can be characterised as *evaluativist* with

moderate multiplism and low absolutism. For simplicity, the clusters have been named *absolutist*, *multiplist*, and *evaluativist* respectively, and have been ordered in line with prior theory ranking *EO* sophistication from *absolutist*, through *multiplist*, to *evaluativist* (Hofer & Pintrich, 1997). However, these clusters are not precisely the same as those conceptual groupings. The subtleties of their makeup must be kept in mind, and will be considered in the discussion.

Cluster 1, (*absolutist*) contained 59 participants (19% of the sample), cluster 2 (*multiplist*) contained 89 (29%), and cluster 3 (*evaluativist*) contained 163 (52%). The total variance in the data (the total sum of squares) was 854.58. Within-cluster sum of squares was 105.06 in cluster 1, 140.39 in cluster 2, and 172.97 in cluster 3. The total within-cluster sum of squares was 418.42. The total between-cluster sum of squares was 436. Overall, 17.51% of variance within the data set was explained by the clustering. Figure 4 shows a cluster plot of how these clusters fit the data. This shows a stronger overlap between cluster 2 (*multiplist*) and cluster 3 (*evaluativist*), than the overlap between cluster 3 (*evaluativist*) and cluster 1 (*absolutist*). This is in keeping with the theoretical framing of *EO*, where multiplism represents a transitional phase between absolutism and evaluativism (Kuhn et al., 2000).

Figure 4: Cluster plot by principle components



Note. Cluster 1 = absolutist; 2 = multiplist; 3 = evaluativist

4.2.2. Modelling critical thinking disposition

As a result of a model comparison process detailed in Appendix A under CTD Model Selection, the model which best fit the data included *fiction score* interacting with *NT*; *nonfiction score* interacting with *NT*; *fiction score* interacting with *educational level*; *nonfiction score* interacting with *educational level*. As *gender* was not included as a variable, there was no need to remove participants who did not disclose their *gender* or were neither male nor female. Therefore all participants who completed the *CTD* scale and were not excluded for ART-G foil checking were included in the model ($n = 321$). The results of this model are shown in Table 3.

Table 3: Results summary for cumulative link model for CTD

Variable	<i>B</i>	<i>SE</i>	<i>p</i> value	Lower 95% CI	OR	Upper 95% CI
Fiction	0.19	0.05	< .001***	1.1	1.21	1.37
Moderately Low NT	1.79	0.67	.007**	1.66	6.01	22.96
Moderately High NT	1.75	0.66	.008**	1.62	5.78	21.69
High NT	2.00	0.99	.043*	1.09	7.42	55.15
Nonfiction	-0.01	0.03	.74	0.92	0.99	1.06
A level	3.07	1.34	.022*	1.68	21.45	351.1
Bachelor's	3.97	1.15	.001***	6.27	52.99	627.26
Master's	4.16	1.17	<.001***	7.21	63.81	784.41
Doctoral	4.51	1.47	.002**	5.67	91.04	1899.96
Fiction:						
Moderately Low NT	-0.13	0.03	<.001***	0.82	0.87	0.93

Fiction:						
Moderately High	-0.06	0.03	.075	0.89	0.95	1
NT						
Fiction: High NT	-0.07	0.06	.194	0.83	0.93	1.05
Nonfiction:						
Moderately Low	0.33	0.09	<.001***	1.19	1.4	1.67
NT						
Nonfiction:						
Moderately High	0.14	0.08	.069	1	1.15	1.35
NT						
Nonfiction: High	0.46	0.25	.064	1.07	1.58	2.96
NT						
Fiction: A level	-0.11	0.06	.082	0.79	0.9	1.01
Fiction:						
Bachelor's	-0.12	0.05	.017*	0.78	0.88	0.97
Fiction: Master's	-0.10	0.05	.046*	0.8	0.91	0.99
Fiction: Doctoral	-0.11	0.06	.051	0.79	0.89	1
Nonfiction: A	-0.18	0.12	.125	0.63	0.84	0.99
level						

Nonfiction: Bachelor's	-0.16	0.09	.076	0.71	0.85	1.02
Nonfiction: Master's	-0.16	0.06	.015*	0.74	0.86	0.96
Nonfiction: Doctoral	-0.09	0.06	.143	0.79	0.91	1.03

Note. CIs are exponentiated to reflect the OR. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

The model was tested for accuracy and generalisability, and results are presented in Appendix A under Model accuracy testing on page 399. The following were the significant effects:

Fiction: *fiction score* predicted increases in *CTD* ($b = 0.19, p < .001$). The odds ratio tells us that per one item increase in *fiction score*, the change in odds of increasing a *CTD* level is 1.21.

Narrative transportation: as compared to *low NT*, *moderately low NT* significantly predicted increase in *CTD* ($b = 1.79, p < .01$), with a change in the odds of increasing a *CTD* level of 6.01. Similarly, being *moderately high* as opposed to *low* on *NT* also significantly predicted increase in *CTD* ($b = 1.75, p < .01$), with a change in the odds of increasing a *CTD* level of 5.78. Finally, having *high NT* compared to *low NT* significantly predicted increase in *CTD* ($b = 2, p < .05$), with a change in the odds of increasing a *CTD level* of 7.42.

Educational level: having any of the higher levels of education as opposed to *GCSE or Vocational* significantly predicted higher *CTD* level. However all

educational level effects had extremely broad 95% CIs. This indicates insufficient information about this effect from this data set.

Fiction interaction with NT: Being *moderately low* on *NT* rather than *low* interacted with *fiction score*, and significantly predicted difference in *CTD* in the opposite direction than *fiction* or *moderately low NT* alone ($b = -0.13, p < .001$), with a change in odds of being in a higher *CTD* level of 0.87. This means that as *fiction score* increases, *moderately low NT* participants become less likely to have a higher *CTD* level than *low NT* participants.

Nonfiction interaction with NT: Being *moderately low* on *NT* rather than *low* interacted with *nonfiction score*, and significantly predicted difference in *CTD* level ($b = 0.33, p < .001$), with a change in odds of being in a higher *CTD* level of 1.4. This means that as *nonfiction score* increases, *moderately low NT* participants become more likely to have a higher *CTD* level than *low NT* participants.

Fiction interaction with educational level: Having a *Bachelor's degree* rather than *GCSE or vocational educational level* interacted with *fiction score*, and significantly predicted decrease in *CTD* level ($b = -0.12, p < .05$), with the change in odds of being in a higher *CTD* level of 0.88. Similarly, having a *Master's degree* rather than *GCSE or vocational educational level* interacted with *fiction score*, and significantly predicted decrease in *CTD* level ($b = -0.1, p < 0.5$), with a change in odds of being in a higher *CTD* level of 0.91. This means that as *fiction score* increases, *bachelor's* and *master's* educated participants become less likely to have a higher *CTD* level than *GCSE or vocational* educated participants.

Nonfiction interaction with Educational level: Having a *Master's degree* rather than *GCSE or vocational educational level* interacted with *nonfiction score*, and significantly predicted decrease in *CTD* level ($b = -0.16, p < .05$), with a change in odds of being in a higher *CTD* level of 0.86. This means that as *nonfiction score* increases, *Master's* educated participants become less likely to have a higher *CTD* level than *GCSE or vocational* educated participants.

Finally, structural equation modelling (SEM) with weighted least squares estimation and bootstrap standard error computation was used to test mediation by *NT* on fiction and nonfiction for an equivalent of the best fitted cumulative link *CTD* model (i.e. a model including *fiction score*, *nonfiction score*, and *educational level* as predictors). In order to be able to use SEM, all categorical variables were recoded to be ordered. *Fiction score* had an insignificant direct effect ($p = .073$), a significant indirect effect via *NT* ($p < .001$), and a significant total effect ($p = .001$). Therefore *NT* does mediate the relationship between *fiction score* and *CTD*. *Nonfiction score* had an insignificant direct effect ($p = .436$), a significant indirect effect via *NT* ($p < .001$), and a significant total effect ($p = .012$). Therefore *NT* also mediates the relationship between *nonfiction score* and *CTD*.

4.2.3. Modelling epistemological orientation

As a result of a model comparison process detailed in Appendix A under EO Model Selection, the model which best fit the data included *fiction score* interacting with *NT*, *gender*, and *nonfiction score*. The model includes only those participants who completed the JCI ($n = 311$). The results of this model comparing *multiplist* against *evaluativist* are shown in Table 4.

Table 4: Summary of results from generalised linear model of EO multiplist vs evaluativist

Variable	<i>B</i>	<i>SE</i>	<i>p</i> value	Lower 95% CI	OR	Upper 95% CI
Fiction	0.02	0.03	.434	0.97	1.02	1.08
Nonfiction	1.12	0.05	.010**	1.03	1.13	1.24
Male	0.59	0.38	.122	0.87	1.8	3.90
Moderately Low NT	0.98	0.70	.165	0.68	2.66	10.91
Moderately High NT	0.13	0.71	.859	0.28	1.14	4.72
High NT	-0.36	1.02	.725	0.08	0.70	4.90
Fiction: Moderately Low NT	-0.03	0.03	.247	0.91	0.97	1.02
Fiction: Moderately High NT	0.01	0.03	.658	0.95	1.01	1.07
Fiction: High NT	0.05	0.05	.285	0.96	1.05	1.18

Note. CIs are exponentiated to reflect the OR. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Nonfiction: The only significant effect found for the difference between being *multiplist* vs *evaluativist* was that of nonfiction. *Nonfiction score* significantly predicted being an *evaluativist* rather than a *multiplist* ($b = 1.12, p < .01$). Per one item increase in *nonfiction score*, the change in odds of being an *evaluativist* rather than a *multiplist* was 1.13. Therefore as nonfiction reading increases, it becomes more likely that one will fall into the *evaluativist* category rather than *multiplist*.

The same model that was selected based upon fitting the *multiplist* vs. *evaluativist* data was then applied to the other *EO* levels. The results for *absolutist* vs. *evaluativist* are shown in Table 5 and for *absolutist* vs. *multiplist* in Table 6.

Table 5: Summary of results from generalised linear model of EO Absolutist vs Evaluativist

Variable	<i>B</i>	<i>SE</i>	<i>p</i> value	Lower 95% CI	OR	Upper 95% CI
Fiction	0.07	0.04	.044*	1.01	1.07	1.16
Nonfiction	-0.07	0.03	.010*	0.88	0.93	0.98
Male	0.01	0.37	.972	0.49	1.01	2.15
Moderately Low NT	1.20	0.82	.145	0.67	3.32	17.39
Moderately High NT	1.80	0.85	.033*	1.17	6.05	33.28
High NT	0.66	1.26	.601	0.17	1.93	26.00
Fiction:						
Moderately Low NT	-0.06	0.04	.106	0.86	0.94	1.01
Fiction:						
Moderately High NT	-0.07	0.04	.077	0.86	0.94	1.00
Fiction: High NT	-0.03	0.05	.501	0.88	0.97	1.06

Note. CIs are exponentiated to reflect the OR. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Fiction: *Fiction score* significantly predicted being an *evaluativist* rather than an *absolutist* ($b = 0.07, p < .05$). Per one item increase in *fiction score*, the change in odds of being an *evaluativist* rather than an *absolutist* is 1.07. Therefore as fiction reading increases, it becomes more likely that one will fall into the *evaluativist* category rather than *absolutist*.

Nonfiction: *Nonfiction score* significantly predicted a difference between being *absolutist* and *evaluativist* ($b = -0.07, p < .01$). Per one item increase in *nonfiction score*, the change in odds of being an *evaluativist* rather than *absolutist* is 0.93. Therefore, as nonfiction reading increases, it becomes less likely that one will fall into the *evaluativist* category rather than be an *absolutist*.

NT: Being *moderately high* on *NT* rather than *low* significantly predicted a difference between being an *absolutist* or *evaluativist* ($b = 1.80, p < .05$), with a change in odds of being an *evaluativist* rather than an *absolutist* of 6.05. Therefore being *moderately high* rather than *low* on *NT* makes being an *evaluativist* more likely than an *absolutist*.

Table 6: Summary of results from generalised linear model of EO absolutist vs multiplist

Variable	<i>B</i>	<i>SE</i>	<i>p</i> value	Lower 95% CI	OR	Upper 95% CI
Fiction	0.03	0.04	.555	0.94	1.03	1.13
Nonfiction	-0.13	0.04	.004**	0.80	0.88	0.95
Male	-0.77	0.49	.116	0.18	0.46	1.21
Moderately Low NT	-0.58	0.93	.535	0.08	0.56	3.40
Moderately High NT	1.09	1.0	.277	0.41	2.97	22.16
High NT	2.80	2.09	.180	0.53	16.52	3291.44
Fiction:						
Moderately Low NT	-0.01	0.05	.876	0.90	0.99	1.08
Fiction:						
Moderately High NT	-0.06	0.05	.184	0.84	0.94	1.03
Fiction: High NT	-0.18	0.10	.057	0.66	0.83	0.98

Note. CIs are exponentiated to reflect the OR. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Nonfiction: The only significant effect for the *absolutist* vs *multiplist* model was that of nonfiction. *Nonfiction score* significantly predicted being an *absolutist* rather than a *multiplist* ($b = 0.13, p < .01$). Per one item increase in *nonfiction score*, the change in odds of being a *multiplist* rather than an *absolutist* is 0.88. Therefore as nonfiction reading increases, it becomes more likely that one will fall into the *Absolutist* category rather than *multiplist*.

All models were tested for accuracy and generalisability, and results are presented in Appendix A under Model accuracy testing (Table 4 model on page 400; Table 5 model on page 401; Table 6 model on page 402)

SEM with weighted least squares estimation and bootstrap standard error computation was used to test mediation by *NT* on fiction and nonfiction for each *EO* model. In order to be able to use SEM, all categorical variables were recoded to be ordered. For *multiplist* versus *evaluativist*, only *nonfiction score* was tested as this had a significant effect in the model (fiction did not). *Nonfiction score* had a significant direct effect ($p = .037$), an insignificant indirect effect via *NT* ($p = .747$), and a significant total effect ($p = .043$). Therefore *NT* does not mediate the relationship between *nonfiction score* and *multiplist* versus *evaluativist EO*. For *absolutist* versus *evaluativist*, both nonfiction and fiction were tested as both had significance in the model. No effect was found for *fiction score* either as a direct effect ($p = .171$) nor as an indirect effect via *NT* ($p = .292$), or a total effect ($p = .063$); for *nonfiction score* the direct effect was significant ($p = .003$), the indirect effect via *NT* was insignificant ($p = .310$), and the total effect was significant ($p < .001$). Therefore *NT* does not mediate the relationships between the *fiction* or *nonfiction score* and *absolutist* versus

evaluativist EO. For *absolutist* versus *multiplist*, again only *nonfiction score* was tested, and this had a significant direct effect ($p = .004$), an insignificant indirect effect via *NT* ($p = .947$), and a significant total effect ($p = .002$). Therefore *NT* does not mediate the relationship between *nonfiction score* and *absolutist* versus *multiplist EO*.

4.3. Discussion

This study investigated the impact of total, fiction, and nonfiction print exposure on CTD and EO, taking account of NT.

It was hypothesised that print exposure would predict increased CTD and more sophisticated EO levels, however this hypothesis is not supported by the data in this study as total print exposure was not found to improve model fit when included for either outcome variable. Within the different tested models, fiction and nonfiction had distinctive and sometimes contradictory effects, suggesting that combining them into one overall score would not be appropriate. This finding indicates that fiction and nonfiction are best assessed independently in terms of their relationship with CTD and EO, and thus potentially with CT.

Secondly, it was hypothesised that fiction readers would have higher CTD levels, and more sophisticated EO, and that this relationship would be stronger than for nonfiction readers. In the case of CTD, the hypothesis is supported by the data in this study, as fiction had a significant association with increased CTD level but nonfiction did not. This implies fiction readers are more disposed to think critically, and this may be reflected in their CT ability. In the case of EO, the hypothesis is partially supported. If we assume the developmental trajectory from absolutism, through multiplism, and into evaluativism is a move from less to more sophisticated EO (Hofer & Pintrich,

1997), then evaluativists are the most sophisticated EO category. Evaluativist EO is also considered necessary for CT (Kuhn et al., 2000). Higher fiction scores were associated with increased odds of being an evaluativist rather than an absolutist, suggesting fiction readers are more likely to have the orientation best suited to CT. Conversely, higher nonfiction scores were associated with a decrease in odds of being an evaluativist rather than an absolutist. These findings support the hypothesis. However, fiction scores were not significantly different between multiplists and the other EO categories.

Furthermore, nonfiction readers were more likely to be absolutist than multiplist, but also more likely to be evaluativist than multiplist. In the latter case, nonfiction therefore predicted more sophisticated EO where fiction did not. Therefore the relationship between nonfiction reading and EO is more complex, and only partially fits the hypothesis.

Finally, it was hypothesised that NT would mediate the effects of fiction and nonfiction on CTD and EO. In the case of CTD, the hypothesis is supported. However, for EO the hypothesis is not supported, as no mediation effect by NT was shown. This shows an interesting difference in the role played by NT in the two outcome variables. In the case of CTD, how transported one is by reading alters how their reading influences their CTD. In the case of EO, what one reads is influential regardless of how transported one is. In addition to the mediation effects, the models that best fit the data for both CTD and EO as outcomes showed significant effects of NT. Compared with having low NT, having higher NT levels increases the odds of having higher CTD. Moving from low to moderately low NT also interacts with both fiction and nonfiction scores to yield increased odds of higher CTD. Additionally,

being moderately high rather than low in NT also increases the odds of being an evaluativist rather than an absolutist. Therefore NT plays a meaningful role in and of itself as a variable, not only as a mediator for reading.

4.3.1. Limitations

Participants in this study do not represent a general population. Firstly, the sample was 70% female, meaning the results may be more generalisable to women than to men. This is not surprising, given a broader trend in women being more likely to respond to online surveys (Cheung et al., 2017; Porter & Whitcomb, 2005). Furthermore, the sample was highly educated, with 85% holding a degree; in contrast, 27% of the UK population held a degree or above qualification at the last census (Office for National Statistics, 2011). Therefore the findings from this study are likely to only generalise to more educated populations. Again, this is in keeping with known bias in survey responses, where individuals with higher educational levels are more likely to volunteer (Cheung et al., 2017). Interestingly, the age of participants in the sample was wide ranging (18-77), and the mean age of 35 implies the survey bias towards older people being more likely to volunteer (Cheung et al., 2017) did not apply in this case. Overall, the findings of this study must be interpreted with the demographics of participants in mind, and thus cannot be decontextualised.

This limitation in participant diversity in terms of educational level plays into the findings pertaining to education and CTD. Increased educational level was significantly associated with increased CTD, but with wide CIs that render this finding insufficiently supported by the data in this study. If this finding could be supported by the data, it would be in keeping with arguments that a key aim of higher education is

to promote CT, and with framings of CT as something to be taught (e.g. Bailin et al., 1999). However, further investigation would be required to establish this relationship, with a more representative sample. Furthermore, a somewhat paradoxical effect of educational level interacting with fiction exposure was shown. While fiction exposure alone was positively associated with CTD ($b = 0.19$), this effect was reversed in participants with a bachelor's degree ($b = -0.12$) and master's degree ($b = -.10$) in contrast to those without degrees. While nonfiction alone was not significantly associated with CTD, nonfiction exposure in those with a master's degree had a negative association with CTD ($b = -0.16$). It would be possible to offer explanations of these findings, for example it may be a case of lower motivation in those who read more, who also exerted a lot of energy in their educational programmes, and thus perhaps feel disinclined to think critically in addition to those already mentally taxing activities. However, any such explanation is purely speculative as this study was purely cross-sectional and gathered no data on motivation or other possible factors. As the reference category of GCSE or Equivalent and Vocational was so small ($n = 26$, 8% of the sample), these contrasts against that category cannot be viewed as robust. Overall, the findings pertaining to educational level are useful indicators that education may play a role in CTD, but no more than that.

ART-G scores give an indication of the extent to which participants are exposed to author names, but do not measure actual reading levels, and thus the results of this study do not capture actual reading behaviour. Furthermore, the ART-G was designed and validated in Canada, and therefore the authors used in it may be more familiar to a Canadian audience than participants from other countries; this may be one

explanation for distributions of scores skewing low (as shown in figures 1-3). The ART-G contains many more fiction than nonfiction authors, and though it permits division into subscales for fiction genres it does not have subscales for nonfiction topics. Thus the measure is arguably more sensitive for fiction than nonfiction. This also impacts the ART-G total score, which combines both fiction and nonfiction and due to the higher number of possible fiction authors will favour fiction readers; someone who recognises half of the fiction authors (55) but only a quarter of the nonfiction authors (13) will perform better in terms of their total score (68), than someone who recognises half of the nonfiction authors (25) but just a quarter of the fiction authors (28) in their total (53). Overall, despite these limitations, it is interesting that total reading amount as measured by the ART-G did not improve model fit for CTD or EO. While it is not possible to generalise the performance on the ART-G to all readers, and the instrument does have some issues, it does nonetheless offer a useful indicator of fiction and nonfiction exposure as a preliminary means of tapping differences between readers of these different types of material.

EO is typically conceptualised in developmental terms as a progression from absolutism, through multiplism, to evaluativism (Hofer & Pintrich, 1997). However, given variation in adult EO (Kuhn et al., 2000), it may be better thought of as distinct, but not necessarily ordered categories. The latter conceptualisation is better supported by the cluster analysis (Table 2) used to group participants into EO categories in this study. It is also notable that clustering in this study differed from McGinnis' (2016) validation of the scale, which found four clusters to be the best fit. Based upon this study data, three clusters were found to fit best, and for brevity have been named

absolutist, multiplist, and evaluativist. However, the absolutist and evaluativist categories have cluster centres that are almost identical for evaluativism (5.28 and 5.25 respectively); what distinguishes these two categories are their levels of absolutism and multiplism. Therefore, the finding that fiction readers are more likely to be evaluativists than to be absolutists in fact tells us they are more likely to occupy a space with low absolutism, moderate multiplism, and high evaluativism. Accounting for this nuance may also shed some light on the finding that nonfiction readers are more likely to be absolutist than either other category, but also more likely to be evaluativist than multiplist. Figure 4 demonstrates an overlap between the evaluativist and multiplist categories, and some overlap between the evaluativist and absolutist categories, and the comparative distance to the absolutist category from the multiplist category; it may be that in being strongly unlikely to be multiplists and more likely to be absolutists, nonfiction readers are more likely to be evaluativists rather than multiplists by falling into a space within the evaluativist category that overlaps with the absolutist category. Furthermore, multiplists have higher levels of absolutism than evaluativists do (cluster centres 2.97 versus 1.82 respectively), which may seem paradoxical given that multiplism and absolutism can be viewed as mutually exclusive stances. However, it may be that multiplists who take a position of multiple realities can view individuals as having absolutely valid knowledge or beliefs within their own unique reality, and this may bridge that paradox. In this sense, if all possible positions are equally valid (a multiplist stance), then they are all absolutely correct (an absolutist stance), and evaluation is rendered meaningless; this would support the view that evaluativism is the EO needed for CT (Kuhn, 1991). This highlights the nuance within these categories, and demonstrates that they should not be interpreted as straightforwardly as their names

imply. Rather, this study shows that reading fiction and nonfiction are associated with different positions in an interwoven EO space.

The outcome variables in this study were measured using self-report Likert scales, and as such must be interpreted as subjective to participants' perceptions of themselves. It is likely that desirability effects played a role in responses to both scales. The CTDS has items such as "I am often on the lookout for new ideas" (Sosu, 2013), which carry a clearly positive connotation. The JCI is perhaps less prone to desirability effect as the statements such as "Most debatable issues have one right conclusion" (McGinnis, 2016), are less normative. However, both scales are validated, and participants in the sample had diverse responses to both. Therefore, although these reflect subjective self-assessments, they are still useful indicators of differences between participants.

It is not possible to determine a causal direction based upon these findings. It may be that people with higher CTD chose to read more fiction, rather than that reading improves CTD. Similarly, people with more evaluativist rather than absolutist stances may be more drawn to fiction reading, while people who are more absolutist may be more attracted to nonfiction. The findings from this study merely demonstrate a relationship between these factors, but the direction of causal influence requires further investigation.

The data analysis in this study diverged from the pre-registered plans in several important ways. In the pre-registered plans, hierarchical models were specified. However, in comparing different models single level models were found to fit better for both CTD and EO as outcomes. Thus the models adopted for both outcome

variables diverged from the pre-registered plans. In retrospect, my plan to use hierarchical models was based on the assumption that educational level, or NT, could be deployed as random incepts, however upon further consideration these are not conceptually cogent as clustering variables. NT, being subjectively rated on a Likert scale, would not be a suitable clustering variable as how different participants interpret the scale is not stable. Educational level, although not subjectively defined, in fact is more logical as a predictor interacting with other variables than a clustering variable, as this was not accounted for in the sampling. Furthermore, no variable that could have been used as the random intercept would meet the large number of categories necessary for meaningful results (Bryan & Jenkins, 2016). Overall, my initial plan to use hierarchical models was flawed, and the resulting decision to favour single level models better fit the data, and made better sense conceptually. Additionally, in the pre-registration EO was considered to be an ordinal variable. However, upon reviewing the results of the cluster analysis for EO, I shifted my understanding of EO towards being distinct, but not necessarily ordered categories, and thus a nominal variable. This was further supported by a closer scrutiny of the JCI scale (McGinnis, 2016), which does not quantify degrees of agreement with similar statements but presents divergent epistemological approaches, implying different positions not different levels. This resulted in a change from the pre-registered study plan. Treating EO as nominal, a more complex view of EO emerges in which the different categories do not represent more or less of the same construct, but are rather different spaces in a nuanced epistemological landscape. These divergences from the pre-registered study plans were deemed appropriate in better suiting the data.

4.3.2. Relationship with the literature

The findings from this study are in line with prior research that found distinct impacts of fiction and nonfiction reading (e.g. Mar et al., 2006), as for both outcome variables the effects of fiction and nonfiction exposure differed both in direction and significance. However, these findings contravene earlier research that took print exposure overall to be a single variable (Stanovich & Cunningham, 1992), as total print exposure was not a relevant contributor to either outcome measured in this study. Indeed, the difference in direction of effect seen between fiction and nonfiction exposure implies combining them would be counterproductive. These findings suggest that fiction and nonfiction are best treated as separate predictors of different outcomes associated with CT.

One way of casting the difference between fiction and nonfiction is in distinguishing “efferent” or “informational reading” in contrast with “experiential reading” (Hampson Lundh et al., 2018; Rosenblatt, 1982). The association between nonfiction exposure and the increased likelihood of being an absolutist rather than a multiplist is congruent with the association of nonfiction and informational reading; if one is oriented towards finding concrete information it is unsurprising that a clear-cut epistemology of truth versus falsehood would align with this. If we interpret the greater likelihood of being an evaluativist rather than multiplist associated with nonfiction reading as a strong rejection of multiplism, acknowledging the overlap between evaluativism and absolutism in the data (Figure 4), then this also fits this distinction. Taking fiction exposure to be more associated with experiential reading, we might then expect a reduced likelihood of absolutism by the same logic; this was the case

when comparing evaluativism with absolutism. Additionally, it could also be argued that experiential reading would lend itself to a multiplist epistemology, as experiencing multiple different viewpoints in a non-judgemental stance may be seen as part of the aesthetic reading experience. However, no association between fiction reading and multiplism was shown in this study. Further research into approaches to reading, and EO, would be worth pursuing to elucidate these potential relationships.

As was explored in the literature review, there is conflicting evidence on the influence of NT upon CT. Some have argued NT inhibits CT (Busselle & Bilandzic, 2008), while others have found CT taking place during transportation (Hoeken & Fikkers, 2014). The findings of this study offer some support for the latter view. In the case of CTD, falling into a higher level of NT rather than having low NT predicted increased CTD; people who tend to be more transported by narratives also tend to be more disposed towards CT. Furthermore, having moderately high rather than low NT was predictive of being an evaluativist rather than an absolutist; argued to be a more conducive EO for CT (Kuhn et al., 2000). Though this is not direct evidence of a positive association between NT and CT, it does demonstrate a positive relationship between NT and factors that arguably feed into CT.

A further finding in terms of NT was the paradoxical interaction NT had with fiction and nonfiction in the CTD model. Moving from low to moderately low NT interacted with both fiction and nonfiction scores, and changed the direction of the fiction and nonfiction effects; fiction readers with moderately low NT become less likely to have a higher CTD level ($b = -0.13$), while nonfiction readers with moderately low NT become more likely to have a higher CTD level ($b = 0.33$).

Interestingly, this paradoxical effect somewhat aligns with findings from Bal and Veltkamp (2013), who found that reading fiction with low NT predicted decreased empathy while the opposite was true for nonfiction. It may be that moderately low NT has an effect in comparison to low NT as it represents a different way of reading, i.e. some transportation is present, while low NT may suggest fully un-transported, perhaps more externally aware reading. Yet this moderately low level of transportation may yield the kind of frustration Bal and Veltkamp suggest made their fiction readers less empathetic as a result. Conversely, having a little transportation while reading nonfiction may increase engagement, but without the feelings of obligation Bal and Veltkamp suggest can arise from high NT nonfiction reading. However, Bal and Veltkamp's study addressed state NT for a specific piece read, while this study looked at overall trait NT and print exposure. Bal and Veltkamp also looked at empathy, which is a different outcome variable to CTD even if there may be some interplay between them. Therefore further investigation would be needed to establish how NT interacts with reading of different types to influence outcomes.

4.3.3. Implications for the next studies

Drawing from the literature, it was hypothesized that overall reading would increase CT, within which fiction would have a greater impact on CT than nonfiction, as fiction has been shown drive a more diverse array of abilities that arguably feed into CT. Though this study did not measure CT, it did find an association between increased fiction exposure and increased CTD, and this association was not present for nonfiction. This finding, in conjunction with the association between fiction exposure and evaluativism, rather than absolutism, supports this hypothesis and indicates the

proposed research into a connection between fiction reading and CT is worth pursuing. However, as total print exposure did not improve model fit, this indicates that fiction and nonfiction ought to be assessed independently of one another and not taken in aggregate. This implies the hypothesis that reading in total will increase CT may be flawed. However, given the limitations of the ART-G, this requires further investigation. These findings suggest further investigation into fiction and nonfiction reading and CT is warranted, as is an exploration of whether they can be treated in aggregate as total reading or not.

The survey response bias demonstrated in the higher levels of women, and people with higher education level, in line with wider trends (Cheung et al., 2017), is a cause for concern. In consequent studies, measures will be taken to avoid highly skewed samples in terms of gender and education. However, as age was broadly represented in this study, this suggests this variable is of less concern. This may suggest the advertising of the study both online and physically in public libraries reached age-diverse participants, and ought to be replicated in the next study. Additionally, only educational level and gender improved model fit for either outcome variable, implying these are the only relevant demographic variables. Only educational level showed any significant effect, but the skewed sample, and wide CIs in the model, made this insufficiently supported from this data; therefore further investigation into the role of educational level within models testing reading and CT is warranted.

NT was found to mediate the relationship between fiction reading and CTD. However, no mediation effect was found for EO. Additionally, NT has been shown to be predictive of higher CTD, and to impact EO, as a predictor variable in its own

right. Having moderate rather than low NT also had a paradoxical interaction with both fiction and nonfiction scores, suggesting a meaningful difference between not being transported at all and being moderately transported. Therefore NT should be included as a variable for further investigation in following studies.

Given the nuance in EO demonstrated in the cluster analysis and findings of this study, EO is a variable that would be interesting to explore through qualitative means. The complex interplay of the different EO dimensions draws into question how helpful grouping EO into discrete categories in fact is. This is perhaps a variable where individual experiences are best treated separately, and allowed space for full articulation of complex epistemological positions. Therefore in study three, interviews were used to explore participant's perceptions of the different EO dimensions, and how they may experience these in relation to reading.

4.4. Conclusions

The main aim of this study was to identify whether or not there was a relationship between reading engagement and factors associated with CT, and whether fiction and nonfiction had distinct associations. As total reading engagement measured was not associated with differences in either CTD or EO, but fiction and nonfiction taken separately did have distinctive relationships, this implies that treating fiction and nonfiction reading separately from one another is appropriate. Increased fiction exposure was associated with increased CTD, suggesting fiction readers are more disposed towards CT, while no such association was found for nonfiction exposure. Fiction and nonfiction exposure were associated with different EO categories, suggesting that readers of these different types of materials have different

epistemological beliefs that may feed into their CT approaches. Therefore the findings of this study imply that reading fiction and nonfiction have different relationships with factors associated with CT, thus providing support for the planned direction of this research project.

5. Study two: A reading log study investigating the impact of fiction and nonfiction reading on change in critical thinking

This study measured change in CT across a two week period, during which participants logged their reading. A summary of the study is given in Box 4. In order to test for an association between reading and changes to CT, specifically aiming to identify possible improvements to CT associated with fiction reading, it was necessary to design a study such that CT could be measured (rather than relying upon self-report), and that this measurement could take place more than once in order to identify change. Furthermore, this study aimed to assess reading behaviour, not merely print exposure. In these ways, this study builds upon the prior cross-sectional, observational study which yielded insight into some associations between reading and factors associated with CT (CTD and EO), but did not test CT itself nor provide any information about a causal direction. As well as furthering the findings of study one, this study also sought to identify trends that could be further interrogated in the following qualitative studies, and thus provide findings from which to develop the further research design of the overall project.

Box 4: Study two summary

This study aimed to test causal relationships between reading fiction and nonfiction over a two week period and CT test score change. Self-selecting participants ($N = 335$) were assigned to randomised blocks, forming a control group and experimental groups assigned either fiction or nonfiction reading. Participants logged their reading on a daily basis. Fiction reading was associated with increased CT change. However, fiction reading measured in entries was associated with greater odds of improvement in CT score, but not when measured in reading time. Nonfiction reading was not associated with increased CT change, and nonfiction entries were associated with lower odds of CT improvement while nonfiction time was associated with improvement. Assigning fiction reading to nonfiction readers had a positive association with both increasing CT change, and the odds of CT improvement.

This study combined experimental and observational components. The experimental component of the study was the assignment of fiction or nonfiction readings. This experimental manipulation was included in the study design to better enable causal tracing, as differences in CT change between the experimental groups can indicate the effects of the assigned reading upon CT. A second predictive component in the study was observational: the daily reading log conducted over the two week study period, into which participants input what they themselves chose to read in addition to the assigned texts. This observational element was included in the study design to provide ecological validity, as participants read what they wished to read and

reported their reading behaviours as they themselves saw fit. The outcome variable in the study was change in CT score, derived from the CT test administered before, and again after, the two week reading log period. By using the test and re-test paradigm, change in CT could be measured, which is essential for drawing causal inferences. Finally, as a further observational element to the study participants were also asked to list their favourite authors, as this has been found to be an effective measure of reading in previous studies (Stanovich & Cunningham, 1992; Stanovich & West, 1989). This enabled correlational analyses with participants' baseline reading engagement at the start of the study.

People who identified as readers were the target population for participant recruitment in this study. As such, this study marked a departure from the wider aims of the prior study in comparing amounts of reading in a range of participants including non-readers, towards a focusing-in on differences in readers. This enabled a shift towards a more detailed focus on the effects of reading, narrowing into a comparison of fiction and nonfiction rather than addressing lower or higher amounts of reading in general.

Participants in the study were put into either the fiction or nonfiction experimental group, or the control group, which determined the assigned readings they were given. In assigning participants to each of these groups, the aim was for each group to contain half participants who identified themselves as fiction readers, and half nonfiction readers. It was also aimed to balance demographics across these groups so as to control for them. However, it should be noted that due to varied numbers of different types of participants signing up for the study, a perfectly even balancing

between groups was not achievable. There were six categories of readers in the study: fiction readers assigned fiction; fiction readers assigned nonfiction; nonfiction readers assigned fiction; nonfiction readers assigned nonfiction; fiction readers assigned to the control group; nonfiction readers assigned to the control group. This permits for a comparison of the differences between the fiction and nonfiction experimental group, and the control group. It also permits for additional exploration of whether there are any differences between readers assigned congruent or incongruent material with their existing preferences.

A control group was included in the experimental design so as to permit for comparisons not only between the effects of being assigned fiction or nonfiction, but also for the effects of being assigned either type of reading at all. In this study, the control group were not assigned any reading. Participants in the manipulated conditions were assigned a selection of short readings (five texts, either fiction or nonfiction). The five texts in each condition added up to approximately 11,000 words. This is a substantial reading volume, so as to make the difference in reading between the two groups more likely to be meaningful. The fiction texts were short stories, and the nonfiction texts were news or magazine articles, so as to be complete and real works with ecological validity.

Table 7 details the texts used. Fiction and nonfiction texts were selected to be matched for theme, to total the same length, and to be diverse in their genres and topics as well as being diverse in their authors. However, it cannot be said that they are equivalent to each other (e.g. Harvey's article is not the nonfiction version of Winterson's story). This places a limitation on the conclusions that can be drawn from comparing the groups assigned either text selection. An alternative would have been to create texts that were analogous except for one being a fictional and the other a nonfictional account, which is an approach taken in other research (e.g. Djikic et al., 2009b). However, this comes with other limitations due to poor ecological validity, and potentially skewed ecology validity if only one text is authentic. Ultimately, the use of authentic texts was deemed more in keeping with the realism of the reading log, and thus ecological validity was maximised with the consequent limitations. All texts were cleared for copyright compliance, and given to participants in accessible word and PDF formats.

Table 7: Assigned texts

Fiction text	Nonfiction text	Theme	Genre/Topic
<i>The 24 Hour Dog</i> by Jeanette Winterson	<i>I gave a blueberry facial to a vegan dog!: inside the pet pampering boom</i> by Lisa Harvey in The Guardian	Human-pet relationships	Literary fiction / Lifestyle
<i>Bella Fleace Gave A Party</i> by Evelyn Waugh	<i>Diary</i> by James Wood in The London Review of Books	Social class	Classic fiction / Politics
<i>Monday or Tuesday</i> by Virginia Woolf	<i>The Depressing Science of What Living in a City Does to Your Brain</i> by Princess Ojiaku in How We Get To Next	Urban versus natural landscapes	Prose poetry / Psychology
<i>The Nine Billion Names of God</i> by Arthur C. Clarke	<i>Will Cern's Large Hadron Collider finally unlock the mysteries of the universe?</i> by Brian Appleyard in The Sunday Times	Human understanding of the universe	Science fiction / Physics
<i>Loneliness Is in Your Blood</i> by Cadwell Turnbull	<i>The ambitious quest to cure aging like a disease</i> by Britt Wray in BBC Future	Extended life	Horror / Medicine

The hypotheses for this study were as follows:

H₀: Reading (quantity nor type) will have no significant association with critical thinking score either at t⁴1, or with the difference between t1 and t2 score.

H₁: Higher reading engagement will have a significant association with higher baseline critical thinking score at t1.

H₂: Higher amounts of reading recorded over the reading log period will have a significant association with increases in critical thinking score from t1 to t2.

H₃: Higher amounts of fiction reading recorded over the reading log period will have a stronger association with increases in critical thinking score from t1 to t2 than nonfiction.

H₄: Narrative transportation will mediate the relationships predicted in H₂ and H₃.

H₅: Participants assigned to the fiction experimental group will have greater increases in their critical thinking test scores from t1 to t2 than either the nonfiction group or the control group.

This study was preregistered: DOI 10.17605/OSF.IO/R8WMT

5.1. Method

This study had a three-group, between-subjects (unpaired), repeated measures design. The study combined experimental and observational components to investigate the relationship between fiction and nonfiction reading, and change in CT test scores,

⁴ t1 refers to “time 1”, i.e. the first questionnaire administered to participants prior to the reading log; t2 refers to “time 2”, i.e. the post-log questionnaire.

over a two week period. For the experimental component of the study, participants were assigned readings; either 5 nonfiction texts for the nonfiction group, or 5 fiction texts for the fiction group, and asked to read these over the two week study period. The control group were not given any assigned reading. For the observational component of this study, participants were asked to log their reading over the two week study period (in addition to their assigned texts). Participants were also asked to list up to 10 of their favourite authors at t1 as a measure of reading engagement. They were given a critical thinking test before and after the two week diary period, using a different test version. An online survey was used to collect responses.

5.1.1. Participants

Participants were recruited using posters in London public and academic libraries, and bookshops, as well as using online calls for participation (callforparticipants, Twitter, reddit). A £15 National Book Token voucher was offered as a reward to each participant who completed the study, as an incentive for participation.

The G*Power statistical software was used to compute required sample sizes for the analyses to be conducted (Faul et al., 2009). Assuming a medium effect size (f^2 0.15), a power of 80%, and an α of .05, it was calculated that 77 participants were required for a multiple regression analysis with 3 predictor variables (total, fiction, and nonfiction reading) upon the outcome variable of CT change. If we increase the number of predictor variables to 6, thus allowing for the inclusion of controls (*age, gender, educational level*), or for the 6 experimental groups to be used as predictors in the regression, the required sample size is 98. Based upon this, I opted for a target

sample size of 120 as this was in excess of the minimum requirement, and permitted for 40 participants in each experimental condition. A sample size of 120 permits for a maximum 10 predictor variables to be used in a multiple regression while maintaining power in the analysis.

Initially, self-selecting interested participants signed up to a participant pool ($N = 423$). Of these, 257 identified as mainly fiction readers, and 166 as mainly nonfiction readers. Of these 296 were female, 108 male, 6 non-binary, and 13 did not state their *gender*. Highest obtained *educational level* was divided into three groups; no degree (*ND*) ($n = 116$), undergraduate (*UG*) ($n = 158$), postgraduate (*PG*) ($n = 149$).

The characteristics of the participants in the pool were used for selection into the main study, using a randomized blocking paradigm. The aim was to recruit 60 fiction readers and 60 nonfiction readers, with an equal *gender* (half male, half female) and *educational level* (half with no degree, half with a degree) balance. However, given the characteristics of the participants in the pool such a perfect division was not possible, and therefore as close to such a balance as possible was made. A random number generator was used to select participants with these characteristics from the pool. There were then allocated to the nonfiction, fiction, and control experimental groups. Invitations to participate were sent out continuously throughout the study period, replacing participants who withdrew or ceased to participate.

Of the total invited participants ($N = 238$), 121 were female, 101 male, 6 non-binary, and 10 did not state their *gender*. In terms of *educational level*, 91 had *ND*, 65 *UG*, and 82 *PG*. There were 111 fiction readers and 127 nonfiction readers invited. Of these, 85 were allocated to the control group, 74 to the fiction group, and 79 to the

nonfiction group. Of the total invited participants, 121 completed the study, and 117 did not. Of those who did not complete the study, 106 did not commence the study; 10 completed some diary days but withdrew before the end of the 14 day period; 1 completed all stages but did not enter answers for the CT test and was therefore removed.

In total 121 participants completed the study: fiction readers ($n = 62$); nonfiction readers ($n = 59$). They were assigned to experimental groups: control ($n = 41$), fiction ($n = 40$), nonfiction ($n = 40$). Complete reading logs were checked to see if participants who were assigned to an experimental group completed their readings, or if they missed any; participants who missed half or more of the assigned readings (≤ 3 texts) were re-categorised into the control group. This was done to ensure that the experimental groups contained only participants who had received the experimental manipulation, and could therefore be used to draw inferences about the effects of the experimental manipulation. Of the participants assigned fiction, 4 missed texts and were therefore reassigned to the control group; of participants assigned nonfiction, 6 missed texts and were therefore reassigned to the control group. Therefore the assigned groups were finally: control ($n = 51$); fiction ($n = 36$); nonfiction ($n = 34$).

The final experimental groups created six blocks; control group fiction readers Cf ($n = 28$); control group nonfiction readers Cn ($n = 23$); fiction group fiction readers Ff ($n = 16$); fiction group nonfiction readers Fn ($n = 20$); nonfiction group fiction readers Nf ($n = 18$); and nonfiction group nonfiction readers Nn ($n = 16$).

Of the total sample, 64 were female, 52 male, 3 non-binary, and 2 did not state their *gender*. Highest obtained *educational level* in the total sample broke down as:

ND ($n = 55$); *UG* ($n = 24$); *PG* ($n = 42$). The mean participant *age* was 38 ($SD = 15.35$), median 33, with a range of 62 from 18 to 80. It should be noted that *gender* and *educational level* was balanced so far as possible in the randomized blocking paradigm, but *age* was not. This decision was taken based upon findings from study one which suggested that *gender* and *educational level* may be associated with differences in CT disposition and EO, and ought therefore to be controlled for. Table 8 shows a breakdown of participant demographics by experimental block.

Table 8: Participant blocks

	Control Group	Fiction Group	Nonfiction Group
	Mean age = 36.2	Mean age = 40.4	Mean age = 42
	Female = 16	Female = 8	Female = 8
	Male = 11	Male = 7	Male = 9
Fiction	Non-Binary = 1	Non-Binary = 0	Non-Binary = 1
readers	Not-Notated = 0	Not-Notated = 1	Not-Notated = 0
	No degree = 14	No degree = 7	No degree = 8
	Degree = 14 (UG = 3, PG = 11)	Degree = 9 (UG = 3, PG = 6)	Degree = 10 (UG = 1, PG = 9)
	Mean age = 36.6	Mean age = 34.2	Mean age = 39.8
	Female = 13	Female = 11	Female = 8
	Male = 8	Male = 9	Male = 8
Nonfiction	Non-Binary = 1	Non-Binary = 0	Non-Binary = 0
readers	Not-Notated = 1	Not-Notated = 0	Not-Notated = 0
	No degree = 9	No degree = 10	No degree = 8
	Degree = 14 (UG = 5, PG = 9)	Degree = 10, (UG = 3, PG = 7)	Degree = 8 (UG = 1, PG = 7)

5.1.2. Measures

International Critical Thinking Essay Test form A (ICTET-A) Short Form.

The ICTET-A (The Foundation for Critical Thinking, 2019) was used to test participants' CT, as at the time of research this was the test of CT found to best fit the conception of CT used in this research project. The test was designed in line with the Paul and Elder CT framework (Paul & Elder, 2005), which includes a broad range of CT facets. The short form of the test was used to keep participant fatigue to a minimum, given the overall lengthy time commitment of the study. The test took approximately 30 minutes to complete. In this test, participants were given a text to read, and then answered seven questions identifying the Purpose, Evidence, Conclusions, Assumptions, Concepts, and Implications. Each question was graded on a scale of 1-10, resulting in a possible maximum score of 70. There were two sample texts used in this study so that each participant was given a different text at t1 and t2. Text A was an excerpt from Erich Fromm's *The Art of Loving*; the ICTET-A using text A has been validated (Hollis et al., in press). Text B was an excerpt from George Orwell's *The Sporting Spirit*; the test with this text has not been previously validated, thus validation was conducted as part of this study (results are presented in Appendix C: Validation of ICTET-A with text B).

These texts were selected as the topics of "love" and "sports" ought to be familiar to a majority of participants. Half of the participants were given text A at t1 and half were given text B; at t2 participants were given the alternate text. By distributing text A and text B in this way, text B could be checked for consistency against the previously validated text A. Table 54 in Appendix B shows the ICTET-A

questions including explanatory notes, as presented to participants. Participants were instructed to write two to four sentences for each question.

Results from the ICTET-A yielded the following variables:

- *CT score*: This was the total score each participant was given on the test at t1, and at t2, i.e. each participant had 2 CT scores.
- *CT change*: This was each participants' t2 score minus their t1 score, i.e. the difference in their performance from before to after the reading log period.
- *CT improvement*: This was a binary category created by grouping participants into the *improved* (*CT change* of 1 or higher) or *no improvement* (*CT change* of 0 or lower) groups.

Reading log

Participants were asked to log their reading each day for a 14 day period. Entries did not have to be made on the day, but could be completed later to permit more flexibility for participants. The reading log form is presented in Appendix B under Measures. There was a maximum of 14 items logged each day, and no minimum. A maximum of 3 missed days was permitted out of the total 14 day period. Entries in the reading log were classified into fiction and nonfiction, using the Goodreads platform and applying the top genre to each item. Reading logs for each participant were summed to yield the following variables:

- *Entries made*: this is the total figure for the number of entries made in the reading log over the 14 day period. This is not necessarily the

number of texts read, as one text could be read and logged several times (e.g. one book of which some is read each day), thus constituting several entries. *Total entries* were further subdivided into *fiction* and *nonfiction entries*.

- Reading time: total reading logged in 30 minute units (half-hours). Participants logged their reading time for each entry in minutes, as this was the simplest unit of measurement for participants to use. However, as one minute of reading is a very small unit unlikely to be impactful, and using hours would be a very large unit potentially missing granularity, 30 minute increments were used as the unit of measure for reading time; time logged in minutes was transformed into half-hours (i.e. /30). Half-hours of reading over the log period can be further subdivided into *total*, *fiction*, and *nonfiction time*.
- Reading time per entry: The average time spent per entry made by each participant (in half-hours), i.e. reading time divided by entries made, subdivided into *total*, *fiction*, and *nonfiction t/e*.
- Average NT level: the mean transportation rating for each participant over the log period. As each entry of reading was given a transportation score from 1-5, this constitutes the average transportation of each participant for their reading over the whole log period.
- Peak and minimum NT level: the highest and lowest transportation score used by each participant, for *fiction*, *nonfiction*, and in *total*. As there is individual variation in trait transportation (Green & Brock, 2013), some participants are expected to be more or less transported than others in

general over whole two week period. By identifying each participants' peak and minimum transportation rating, state transportation during the log period can be compared. As this is further divided into fiction and nonfiction, it is also possible to compare the different transportation levels these two forms of reading may elicit. The following variables were derived from reading amounts at the different NT levels: *Peak NT time; peak NT entries; minimum NT time; minimum NT entries.*

Reading engagement

Participants were asked to list up to ten of their favourite authors (with no minimum number), to establish their reading engagement level at t1. The question phrasing invited participants to list fiction, nonfiction or both kinds of author. Asking participants to list favourite authors has been found to be an effective measure of reading in previous studies (Stanovich & Cunningham, 1992; Stanovich & West, 1989). In order to classify authors into fiction and nonfiction, their top rated book on the Goodreads platform was used, and its type applied to the author. Responses were classified into *total, fiction, and nonfiction engagement.*

Outcome variables:

- CT score: score at t1, and at t2
- CT change: score at t2 minus score at t1
- CT improvement:
 - Improved / No improvement

Predictor variables and mediators:

- Experimental group:
 - Cf: fiction readers assigned to the control group
 - Cn: nonfiction readers assigned to the control group
 - Ff: fiction readers assigned fiction
 - Fn: nonfiction readers assigned fiction
 - Nf: fiction readers assigned nonfiction
 - Nn: nonfiction readers assigned nonfiction
- Reading engagement: Fiction / Nonfiction / Total
- Entries:
 - Fiction / Nonfiction / Total
- Time (in half-hours):
 - Fiction/ Nonfiction / Total
- Time per entry (t/e):
 - Fiction / Nonfiction /Total
- NT:
 - Average (mean)
 - Peak (highest rating per participant):
 - Peak NT entries and time: Fiction / Nonfiction / Total
 - Minimum (lowest rating per participant):
 - **Min** NT entries and time: Fiction / Nonfiction / Total

Control variables:

- Age
- Gender
- Educational level

5.1.3. Procedure

Responses were collected between 06/01/2020 and 27/03/2020. The study was conducted online using the Opinio survey platform. In the registration of interest, after obtaining informed consent, basic demographic information (fiction or nonfiction preference; *gender*; highest obtained *educational level*) was collected, as well as participant email addresses. Participants who were selected into the main study were then emailed with their group allocation (fiction, nonfiction, or control) and the respective set texts in accessible word and PDF formats. Participants in the fiction and nonfiction group were instructed to read their set texts whenever they wished within the 14 day log period. The email also contained the link to the first study questionnaire. In the first study questionnaire, further informed consent was obtained and *age* was asked for to complete the demographic information. Participants were then asked to list up to ten of their favourite authors. Next, participants were then given instructions on how to answer the CT test, including information on how they would be graded. They were given the test text to read and questions to answer underneath. The test used open text entry spaces for each question. At the end of the test, participants were asked to confirm that they answered every question and were satisfied with their answers. Once this form was completed, participants received an email reminder with a link to the daily reading log form each evening after 20:00 for the next 14 days. Upon completing the 14th log entry, participants were emailed the link to the final CT test form, which was presented in the same way as at t1 but with the alternate text given. Participants were then given debriefing information. Reward vouchers were emailed to participants upon completion.

I graded the anonymised CT tests as the sole grader following the ICTET-A rubric. I also classified reading engagement authors and reading log entries into genre as the sole assessor. Data was analysed using R 4.0.4. in R Studio version 1.4.1103.

5.2. Results

Tests of equality of variance, and normality of distribution, are listed in Appendix B under Assumption checking.

5.2.1. Descriptive statistics

ICTET-A

First, the time periods taken by participants to complete the test at t1 and t2 were assessed, as this helps verify that there was a relatively stable timeframe across participants in the study. The mean number of days between participants completing the ICTET-A at t1 and t2 was 16.79 ($SD = 3.06$), median 16, with a range of 17 from 13 to 30. The mean number of days from the final reading log entry completed by participants to the completion of the CT test at t2 was 2.81 ($SD = 2.56$), median 2, with a range from 0-13. There was no deadline set for participants for complete the test after the end of the reading log, and though most completed it soon after it should be noted that in some cases there was a lag between the final log entry and the CT test at t2.

ICTET-A scores compared across t1 and t2

As the test was administered twice (though using the two different texts), there could be an effect of the experience of taking the test at t1 upon the way participants performed at t2. We may expect performance at t2 to be generally better than at t1, as the t1 experience of test could provide practice. Therefore comparisons were made of

performance at t1 and at t2, with the two texts, to identify any significant differences.

Figures 5 and 6 show the score distributions.

Student's two sample t-test was used to test for difference in text A scores between t1 and t2, and found no significant difference ($t(119) = 0.45, p = .656$). Due to unequal variance, Welch's unequal variance t-test was used, and found no significant difference for text B scores at t1 and t2 ($t(119) = -1.19, p = .235$). This suggests no significant difference in performance between the two test times. Combining text A and B scores, Welch's unequal variance t-test was used to test for a significant difference in t1 and t2 scores, and found no significant difference ($t(229.55) = -0.54, p = .593$). This suggests that overall there was no difference in performance on the ICTET-A between t1 and t2.

A paired t-test was used to test whether there was any significant difference in participants' performance at t1 and at t2; no significant difference was found ($t(120) = -0.76, p = .449$). This suggests that participants' performance on the ICTET-A did not significantly change from t1 to t2. Therefore taking the test at t1 did not provide practice experience leading to significant improvement in performance at t2, and this does not need to be controlled for in further analyses.

Figure 6: Histogram of CT scores at t1

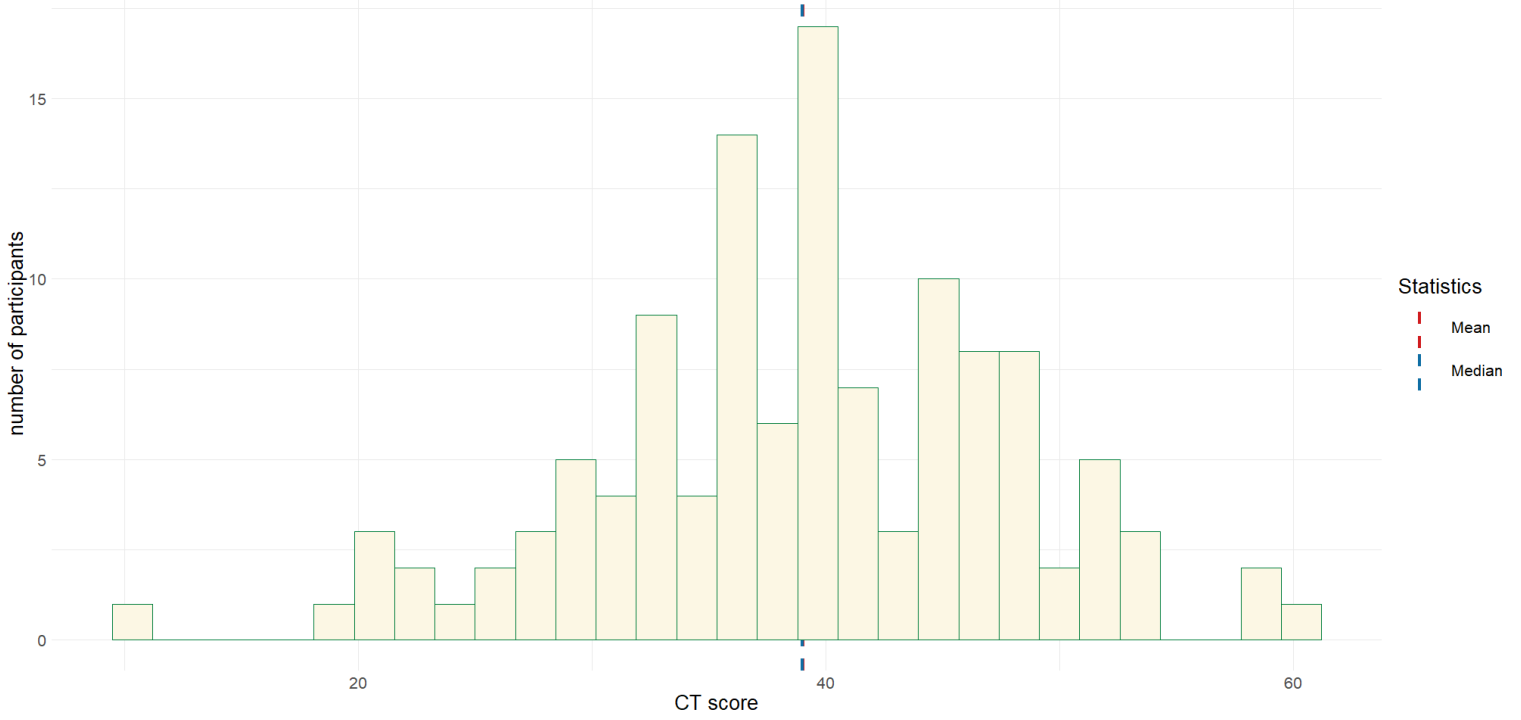
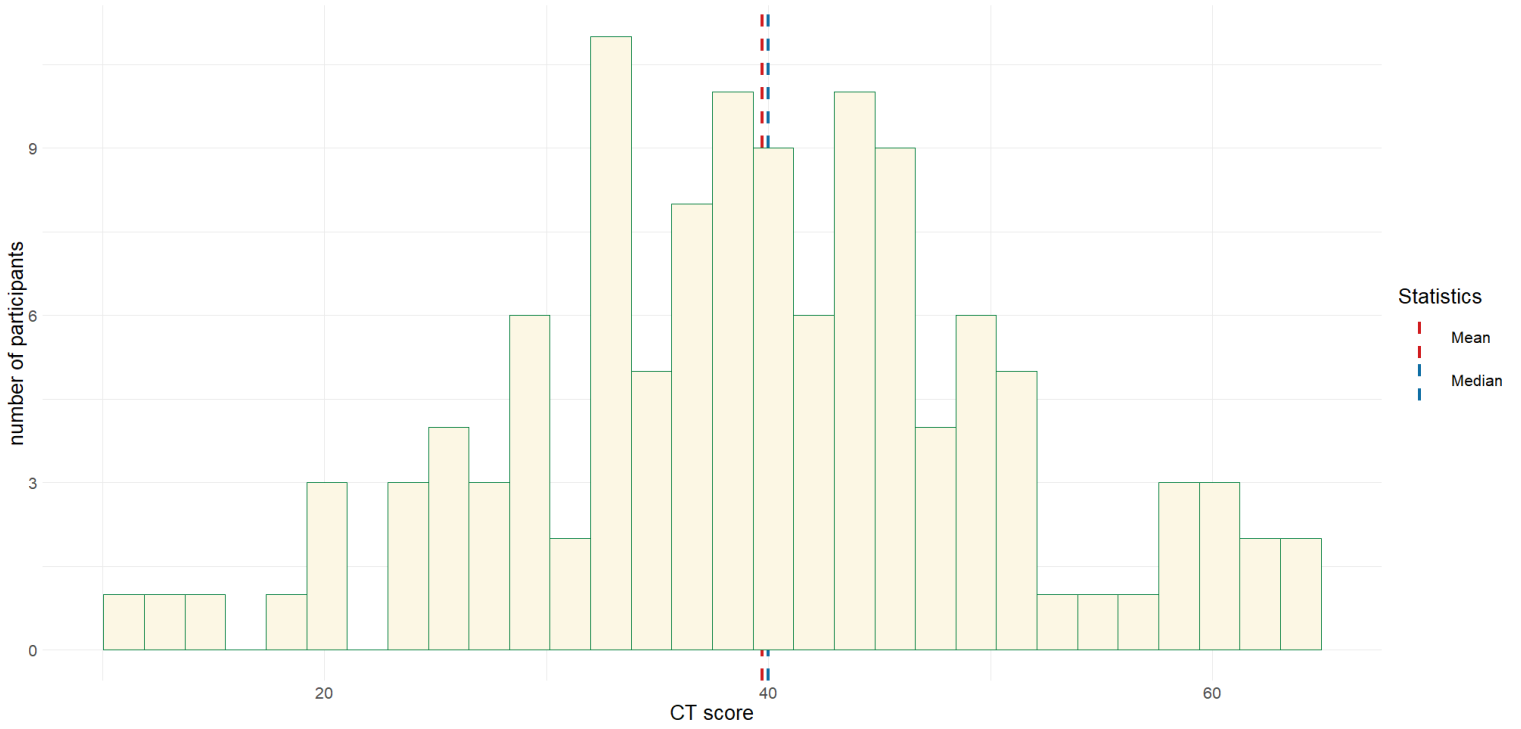


Figure 5: Histogram of CT scores at t2



ICTET-A performance across demographic variables

Next, performance on the ICTET-A was assessed for differences between demographic groups. This was done to determine whether any demographic variables need to be controlled for in further analyses.

Firstly, *gender* differences were tested. Participants who specified a *gender* other than male or female, and those who chose not to state their *gender*, were combined so as to have sufficient numbers for groups comparisons. A one-way ANOVA was carried out, which showed no significant difference in scores between *gender* groups on the ICTET-A at t1 ($F(2,118) = 0.24, p = .786$). A one-way analysis of means was carried out, which showed no significant difference in scores between *gender* groups on the ICTET-A t2 ($F(2,118) = 0.68, p = .528$). This suggests that participants of different genders did not perform differently on the test, and *gender* does not therefore need to be controlled for in further analyses of t1 and t2 *CT scores*.

Differences in test performance across *educational level* were assessed. A one-way ANOVA was carried out, which found no significant difference in t1 scores across *educational level* ($F(2, 118) = 2.28, p = .107$), nor in t2 scores ($F(2, 118) = 1.52, p = .222$). This suggests that participants with different *educational level* scored similarly on the test. *Educational level* therefore does not need to be controlled for in further analyses of t1 and t2 *CT scores*.

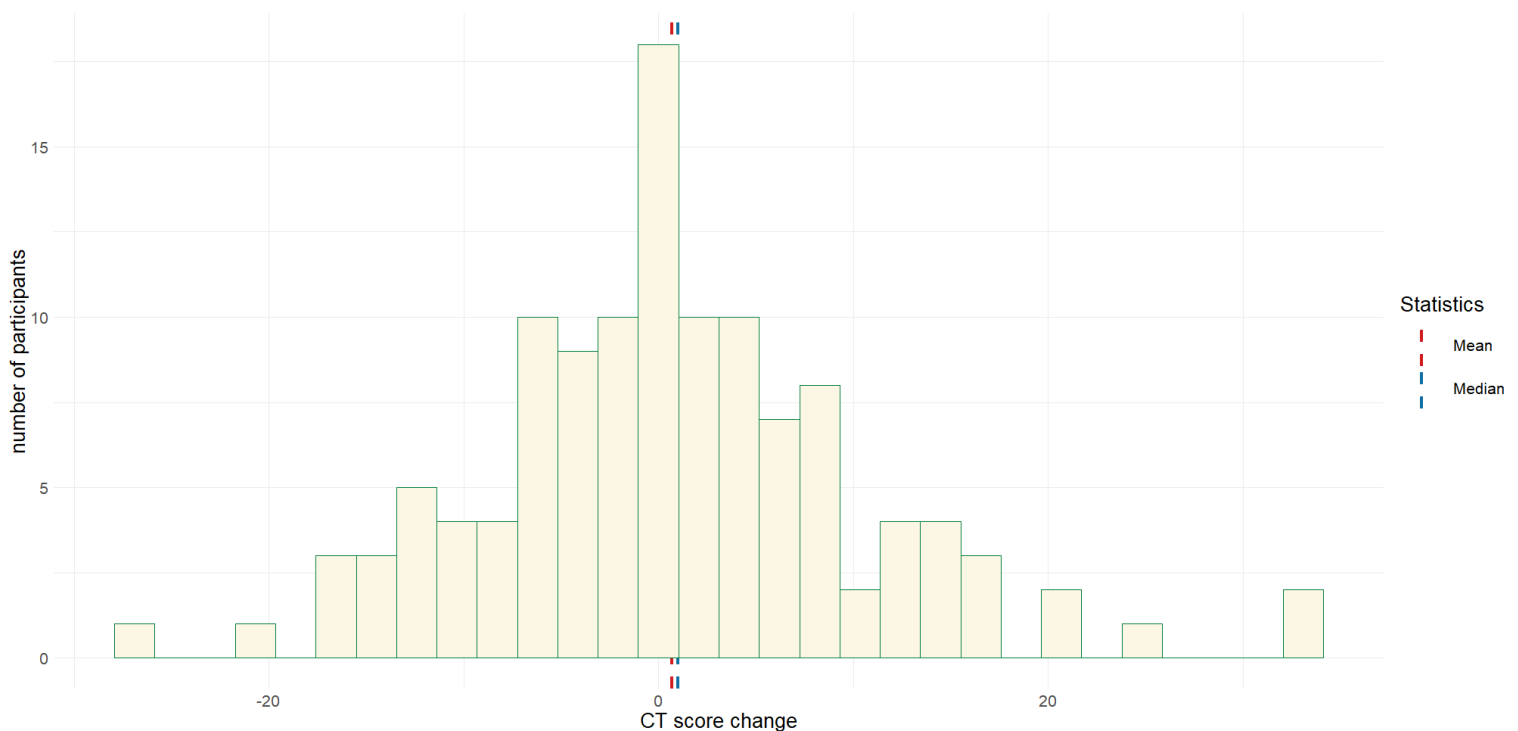
Finally, an association between *age* and test performance was tested. A Shapiro-Wilk normality test found participant *age* not to be normally distributed ($W = 0.92, p < .001$). Therefore Kendall's rank correlation was used and showed no significant correlation between ICTET-A scores at t1 and *age* ($r\tau = -0.03, p = .607$), nor

between t2 scores and *age* ($r\tau = 0.07, p = .276$). This suggests that *age* and test performance were not related, and *age* therefore need not be controlled for in analyses of t1 and t2 *CT scores*.

Change in CT score from t1 to t2

For the purposes of testing the study hypotheses, the change in test score from t1 to t2 is a key outcome variable. *CT change* was derived by subtracting t1 score from t2 score. Mean *CT change* was 0.69 ($SD = 9.93$), median 1, with a range of 60 from -27 to 33. A Shapiro-Wilk normality test found *CT change* in scores to be normally distributed ($W = 0.98, p = .067$). Figure 7 shows the distribution of *CT change*.

Figure 7: Histogram of *CT score change* from t1 to t2



CT change was compared across demographic variables to determine whether any of these need to be controlled for in further analyses.

First, *gender* differences in *CT change* were assessed. A one-way ANOVA found no significant difference in *CT change* between genders ($F(2,118) = 0.24, p = .786$). This suggests that changes in test performance did not differ between participants of different genders, and *gender* therefore does not need to be controlled for in analyses of *CT change*.

Differences in test *CT change* across *educational level* were assessed. A one-way ANOVA found no significant difference in *CT change* across *educational level* ($F(2,118) = 0.64, p = .528$). This suggests that changes in test performance did not differ between participants of different *educational level*, and *educational level* therefore does not need to be controlled for in analyses of *CT change*.

Due to the non-normal distribution of *age* in the sample, Kendall's rank correlation was used to test for a relationship between *CT change* and *age*; no correlation was found ($r\tau = 0.11, p = .086$). Thus, *age* was not a necessary variable to be controlled for in further analyses of *CT change*.

Reading engagement

Total engagement (i.e. total number of favourite authors) had a mean of 6.24 ($SD = 2.72$), median 6, with a range from 0-10. *Fiction engagement* had a mean of 4.86 ($SD = 2.85$), median 4, range from 0-10. *Nonfiction engagement* had a mean of 1.38 ($SD = 1.83$), median 1, and a range from 0-8.

As participants were initially asked to self-identify as either being fiction readers or nonfiction readers, comparing this self-identification with the participants' favourite authors listed allows for a check of consistency between participants' self-identification and their reading engagement. A one-way ANOVA showed a significant difference in

fiction engagement scores between reader types ($F(1,119) = 25.22, p < .001$), and post-hoc Tukey's comparison showed that participants who identified as fiction readers had an average increase of 2.37 (95% CI 1.44, 3.31) in *fiction engagement* scores compared to nonfiction readers ($p < .001$). A one-way analysis of means showed a significant difference in *nonfiction engagement* scores between participant types ($F(1,119) = 41.04, p < .001$), and a Games-Howell post-hoc test showed that participants who identified as nonfiction readers has an average increase of 1.87 [95% CI 1.29, 2.45] in *nonfiction engagement* scores compared to fiction readers ($p < .001$). Therefore how participants self-identified (as fiction or nonfiction readers) corresponded to their engagement with fiction and nonfiction as measured by the number of favourite authors of either type they listed.

Reading engagement across demographic variables

Reading engagement scores were compared across demographic variables.

First, *gender* differences in *reading engagement* were assessed. A one-way ANOVA was carried out, which found no significant difference in *total engagement* scores across genders ($F(2, 118) = 0.40, p = .673$), nor *fiction engagement* ($F(2, 118) = 0.40, p = .673$). A significant difference was found for *nonfiction engagement* across different genders ($F(2, 118) = 3.13, p = .047$); however Tukey's Post-hoc comparisons showed no significant differences. This suggests that *gender* has no relationship with *reading engagement*.

Differences in *reading engagement* across *educational level* were assessed. A one-way ANOVA was carried out, which found a significant difference in *total engagement* scores across *educational level* ($F(2, 118) = 7.19, p < .001$). Post-hoc

Tukey's comparisons showed that the significant difference ($p < .001$) was between *ND* ($M = 5.44$, $SD = 2.36$) and *PG* ($M = 7.43$, $SD = 2.86$) levels, with the *total engagement* increasing on average from *ND* to *PG* by 1.99 (95% CI 0.73, 3.25). The difference in *total engagement* scores for participants in the *ND* category and *UG* ($M = 6$, $SD = 2.59$) was not significant ($p = .647$), and neither was *UG* to *PG* ($p = .083$). A one-way ANOVA showed a significant difference in *fiction engagement* across education levels ($F(2, 118) = 6.78$, $p < .001$). Post-hoc Tukey's comparisons showed that the only significant difference ($p < .001$) was between *ND* ($M = 4$, $SD = 2.62$) and *PG* levels ($M = 6.05$, $SD = 3$), with *fiction engagement* increasing on average from *ND* to *PG* by 2.05 [95% CI 0.73, 3.37]. The difference in *fiction engagement* scores for participants in the *ND* category and *UG* ($M = 4.75$, $SD = 2.40$) was not significant ($p = .499$), and nor between *UG* and *PG* ($p = .154$). A one-way ANOVA showed no significant difference in *nonfiction engagement* across education levels ($F(2, 118) = 0.09$, $p = .918$). This suggests that postgraduates had greater *total engagement*, and greater *fiction engagement*, than participants without degrees. *Nonfiction engagement* was similar across participants of different *educational level*.

As *age* and *reading engagement* were not normally distributed, Kendall's rank correlation was used to test whether *total*, *fiction* and *nonfiction engagement* was correlated with *age*. *Total engagement* had a slight positive correlation with *age* ($r\tau = 0.21$, $p = .001$). *Fiction engagement* had a very slight positive correlation with *age* ($r\tau = 0.17$, $p = .016$). *Nonfiction engagement* had no significant correlation with *age* ($r\tau = 0.05$, $p = .597$). This suggests that older participants in the sample had slightly higher

reading engagement, and very slightly higher *fiction engagement*, than younger participants. *Nonfiction engagement* was similar across ages.

Reading log

Firstly, reading logs were assessed for the timeframes in which participants' completed their logs. The mean number of days between starting and finishing the reading log was 13.14 ($SD = 1.25$) median 13, with a range of 13 from 11 to 24. Although participants were asked to complete 14 consecutive days, this range accounts for the permitted 3 missing days from a log, and for some participants pausing their log for some interval and then resuming making entries (which was permitted so long as 14 days were ultimately completed).

Over the 14 day log period, participants logged a mean 16.99 *total entries* ($SD = 9.64$), median 16, with a range from 0 to 55. The *total time* had a mean of 24.75 ($SD = 23$), median 19.17, with a range from 0 to 143. In terms of *fiction entries*, the mean was 7.1 ($SD = 6.19$), median 5, with a range from 0 to 28. *Fiction time* had a mean of 12.04 ($SD = 17.16$), median 5.67, with a range from 0 to 108. For *nonfiction entries*, the mean was 9.89 ($SD = 8.84$), median 8, with a range from 0 to 42. *Nonfiction time* had a mean of 12.71 ($SD = 15.15$), median 7.67, with a range from 0 to 82. The mean *NT* score was 3.6 ($SD = 0.71$), median 3.66, with a range from 0 to 5.

Reading amounts logged across experimental groups

Differences in reading logged between the different experimental groups were assessed, in order to determine whether group assignment was associated with greater amounts of reading of the type matching the group. Table 9 shows the *total* reading in

entries and *time* logged by the different experimental groups. Table 10 shows the *fiction* reading logged by the different groups, and Table 11 the *nonfiction* reading.

Table 9: Description of total reading logged by experimental group

Group	Total entries M	SD	median	Total time M	SD	median
Cf	13.82	7.93	13.00	26.03	28.10	23.67
Cn	17.43	10.01	15.00	23.17	20.13	13.33
Ff	18.00	8.16	17.50	25.24	21.76	19.92
Fn	16.30	8.80	17.50	26.99	28.80	14.97
Nf	20.06	11.22	17.50	23.75	18.60	22.19
Nn	18.31	11.92	18.50	22.61	16.76	18.50

Table 10: Description of fiction reading logged by experimental group

Group	Fiction entries M	SD	median	Fiction time M	SD	median
Cf	8.04	6.83	6.50	17.62	23.23	7.67
Cn	3.43	4.58	1.00	5.38	8.66	1.17
Ff	13.25	5.88	12.00	19.06	21.06	12.55
Fn	7.75	3.11	7.00	9.20	10.02	5.59
Nf	7.94	6.18	8.00	15.11	18.61	8.00
Nn	2.81	4.23	0.00	4.90	7.43	0.00

Table 11: Description of nonfiction reading logged by experimental group

Group	Nonfiction entries M	SD	median	Nonfiction time M	SD	median
Cf	5.79	4.80	6.00	8.41	9.67	4.59
Cn	14.00	11.09	11.00	17.79	19.06	11.07
Ff	4.75	5.71	2.00	6.18	7.16	2.50
Fn	8.55	7.72	9.50	17.78	23.37	6.72
Nf	12.11	8.59	8.00	8.64	5.97	7.85
Nn	15.50	9.19	14.00	17.71	12.32	17.50

First, *total entries* made and *time* logged were compared across experimental groups. A one-way ANOVA found no significant difference in *total entries* logged between participants in different experimental groups ($F(5,115) = 1.01, p = .365$), nor for *total time* ($F(5,115) = 0.11, p = .99$). This suggests that in total, participants in the different experimental and control groups made a similar number of entries and spent a similar time reading.

Next, *fiction entries* logged were compared across experimental groups. A one-way analysis of means found a significant difference in *fiction entries* logged between participants in different experimental groups ($F(5,115) = 8.63, p < .001$), and a Games-Howell post-hoc test showed the significant differences to be: *fiction entries* logged increased on average between *Cn* and *Ff* by 9.82 ([95% CI 4.44, 15.19] $p < .001$); *fiction entries* logged decreased on average between *Cf* and *Nn* by -5.22 ([95% CI -10.21, -0.24] $p = .035$); *fiction entries* logged increased on average between *Cn* and

Fn by 4.32 ([95% CI 0.77, 7.86] $p = .009$); *fiction entries* logged decreased on average between *Ff* and *Fn* by -5.50 ([95% CI -10.58, -0.42] $p = .029$); *fiction entries* logged decreased on average between *Ff* and *Nn* by -10.44 ([95% CI -15.98, -4.89] $p < .001$); *fiction entries* logged decreased on average between *Fn* and *Nn* by -4.94 ([95% CI -8.8, 2 -1.06] $p = .007$); all other differences between groups were not significant. This suggests that more *fiction entries* were logged by *Ff* and *Fn* participants than by *Cn* participants; i.e. control group nonfiction readers did not log as many *fiction entries* as readers of either kind in the fiction group. *Cf*, *Ff* and *Fn* participants logged more *fiction entries* than *Nn* participants; i.e. nonfiction readers assigned to the nonfiction group did not log as many *fiction entries* as fiction readers in the control or fiction condition, nor as many as nonfiction readers assigned fiction. *Ff* participants also logged more *fiction entries* than *Fn* participants; i.e. though they were assigned the same fiction reading, fiction readers logged more *fiction entries* than nonfiction readers in this condition.

Additionally, *fiction time* logged was compared across experimental groups. A one-way analysis of means found a significant difference in *fiction time* logged between participants in different experimental groups ($F(5,115) = 2.80, p = .02$). However, a Games-Howell post-hoc test showed no significant differences between specific experimental groups. This suggests that there was no specific significant difference between experimental groups in terms of *fiction time*.

Next, *nonfiction entries* made were compared across experimental groups. A one-way ANOVA found a significant difference in *nonfiction entries* logged between participants in different experimental groups ($F(5,115) = 5.89, p < .001$), and a

Tukey's post-hoc test showed the significant differences to be: *nonfiction entries* logged increased between *Cf* and *Cn* on average by 8.23 ([95% CI 1.64 14.79] $p = .006$); *nonfiction entries* logged increased between *Cf* and *Nn* on average by 9.71 ([95% CI 2.39, 17.04] $p = .003$); *nonfiction entries* logged decreased on average between *Ff* and *Cn* by -9.25 ([95% CI -16.86, -1.64] $p = .008$); *nonfiction entries* logged increased between *Ff* and *Nn* on average by 10.75 ([95% CI 2.49, 19.01] $p = .003$); no other differences between groups were significant. This suggests that more *nonfiction entries* were logged by *Cn* and *Nn* participants than by *Cf* participants; *Cn* and *Nn* participants logged more nonfiction than *Ff* participants; i.e. control and fiction group fiction readers did not log as many *nonfiction entries* as nonfiction readers assigned to either the control or nonfiction conditions.

Lastly, *nonfiction time* logged was compared across experimental groups. A one-way analysis of means found a significant difference in *nonfiction time* logged between participants in different experimental groups ($F(5,115) = 2.82, p = .019$), and a Games-Howell post-hoc test showed the only significant difference to be between *Ff* and *Nn*, with *nonfiction time* increasing from *Ff* to *Nn* on average by 11.53 ([95% CI 0.52, 22.54] $p = .036$). Therefore more *time* was spent reading nonfiction by nonfiction readers in the nonfiction group than fiction readers in the fiction group.

Figure 8 shows differences in *fiction* and *nonfiction entries* logged by the different experimental groups, and Figure 9 shows time (in half-hours) logged.

Figure 8: Box plots of fiction and nonfiction items logged by different experimental groups

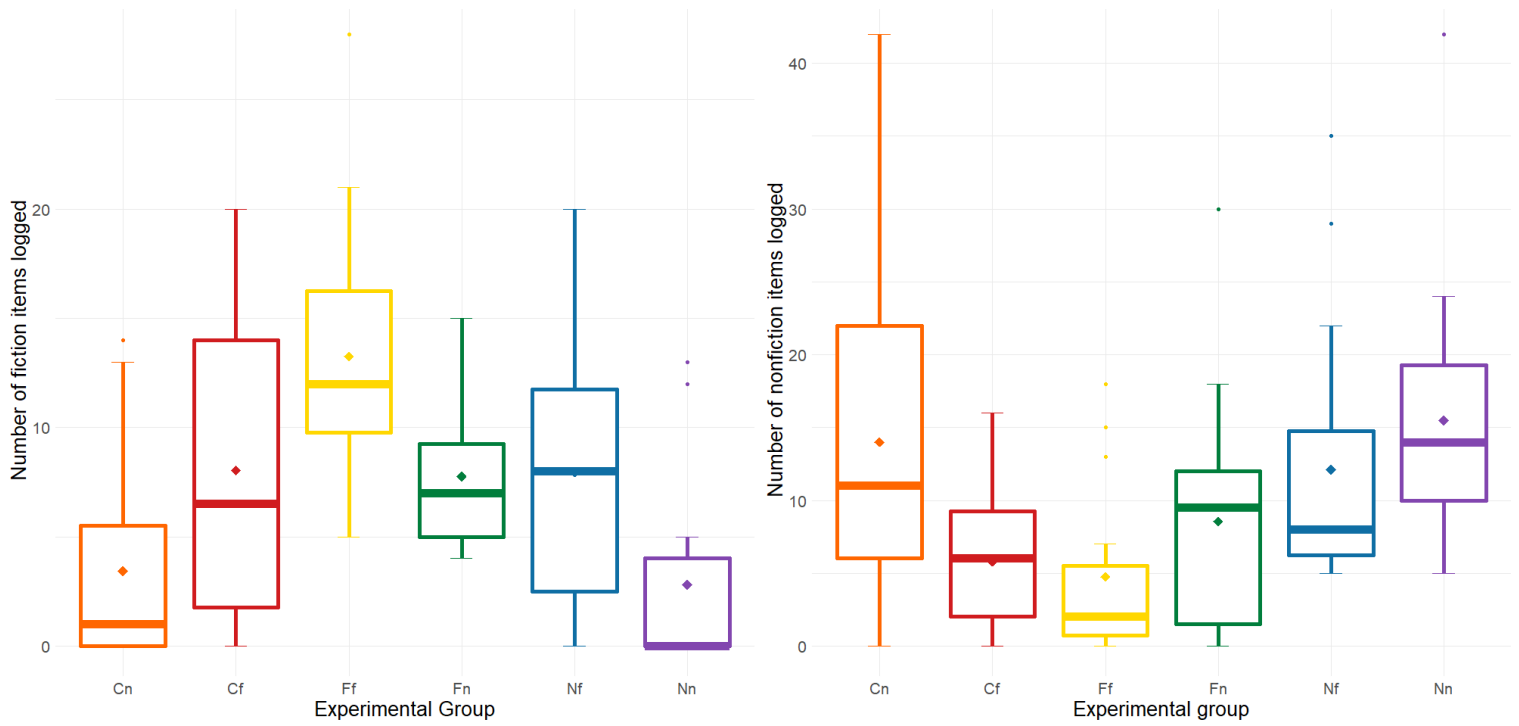
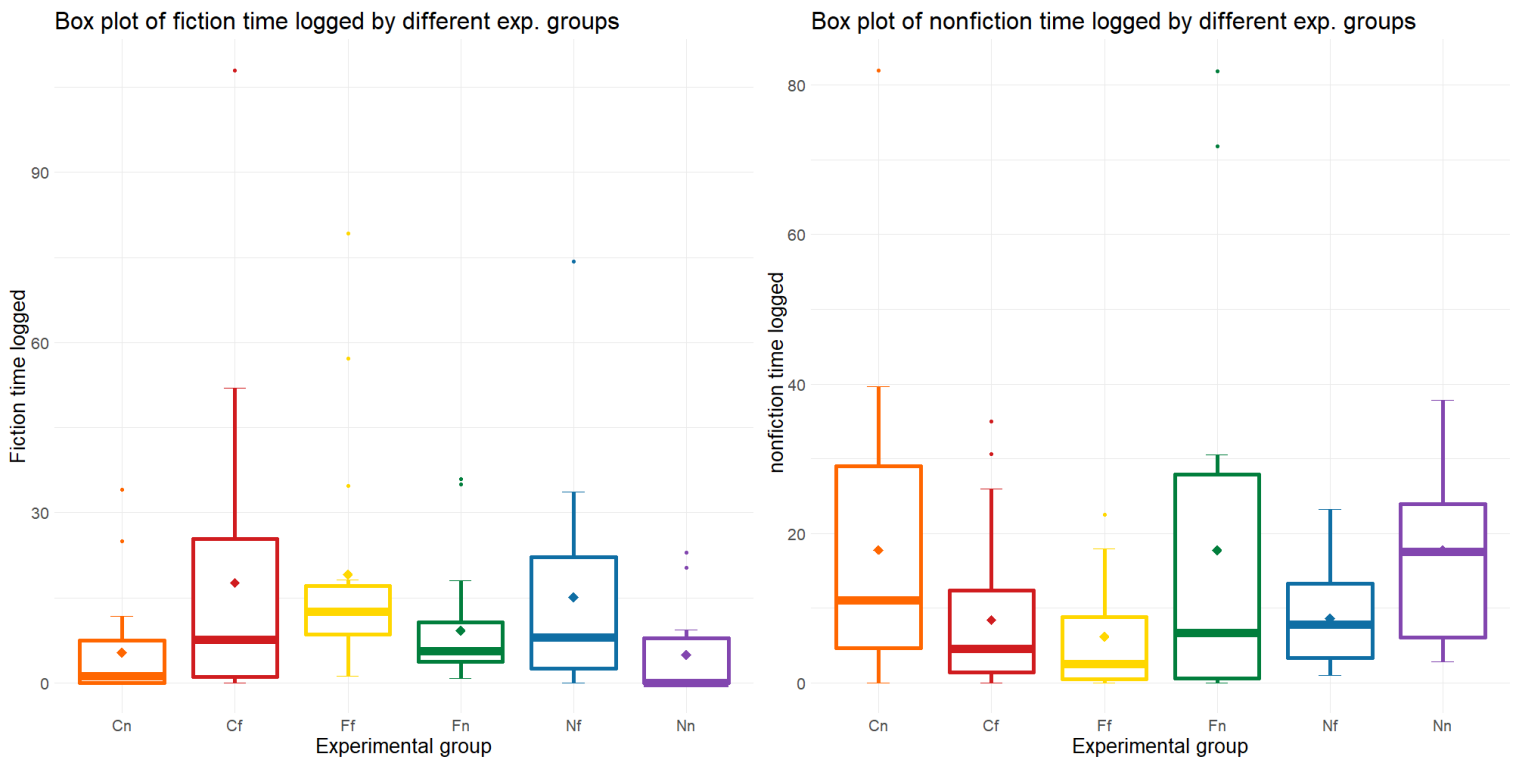


Figure 9: Box plots of fiction and nonfiction time logged by different experimental groups



Next, *average NT* scores for participants in the different experimental groups were compared. As some participants were assigned reading that was incongruous with their usual reading preferences (e.g. fiction readers being assigned nonfiction reading), this could impact the reading experience and thus there could be differences in how transported by their reading the different experimental groups were.

Table 12 shows a description of *average NT* score for each experimental group.

Table 12: Description of average NT by experimental group

Group	<i>M</i>	<i>SD</i>	Median
Cf	3.74	0.96	3.96
Cn	3.58	0.77	3.50
Ff	3.68	0.56	3.91
Fn	3.45	0.61	3.47
Nf	3.45	0.51	3.60
Nn	3.65	0.53	3.62

Note. These values refer to a grand mean across participants, each of whom had a mean *NT* value.

A one-way ANOVA showed no significant difference in *average transportation* scores between different experimental groups ($F(5,115) = 0.60, p = .703$). This suggests that being placed into different experimental groups (i.e. being assigned different reading or not) did not alter how transporting participants found their reading.

Reading amounts logged and reading engagement

Next, reading logged was assessed for a correlation with *reading engagement*, to test whether participants who had a higher *reading engagement* (i.e. listed more

favourite authors) had any difference in the amount they read, both in *total* and in terms of *fiction* and *nonfiction*.

As the reading log data (i.e. reading *entries* or *time*) were not normally distributed, Kendall's rank correlation tau was used to test whether *total*, *fiction* and *nonfiction entries* and *time* logged were correlated with *reading engagement*. *Total entries* logged had a slight positive correlation with *total engagement* ($r\tau = 0.27, p < .001$), as did *total time* ($r\tau = 0.22, p = .001$). *Fiction entries* logged had a positive correlation with *fiction engagement* ($r\tau = 0.36, p < .001$), as did *fiction time* ($r\tau = 0.29, p < .001$). *Nonfiction entries* logged also had a slight positive correlation with *nonfiction engagement* ($r\tau = 0.25, p < .001$), likewise *nonfiction time* ($r\tau = 0.25, p < .001$). This suggests that *reading engagement* as measured by numbers of favourite authors listed by participants did relate to reading behaviour.

Reading logged across demographic variables

Next, reading logged in terms of *total*, *fiction* and *nonfiction entries* and *time*, as well as *average transportation*, were compared across demographic variables to ascertain whether there were any trends in reading between different demographic groups.

First, *gender* differences were assessed. A one-way ANOVA showed no significant difference in *total entries* logged between *gender* groups ($F(3,117) = 0.36, p = .782$), and no significant difference in *total time* between *gender* groups ($F(3,117) = 0.53, p = .661$). There was no significant difference between *gender* for *fiction entries* logged ($F(3,117) = 0.62, p = .603$), nor *fiction time* ($F(3,117) = 0.09, p = .964$).

There was also no significant difference between *gender* for *nonfiction entries* logged

($F(3,117) = 0.31, p = .82$), nor *nonfiction time* ($F(3,117) = 0.58, p = .629$). This suggests that participants with different *gender* did not have different reading habits over the 14 day log period. Furthermore, a one-way ANOVA showed no significant difference in *average transportation* between genders ($F(3,117) = 0.67, p = .57$).

Next, differences in reading across *educational level* were tested. A one-way ANOVA showed no significant difference in *total entries* logged across *educational level* ($F(2,118) = 1.59, p = .21$), and no significant difference in *total time* across *educational level* ($F(2,118) = 0.91, p = .407$). There was no significant difference across *educational level* for *fiction entries* logged ($F(2,118) = 0.87, p = .42$), nor *fiction time* ($F(2,118) = 0.14, p = .867$). A one-way analysis of means showed no significant difference across *educational level* for *nonfiction entries* logged ($F(2,118) = 2.12, p = .128$), and an ANOVA showed no significant difference for *nonfiction time* ($F(2,118) = 1.53, p = .222$). This suggests that participants with different educational qualifications did not have different reading habits over the 14 day log period. Additionally, a one-way ANOVA showed no significant difference in *average transportation* across *educational level* ($F(2,118) = 0.77, p = .926$).

As the reading data (i.e. *entries* and *time* logged) were not normally distributed, Kendall's rank correlation tau was used to test whether *total, fiction* and *nonfiction entries* and *time* logged were correlated with *age*. *Total entries* logged had a slight positive correlation with *age* ($r\tau = 0.29, p < .001$), as did *total time* ($r\tau = 0.28, p < .001$). *Fiction entries* logged had a very slight positive correlation with *age* ($r\tau = 0.15, p = .018$), as did *fiction time* ($r\tau = 0.20, p = .002$). *Nonfiction entries* logged also had a slight positive correlation with *age* ($r\tau = 0.19, p = .003$), likewise *nonfiction time* ($r\tau$

= 0.20, $p = .002$). This suggests that there is a slight increase in reading of all kinds logged over the 14 day period with increase in participant *age*. Finally, Kendall's rank correlation also found *average transportation* scores to have a very slight positive correlation with *age* ($r\tau = 0.15$, $p = .018$).

5.2.2. Inferential statistics

Baseline CT scores

In this section the following hypothesis will be assessed:

H₁: Higher reading engagement will have a significant association with higher baseline critical thinking score at t1.

Furthermore, *reading engagement* will be broken down into *fiction* and *nonfiction engagement* as a further exploratory avenue of analysis.

In order to test H_1 , multiple regression analysis was used to test the association between the measured variables prior to the reading log, and *t1 CT scores*. However, prior to conducting multiple regression analyses, basic correlational analyses were first conducted to ascertain relationships between individual predictor variables and the outcome variable of *t1 CT score*.

As the *reading engagement* data were not normally distributed, Kendall's rank correlation tau was used to test whether *total*, *fiction* and *nonfiction engagement* was correlated with *t1 CT score*. *Total engagement* had a very slight positive correlation with *t1 CT score* ($r\tau = 0.18$, $p = .006$). *Fiction engagement* had a very slight positive correlation with *t1 CT score* ($r\tau = 0.19$, $p = .005$). *Nonfiction engagement* had no significant correlation with *t1 CT score* ($r\tau = 0.01$, $p = .94$). This suggests that

participants with higher *t1 CT score* had slightly higher *reading engagement*, and slightly higher *fiction engagement*, than lower scoring participants. *Nonfiction engagement* was similar for participants with different *t1 CT score*. This suggests that a further regression analysis is warranted.

As in the descriptive analyses *age* and *gender* were not found to be associated with differences in *t1 CT scores*, these variables were not included in the multiple regression analysis. *Educational level* was also not found to be associated with differences in *t1 CT scores*, however as this variable is conceptually relevant, it was included as a control. *Reading engagement* in terms of *total* and *fiction engagement* was significantly associated with *t1 CT score*, however *nonfiction engagement* was not; however as these are all important variables for hypothesis testing, they were included in the multiple regression.

Regression 1 included only *educational level* as the predictor variable, as this is the variable to be controlled for (or partialled out) in the hierarchical analysis. In regression 2, *total engagement* was added. In regressions 3 and 4, *total engagement* was not included but *nonfiction* and *fiction engagement* were added respectively; *total engagement* was removed to avoid multicollinearity (as this is the sum of nonfiction and fiction scores). Regression 5 contained both *nonfiction* and *fiction engagement* together.

Table 13 shows a comparison of regression models, displaying the significance of the different predictor variables, and showing a comparison of R^2 values across the different models.

Table 13: Comparison of regression models for CT t1 score and reading engagement

Predictor Variables	Regression Number				
	1	2	3	4	5
Educational level					
ND – UG	.187	.246	.190	.264	.266
ND – PG	.044 *	.214	.045 *	.199	.247
Total	-	.032 *	-	-	-
Nonfiction	-	-	.974	-	.372
Fiction	-	-	-	.04*	.026 *
R^2	.04	.07	.04	.07	.08
ΔR^2	0.04	0.04	-	-	-
	0.04	-	0	-	0.04
	004	-	-	.03	-

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

This comparison shows us that regression 5 had the best fit based on R^2 value. We can also see that having a *PG* degree rather than *ND* was a significant predictor of *t1 CT score* only when *total* and *fiction engagement* were not included in the regression. Both *total engagement* and *fiction engagement* were significant predictors when added to the regression, and their inclusion also improved model fit (i.e. increased R^2).

Table 14 describes the results from regressions 1 and 2 as steps in the hierarchical regression logic.

Table 14: Hierarchical regression results for t1 CT score with total engagement

Predictor Variable	$B(SE)$	95% CI for B		β	R^2	ΔR^2
		LL	UL			
Step 1 (regression 1)					.04	.04
Educational level						
ND – UG	2.84 (2.14)	-1.40	7.08	.187		
ND – PG	3.66 (1.79)	0.10	7.21	.044 *		
Step 2 (regression 2)					.07	.04*
Educational level						
ND – UG	2.47 (2.12)	-1.72	6.66	.246		
ND – PG	2.33 (1.87)	-1.37	6.04	.214		
Total reading	0.66 (0.31)	0.06	1.27	.032*		

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

This hierarchical regression shows that *educational level* alone in step 1 has a significant relationship with CT score at t1, with participants with *PG* degrees having significantly higher scores than participants with *ND*. However, when *educational level* is controlled for and *total engagement* is added to the regression in step 2, *educational level* no longer has a significant relationship with CT score, but *total engagement* does.

Regression 2 was assessed for accuracy, and the results can be found in Appendix B under the Model testing heading of Baseline CT scores on page 422. Based upon this model, when *educational level* is controlled for *total engagement* is significantly associated with increased *t1 CT score*; per additional author listed as a favourite, *t1 CT score* increased by 0.66 ($p < .05$). The model had an R^2 of .07, thus

explaining 7% of the variance in *t1 CT score*. The results of this model therefore support H₁: higher *reading engagement* did have a significant association with higher baseline critical thinking score at t1.

Fiction and Nonfiction engagement and t1 CT score

Next, *total engagement* was broken down into *nonfiction* and *fiction engagement* to explore whether either type of engagement would have a stronger association with *t1 CT score*. *Nonfiction* and *fiction engagement* were added in two separate versions of the hierarchical regression. Table 15 describes the results from regressions 1, 3 and 4.

Table 15: Hierarchical regression results for *t1 CT score* with reading engagement type

Predictor Variable	<i>B</i> (<i>SE</i>)	95% CI for <i>B</i>		β	<i>R</i> ²	ΔR^2	
		LL	UL				
Step 1 (regression 1)						.04	.04
Educational level							
ND – UG	2.84 (2.14)	-1.40	7.08	.187			
ND – PG	3.66 (1.79)	0.10	7.21	.044 *			
Step 2a (regression 3)						.04	0
Educational level							
ND – UG	2.83 (2.15)	-1.42	7.10	.19			
ND – PG	3.66 (1.80)	0.09	7.22	.045*			
Nonfiction	-0.01 (0.43)	-0.88	0.85	.974			
Step 2b (regression 4)						.07	.03*
Educational level							
ND – UG	2.39 (2.12)	-1.82	6.59	.264			
ND – PG	2.41 (1.89)	-1.29	6.59	.199			
Fiction	0.61 (0.29)	0.03	1.19	.04*			

Note. **p* < .05 ***p* < .01 ****p* < .001

This shows that when *nonfiction engagement* is added in step 2a, it does not have a significant relationship with *t1 CT score*. However, when *fiction engagement* is added in step 2b, it does have a significant positive relationship with *t1 CT score*.

Therefore for the next regression analysis, *nonfiction* was added was in a step before *fiction*, to control for it and to determine if *fiction engagement* would then still have a significant positive relationship with CT score. This forms the final hierarchical regression chain, taking steps from including only *educational level* as a control variable, next controlling for *nonfiction engagement*, and finally adding *fiction engagement* as the final regression step. Table 16 describes the results from regressions 1, 3 and 5.

Table 16: Hierarchical regression results for CT t1 score with both fiction and nonfiction

Predictor Variable	$B(SE)$	95% CI for B		β	R^2	ΔR^2
		LL	UL			
Step 1 (regression 1)					.04	.04
Educational level						
ND – UG	2.84 (2.14)	-1.40	7.08	.187		
ND – PG	3.66 (1.79)	0.10	7.21	.044 *		
Step 2 (regression 3)					.04	0
Educational level						
ND – UG	2.83 (2.15)	-1.42	7.10	.190		
ND – PG	3.66 (1.80)	0.09	7.22	.045*		
Nonfiction	-0.01 (0.43)	-0.88	0.85	.974		
Step 3 (regression 5)					.08	.04*
Educational level						
ND – UG	2.38 (2.13)	-1.83	6.59	.266		
ND – PG	2.19 (1.89)	-1.54	5.93	.247		
Nonfiction	0.42 (0.47)	-0.51	1.36	.372		
Fiction	0.73 (0.32)	0.09	1.36	.026*		

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

Based upon this comparison of models, regression 2 was found to be a better fit for the data than regression 1; i.e. including *total engagement* in addition to *educational level* improved model fit for *t1 CT score*. Additionally, regression 5 was found to be a better fit for the data than regression 1, 3, or 4; i.e. adding *fiction engagement* in addition to *educational level* and *nonfiction engagement* improved model fit for *t1 CT score*. Regression 5 is the best fitting model overall, based on having the highest R^2 value. This model was found to be significantly different to the prior models ($F(1, 116) = 5.12, p = .025$), with an R^2 of 0.8.

Regression 5 was found to be the model that best fit the data, and it was also found to be accurate and generalisable (see Appendix B Model testing; Baseline CT scores, page 424). Based upon this model, when *educational level* is controlled for *nonfiction engagement* does not significantly predict any difference in *t1 CT score*; however, per additional fiction author listed as a favourite, *t1 CT score* increased by 0.73 ($p < .05$). The model had an R^2 of .08, thus explaining 8% of the variance in *t1 CT score*.

Summary of associations with baseline CT score

In this section, the following hypothesis was tested: H₁: Higher reading engagement will have a significant association with higher baseline critical thinking score at t1.

Total engagement was found to be associated with higher *t1 CT score*, thus H₁ was supported. Furthermore, exploratory analysis found that higher *fiction engagement* was significantly associated with higher *t1 CT score*, while *nonfiction engagement* was not.

CT change from t1 to t2

In this section the following hypotheses will be tested:

H₂: Higher amounts of reading recorded over the reading log period will have a significant association with increases in critical thinking score from t1 to t2.

H₃: Higher amounts of fiction reading recorded over the reading log period will have a stronger association with increases in critical thinking score from t1 to t2 than nonfiction.

Before fully testing these hypotheses, basic correlation analyses were conducted to test for a significant association between reading amounts and *CT change*. As reading log data (i.e. *entries* and *time*) were not normally distributed, Kendall's rank correlation was used. This found *total entries* logged to have a very slight positive correlation with *CT change* ($r\tau = 0.14, p = .024$), and a slightly more positive relationship between *total time* and *CT change* ($r\tau = 0.21, p = .001$). There was a stronger positive relationship between *fiction entries* logged and *CT change* ($r\tau = 0.38, p < .001$), with the same relationship for *fiction time* ($r\tau = 0.37, p < .001$). There was no significant relationship between *nonfiction entries* logged and *CT change* ($r\tau = -0.10, p = .104$), nor for *nonfiction time* ($r\tau = -0.02, p = .805$). This suggests that as reading amount in terms of *entries* and *time* increases, *CT change* increases, and this is particularly the case for fiction reading, but nonfiction reading has no relationship with *CT change*. This suggests the hypotheses may be supported, and warrant further investigation through regression analyses.

Reading entries

In order to assess how *entries* recorded over the 14 day log period related to *CT change*, hierarchical multiple regression was used to partial out relevant covariates and test the effects for the predictor variables.

The prior descriptive analyses showed no significant relationship between *CT change* and *gender*, *educational level*, or *age*. Conceptually, these variables are not expected to play a role as they are stable over the 14 day period. Therefore these variables were not included in the regression analysis.

Regression 1 included only *total entries* as the predictor variable. In regressions 2 and 3, *total entries* was not included but *nonfiction*, and then both *nonfiction* and *fiction*, *entries* were added respectively; *total entries* was removed to avoid collinearity (as this is the sum of *nonfiction* and *fiction entries*). *Nonfiction* was added prior to *fiction* in the hierarchical regression steps as the latter is hypothesised to have the greater effect, and thus first explaining variance in *CT change* with *nonfiction entries* was the most conservative approach (Stanovich & Cunningham, 2004).

First, *total entries* logged was assessed as the predictor variable of interest. To this end, a hierarchical multiple regression was used (regression 1). Table 17 shows the results.

Table 17: Hierarchical regression results for *CT change* with *total read entries*

Predictor Variable	<i>B</i> (<i>SE</i>)	95% CI for <i>B</i>		β	<i>R</i> ²	ΔR^2
		LL	UL			
Step 1 (regression 1)					.04	.04*
Total entries	0.20 (0.09)	0.02	0.39	.029*		

Note. **p* < .05 ***p* < .01 ****p* < .001

Regression 1 was tested for accuracy (see Appendix B Model testing; *CT change* from t1 to t2, page 426) and was found to be accurate and generalisable. Based upon this model, *total entries* logged over the course of the 14 day reading log period was significantly associated with *CT change*; per additional entry logged, *CT change* increased by 0.20 (*p* < .05). The model had an *R*² of .04, thus explaining 4% of the variance in *CT change*. The results of this model therefore do support H₂: higher amounts of reading recorded over the reading log period in number of *entries* logged had a significant association with increases in *CT change* from t1 to t2.

Nonfiction and fiction reading entries as predictors for CT change

The next analyses, testing H₃, added *nonfiction* and *fiction entries* as predictors in respective steps of hierarchical regression in addition to *age*. Table 18 shows the results from regressions 2 and 3.

Table 18: Hierarchical regression results for CT change with read entries by type

Predictor Variable	<i>B</i> (<i>SE</i>)	95% CI for <i>B</i>		β	<i>R</i> ²	ΔR^2	
		LL	UL				
Step 1 (regression 2)						.001	.001
Nonfiction entries	-0.09 (0.10)	-0.30	0.11	.37			
Step 2 (regression 3)						.18	.18***
Nonfiction entries	0.01 (0.10)	-0.18	0.20	.906			
Fiction entries	0.69 (0.14)	0.42	0.96	<.001***			

Note. **p* < .05 ***p* < .01 ****p* < .001

This shows that *nonfiction entries* did not have a significant relationship with *CT change*, and the regression with *nonfiction entries* alone was not significantly different from an empty model. However, when *fiction entries* was added in step 2, it did have a significant positive relationship with *CT change*, and the model was significantly different from model 2. Based upon this comparison of models, regression 3 was found to be a better fit for the data than regression 2.

Regression 3 was tested for accuracy (see Appendix B Model testing; CT change from t1 to t2, page 428) and was found to be accurate and generalisable. Based upon this model, when *nonfiction entries* are controlled for, *fiction entries* were significantly associated with increased *CT change*; per additional *fiction entry* logged, *CT change* increased by 0.69 (*p* < .01). The model had an *R*² of .18, thus explaining 18% of the variance in *CT change*. The results of this model therefore support H₃:

higher amounts of fiction reading in *entries* made during the reading log period did have a stronger association with increases in *CT change* from t1 to t2 than nonfiction, as indeed only *fiction entries* were associated with increases in *CT change*, while *nonfiction entries* had no significant association.

Reading time

In order to assess how reading time (in half-hours) recorded over the 14 day log period related to *CT change* from t1 (before the log) to t2 (after the log), hierarchical multiple regression was used in the same manner as for reading *entries*.

Regression 1 included only *total time*. In regressions 2 and 3, *total time* was not included but *nonfiction*, and then both *nonfiction* and *fiction*, *time* were added respectively; *total time* was removed to avoid collinearity (as this is the sum of *nonfiction* and *fiction time*).

Table 19 describes the results from regression 1, predicting *CT change* from *total time*.

Table 19: Hierarchical regression results for *CT change* with *total time*

Predictor Variable	<i>B</i> (<i>SE</i>)	95% CI for <i>B</i>		β	<i>R</i> ²	ΔR ²
		LL	UL			
Step 1 (regression 1)					.14	.14***
Total time	0.16 (0.04)	0.90	0.24	<.001***		

Note. **p* < .05 ***p* < .01 ****p* < .001

Regression 1 was tested for accuracy (see Appendix B Model testing; *CT change* from t1 to t2, page 430) and was found to be accurate and generalisable. Based upon this analysis, *total time* was a significant predictor of *CT change*; per additional half-hour of reading logged, *CT change* increased by 0.16 (*p* < .001). The model had

an R^2 of .14, thus explaining 14% of the variance in *CT change*. The results of this model therefore support H₂: higher amounts of reading recorded in half-hour increments over the reading log period did have a significant association with increases in *CT change* from t1 to t2.

Nonfiction and fiction reading time as predictors for CT change

The next analyses, testing H₃, included *nonfiction* and *fiction time* as predictors in respective steps of the hierarchical regression. Table 20 shows the results from regressions 2 and 3.

Table 20: Hierarchical regression results for CT change with reading time by type

Predictor Variable	$B(SE)$	95% CI for B		β	R^2	ΔR^2	
		LL	UL				
Step 1 (regression 2)						.03	.03
Nonfiction time	0.11 (0.06)	-0.01	0.22	.078			
Step 2 (regression 3)						.16	.13***
Nonfiction time	0.10 (0.06)	-0.01	0.21	.065			
Fiction time	0.18 (0.05)	0.08	0.29	.001***			

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

Based upon this comparison of models, regression 3 was found to be a better fit for the data than regression 2; i.e. including *fiction time* in addition to *nonfiction* improved model fit for *CT change*. Regression 2 did not improve model fit in comparison with an empty model.

Regression 3 was tested for accuracy (see Appendix B Model testing; CT change from t1 to t2, page 432) and was found to be accurate and generalisable. Based upon this model, when *nonfiction time* is controlled for, *fiction time* was significantly

associated with increased *CT change*; per additional fiction half-hour of reading logged, *CT change* increased by 0.18 ($p < .001$). The model had an R^2 of .16, thus explaining 16% of the variance in *CT change*. The results of this model therefore support H₃: higher amounts of *fiction time* in half-hours logged over the course of the 14 day log period did have a stronger association with increases *CT change* than *nonfiction time*, as indeed only *fiction time* was associated with increases in *CT change*, while *nonfiction time* had no significant association.

Summary of associations with CT score change

In this section, the following hypotheses were tested:

H₂: Higher amounts of reading recorded over the reading log period will have a significant association with increases in critical thinking score from t1 to t2.

H₃: Higher amounts of fiction reading recorded over the reading log period will have a stronger association with increases in critical thinking score from t1 to t2 than nonfiction.

Total entries logged over the course of the 14 day reading log period, and *total time*, were significantly associated with *CT change*; thus H₂ was supported.

When *nonfiction entries* were controlled for, *fiction entries* were significantly associated with increased *CT change*, and likewise for *fiction time*. These results therefore support H₃. It should be noted that *nonfiction* logged both in terms of *entries* and *time* had no significant association with *CT change*.

Mediation by narrative transportation

H₄: Narrative transportation will mediate the relationships predicted in H₂ and H₃.

Kendall's rank correlation found no significant correlation between *average NT* and *CT change* ($r\tau = 0.10$, $p = .124$). This suggests that how transported participants were overall by their reading was not related to changes in their CT test performance.

However, as conceptually it was hypothesised that how transported one is by reading will influence the impact of the reading, the effects of both reading *entries* and reading *time* logged were tested for mediation by *average NT*. The mediation analyses can be found in Appendix B: Study Two under Mediation analysis results. To summarise these, the participants' *average NT* was not found to mediate the relationships between *total*, *nonfiction*, or *fiction entries or time*, and *CT change*. H₄ is therefore not supported; *average NT* did not mediate the relationships predicted in H₂ or H₃.

Experimental Groups

In this section the following hypothesis will be assessed:

H₅: Participants assigned to the fiction experimental group will have greater increases in their critical thinking test scores from t1 to t2 than either the nonfiction group or the control group.

Multiple regression analyses were conducted to test the hypothesis by determining whether belonging to any of the experimental groups was significantly associated with increased *CT change*.

Prior to conducting the multiple regressions, basic analyses were conducted to determine if there was any significant association between assigned reading group and *CT change*. To explore the effects of the assigned readings on different reader types, the experimental groups were tested at the level of the six assigned blocks: control group fiction readers (*Cf*) ($n = 28$), control group nonfiction readers (*Cn*) ($n = 23$), fiction group fiction readers (*Ff*) ($n = 16$), fiction group nonfiction readers (*Fn*) ($n = 20$), nonfiction group fiction readers (*Nf*) ($n = 18$), and nonfiction group nonfiction readers (*Nn*) ($n = 16$). Table 21 shows a description of *CT change* across the experimental groups.

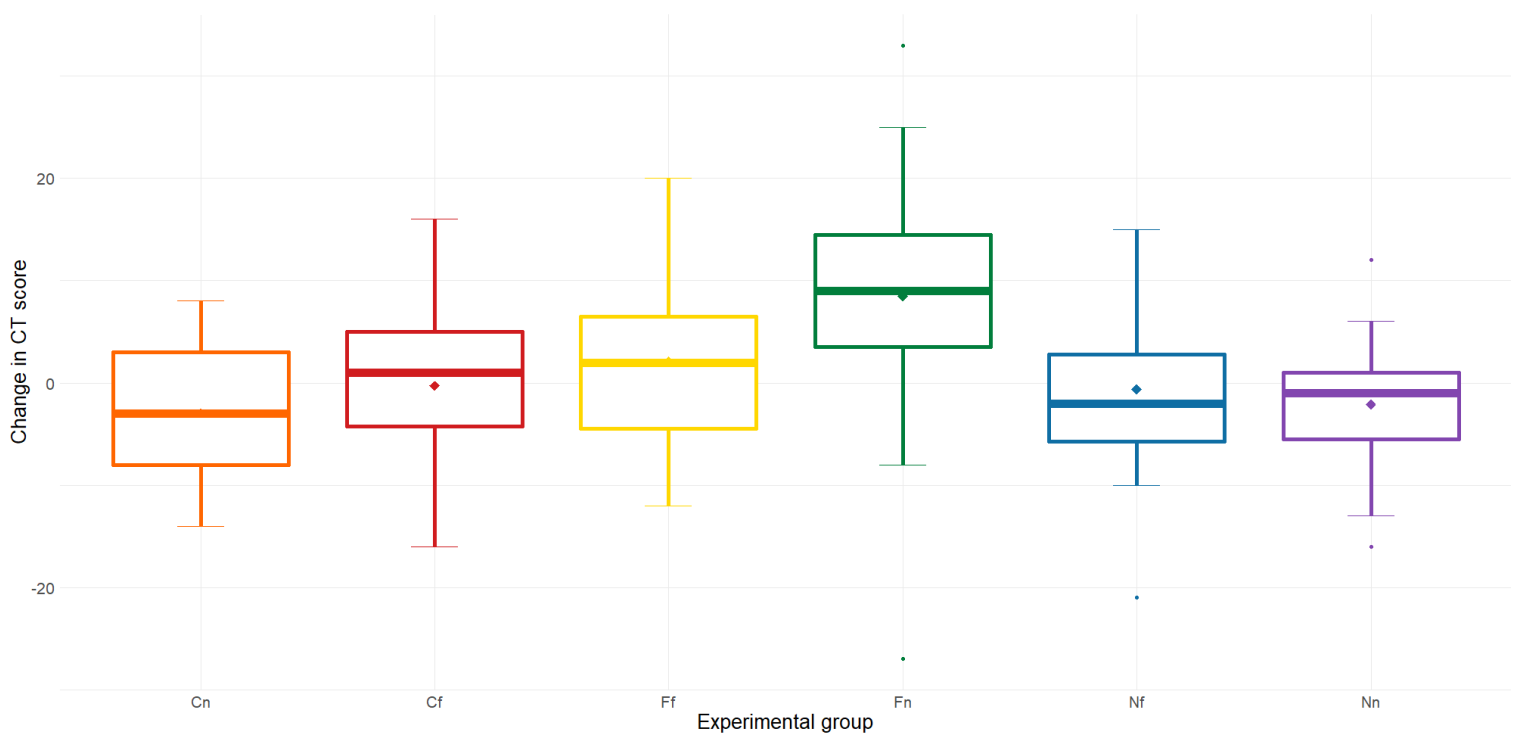
Table 21: Description of *CT change* across experimental groups

Group	<i>M</i>	<i>SD</i>	SE	Median	Min	Max	Range
Cf	-0.29	8.26	1.56	1	-16	16	32
Cn	-2.96	7.20	1.50	-3	-14	8	22
Ff	2.13	9.11	2.28	2	-12	20	32
Fn	8.50	13.91	3.11	9	-27	33	60
Nf	-0.61	9.04	2.13	-2	-21	15	36
Nn	-2.13	7.29	1.82	-1	-16	12	28

Prior to conducting a multiple regression, an initial basic analysis with no additional predictor variables was conducted to see whether membership of any experimental condition was associated with greater *CT change*. A one-way ANOVA found a significant difference in *CT change* between different experimental groups ($F(5,115) = 3.97, p = .002$). Tukey's multiple comparison of means found that the significant differences were: *CT change* increased between *Cn* and *Fn* participants on average by 11.46 ([19%CI 3.16, 19.76] $p = .002$); *CT change* decreased between *Fn* and *Cf* on average by 8.79 ([19%CI 0.84, 16.73] $p = .021$); *CT change* decreased between *Fn* and *Nf* on average by -9.11 ([95%CI -17.93 -0.29] $p = .039$); *CT change* decreased between *Fn* and *Nn* on average by -10.63 ([95% CI -19.73, -1.52] $p = .012$); differences between all other groups were not significant. This suggests that *Fn* participants had more positive *CT change* than *Cn*, *Cf*, *Nn*, and *Nf* participants; nonfiction readers placed in the fiction group had higher *CT change* than nonfiction readers in either other condition, and these participants also had higher *CT change*

than fiction readers who were assigned nonfiction or placed in the control group. This indicates an association to be further tested in a multiple regression analysis controlling for additional predictor variables. Figure 10 shows the distribution of *CT change* within the different experimental groups.

Figure 10: Box plot of *CT score change* in different experimental groups



Multiple regression analyses were used to assess the influence of experimental group assignment on *CT change*. As no demographic variables were found to have a relationship with *CT change* in the basic analysis, and since the randomised blocking design controlled for *educational level* and *gender*, no demographic variables were included in the regressions. Reading logged over the 14 day period in *entries*, and in *time*, were included as covariates to test whether group assignment played any role independently of the reading participants logged. This was done because participants

reading could vary significantly in addition to what they were assigned. Reading *entries* and reading *time* were included in separate models in order to avoid exceeding the number of predictor variables permitted to maintain statistical power, and to avoid covariance. The reference category for the experimental groups was nonfiction readers assigned to the control group (*Cn*).

Table 22: Multiple regression results for *CT change* and experimental group with reading *entries*

Predictor Variable	<i>B</i>	<i>SE</i>	95% CI for <i>B</i>		β	<i>R</i> ²
			LL	UL		
Nonfiction entries	0.01	0.10	-0.19	0.20	.937	.30
Fiction entries	0.73	0.15	0.44	1.03	<.001***	
Group:						
Cf	-0.64	2.63	-5.85	4.57	.807	
Ff	-2.05	3.27	-8.53	4.42	.531	
Fn	8.33	2.75	2.89	13.77	.003**	
Nf	-0.95	2.78	-6.46	4.56	.733	
Nn	1.28	2.79	-4.26	6.81	.649	

Note. **p* < .05 ***p* < .01 ****p* < .001

Based upon this model, *fiction entries* made, and being a nonfiction reader assigned to the fiction group (*Fn*) in comparison with a control group nonfiction reader (*Cn*), significantly predicted difference to *CT change*. Per fiction entry made, *CT change* increased on average by 0.62; being in the *Fn* group opposed to the *Cn* group yielded an average *CT change* of 8.76. The model had an *R*² of .30, thus explaining 30% of the variance in *CT change*.

The model was found to be accurate and generalisable (see Appendix B Model testing; Experimental Groups, page 435). The results of this model therefore partly support H₄: in the case of nonfiction readers, being assigned to the fiction

experimental group was associated with greater increases in *CT change* than any type of participants assigned to any other group. However, fiction readers assigned to the fiction group did not show the same effect, and therefore this hypothesis is only supported in the case of nonfiction readers.

Next, the same multiple regression analysis was conducted, this time including *nonfiction* and *fiction time* in place of *entries* logged. Table 23 shows the results for this regression.

Table 23: Multiple regression results for *CT change* and experimental group with reading time

Predictor Variable	<i>B</i>	<i>SE</i>	95% CI for <i>B</i>		β	<i>R</i> ²
			LL	UL		
Nonfiction time	0.08	0.06	-0.03	0.19	.159	.30
Fiction time	0.22	0.05	0.12	0.31	<.001***	
Group:						
Cf	0.77	2.56	-4.31	5.84	.765	
Ff	3.04	2.97	-2.83	8.92	.307	
Fn	10.63	2.64	5.41	15.85	<.001***	
Nf	0.97	2.80	-4.59	6.52	.731	
Nn	0.94	2.80	-4.60	6.49	.737	

Note. **p* < .05 ***p* < .01 ****p* < .001

Based upon this model, *fiction time* and being a nonfiction reader assigned to the fiction group (*Fn*) in comparison with a control group nonfiction reader (*Cn*), significantly predicted difference to *CT change*. Per fiction half-hour logged *CT change* increased on average by 0.18; being in the *Fn* group opposed to the *Cn* group yielded an average *CT change* of 11.02. The model had an *R*² of .30, thus explaining 30% of the variance in *CT change*.

The model was found to be accurate and generalisable (Appendix B Model testing; Experimental Groups, page 437). The results of this model therefore partly support H₄: in the case of nonfiction readers, being assigned to the fiction experimental group was associated with greater increases in *CT change*, and no such association was found for any type of participant assigned to any other group. However, fiction readers assigned to the fiction group did not show the same effect, and therefore this hypothesis is only supported in the case of nonfiction readers.

Summary of associations between experimental group and CT score change

The results of these analyses partly support H₄. When nonfiction and fiction reading (in *entries* made and in *time* logged) are controlled for, it is only in the case of nonfiction readers that being assigned to the fiction experimental group was associated with greater increases in *CT change*. No other experimental group had any significant association with increased *CT change*. However, the hypothesis is not fully supported as fiction readers assigned to the fiction group did not show the same effect, and it is therefore not simply the assignment of fiction that yields this effect.

Distinguishing improvement in CT score from t1 to t2 from decrease

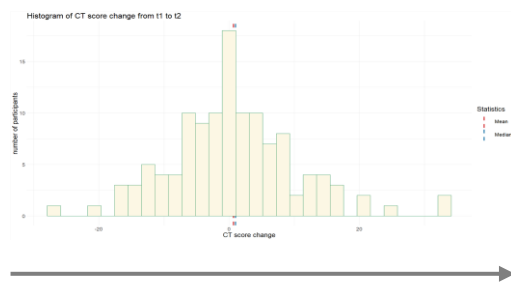
The previous analyses looked at the *CT change*, however in some cases this change was negative (i.e. participants performed worse at t2) and in some cases it was positive (i.e. participants improved at t2). As the hypotheses of this study stipulated improvement in CT score, it is important to distinguish between negative and positive change in score. Participants were therefore grouped into *no improvement* (*CT change* score ≤ 0 , $n = 60$), and *improved* (*CT change* score ≥ 1 , $n = 61$) categories. The

mean *CT change* for the *improved* group was 8.11 ($SD = 7.23$), median 6, with a range of 1 to 33. The mean *CT change* for the *no improvement* group was -6.87 ($SD = 5.68$), median -6, with a range from 0 to -27.

Box 7 summarises *CT change* and *CT improvement* as distinct predictor variables.

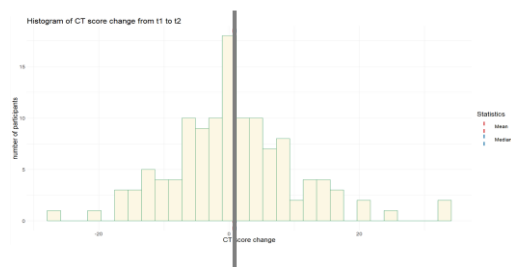
Box 7: Summary of CT change vs. CT improvement

CT change = difference in *CT score* from t1 to t2, yielding a range of values from negative through to positive. Increasing *CT change* implies increasing positivity in score.



Increasing *CT change*

CT improvement = *improved* vs. *no improvement* categories, permitting a differentiation between increased CT test performance or not.



Did not improve

Improved

Demographics across improved vs no improvement groups

First, basic analyses were conducted to determine whether any of the demographic variables (*age, gender, educational level*) were associated with being in the *improved* or *no improvement* group, and would thus need to be controlled for in further analyses.

Firstly, *gender* differences in the *improved* versus *no improvement* groups were assessed; participants who did not state their *gender* and those who were neither male nor female were combined to create a sufficient group size for testing. A chi squared analysis was to test whether there was a significant difference in the *gender* of the *improved* and *no improvement* groups; no significant difference was found ($\chi^2(2,118)=2.04, p = .360$). Therefore *gender* was balanced between the two groups, and did not need to be controlled for.

Similarly, a chi squared analysis was to test whether there was a significant difference in the *educational level* of the *improved* and *no improvement* groups; no significant difference was found ($\chi^2(2,118)= 0.42, p = .812$). Therefore *educational level* was similar between the two groups, and did not need to be controlled for.

Finally, *age* differences across the two groups were tested. Wilcoxon two sample rank sum test with continuity correction was used to assess whether there was any significant difference in the ages of participants in the *improved* and *no improvement* groups; no significant difference was found ($W = 1586.5, p = .207$), and therefore *age* did not need to be controlled for.

Reading logged across improved vs no improvement groups

Next, basic analyses were conducted to test for any significant associations between reading logged, and being in the *improved* or *no improvement* groups. This was done to gain an indication of any possible association that may warrant further investigation.

First, *total* reading both in *entries* and in *time* was assessed. Wilcoxon two sample rank sum test with continuity correction was used to assess whether there was any significant difference in *total entries* made between the groups, and found no significant difference ($W = 1503, p = .090$). However, for *total time* a significant difference was found ($W = 1270.5, p = .004$). Therefore participants in the *improved* and *no improvement* groups made the same number of *total entries* on average, those in the *improved* group spent a significantly longer *total time* reading.

Wilcoxon two sample rank sum test with continuity correction was used to assess whether there was any significant difference in *fiction entries* made between the groups, and found a significant difference ($W = 957, p < .001$); likewise for *fiction time* ($W = 1024.5, p < .001$). Therefore participants in the *improved* group read more fiction than those with *no improvement*.

Wilcoxon two sample rank sum test with continuity correction was used to assess whether there was any significant difference in *nonfiction entries* made between the groups, and found no significant difference ($W = 2091.5, p = .175$); likewise for *nonfiction time* ($W = 1758.5, p < .713$). Therefore participants in the *improved* group read similar amounts of nonfiction as those with *no improvement*.

These basic analyses suggest that there is a significant association between *total* and *fiction* reading, for both as measured by both *entries* logged and for the latter also in reading *time*, and being in the *improved* group. This suggests that further investigation is justified.

Logistic regression analyses of improved vs no improvement in CT score

In order to assess the impact of reading logged over the 14 day period upon either improving or not improving one's CT score, a logistic regression was carried out. First, *total* reading logged assessed, both in terms of *entries* and *time*. Models with each predictor separately, and combined, were compared.

Table 24: Comparison of models for CT improvement and total reading

Model	Variables	AIC	Pseudo R^2
glm0	Improved ~ 1	169.73	0
glm1	Improved ~ Total entries	169.98	0.02
glm2	Improved ~ Total time	159.03	0.13
glm3	Improved ~ Total entries + Total time	159.43	0.15

Note: Nagelkerke's Pseudo R^2

Based upon the AIC, model 2 was best fitted to the data. However, based on the pseudo R^2 values, model 3 was the best fitted to the data. As the difference in both AIC and pseudo R^2 between models 2 and 3 are so small, model 3 was selected as being most conceptually relevant. Thus the selected model included both *total entries* logged and *total time* logged as predictors of belonging to the *improved* group. Table 25 shows the results for the logistic regression.

Table 25: Results for logistic regression model of total entries and time on CT improvement

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
Total entries	-0.03	0.03	.213	0.92	0.97	1.02
Total time	0.05	0.02	.003**	1.02	1.05	1.08

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

Total time significantly predicted being in the *improved* rather than *no improvement* group, $b = 0.05$, $p < .01$; per half-hour increase in *total time*, the change in odds of being in the *improved* group rather than the *no improvement* group was 1.05. *Total entries* did not have a significant impact.

The model was tested for accuracy and generalisability (see Appendix B Model testing; Distinguishing improvement in CT score from t1 to t2 from decrease, page 440). Based upon this model, there is partial support for H₂: higher amounts of reading recorded in terms of *time* logged over the reading log period had a significant association with having an increase in CT score from t1 to t2. However, when reading amounts are measured in the number of *entries* made in the reading log, the hypothesis is not supported.

Next, *fiction* and *nonfiction entries* and *time* were assessed as predictors of being in the *improved* category. Although *nonfiction* was not predictive of membership of either group in the basic analyses, it was included as it is conceptually important for the hypotheses of this study. First, models with *entries* alone, and *time* alone, were tested and then compared to a model including all predictors.

Table 26: Comparison of models for CT improvement and total reading

Model	Variables	AIC	Pseudo R^2
glm0	Improved ~ 1	169.73	0
glm1	Improved ~ Nonfiction entries + Fiction Entries	155.11	0.19
glm2	Improved ~ Nonfiction time + Fiction time	156.85	0.17
glm3	Improved ~ Nonfiction entries + Nonfiction time + Fiction entries + Fiction time	148.21	0.29

Note: Nagelkerke's Pseudo R^2

Based upon the AIC and pseudo R^2 values, model 3 was the best fitted to the data; including both *entries* logged and *time* logged improved model fit. Table 27 shows the results from this model.

Table 27: Results for logistic regression model of fiction and nonfiction entries and time on CT improvement

Variable	B	SE	p	Lower 95% CI	OR	Upper 95% CI
Nonfiction entries	-0.09	0.04	.016*	0.84	0.91	0.98
Fiction entries	0.13	0.06	.018*	1.02	1.14	1.28
Nonfiction time	0.07	0.03	.01**	1.02	1.07	1.13
Fiction time	0.01	0.02	.673	0.97	1.01	1.06

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

The number of *nonfiction entries* significantly predicted being in the *no improvement* rather than *improved* group, $b = -0.09$, $p < .05$; per *entry* increase in *nonfiction* reading, the change in odds of being in the *improved* group rather than the *no improvement* group was 0.91. Conversely, *nonfiction time* significantly predicted being in the *improved* rather than *no improvement* group, $b = 0.07$, $p < .05$; per half-hour increase in *nonfiction* reading, the change in odds of being in the *improved* group rather than the *no improvement* group was 1.07. Logging more *fiction entries* had a positive relationship with being in the *improved* group, $b = 0.13$, $p < .05$; per *entry* of *fiction* read the odds of being in the *improved* rather than the *no improvement* group increased by 1.14. *Fiction time* had no significant relationship with being in either category.

The model was found to be accurate and generalisable (see Appendix B Model testing; Distinguishing improvement in CT score from t1 to t2 from decrease, page 441). Therefore based upon this model, there is only limited support for H₃: higher amounts of *fiction* reading in terms of *entries* logged did have a stronger association with *improved* CT score from t1 to t2 than *nonfiction entries*, which in fact had a negative association with *improved* CT score. However, when taking *time* as the measure of reading amount, *nonfiction* reading in half-hours was significantly associated with *improved* CT score while *fiction* reading had no significant association. Therefore based on this model H₃ is supported in terms of reading *entries* made, but not supported in terms of *time* spent reading. Furthermore, the paradoxical finding that logging more *entries* of nonfiction is associated with a reduced likelihood of *improved*

CT score, while spending more *time* reading nonfiction is associated with a greater likelihood of *improved* CT score, warrants further investigation.

Experimental groups: improved vs. no improvement

In addition to the finding that being a nonfiction reader assigned to the fiction group (*Fn*) was associated with increased *CT change*, it was further tested whether there was any association between experimental group and having *improved* CT score vs no change or decreased score. To this end, logistic regression analyses were conducted controlling for *nonfiction* and *fiction* reading (first in *entries*, and then in *time*), to determine whether there was any effect of experimental group in addition to what kind of reading participants logged. Table 28 shows the results for the logistic regression with reading included in *entries*.

Table 28: Results of logistic regression for improved or no improvement group with reading entries

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
Nonfiction entries	-0.01	0.03	.788	0.94	0.99	1.05
Fiction entries	0.16	0.04	.001***	1.08	1.18	1.29
Group:						
Cf	0.26	0.68	.699	0.34	1.3	4.99
Ff	-0.66	0.84	.436	0.1	0.52	2.67
Fn	1.75	0.81	.031*	1.28	5.76	32.7
Nf	-1.21	0.79	.127	0.06	0.3	1.34
Nn	0.26	0.72	.717	0.31	1.3	5.41

* $p < .05$ ** $p < .01$ *** $p < .001$

Based upon this model, *fiction entries* had a significant association with being in the *improved* group, $b = 0.16$ ($p < .001$); the odds ratio indicated that per additional fiction entry logged, the odds of being in the *improved* group increased by 1.18. Being a nonfiction reader assigned fiction (*Fn*) as opposed to a nonfiction reader in the control group (*Cn*), was also significantly associated with being in the *improved* group,

$b = 1.75$ ($p < .05$); the odds ratio tells us that being in the *Fn* rather than the *Cn* group yielded a change in odds of 5.76 in favour of being in the *improved* group. See Appendix B Model testing; Distinguishing improvement in CT score from t1 to t2 from decrease, page 443, for model accuracy and generalisability.

Next, the same logistic regression was conducted, this time using *fiction* and *nonfiction time* as a control. Table 29 shows the regression results.

Table 29: Results of logistic regression for improved or no improvement group with reading time

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
Nonfiction time	0.02	0.02	.288	0.99	1.02	1.05
Fiction time	0.07	0.02	.001**	1.03	1.07	1.12
Group:						
Cf	0.56	0.66	.4	0.48	1.75	6.58
Ff	0.43	0.76	.569	0.35	1.54	6.99
Fn	2.36	0.80	.003**	2.44	10.61	60.53
Nf	-0.94	0.81	.25	0.07	0.39	1.83
Nn	0.15	0.71	.828	0.28	1.17	4.72

* $p < .05$ ** $p < .01$ *** $p < .001$

Based upon this model, *fiction time* had a significant association with being in the *improved* group, $b = 0.07$ ($p < .001$); the odds ratio indicated that per additional half-hour of fiction logged, the odds of being in the *improved* group increased by 1.07. Being a nonfiction reader assigned fiction (*Fn*) as opposed to a nonfiction reader in the control group (*Cn*), was also significantly associated with being in the *improved* group, $b = 2.36$ ($p < .01$); the odds ratio tells us that being in the *Fn* rather than the *Cn* group yielded a change in odd of 10.61 in favour of being in the *improved* group. See Appendix B Model testing; Distinguishing improvement in CT score from t1 to t2 from decrease, page 443, for model accuracy and generalisability.

Summary of findings for improved vs. no improvement groups

Overall, these models indicate that total reading logged in *time*, though not in *entries*, had a significant association with being in the *improved* group. Furthermore, *fiction* and *nonfiction* have more complex and somewhat contradictory effects. *Fiction* logged in *entries*, but not in *time*, had a significant association with being in the *improved* group. *Nonfiction* logged in *entries* had a significant negative relationship with being in the *improved* group, but in terms of *time* a significant positive relationship. Therefore, spending more *time* reading in total, and more *time* reading *nonfiction*, was associated with *improved* CT score. However, having more *entries* of reading was only relevant for *fiction* as a positive predictor, but for *nonfiction* this was a negative predictor of *improved* CT. These findings somewhat support H₃, but also portray a more complex relationship. In addition to this, when *fiction* and *nonfiction* reading are controlled for both in terms of *entries* logged and *time*, being a nonfiction reader in the fiction group (*Fn*) in contrast with nonfiction readers in the control group

(*Cn*) is significantly associated with being in the *improved* CT group. These findings particularly support H₄: in the case of nonfiction readers, being assigned to the fiction experimental group was significantly associated with having an improvement in CT scores from t1 to t2, while no other group had such an association. However, for fiction readers assignment to the fiction group had no such effect, and thus the hypothesis is not supported in their case.

5.2.3. Exploratory Analyses

Average hours per entry made

As the difference between *nonfiction entries* made and *nonfiction time* suggests that reading smaller amounts but with more *time* spent is associated with being in the *improved* CT score category, further exploratory analyses were conducted using the average time spent reading per entry (i.e. reading *time* in divided by *entries*). This created new predictor variables of *total time per entry* (*T t/e*), *fiction time per entry* (*F t/e*), and *nonfiction time per entry* (*NF t/e*), which were used in the same regressions as the prior analyses.

The *total time per entry* had a mean of 1.35 (*SD* = 0.88), median 1.13, with a range from 0 – 4.43. *Nonfiction time per entry* had a mean of 1.12 (*SD* = 0.88), median 1, with a range from 0 – 4.79. *Fiction time per entry* had a mean of 1.18 (*SD* = 1.16), median 0.9, with a range from 0 – 6.19.

Improved vs no improvement categories

A logistic regression was carried out with *total time per entry* as a predictor for being in the *improved* CT score group. Table 30 shows the regression results.

Table 30: Results of logistic regression with total reading time per entry upon CT improvement

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
T t/e	0.79	0.26	.002**	1.37	2.19	3.78

* $p < .05$ ** $p < .01$ *** $p < .001$

This model suggests that increased *total time per entry* significantly predicted being in the *improved* rather than the *no improvement* group $b = 0.79$, $p < .01$. The odds ratio tells us that per half-hour more spent per entry, the change in odds of being in the *improved* group rather than the *no improvement* group was 2.19. Therefore spending more time reading each entry was associated with having an improvement in CT score from t1 to t2. See Appendix B Model testing; Average hours per entry made on page 445, for model accuracy and generalisability.

Next, *nonfiction* and *fiction time per entry* were used as predictor variables for being in the *improved* CT score group. Table 31 shows the results from the logistic regression.

Table 31: Results of logistic regression with fiction and nonfiction reading time per entry upon CT improvement

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
NF t/e	0.55	0.26	.033*	1.07	1.74	2.98
F t/e	0.47	0.21	.022*	1.10	1.61	2.49

* $p < .05$ ** $p < .01$ *** $p < .001$

Nonfiction time per entry significantly predicted being in the *improved* CT group, $b = 0.55$, $p < .05$. The odds ratio tells us that per half-hour increase on average reading each nonfiction entry logged, the change in odds of being in the *improved* group rather than the *no improvement* group is 1.74. Increased *fiction time per entry* significantly predicted being in the *improved* group, $b = 0.47$, $p < .05$; per half-hour increase reading each fiction entry logged, the change in odds of being in the *improved* group rather than the *no improvement* group was 1.61. This suggests that spending more time on each nonfiction and fiction entry made is associated with *improved* CT scores from t1 to t2. See Appendix B Model testing; Average hours per entry made on page 446 for model accuracy and generalisability.

Overall, these models indicate that increased time spent per entry made in the reading log was associated with being in the *improved* CT group. However, membership of the *improved* vs *no improvement* group is a broad distinction, and therefore to explore the role of these variables further they were used as predictors in regressions testing for an association with *CT change*.

CT change

Linear regression was conducted to test for an association between *total time per entry* and *CT change*. Table 32 shows the results for the regression.

Table 32: Linear regression results for CT change and total time read per entry

Predictor variable	<i>B</i>	<i>SE</i>	95% CI for <i>B</i>		β	<i>R</i> ²
			LL	UL		
Total t/e	3.56	0.98	1.61	5.50	<.001***	.10

p* < .05 *p* < .01 ****p* < .001

Based upon this model, *total time per entry* was significantly associated with *CT change*. Per half-hour spent reading each entry logged, *CT change* increased by 3.56 (*p* < 0.01). As the model has an *R*² .10, it accounts of 10% of the variation in *CT change*. See Appendix B Model testing; Average hours per entry made on page 446 for model accuracy and generalisability.

Finally, the predictor variables of *nonfiction* and *fiction time per entry* were used in a multiple regression upon *CT change* as the outcome variable. Table 33 shows the regression results.

Table 33: Linear regression results for CT change with NF t/e and F t/e

Predictor variable	<i>B</i>	<i>SE</i>	95% CI for <i>B</i>		β	<i>R</i> ²
			LL	UL		
NF t/e	1.21	1.00	-0.76	3.19	.227	.16
F t/e	3.01	0.76	1.51	4.52	<.001***	

p* < .05 *p* < .01 ****p* < .001

Based upon this model, only *fiction time per entry* was significantly associated with *CT change*. Per additional half-hour spent on each fiction entry logged, *CT change* increased by 2.76 (*p* < .001). As the model has an *R*² .16, it accounts of 16% of

the variation in *CT change*. See Appendix B Model testing; Average hours per entry made on page 448, for model accuracy and generalisability.

In summary, these models tell us that spending more time per each entry made of reading overall is associated with greater *CT change*. However, when looked at with *fiction* and *nonfiction* separated, it is only in the case of *fiction* that spending more time per entry had a significant association with *CT change*. *Nonfiction time per entry* had no significant association with *CT change* in this case.

Narrative Transportation

NT was found not to mediate the effects of reading logged either in *entries* or in *time*, and thus H₄ is not supported. However, the mediation analyses used the *average NT* score per participant; i.e. their average transportation level across the 14 day log period. As 14 days is a fairly long period, this is essentially a measure of participants' trait *NT*. However, the reading log involved logging a transportation score for each entry made, thus also providing state *NT* measurements across the log period. It is therefore possible to further investigate the role of *NT* in a more fine-grained manner, by using *NT* variables other than only the mean score. To this end, *peak* and *minimum NT* levels were identified for each participant (i.e. highest and lowest transportation rating given), for *fiction*, *nonfiction*, and in *total*. As reading was measured both in *entries* and in *time*, this yielded four additional variables: *peak NT time*; *peak NT entries*; *minimum NT time*; *minimum NT entries*. These variables can be further subdivided into *fiction*, *nonfiction*, and *total*. The relationships between these additional transportation variables and *CT change* will be explored here.

Table 34 shows a description of the *peak* and *minimum NT* levels across participants (i.e. highest and lowest *NT* rating given). It should be noted that participants who did not log any reading either in total, or did not log any fiction or any nonfiction, are not included as they provided no *NT* scores.

Table 34: Description of minimum and peak *NT* levels

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Median	Min	Max	Range
Total Min	120	2.07	0.92	2	1	5	4
Total Peak	120	4.74	0.51	5	3	5	2
Fiction Min	95	3.01	1.17	3	1	5	4
Fiction Peak	95	4.64	0.54	5	3	5	2
NF Min	106	2.21	1.04	2	1	5	4
NF Peak	106	4.42	0.75	5	2	5	3

Table 35 shows a description of the amount of reading *entries* made with a *peak NT* rating, and with a *minimum NT* rating.

Table 35: Description of reading entries at minimum and peak *NT* levels

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Median	Min	Max	Range
Total Min	118	2.69	2.54	2	1	19	18
Total Peak	120	5.38	4.53	4	1	20	19
Fiction Min	95	2.93	2.99	2	1	19	18
Fiction Peak	95	4.27	3.5	3	1	19	18
NF Min	106	2.56	2.27	2	1	15	14
NF Peak	106	3.24	2.74	2	1	15	14

Table 36 shows a description of the amount of *time* spent reading at *peak NT* level, and at *minimum NT* level.

Table 36: Description of reading time at minimum and peak NT levels

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Median	Min	Max	Range
Total Min	120	3.12	5.3	1.33	0.07	41.8	41.73
Total Peak	120	9.94	12.51	5.36	0.17	65.33	65.16
Fiction Min	95	4.76	8.84	1.83	0.03	56	55.97
Fiction Peak	95	8.31	10.94	4.33	0.17	54.83	54.66
NF Min	106	3.53	7.83	1.33	0.07	71.83	71.76
NF Peak	106	5.55	8.76	2.5	0.1	71.83	71.73

Analyses were conducted to test for significant differences between *fiction* and *nonfiction* reading logged at the *peak* and the *minimum NT* level. As the *peak NT* value for *fiction* and *nonfiction* is distinct, this is not a comparison of how much *fiction* or *nonfiction* was read at *NT* level 5, for instance, but rather a comparison of how much reading was logged by participants at their own peaks and minimums for *fiction* and *nonfiction* reading.

Kendall's rank correlation showed that *peak NT fiction entries* had no significant association with *peak NT nonfiction entries* ($r\tau = 0.11$, $p = .217$); however, *peak NT time* had a significant association between *fiction* and *nonfiction* ($r\tau = 0.19$, $p = .016$). Kendall's rank correlation showed that *minimum NT fiction entries* had no significant association with *minimum NT nonfiction entries* ($r\tau = 0.11$, $p = .224$); however, again *peak NT time* had a significant association between *fiction* and

nonfiction ($r\tau = 0.19, p = .013$). Thus, for *peak* and *minimum NT* level *fiction* and *nonfiction* were different in terms of *entries* logged, but aligned in terms of *time* spent reading.

Reading at minimum and peak NT levels and CT change

Prior to testing these variables as predictors of *CT change* in multiple regression analyses, basic correlation analyses of direct relationships were first undertaken. As a Shapiro-Wilk normality tests found all of the *peak* and *minimum NT* reading amount variables not to be normally distributed, non-parametric tests were used.

Kendall's rank correlation showed that *minimum NT fiction entries* had no significant association with *CT change* ($r\tau = -0.03, p = .703$). However, *peak NT fiction entries* had a significant positive association with *CT change* ($r\tau = 0.27, p < .001$). Likewise, *minimum NT fiction time* had no significant association with *CT change* ($r\tau = -0.01, p = .864$), but *peak NT fiction time* had a significant association with *CT change* ($r\tau = 0.25, p = .001$). This suggests that reading fiction at a *peak NT* level may be associated with increased *CT change*.

Kendall's rank correlation *minimum NT nonfiction entries* had no significant association with *CT change* ($r\tau = 0.07, p = .351$). Similarly, *peak NT nonfiction entries* had no significant positive association with *CT change* ($r\tau = -0.06, p = .413$). Furthermore, *minimum NT nonfiction time* had no significant association with *CT change* ($r\tau = 0.10, p = .123$), and *peak NT fiction time* had no significant association with *CT change* ($r\tau = 0.04, p = .566$). This suggests that reading nonfiction at either *peak* or *minimum NT* does not have a relationship with *CT change*.

Overall, this suggests that reading more fiction at *peak NT* is significantly associated with *CT change*. However, reading fiction at *minimum NT* had no such association, and *nonfiction* read at either a *peak* or a *minimum NT* had no association with *CT change* either. Based upon these findings, reading with a minimum *NT* level was not pursued in further analyses, but reading with a *peak NT* level was further tested as a predictor of *CT change*.

Multiple regression analyses of CT improvement and peak NT reading

A hierarchical multiple regression analysis was conducted to test for a relationship between *CT change*, and *peak NT fiction* and *nonfiction entries*. As *peak NT nonfiction entries* were not found to have a significant relationship with *CT change* in the direct correlation analysis, this variable was added in step 1 prior to then adding *peak NT fiction entries* in step 2. This allows for a conservative approach avoiding type I error in assessing the impact of *fiction* reading at *peak NT* (Stanovich & Cunningham, 2004). As not all participants logged *fiction* and/or *nonfiction*, not all participants had *peak* or *minimum NT entries*, therefore these regressions include only those who had at least one *entry* of both *fiction* and *nonfiction* ($n = 79$). Table 37 shows the regression results.

Table 37: Hierarchical regression results for CT change with peak NT reading entries

Predictor Variable	$B(SE)$	95% CI for B		β	R^2	ΔR^2	
		LL	UL				
Step 1						.03	.03**
Peak nonfiction	0.59 (0.41)	-0.23	1.41	.156			
Step 3						.14	0.11**
Peak nonfiction	0.41 (0.39)	-0.37	1.20	.299			
Peak fiction	1.01 (0.33)	0.36	1.66	.003**			

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

Based upon these analyses, step 2 significantly increased model fit (i.e. ΔR^2), suggesting this is the best model for the data. Based on this model, *peak NT nonfiction entries* had no significant association with *CT change*. However, controlling for *peak NT nonfiction entries*, *peak NT fiction entries* were significantly associated with increased *CT change*; per additional *entry* of *fiction* made with a *peak NT* score, *CT change* increased by 1.01 ($p < .01$).

Next, the same hierarchical multiple regression was conducted for *peak NT time*. Table 38 shows the regression results.

Table 38: Hierarchical regression results for CT change with peak NT reading time

Predictor Variable	$B(SE)$	95% CI for B		β	R^2	ΔR^2	
		LL	UL				
Step 1						.07	.07**
Peak nonfiction	0.29 (0.12)	0.06	0.52	.015*			
Step 2						.17	.10**
Peak nonfiction	0.23 (0.11)	0.01	0.45	.042*			
Peak fiction	0.29 (0.10)	0.10	0.48	.004**			

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

Based upon these analyses, step 2 significantly increased model fit (i.e. ΔR^2). Based on the R^2 , step 2 is the best model for the data. This model was assessed for accuracy and was found to be accurate and generalisable (see Appendix B Model testing; Narrative Transportation, page 450). The R^2 tells us that the model explains 17% of the variation in *CT change*. *Peak NT nonfiction time* had a significant association with *CT change* both when added in a step alone ($b = 0.29, p < .05$), and when *fiction* was also added ($b = 0.23, p < .05$). Controlling for *nonfiction*, *peak NT fiction time* was significantly associated with increased *CT change* ($b = 0.29, p < .01$). Therefore, spending more *time* reading both *nonfiction* and *fiction* with *peak NT*, predicted increased *CT change*.

Overall, *fiction entries* made and *fiction time* spent reading with a *peak NT* score were significantly associated with increased *CT change*. In terms of nonfiction, *entries* made were not associated with *CT change*, *peak NT nonfiction time* had a significant association with *CT change*.

Reading at minimum and *peak NT* levels and improved vs. no improvement in CT score

As *CT change* incorporates both negative and positive values (i.e. participants whose scores decreased from t1 to t2, as well as those who improved), further analyses were conducted to identify possible associations between reading at *peak* or *minimum NT* level and being in the *improved* or *no improvement* CT score category. First, basic analyses of the individual predictor variables were conducted, to identify potential trends for further investigation. Table 39 shows a description of *fiction* and *nonfiction*

reading logged both in *entries* and in *time* with a *minimum NT* score, by participants in the *improved* and *no improvement* categories.

Table 39: Description of minimum NT reading logged by CT improvement category

Variable	Improved			No improvement		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
NF entries	52	2.96	2.86	54	2.17	1.42
NF time	52	5.11	10.78	54	2.01	2.20
F entries	55	3.09	3.42	40	2.70	2.30
F time	55	5.78	11.21	40	3.35	3.30

Table 40 shows a description of *fiction* and *nonfiction* reading logged both in *entries* and in *time* with a *peak NT* score, by participants in the *improved* and *no improvement* categories.

Table 40: Description of peak NT reading logged by CT improvement category

Variable	Improved			No improvement		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
NF entries	52	3.10	2.72	54	3.37	2.77
NF time	52	6.79	11.15	54	4.36	5.43
F entries	55	5.15	3.67	40	3.08	2.89
F time	55	10.70	12.61	40	5.02	7.01

First, associations between reading at a *minimum NT* level and CT score *improvement* were tested. A one-way analysis of means (not assuming equality of

variance) was carried out, which showed no significant difference in *minimum NT nonfiction entries* between *improved* and *no improvement* groups ($F(1,104) = 3.35$ $p = .075$). Likewise, a one-way analysis of means (with equality of variance) showed no significant difference in *minimum NT fiction entries* between *improved* and *no improvement* groups ($F(1,93) = 0.39$, $p = .532$). This suggests that in terms of *entries* logged with a *minimum NT* score, there was no significant association with being in the *improved* score category.

The same analysis was repeated for *time* logged at a *minimum NT* level. A one-way analysis of means (assuming equal variance) was carried out, which showed a significant difference in *minimum NT nonfiction time* between *improved* and *no improvement* groups ($F(1,104) = 4.28$, $p = .041$). Conversely, no significant difference was shown in *minimum NT fiction time* between *improved* and *no improvement* groups ($F(1,93) = 1.77$, $p = .186$). This suggests that in terms of *time* logged with a *minimum NT* score, only nonfiction was associated with being in the *improved* score category.

Next, reading at a *peak NT* level was assessed for associations with CT *improvement*. A one-way analysis of means (assuming equal variance) was carried out, which showed no significant difference in *peak NT nonfiction entries* between *improved* and *no improvement* groups ($F(1,104) = 0.26$, $p = .609$). However, a significant difference in *peak NT fiction entries* between *improved* and *no improvement* groups was found ($F(1,93) = 8.78$, $p = .004$). This suggests that in terms of *entries* logged with a *peak NT* score, only *fiction* had a significant association with being in the *improved* score category.

The same analysis was repeated for *time* logged at a *peak NT* level. A one-way analysis of means (assuming equal variance) was used, which showed no significant difference in *peak NT nonfiction time* level between *improved* and *no improvement* groups ($F(1,104) = 2.05, p = .155$). However, a one-way analysis of means (not assuming equality of variance) showed a significant difference in *peak NT fiction time* between *improved* and *no improvement* groups ($F(1,93) = 7.85, p = .006$). This suggests that in terms of *time* logged with a *peak NT* score, only *fiction* had a significant association with being in the *improved* score category.

Overall, these analyses suggest that reading at a minimum *NT* level had a significant association with being in the *improved* CT group in the case of *nonfiction time*, though not for *nonfiction entries*. Fiction reading at a minimum *NT* level had no association with being in the *improved* CT group. Conversely, reading at a *peak NT* level was significantly associated with being in the *improved* CT group in the case of *fiction* reading (as measured both by *entries* and *time*). *Peak NT nonfiction* reading had no significant association with being in the *improved* CT group. These analyses suggest that both minimum and *peak NT* reading are worth further investigation in relation to having an *improved* CT score or *no improvement*.

Logistic regression analyses of CT improvement and minimum/*peak NT* reading

Fiction and *nonfiction entries* and *time* at both *minimum* and *peak NT* levels were assessed as predictors of being in the *improved* score category. First, logistic regression models with *minimum NT* alone, and with *peak NT* alone, were tested and

then compared to a model including all predictors. Table 41 shows a comparison of models.

Table 41: Comparison of models for CT improvement with peak and minimum NT reading

Model	Variables	AIC	Pseudo R^2
glm0	Improved ~ 1	110.49	0
glm1	Improved ~ Min NF entries + Min F entries	110.77	.06
glm2	Improved ~ Min NF time + Min F time	108.17	.10
glm3	Improved ~ Peak NF entries + Peak F entries	104.25	.16
glm4	Improved ~ Peak NF time + Peak F time	100.85	.21
glm5	Improved ~ Min NF entries + Min F entries + Peak NF entries + Peak F entries	103.83	.23
glm6	Improved ~ Min NF time + Min F time + Peak NF time + Peak F time	104.32	.22

Note: Nagelkerke's Pseudo R^2

Based upon the AIC values, model 4 was the best fitted. However, the pseudo R^2 values of models 5 and 6 suggested a slightly better fit. As the differences in AIC and R^2 are so small, these models all have a very similar fit to the data. As models 5 and 6 are conceptually most interesting, permitting an analysis of *minimum* and *peak NT* reading combined, I chose to use both models to test reading measured in both *entries* and *time*.

Table 42 shows the results from model 5, assessing reading as measured in *entries* logged.

Table 42: Results for logistic regression model of fiction and nonfiction minimum and peak entries on CT improvement

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
Min F entries	-0.06	0.12	.623	0.75	0.94	1.20
Min NF entries	0.27	0.15	.066	1.01	1.31	1.82
Peak F entries	0.28	0.09	.003**	1.11	1.33	1.62
Peak NF entries	-0.05	0.11	.653	0.76	0.95	1.19

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

The model was found to be accurate and generalisable (see Appendix B Model testing; Narrative Transportation, page 452). *Peak NT fiction entries* significantly predicted being in the *improved* CT score group, $b = 0.28$, $p < .01$; per additional *peak NT fiction entry*, the change in odds of being in the *improved* group rather than the *no improvement* group was 1.33. Table 43 shows the results from model 6, assessing reading as measured in *time*.

Table 43: Results for logistic regression model of fiction and nonfiction minimum and peak time on CT improvement

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	Lower 95% CI	OR	Upper 95% CI
Min F time	-0.02	0.05	.673	0.88	0.98	1.09
Min NF time	0.08	0.12	.5	0.88	1.08	1.40
<i>Peak</i> F time	0.07	0.04	.071	1.00	1.08	1.18
<i>Peak</i> NF time	0.09	0.07	.192	0.97	1.1	1.29

p* < .05 *p* < .01 ****p* < .001

The model was found to be accurate and generalisable (see Appendix B Model testing; Narrative Transportation, page 453). None of the predictor variables in this model had a significant association with being in the *improved* rather than *no improvement* CT group.

Overall, findings from these models suggest that logging more *fiction entries* with a *peak NT* level was associated with a greater likelihood of being in the *improved* CT group, rather than having *no improvement* or decrease in CT score. However, when using *time* as the measurement rather than *entries* this effect was not found. *Minimum NT* level reading was not found to have any significant effect on being in the *improved* or *no improvement* groups. Reading *nonfiction* at either *minimum* or *peak NT* also had no effect.

Summary of minimum and peak NT level reading findings

These analyses suggest that reading amount with a *peak NT* level has some significant associations with difference in *CT change* from t1 to t2. *Minimum NT* level reading was not found to have any significant association with *CT change* in the basic analyses, and was therefore not pursued further. Based on multiple regression analyses, both *peak NT fiction entries* and *time* were significantly associated with increased *CT change*. *Peak NT nonfiction time* had a significant association with *CT change*, though *nonfiction entries* did not. Additionally, logistic regression analysis found that *peak NT fiction entries* was associated with a greater likelihood of being in the *improved CT* group, rather than having *no improvement* or decrease in CT score, although for *time* there was no such association. Overall, this suggests some association with increased *CT change*, but a more limited association with that change being positive (i.e. being in the *improved* group).

5.3. Discussion

Overall, the hypotheses for this study were broadly supported, though not entirely, and not without importance nuance. In this discussion I will firstly provide a brief summary of the extent to which each hypothesis was supported or not, and then expound upon each hypothesis in more detail, also discussing broader findings and issues. As the hypotheses are answered using both CT change and improvement, Box 7 may offer useful clarification of these variables.

5.3.1. Hypotheses testing summary

This study tested the following hypotheses:

H₁: Higher reading engagement will have a significant association with higher baseline critical thinking score at t₁.

H₂: Higher amounts of reading recorded over the reading log period will have a significant association with increases in critical thinking score from t₁ to t₂.

H₃: Higher amounts of fiction reading recorded over the reading log period will have a stronger association with increases in critical thinking score from t₁ to t₂ than nonfiction.

H₄: Narrative transportation will mediate the relationships predicted in H₂ and H₃.

H₅: Participants assigned to the fiction experimental group will have greater increases in their critical thinking test scores from t₁ to t₂ than either the nonfiction group or the control group.

H₁ was supported; higher reading engagement (i.e. listing more authors as favourites) was significantly associated with a higher CT test score at t₁.

H₂ was partially supported; when measuring reading amount in the number of entries made over the log period, there was no significant association between the total amount of reading and CT score change from t₁ to t₂. However, when measuring reading amount in time recorded over the log period, there was a significant association between the amount read and CT score change from t₁ to t₂. Therefore it is the total time spent reading, not the total number of cases of reading, that is associated with greater change in CT. Furthermore, the total number of entries made over the log period had no significant association with the likelihood of having an

improved CT score at t2 than at t1, rather than having no change or a decrease in score. However, total time spent reading was significantly associated with having an improved score. Thus time spent reading, and not the instances of reading, is significantly associated with CT score improvement. Overall, these findings suggest that spending more time reading is associated both with having an improvement in CT, and with a greater magnitude of the change in CT.

H₃ was partially supported; higher amounts of fiction reading, as measured both in entries made and in reading time recorded, did have a stronger association with increases in CT score change from t1 to t2 than nonfiction. In fact, nonfiction logged both in terms of entries and time had no significant association with CT score change. Thus the hypothesis is supported in terms of the magnitude of CT score change. However, when it came to distinguishing between having an improvement in CT score from t1 to t2, as opposed to no change or decrease in CT score, the hypothesis is not fully supported. Higher amounts of fiction entries logged had a significant association with a greater likelihood of having an improved CT score, while conversely more nonfiction entries had a negative association with improved CT score. However, when taking time as the measure of reading amount, nonfiction reading was significantly associated with improved CT score while fiction reading had no significant association. Therefore this hypothesis is supported in terms of the instances of reading that participants reported over the log period, but not supported in terms of the time they spent reading. Overall, this suggests that while fiction reading has a stronger association with the magnitude of CT change, it does not have as clear an association with that change being positive.

H₄ was not supported; no significant mediation effect of *NT* upon CT score change was found for total, fiction, or nonfiction reading as measured either in entries made or time spent reading.

H₅ was partially supported; being a nonfiction reader assigned to the fiction group was significantly associated with increased CT score change, and having improved CT score rather than no change or decreased score. However, fiction readers who were assigned fiction did not have a significant difference in CT change, nor were they more likely to have improved CT score. Therefore the hypothesis only supported in the case of nonfiction readers.

H₀ can be rejected based upon these findings, as reading both in terms of total amount and in terms of fiction and nonfiction type has been shown to have some significant associations with CT score both at the baseline measurement at t₁, and with the difference between t₁ and t₂ score.

5.3.2. Broader discussion of hypotheses, wider findings, and relation to literature

Associations with baseline CT test score at t₁

While H₁ was supported (higher reading engagement had a positive significant association with higher CT score at t₁), the exploratory analyses around the baseline CT scores yielded further nuanced insights.

Firstly, it is noteworthy that when reading engagement was included in the regression, no association between educational level and CT scores at t₁ was found. At an intuitive level, we may have expected to find higher educational level to be

predictive of higher CT scores, not in the least because teaching CT is often considered to be a key aim in HE (e.g. hooks, 1994; Kuhn, 2005). Although the participants in the study were more highly educated than we would expect in a wider population (over half had a degree), there was nonetheless a fair balance: no degree ($n = 55$); undergraduate ($n = 24$); postgraduate ($n = 42$). This ought to have therefore permitted for the effects of educational level to be shown. It should be noted however that the sample recruited for this study were readers; therefore a wider conclusion about education and CT cannot be drawn, as there may be different associations between these variables in other populations.

Secondly, we may have expected to see an association with age and baseline CT score, as prior research has found some gains in cognitive capacities with age and also some decline of capacities in advanced ageing (Denney, 1995; C. M. Friend & Zubek, 1958; Stanovich et al., 1995); however a Kendall's correlation analysis showed there was no association between age and t1 CT score. The mean participant age was 38 ($SD = 15.35$), median 33, with a range of 62 from 18 to 80; this is therefore a fair spectrum of ages, and though not fully representative of a wider population this should nonetheless give enough scope for age effects to be shown. As with educational level, it is important to remember that the sample was drawn from people who identify themselves as readers. As research conducted by Stanovich and colleagues (1995, 1998) found reading to have a protective effect upon some indicators of cognitive decline in aging, this may be partly responsible for the lack of difference in baseline CT score across ages in the study.

Exploratory analysis found a distinction in fiction and nonfiction engagement and baseline CT score; fiction engagement was significantly associated with higher CT score at t1, but nonfiction engagement was not. This is in keeping with findings from the previous observational study, which showed that fiction engagement (as measured with the ART-G) was associated with CT disposition, while nonfiction engagement was not. However, engagement was measured in a different manner in this study, as an author recognition test would have been time consuming for participants and also comes with limitations. In this study, reading engagement was measured by asking participants to list their favourite authors, up to 10. As an ANOVA showed a significant difference between fiction and nonfiction engagement between participants who self-identified as fiction and nonfiction readers, this corroborates the engagement scores. However, when we look at the descriptive statistics we can see that participants overall included more fiction than nonfiction authors as favourites: total reading engagement (i.e. total number of favourite authors) had mean of 6.24 ($SD = 2.72$), median 6, with a range from 0-10. Fiction engagement had a mean of 4.86 ($SD = 2.85$), median 4, range from 0-10. Nonfiction engagement had a mean of 1.38 ($SD = 1.83$), median 1, and a range from 0-8. This means that while nonfiction readers did list more nonfiction authors than fiction readers, overall there was still a tendency to list more fiction than nonfiction authors. One explanation may be in the phrasing of the question:

“Please list your favourite authors (fiction and/or nonfiction). You may include up to 10 names maximum. There is no minimum number required. Please list as many as you consider to be your favourite. The order does not matter.”

It may be that including fiction first prompted more fiction author names to come to mind, or it may be that in general when asked about favourite authors people are more likely to consider fiction; it may simply be more intuitive to think of authors in terms of fiction writing. This may be worth further investigation. Nonetheless, based upon the data from this study, there is an indication that greater fiction engagement may be associated with greater baseline CT test score, while this is not the case for nonfiction.

Finally, it is interesting to return to the finding that H_1 was supported, and that higher overall reading engagement had a positive significant association with higher baseline CT score; however this regression model only accounted for 7% of the variability in t1 CT score. This means that the variation in CT score at t1 is most likely best explained by other variables which were not measured in the study. For example, intelligence (i.e. IQ) was not measured in the study, which is one variable likely to play a role in CT test performance (Halpern, 2008). Other variables such as those associated with emotional intelligence may also be relevant predictors of CT (Thayer-Bacon, 2000). This study has shown reading engagement to have a small effect, but an effect nonetheless. Though modest in scope, this finding has relevance for my overall research hypotheses due to the way it fits in to the emerging pattern of benefits associated with higher engagement with reading. Just as correlations have been shown between increased reading engagement and cognitive capacities (e.g. Cunningham & Stanovich, 1998), and interpersonal capacities (e.g. Mar et al., 2006), this finding presents a correlation between reading engagement and an outcome that taps into

both; CT. This correlation also fits with those found in study one, building a stronger case for there being a relationship between reading and CT.

Associations with changes in CT score from t1 to t2

While correlations with baseline CT scores offer some interesting insights, the main purpose of this study was to move beyond correlation and to enable some indication of causal direction between predictor variables and CT. To this end, the analyses accounting for how CT test performance changed from t1 to t2 yielded some interesting insights. The change in CT score from before to after the log period was addressed in two ways: CT score change was derived by subtracting t1 score from t2 score; two categories were created by dividing participants into improved (CT change score > 0) and no improvement (CT change score ≤ 0) groups (see Box 7). This allowed for analyses both on magnitude of change in CT between the two tests, and also testing whether that change was or was not positive.

As a 14 day period is a short time frame, very little change could take place in demographic variables such as educational level, gender, and age. It is therefore not surprising that none of these were found to have any association with change in CT score from t1 to t2.

H₂ was supported in terms of time spent reading; higher amounts of reading time recorded over the reading log period had a significant association with increases in CT score change score from t1 to t2. Likewise, a logistic regression analysis found that more time spent reading in total over the log period was significantly associated with being in the improved CT score category, rather than having no change or a decrease in CT score. It is interesting that this was not the case when reading was measured in

the total entries logged over the 14 day period; total entries had no significant association either with CT score change or improved/no improvement category membership. This divergence between the number of entries logged, and the time logged, highlight potential issues in measuring reading quantity using the reading logs. Firstly, it should be noted that reading entries do not represent whole texts read; one book, for example, could be read over several days and therefore recorded in several entries in a participants' log (indeed, books commonly appear in the participants' logs repeated over multiple entries in this way). Entries also do not necessarily equal discrete sessions of a participant reading; a participant could read one text in 3 discrete sessions in one day, taking breaks and undertaking other activities between reading, for instance, but at the end of the day when logging their reading they could chose to make one entry for this text and total up the time spent over those 3 distinct sessions into one. Finally, how participants conceptualised texts impacts how many entries they may have logged; some participants logged articles in a newspaper as distinct entries, for example, while others logged the entire newspaper as an entry. In contrast to this, reading time is a far less flexible construct, and offers an arguably clearer measure of reading than entries do. Therefore when assessing H₂, the findings in support of the hypothesis in terms of time are arguably stronger than those against the hypothesis in terms of entries.

Reading time and entries logged also yielded divergent results in terms of assessing H₃. Taking entries as the measure of reading, fiction had a significant positive association with being in the improved CT category, while nonfiction had a significant negative association with being in the improved category. This supports the hypothesis,

and indeed goes beyond what the hypothesis stipulates as this is not only a greater effect for fiction but an opposite effect entirely. Fiction entries also had a significant positive relationship with CT score change, while nonfiction entries had no significant association, also supporting the hypothesis. However, as time is the clearer measure of reading amount, it is more interesting to assess the hypothesis in terms of the time participants logged reading fiction and nonfiction, and this yields a more complex set of results. Nonfiction reading time was significantly positively associated with being in the improved CT score category, while fiction reading time had no significant association with being in either category. Conversely, fiction reading time was significantly positively associated with increased CT score change, while nonfiction reading time had no significant association. This suggests that spending more time reading nonfiction made it more likely that one's CT score would improve from t1 to t2, but had no effect on the magnitude of the change. Spending more time reading fiction was associated with a greater magnitude of CT score change, but didn't increase the likelihood that one's score would improve. Therefore the hypothesis is not fully supported by these findings. Rather these findings reflect a more nuanced relationship between fiction and nonfiction reading, and increased CT.

Because of the divergence in findings between entries logged and time spent reading, further exploratory analyses were conducted. By dividing the time (in half-hours) spent reading by the entries logged, for total, fiction, and nonfiction read, new predictor variables of half-hours per entry were generated. These variables were then used in replications of the multiple regressions that best fit the data for change in CT scores as the outcome variable (i.e. including total reading, and then including

nonfiction prior to fiction in the hierarchical regression logic). Spending more half-hours per entry logged overall had a significant positive association with CT score change. Likewise, spending more half-hours per entry of fiction logged was significantly associated with increased CT change. There was no significant association between nonfiction half-hours spent per entry and CT change. These exploratory findings lend further support to the hypotheses (H₂ and H₃).

It is important to note the small effect sizes found. These models explain between 4% and 18% of the variation in CT score change from t₁ to t₂. These relatively low R^2 values suggest that other variables that were not measured in this study play an important role in how CT changes over a 14 day period. Nonetheless, these findings do show significant associations between what participants read in between taking the two versions of the CT test, and how their test scores changed.

Narrative transportation

NT was not found to mediate the relationships predicted in H₂ and H₃; H₄ therefore is not supported. However, the measure of NT used in the mediation analyses did not take in account potential differences in the way that participants may be transported by fiction, and by nonfiction; i.e. this was a mean score for NT across all types of reading. It may be the case that being highly transported by fiction is different to being highly transported by nonfiction, thus yielding differences in the ways in which participants ranked their NT on the Likert scale. Furthermore, because NT was measured with a Likert scale it is not a true numeric variable (Jamieson, 2004). The way that each participant interpreted the scale and applied it to their reading experiences is likely to be somewhat different. This means that average NT score for

each participant misses nuance in types of NT, and individual differences in the interpretation of NT. NT was therefore explored further with additional variables derived from the reading logs. Minimum and peak NT levels were identified for each participant for their fiction reading and their nonfiction reading; i.e. of all the fiction entries made by one participant, the highest and lowest NT score given was found, etc. The time each participant spent reading fiction, and nonfiction, at their minimum and peak NT levels was then totalled, and likewise for entries made with minimum and peak NT.

Peak NT time and entries were used in regression analyses to identify any significant associations with CT change. Peak NT fiction entries had a significant positive association with CT score change, though peak NT nonfiction entries did not. However, when looking at time spent reading at peak NT, both fiction and nonfiction were significantly associated with increased CT change score. This suggests that more instances of reading at a peak NT is only important to CT when it comes to fiction, but spending more time reading at a peak NT level of both fiction and nonfiction has a relationship with CT change.

Next, these variables were used as predictors of being in the improved rather than no improvement in CT score groups. Both minimum and peak NT fiction and nonfiction reading were used in logistic regression analyses, firstly with entries and then with time as the measures of reading. In these analyses, only peak NT fiction entries had a significant positive impact on the likelihood of having an improved CT score.

While NT was not found to be a mediator of CT change in the way that was posited in the study hypotheses, the exploratory analyses do suggest that spending more

time reading while being highly transported is connected to greater change in CT. This is distinct from the finding that reading more fiction was associated with greater CT score change because this association was not shown for minimum NT level fiction reading; i.e. it is not simply the case that more reading drives the association, it is specific to peak NT. It is also interesting to note that overall nonfiction reading was not associated with increased CT score change, but peak NT nonfiction reading time was, implying that generally reading more nonfiction may not play a role but the time one spends very highly transported by nonfiction does. It is also interesting that peak NT reading impacted the magnitude of change in CT scores, but did not have a clear connection with the likelihood the change being positive. In this sense these findings align with those from study one, which found NT mediated the effects of fiction engagement on CTD, but found no mediation of EO. Taken together, the findings from both studies imply that NT plays a role in influencing how reading impacts CT, but this role is not straightforward. This is also commensurate with the wider literature, which encompasses findings suggesting NT is negatively associated with more cognitive and logically evaluative factors of CT (Dal Cin et al., 2004; Escalas, 2004), as well as supporting more interpersonally evaluative aspects of CT such as empathy (Hoeken & Fikkers, 2014; van Laer et al., 2013). Thus it is not possible to give a neat unidirectional account of the relationship NT has in influencing how what we read may influence the ways we think critically.

Finally it should be noted that all of these effects are fairly small. These regression analyses only explain 14–23% of the differences in CT. As was the case with the previously discussed findings, these R^2 values tell us that while NT and reading

undertaken between the two tests times do account for some changes in CT, there are other unmeasured variables at play.

Results of the experimental manipulation: assigning readings

H₅ was partially supported, as nonfiction readers assigned to the fiction reading condition were found to have greater change to their CT scores. They were also more likely to be in the improved score category. However, this was not the case for fiction readers assigned to the fiction reading group. It is worth discussing the way in which this hypothesis was tested, before considering wider implications of these findings.

Firstly, multiple regression analysis was used to assess the influence of experimental group assignment on change in CT score from t1 to t2. Fiction and nonfiction reading in both entries and time were also included as covariates to test whether group assignment played any role independently of the reading participants logged. Participants may have chosen to read large volumes of fiction and/or nonfiction in addition to their set texts; I therefore chose to control for fiction and nonfiction reading amount in order to ascertain the specific effect of reading assignment, independently of the overall amount of either type of reading participants engaged in. This permits me to draw conclusions specifically about the impact of what was assigned, rather than how assignment may have contributed to overall reading amounts.

Based upon the regression model including nonfiction entries, fiction entries, and experimental group, the following variables had significant positive associations with CT score change: fiction entries; being a nonfiction reader assigned to the fiction group (Fn) in comparison with a control group nonfiction reader (Cn). Based upon the

same regression model this time using reading time rather than entries, fiction reading time and being a nonfiction reader assigned to the fiction group (Fn) in comparison with a control group nonfiction reader (Cn), significantly predicted difference to change in CT score. Both of these models explained 30% of the variation in CT score change. From these analyses we can conclude that in addition to the volume of fiction read, being assigned fiction reading in the case of nonfiction readers was associated with greater CT change.

Next, a logistic regression model was used to test whether experimental group membership was predictive of having improvement, rather than no improvement or decline, in CT score. Fiction and nonfiction entries, and time, were controlled for in two respective versions of the regression model. Based upon the model controlling for fiction and nonfiction entries, reading more fiction entries had a significant association with being in the improved CT score group, as did being in the Fn group rather than the Cn group. Likewise based on the model controlling for fiction and nonfiction reading time, spending more time reading fiction had a significant association with being in the improved CT score group, and being in the Fn group rather than the Cn group has a significant association with being in the improved score group. These analyses suggest that in addition to reading more fiction, being assigned fiction in the case of nonfiction readers was significantly associated with improvement in CT score.

These findings suggest that in the case of nonfiction readers, being assigned fiction had a positive effect both on the magnitude of CT score change, and on the likelihood of having a positive change in CT score. This effect was significant even when controlling for reading amount, i.e. in addition to reading more fiction (either in

entries or in time logged), being a nonfiction reading assigned fiction was a significant predictor of the CT outcome variables. This is particularly notable because no such effect was found for fiction readers who were assigned fiction. This therefore implies that being assigned fiction reading is uniquely impactful for people who identify themselves as typically being nonfiction readers. Therefore these findings point to fiction text assignment as a potentially useful intervention for prompting change and improvement in CT for people who ordinarily tend to read nonfiction. The use of fiction texts in this manner would be worthy of further investigation.

5.3.3. Limitations

The pre-registration of this study included plans to use multiple regression analyses with CT change score as the outcome variable, but no plans were included in the pre-registration for splitting participants into improved and no improvement CT categories and conducting logistic regression analyses with this as the outcome variable. Before commencing this study, I had assumed that performance on the CT test at t1 and at t2 would stay the same or improve for most participants, and I did not anticipate large numbers of participants having a decline in their score at t2. I assumed that the experience of taking the test at t1 would be likely to boost performance at t2, as the test format would be familiar and participants may have reflected on their answers at t1, thus developing improved approaches for taking the test again at t2. However, my assumptions were shown to be incorrect, as CT change scores in fact had a broad range from negative to positive; the CT change scores were normally distributed, mean change was 0.69 ($SD = 9.93$), median 1, with a range of 60 from -27 to 33. There are many possible reasons why some participants had a decline in score; taking the same

test (though with a different text) again after only a relatively short 14 day period may have resulted in fatigue and disengagement at t2; the fatigue of participating in the study on a daily basis may have built up over the 14 day period leading to disengagement by t2 for some participants; the knowledge that the test at t2 is the final step of the study and upon completion the reward voucher will be issue may have prompted some participants to rush the test at t2. However, these are all speculative explanations. Ultimately, finding that CT score change had a greater range from negative to positive than I had anticipated, I decided that the addition of the binary outcome variable of improved vs. no improvement or decline in CT score was necessary to fully address my hypotheses. This meant going beyond my pre-registered study plans. However, I believe this was a justifiable divergence from the pre-registration given the clearer conclusions these additional analyses have enabled me to draw.

A limitation of this study design was a lack of control for many possible variables impacting CT. Most notably, intelligence has a strong face association with CT. Although intelligence is controversial, and it has been argued that CT is a practical replacement for it (Halpern, 2008), it is hard to ignore robust evidence for intelligence as a distinct and impactful construct (W. Johnson et al., 2004, 2008). It would have been possible to control for intelligence in this study by giving participants a fluid intelligence test. At the time of writing, the shortest robustly validated IQ test identified was the 15 minute version of Raven's progressive matrixes (Raven, 2000). However, adding 15 minutes on to the workload for participants would have been a substantial increase in addition to the 30 minute CT test. Additionally, intelligence tests

can provoke anxiety, and this effects performance (e.g. Moutafi et al., 2006); this may also have had further impact on participants' responses to the rest of the study. In the interests of keeping both workload and anxiety for participants to a minimum, I decided to omit intelligence as a control variable in this study. This poses a limitation for the conclusions that can be drawn from the analysis of CT performance at t1. However, as intelligence is thought to be a stable construct over time (Raven, 2000), it ought not change between t1 and t2, and therefore a failure to control for intelligence ought not to impact findings pertaining to CT change from t1 to t2. There are other potentially relevant variables that could have been controlled for (e.g. socio-economic status, profession, personality, etc.) but by in large these can be presumed to be stable of the two period and thus not relevant for CT change. Some variables such as mood could be relevant to the CT test at t1 and t2, and could have been included for control. However, as the literature review yielded no indication of a previously studied association between such variables and CT, I took the decision to favour brevity for the participants' experience and did not add such further controls. As discussed, the R^2 values of the regressions were all fairly low, generally explaining approximately 20% of the variation in the outcome variables. This suggests that other, unmeasured, variables played an important role in shaping how CT test performance changed from t1 to t2. However, it is nonetheless possible to conclude that what participants read and logged over the 14 days did play a role, as this was significantly associated with the changes in CT. Indeed, a large effect size was never stipulated for this study; reading was never predicted to be the primary driver of CT, rather it was predicted to be a contributing factor playing an important role within a broader conglomeration of influences. This means that even given the small effects found, the aims of this study have nonetheless

been met, even if wider factors contributing to CT change remain to be explored in other studies.

The associations found between reading and change in CT may have alternative explanations. For instance, there may be underlying variables that drive both greater amounts of reading logged and greater changes in CT. One such alternative explanation could be that participants who deployed more effort in participating in the study made a more careful daily accounting of what they read, and thus logged more reading on a daily basis. These participants may then have also spent more time on the CT test, taking it more seriously and working harder to answer the questions. Conversely, some participants may have deployed lower levels of effort, thereby only logging small amounts of reading, and only giving minimal thought to their CT test answers. Thus, effort could be an unmeasured variable in fact responsible for these differences among participants. However, it seems unlikely that this could account for the change in CT score from t1 to t2, as this would require specifically more effort to be deployed at t2. Similarly, this could not account for the divergence between fiction and nonfiction amounts, as more effort in logging reading each day would surely apply to both types of material. Another possible underlying variable could be language ability; participants with greater language abilities would likely have read more, and would also find comprehending the CT test texts easier, and may be better able to write fluent answers. However, again this cannot explain the difference in CT text performance at t1 and t2, nor does it seem likely that language ability would differentially drive fiction or nonfiction reading amounts. Furthermore, some participants did log foreign language texts, thus they would have had had high reading

amounts even if their CT test performance was impaired by their English language skills. Overall, it seems unlikely that an unmeasured variable could be driving both increased reading in the log period, and also the difference in test score at t2.

Language ability may, however, be an alternative explanation for the impact of reading on CT test performance; reading more may have trained and therefore improved participants' language skills, thus making them better able to comprehend and evaluate the text given at t2 and also write clearer answers to the CT test questions. It may be that the language used in the kinds of fiction participants read was richer and more varied than in the nonfiction (this seems plausible if we consider the language of novels in comparison to newspapers), and thus the findings pertaining to fiction benefits could be due to language learning. Indeed, authors from the field of ESL have prominently advocated for the use of fiction in the ESL classroom (e.g. Tabačková, 2015). However, it is unclear that a two week period involving fiction reading would be sufficient to cause such improvement to language capabilities. Further research would be needed to differentiate this explanation to these findings, from that derived from my literature review stipulating that reading fiction would develop interpersonal, imaginative capacities that feed into CT. This would require language testing in addition to CT testing, thus significantly intensifying the workload for participants, and calling for a larger scale study than was achievable within the framing of this research project.

The control group in this study was not assigned any reading, and thus some of the effects seen in the experimental groups may be merely results of reading assignment, rather than its content. An alternative option for a control group would

have been to give assigned readings that were not fiction or nonfiction; however, identifying a clear category of texts between these two types is inherently a subjective and fraught activity (Freeman, 2003; S. Friend, 2008). Furthermore, individual participants may have perceived these texts as being fiction or nonfiction and may have read them as such; in terms of the definitions used in this research project, they would be interpreting the text as an invitation to make-believe or believe (Currie, 1985), and thus they would not have a neutral (neither fiction nor nonfiction) experience of the texts. Another possibility would have been the use of texts that participants ought not be able to understand and thus would not be able to interpret or experience as fiction or nonfiction; e.g. text that is grammatically correct but semantically meaningless in the spirit of Chomsky's (1957) "colourless green ideas" could be used. This could provide an equivalent volume of reading without the presumed effects granted by comprehending what is read. However, prior research has shown that even with low comprehension reading can have an impact (Stanovich & Cunningham, 1993), and so this would not be a true control condition. Furthermore, reading incomprehensible text would also be likely to give participants in the control group a fatiguing and unpleasant experience, which in and of itself may have had an impact on study outcomes. Ultimately, the option of no set reading in the control group was deemed most appropriate, as this permitted a comparison between the un-manipulated behaviours of readers, and the effects of introducing the manipulation of assigned texts. As there were clear differences between fiction and nonfiction assignment, it seems unlikely that the mere fact of being assigned a text could alone be responsible for these findings.

A further limitation to the findings of this study is in the way they should be interpreted. These findings do not show that reading within a two week window changed the abilities of participants to think critically, i.e. no conclusions should be drawn on any long term impact. As has been the case with prior research showing effects from reading emerging immediately after reading has taken place (e.g. Kidd & Castano, 2013), or after short time periods (e.g. Bal & Veltkamp, 2013), conclusions are specific to those short timeframes. Critics have argued that reading fiction cannot improve ToM (for instance) after only one session (Panero et al., 2016), and indeed if one fictional text could result in an enduring increase such capacities this would be highly surprising, and would imply that a prescription of a modest fiction reading list could increase such capacities across the population; if this were true, surely school curricula including literature would have created societies of ToM and empathy savants. This criticism misses the contextualisation to timescale in these studies, and misinterprets their conclusions to extend beyond it. The present study fits with prior research showing impacts of reading after only a brief interval, and concludes that there are impacts of reading upon CT within a two week window, but not necessarily extending beyond it. Perhaps the most plausible explanation for how such time-bounded effects could operate is through a form of priming⁵. Using the term in a more broad sense than its strict deployment in psychology, what one reads can be thought of as priming how one may think shortly after, and given the varied potency of reading experiences this effect can vary dramatically in its persistence (Cupchik et al., 1998). This is in keeping with the way foregrounding is argued to take place in literary

⁵ Priming occurs when a stimulus has an impact upon behaviour or attitude shortly after being presented, typically without a participants' conscious awareness, though there are different forms of priming and much debate around it (Molden, 2014).

fiction, where one part of the text contrasts with the rest and is thus made out stand out, producing defamiliarization as it violates existing schema (Miall, 2008). It may be that fiction entails more foregrounding, prompting more experiences of defamiliarization, and thus has a more potent priming effect. This could be one possible mechanism through which differences in CT test performance could arise after two weeks with different readings. However, whatever the underpinning mechanisms, these findings must be interpreted within the context of the short timeframe of this study. No claims can, or should, be made about the impact of reading within the two log period upon CT at any later date, based on this study. It would be interesting for further research to explore the possible time ranges of such effects, and factors influencing for how long a reading experience may have an impact upon CT.

Finally, the specific characteristics of this study sample must be kept in mind when generalising these results. Although randomised blocking was deployed to balance gender, education, and reading preferences, the sample was overall more highly educated than the wider population, with 55% having a degree in the sample versus 27% of the UK population holding a degree or above qualification at the last census (Office for National Statistics, 2011). There were slightly more women (64) than men (52) in the sample. While there was a broad age range, most participants were in their 30s. Thus, the sampling strategy reduced some of the inequalities found in study one, but did not eliminate them. Results from this study must therefore be viewed in the context of this sample.

5.3.4. Further directions

Suggestions for other research

This study found that assigning fiction texts to nonfiction readers was associated with both greater change in CT score over a two week, and with greater likelihood of that change being positive. This finding would be worth exploring further, utilising different assigned texts, and including participants from broader populations. It would be interesting, for example, to test the effects of assigned reading on non-readers. The use of reading assignment as an intervention for CT change may be a worthwhile wider research direction.

The reading log data included detail that was not captured in the predictor variables derived from the reading logs for the purposes of this study. For example, the genres and topics of the texts participants logged are present in the data, but fell outside of the scope of the hypotheses and interests of this study. Finer-grained analyses would be possible to address any trends in specific types of material read and CT. Similarly, the reading log data contains information about the format (print or electronic) that participants read material in, which also fell outside of the scope of the current study. Patterns in reading over time within the 14 day timeframe of the logs could also be analysed (e.g. comparing readers who read a little every day, to those who read in large volumes but less often). The reading logs also contain data pertinent to the specific time period during which the study was conducted, during which the COVID-19 pandemic was a global issue and many participants logged news reading relating to this, and other current affairs. The reading logs will be published as a freely accessible dataset

for other researchers to make use of, and I am hopeful that further interesting findings will be derived from this data.

In addition to addressing further questions with different variables derived from the logs, some of the variables utilised from the logs in this study could be delved into deeper. For instance, as has been discussed reading entries is not as consistent a variable as reading time, given that participants may have interpreted entries differently (e.g. a whole newspaper could be seen a single entry, as could each individual article). It would be possible to individually code each reading entry logged with a more consistent scheme. For instance, reading material type could be classified into categories such as book, journal, article, etc. These could be further sub-divided into fiction and nonfiction, or fine-grained categories. This would perhaps help to elucidate the differences found in this study between entries logged and time. As prior research has shown book reading to be associated with CT (Hawkins, 2012), differentiating between books and other read materials would be a relevant avenue to further test the hypotheses of this study. The difference between books and other materials may be more pronounced in the case of nonfiction, where non-book materials are arguably more diverse than in the case of fiction, which is a possible explanation for the contradictory findings between nonfiction entries and time. Perhaps reading longer nonfiction materials such as books is beneficial, while reading very short nonfiction pieces such as brief articles is not. This deeper analysis fell outside the scope of the current study, but would be relevant for further investigation.

To be taken further in qualitative analysis

As this study was part of a wider mixed-methods research approach, some of the findings from the quantitative analyses discussed here will be taken forward and explored further in a following qualitative interview study.

Some issues arising from this study were further elucidated in the study three interviews. Firstly, the issue of whether being asked to name favourite authors (as a measure of reading engagement) prompts more fiction authors than nonfiction authors to be named, even for people who tend to primarily read nonfiction, was further explored in discussions with interview participants. Additionally, the issues surrounding entries logged vs. time spent reading was explored by asking participants how their reading experience may be different in many short reading sessions rather than fewer longer sessions, and how they feel they may benefit from their way of reading. These clarifications are presented in Appendix D under Elucidating quantitative findings.

The main findings of this study were also further expanded upon in the study three interviews with readers, and study four diaries. Readers' different experiences of how spending more or less time reading fiction and nonfiction can prompt changes in their thinking were explored in the interviews. Readers' insights into the ways in which NT shapes their reading experiences, and how this may also be contributing to their CT approaches, were sought in both study three and four. The finding that assigning fiction texts to nonfiction readers had an impact on CT change also formed an avenue of conversation in study three; participants were to reflect on whether reading material that is divergent from their usual reading habits results in distinctive experiences for them, and whether they feel this may change their thinking approaches

in any way. They were also asked to provide some thoughts and intuitions on what kinds of texts may promote CT if assigned as readings. Therefore the findings from this study were brought into the design of studies three and four.

5.4. Conclusions

In readers, CT test performance was shown to change over a two-week period, with significant associations found between the amount read over that time in terms of fiction and nonfiction, and the degree and valence of that change. Prior to commencing the 14 day reading diary, participants' pre-existing levels of reading engagement were shown to be significantly associated with a higher CT test score at t1. This supports prior findings of a correlation existing between reading engagement and factors of CT. Moving beyond correlational analysis and into identifying a causal direction for these effects, this study found significant associations between participants' reading and differences in their CT test performance from t1 to t2. There was a significant association between the amount of time participants spent reading and CT score change from t1 to t2, as well as the likelihood of having improved CT score at t2; the total time spent reading in between CT tests was predictive of both greater magnitude in change, and of that change being positive. Next, fiction and nonfiction reading were compared. Higher amounts of fiction reading had a stronger association with increases in CT score change from t1 to t2 than nonfiction (which had no significant association), when measuring reading both in time and in entries. Fiction reading was therefore uniquely associated with increased CT change. In assessing the likelihood of that change being positive, the results were more mixed. Higher amounts of fiction reading were only associated with greater odds of improvement in CT score when measured in entries logged, not in time. Conversely, more nonfiction entries

logged was associated with no change or negative CT change. Contradicting this, higher amounts of nonfiction reading was significantly associated with greater odds of improvement in CT score when measured in time. This suggests that while fiction reading has a stronger association with increase in CT change, fiction and nonfiction have a more complex association with the valence of CT change. As the study included experimental manipulation in the form of assigned readings, it was also shown that assigning fiction reading to nonfiction readers was an effective intervention for both increasing CT change, and for increasing the likelihood of that change being positive. Finally, NT was not shown to mediate the effects of reading on CT, but reading more with a peak NT level was found to have some impact. Therefore the findings from this study show that how much reading people engage in over a two-week period, in total and in terms of fiction and nonfiction, has an impact on changes to their CT. These findings broadly support the wider hypotheses of this research project, and provide insights to be pursued further through following qualitative studies.

6. Study three: **An exploration of reader's experiences of reading and critical thinking**

This study was conducted both as a standalone attempt to address the wider questions of this overall research project, and also as a follow-up to the previous quantitative studies. As such, it aimed to both seek elucidation to some issues raised in the quantitative research, and to explore the experiences of participants to gain a richer insight into the factors being investigated. As such, this study forms a hinge in the mixed methods sequence of this research project, firstly deepening understanding of previously identified trends and shifting the research into broader exploration of experiences. Both reading and thinking critically are deeply personal experiences, and therefore in addition to identifying trends across different participants I also wished to address those experiences at an individual level. Therefore interviews were conducted with people who considered themselves to be readers, and they were asked to articulate what it is like for them to engage in reading, and the ways in which they think critically.

From the prior quantitative studies, complexity was particularly notable in the findings pertaining to EO in study one, and the findings pertaining to NT in both study one and two. Although EO is often treated as a scale from absolutism, through multiplism, to evaluativism (Hofer & Pintrich, 1997), study one found a far more nuanced and overlapping clustering within these categories, and reading fiction and nonfiction were differently associated with different positions in this complex EO space. As such, further exploration of how participants orient themselves epistemologically, and how they locate their reading experiences within this

orientation, is warranted. Therefore the interviews included questions to capture this. Furthermore, the interplay between reading, NT, and CT was also found to be intricate and worth deeper exploration. Mediation by NT was only found in study one for CTD, but not for EO, nor for any outcomes in study two. NT had direct effects for both the outcome variables of study one, but none for study two. However, in study two differentiating between reading at an individual peak NT level, it was found that reading while being maximally transported had an association with increased CT for both fiction and nonfiction (though with inconsistencies between entries and time logged). This invites further probing into the influence of NT in different kinds of reading, upon CT. Participants' transportation experiences, with attention to time spent reading different items, were therefore also sought in the interviews. Box 8 gives a summary of the study.

Box 8: Study three summary

This study aimed to investigate readers' experiences of CT and reading, comparing fiction and nonfiction. Two interviews were conducted with self-selecting participants who self-identified as readers ($N = 12$). An iterative coding process yielded 10 codes from the data, forming 5 categories. These were amalgamated into 3 themes. The findings suggested NT integrates CT into the reading processes, and that fiction and nonfiction were experienced as having distinctive relationships with CT. Fiction was found to give deep understandings of the real world, promoting CT through the circuitous and complex way it presented a range of views.

The research questions guiding this study were:

- a. In what ways do readers consciously experience an influence of what they read on the way they think critically?
- b. In what ways might readers' epistemological orientations relate to their reading and CT experiences?
- c. How do readers' experiences of NT relate to the ways in which they think critically about a text?

In order to ensure transparency, this study was pre-registered. The preregistration was amended due to changes made to the study plans resulting from the COVID-19 pandemic, prior to data collection. The original registration is available here: <https://osf.io/3wjpr>; amendments: <https://osf.io/ncp4d>.

This study has been adapted for a journal article:

Hollis, H. (2021). Readers' experiences of fiction and nonfiction influencing critical thinking. *Journal of Librarianship and Information Science*. 1-15.
<https://doi.org/10.1177/09610006211053040>

The data from this study has been published, with access restrictions in place for safeguarding purposes:

Hollis, H.(2021). Transcripts from interviews on reader's experiences of reading and critical thinking. [Data Collection]. UK Data Service. 10.5255/UKDA-SN-855246

6.1. Method

6.1.1. Participant recruitment

Participants were self-selecting, responding to calls for participation on social media (Twitter and Reddit), and via the study email newsletter. The calls for participation asked for readers or bookworms, reading fiction and/or nonfiction. No other selection criteria were used. Interested participants responded to the recruitment call by email, whereupon they were sent the information sheet and consent form. Every participant who consented to participation was interviewed, up until the point that saturation was reached and no further participation calls were put out. In total, 12 participants were recruited, and all completed the study.

6.1.2. Interviews

I conducted semi-structured one-to-one interviews with each participant. Each participant was interviewed twice, and asked to read a text of their choosing between the two interviews. Interview 1 focused on participants' views on reading and CT in general; interview 2 focused on participants' responses to a specific text they had read in terms of their reading and CT experience. The text to be discussed in interview 2 was entirely of the participants' own choosing, and it was made clear that it could be any type of reading they preferred. A set of prompt questions was provided at the end of interview 1, so that participants' would have some idea of what to expect from interview 2 and would be able to think about their chosen text in terms of the interview topics ahead of the conversation. The scheduling of the interviews was flexible so that participants could spend as much time as they wanted reading their selected text in between the two sessions.

Both interview 1 and 2 contained questions on the topics of CT, NT, and the experience of reading, and how reading may impact CT. The topic guides for each interview are available in Appendix D. Interviews were conducted between 18/06/2020 and 12/09/2020. Each interview lasted between 30–45 minutes, and was conducted online via the Whereby videoconferencing platform. Interviews were recorded, and the recordings transcribed using the automated Otter.ai transcription service. I then checked and corrected transcripts in two phases of listening and correction, and then deleted the recordings in accordance with data protection provisions. All transcripts were anonymised, with participants given letters of the alphabet⁶ as pseudonyms (e.g. Participant B).

6.1.3. Analysis

Participants' answers to questions following up from concrete issues raised in the prior quantitative studies were analysed separately to identify possible elucidation for some of the findings of those studies (see Appendix D Elucidating quantitative findings). Here, analysis of the interview data in relation to the overall aims of the research project and the above stated research questions will be discussed.

Saldaña's (2016, p.70) “pragmatic eclecticism” was adopted as an approach to coding and then categorising the data. As such, I sought to maintain a high level of openness in the initial stages of analysis, and therefore used a constant comparative coding approach, starting with concept coding and then moving into holistic coding (Saldaña, 2016). I commenced coding when 6 out of 12 participants had completed

⁶ The letter ‘I’ was not used so as to avoid confusion with the use of the I pronoun. The letter H was not used so as to avoid confusion with the interviewer’s initials.

both interviews, and continued to code new transcripts as they were completed in an iterative manner. Initially, I used a free coding approach that yielded a very high number of codes (in excess of 80), which allowed me to capture my first detailed interpretations of the data. I then focused on reducing this prolific free coding by focusing on concepts and seeking duplication, and integrating similar codes. Next, I sought to reduce the number of codes into more meaningful, less granular, wider holistic coding. To achieve this, I grouped codes by relevance to my research questions, to facilitate the seeking of broader units of meaning in relation to my research aims. This enabled a re-coding into a smaller number of more relevant codes. Ultimately, I felt the research questions to be too constraining, and therefore I moved to approaching the codes in groupings by topic; CT, NT and reading. From this approach, I arrived at a final set of codes for each topic (6 codes on the topic of CT; 2 codes on the topic of NT; 8 codes on the topic of reading). I then derived 8 categories from the codes which captured wider patterns emerging from the coding.

6.2. Participants

Participants were all speakers/readers of English, and based in the EU. All were adults (i.e. 18 years old or above), and of varied backgrounds, genders and ages.

Table 44 shows participant reader type self-identification, genres or topics identified as favourites or main readings, and a description of chosen text prior to interview 2.

Table 44: Table of Participants

Participant	Reader type	Preferred genres/topics	Chosen text
A	Nonfiction	Current affairs, History, How-to guides	Science article on COVID-19
B	Fiction	Classics, Literary Fiction, Mystery, Biography	Mystery novel
C	Balanced	Horror, Historical Fiction, Literary Fiction, History	Historical fiction novel
D	Balanced	Popular Fiction, Graphic Novels, Fantasy, Science	Science article on pain
E	Balanced	Thriller, Science, Biography	Autobiography book
F	Balanced	Classics, Fantasy, Sci-fi, Politics and Economics, Science	Literary fiction short story collection
G	Fiction	Fantasy, Historical Fiction, Mystery, Culture	Fantasy novel

J	Nonfiction	Classics, History, Philosophy, Science	Popular science book on genetics
K	Fiction	Sci-fi, Literary Fiction, Classics	Science fiction novel
L	Nonfiction	Science, Biography, Professional	Design article on chairs
M	Fiction	Literary Fiction, Fantasy, Mystery	Opinion article on school uniforms
N	Nonfiction	Fantasy, Professional, History, Biography, Science	Short story collection

6.3. Results

The results are presented here first at the finer-grained level of categories derived from the data, and then themes derived from the categories. This permits a full description of each category in detail first, including each code that comprised it, prior to a broader thematic understanding.

6.3.1. Categories

The findings from the interviews will be presented here under the topic headings of CT, Reading, and NT, as these were the topics of the interview questions and the

categories and codes derived from the data fit into these headings. I place emphasis on categories, rather than individual codes, in presenting the results here with an aim to prioritise the implicit interconnections between codes over the explicit codes themselves (Saldaña, 2016). Following on from this, the interconnections between these categories will be described by discussing them together under broader themes. Table 45 provides an overall summary of the codes and categories.

Table 45: Summary of codes and categories

Code	Category
Prioritisation and identifying key information Logic	Critical thinking is procedural not personal
EO as a personal approach to CT Openness and curiosity (Non-conformism)	Critical thinking is a matter of personal characteristics
Inference making and imagining Slow and Reflective	Critical thinking is an experiential process
Expanding or shifting mental capacities Prompting further thought and action.	Reading shifts modes of thinking
Reading as widening exposure (Authors as sources) Reading prompts CT circuitously	Reading material contains thinking prompts
Illuminating other life experiences Source of social or historical information (Issues around veracity prompt CT)	Fiction provides understanding of the real world
Reading to gain information Explicit opinions and questions directly prompt CT	Nonfiction provides direct information
Narrative transportation is active engagement Narrative transportation is connection-making	Narrative transportation integrates CT into reading

Note: sub-codes are presented in brackets.

Critical Thinking

Participants were asked both to describe what they believe CT to be, and also to describe their own applications of CT. The following categories are derived from both ways of talking about CT.

Critical thinking is procedural not personal

Many participants described CT as a dispassionate application of procedure, characterised in the codes of *prioritisation and identifying key information* and *logic*. They viewed CT in a goal-oriented manner, and identified steps to be taken to reach the goal of the thinking.

The application of a procedure was a prevalent concern in this category. Participants emphasised an ordered *logic* to their CT, describing following a “method or a recipe” (Participant C) in a “stepwise” (Participant D) manner. This deployment of logical procedure was seen as a key way in which CT could be reliable:

“deploying more or less the same system, the same things each time, of analysis across different problems and different issues so that their thinking is structured and reliable in some sense.” Participant J

Thus emphasising the goal-oriented focus of the *logic* CT approach. Furthermore, a separation of thinking and emotion was a key element in this code, as an extension of the concern for reliability. Participants described a need for “cool, quite clinical” (Participant F) approaches:

“I think if you are going to put a hierarchy on it, and critical thinking is above an opinion, opinions are quite emotionally based.”
Participant D

Suggesting that deploying a dispassionate CT *logic* procedure was perceived by some participants as superior mode of thinking. This also emphasises the sense in which CT was perceived as impersonal when thought of in this logical, procedural manner.

Furthermore, when describing the procedure of their CT, participants expressed a need for attention to detail in *prioritisation and identifying key information*. This part of the CT procedure entailed “breaking it down into simple elements” (Participant A) in order to “glean the key information off” (Participant C). Again, this approach was highly goal-oriented, with participants focusing on how information could be found and isolated for their purposes. This further de-emphasises personal aspects, such as the role of the authors of the information, in favour of impersonal atomised uses of information.

Finally, *prioritisation and identifying key information* and *logic* were often connected and entwined as part of the procedure of CT:

“Well, there’s critical in the sense of analytical. Or determining accuracy. Or perceiving waffle. The ability to read for those – I’m a Yorkshireman, there’s an old Yorkshire saying ‘happen he talks a lot but he says nowt’ – and a lot stuff I read, particularly in newspapers, are that. Even in books. So they go on and on, and I find their logic foolish, or the initial premise is nonsense.” Participant A

Here participant A describes seeking relevant information and filtering out “waffle” as part of an “analytical” process. The Yorkshire adage cited by Participant A also neatly summarises the impersonal emphasis of this approach to CT, as the focus is on extracting information identified as relevant to one’s goals, and discarding everything else.

Within the category of *critical thinking is procedural not personal*, the codes *logic* and *prioritisation and identifying key information* thus combine to form a specific approach to CT adopted by some participants that emphasises generalisable procedure over personal elements.

Critical thinking is a matter of personal characteristics

Participants articulated a grounding to their CT based in their own personal characteristics, as captured in the codes of *EO as a personal approach to CT* and *Openness and curiosity*. Here, CT was characterised as emerging from this personal trait foundation.

It was striking that when describing their EO, participants did not treat it as separable from their ways of thinking critically, but rather discussed it as the substructure of their CT. This was a unifying characteristic in the way participants talked about their epistemology across different orientations, thus prompting me to adopt the code *EO as a personal approach to CT* rather than maintain separate codes for different orientations. Absolutist examples of this include Participant E's assertion that in the case of the climate emergency "there's only one right" but CT is required as a means of navigating the issues and responses to them in order to "arrive at that point". A multiplist example can be seen in Participant K, who articulates how approaches can be "different but valid" and thus CT is required to fully explore and understand all of those differences. An evaluativist example is Participant B, who advocates "rational analysis" alongside accounting for personal experiences as a CT approach taken in order to bring about equitable outcomes. Thus, each of these example participants has a different EO which provides a grounding from which they

then deploy their CT. Furthermore, *EO as a personal approach to CT* could be seen in the way participants articulated having different EO for different domains, and how this could facilitate different types of CT, for instance:

“But, are there some things that have equal weighting? Probably not. The perfect example currently would probably be the debate, which I don't think it should be a debate at all, about vaccinations. To my mind, you could listen and explore the other side of the argument – the anti-vaccine argument. Then, to me, the fundamental truth if you like is: vaccines do save lives, they do prevent people from getting vaccine preventable disease. So I think you can listen to the other side, perhaps understand how people came to hold such an extreme view – to me personally to be against vaccination is an extreme view. But I would be interested in how did somebody come to believe in such a thing? You need to understand how they got there. [...] It's almost like – I'm not interested in that view at all, I'm more interested in allowing you to speak and show me what you're reading, so I can understand the how or the why. Why do you believe this? How did you come to believe this? If we don't allow them to speak, to share, then you can't understand the how and the why.” Participant D

In this example, Participant D expresses an absolutist approach within the medical domain, while allowing for a multiplist approach to other's perceptions of medical issues, thus opening a window to think critically about the current vaccination debate. Furthermore, participants often identified changes in their EO across their lifespan as leading to changes in their CT, for example: Participant F describes a shift away from absolutism over time alongside more sensitivity to nuance; Participant J describes a shift away from multiplism along with building CT confidence. Therefore *EO as a personal approach to CT* is not stable in the sense that each participant does not always have one EO, but rather there is an overarching sense in which EO is a context-dependent

personal starting point for CT. Finally, CT was seen as a way of feeding back into one's EO, and thus perhaps the relationship may best be described as cyclical:

“But I'd say the flip side of that is sometimes I think when you think more critically, you're also more accepting, which sounds bizarre. It sounds like they should cancel each other out. I think when you think critically, you can understand all the multi-dimensions of any given subject. And you're not so quick to disqualify certain facets of a subject. So you're questioning everything. But you're also accepting that something can be viewed in different ways.” Participant D

In this example Participant D suggests increased CT can increase one's multiplism. This further emphasises how EO and CT were enmeshed in the participants' experiences.

A personal characteristic described as necessary for CT that seemed to cut across differences in EO was a need for *openness and curiosity* in order to engage with issues and commence CT. Many participants articulated a need to challenge oneself in order to be a critical thinker, and this was furthermore viewed as a “the joy” (Participant D) of CT. Part of *openness and curiosity* was often articulated as a need to go beyond what is given, and seek additional information or ideas:

“I would see a critical thinker as someone who is willing to look outside their own perspective, or perhaps the perspective of the author or the text that they're reading, so someone who wants to understand more than what they're reading. So they might use parts of their own experiences to do that, but also might reference back to other things they've read, try and unpick and understand the relevance and the truth. The facts behind what they're reading. So maybe not taking something for face value.” Participant N

Indeed, not taking things at face value was expressed by some participants as a kind of *nonconformism*, as part which participants conveyed a pride in personal characteristics

that may be traditionally associated with negative connotations: “confrontational” (Participant A); “cynicism” (Participant E); “ask uncomfortable or annoying questions” (Participant L). These were often described as characteristics that caused trouble, but which were necessary to be a critical thinker. Thus, participants offered a prescription of challenge; challenging oneself through *openness and curiosity*, and challenging others through *nonconformism*:

“So I guess a critical thinker has to be, to some extent, quite accepting of discomfort. Quite willing to be an outsider, or to step out of line. And, additionally, I think that comes with being open. Open to ideas, open to different points of view, and definitely open to being wrong.” Participant J

In this way *openness and curiosity* were perceived as necessary for CT, from which *EO as a personal approach to CT* could then operate to shape the direction of CT.

Critical thinking is an experiential process

Participants described the experience of thinking critically in terms of the codes *Inference making and imagining* and *Slow and Reflective*, portraying an inner process of gradually altering their experiences and coming to new understandings. They viewed their CT processes as reflective and internal, as well as active in connection building and extrapolating beyond the given text or problem.

The need for the experiential process of CT to extend over time was notable, with participants expressing the importance of a slower pace:

“Careful. The word slow just comes in. I can't imagine being a critical thinker and rushing it somehow. So unhurried, maybe equivalent to slow there, maybe captures it a little better.” Participant F

With this slowness, participants furthermore articulated reflection that extended throughout the course of their CT. Indeed, the slowness was perceived as necessary for reflection, requiring the time course to “let it sink in” (Participant E), and extending indefinitely “it doesn’t have an endpoint” (Participant J). The reflective experiences of CT were very profound: “I completely, completely changed” (Participant D), and this experiential CT process was often described as personally highly meaningful. Thus, the *slow and reflective* nature of CT reveals a gradual, building process of thinking that yields insights and changes.

The contents of the *slow and reflective* CT process were described by participants as entailing *inference making and imagining*. Inference-making was often approached by participants in terms of asking questions, to move from the text or topic at hand to the inferences they could draw from it:

“And you've got to ask yourself first of all, what am I reading? And then why? Why am I seeing it? Why am I reading it? Who wrote it? Why did they write it? What audience did they intend it for? What else was going on at the time that they wrote it? And I think once you've asked all those questions, then it's time to look at the content in more detail and see, does anyone else agree with this? Is this a widely believed opinion? Is it factually backed up by things that you can prove happened? Or is it presenting their opinion as facts? And then you can go even deeper and do things like reverse image searches and stuff on Google to find out where is this image from. Is the caption telling me what it is, what the author wants me to think, or it is actually true?” Participant G

This questioning was integral to the ongoing process of CT, and was described as taking place throughout. Furthermore, imagination was intimately tied to being able to draw inferences, as an imaginative process of filling in the questions raised through

inference making: “you have to kind of, imagine, the chains of their thoughts” (Participant K), such that the CT process was not solely about identifying questions but also entailed creatively seeking different answers to them. This illustrates how *inference making and imagining* were entwined into one experiential process of CT for participants. Additionally, *inference making and imagining* about one’s self was often described, turning the experience of CT inward, and re-emphasising the reflective components. Furthermore, a *slow and reflective* approach was associated with enabling the *inference making and imagining* necessary for CT:

“So, maybe it helps to think of an example, but like with what’s going on in the world today, you can hear something like ‘defund the police’, and you could just think ‘well that’s obviously stupid’. You could make a really quick immediate gut-reaction type response. That wouldn’t be thinking critically, to think critically you would need to pause, and go really slowly, and think ‘okay, what exactly is meant by defund the police’? What are in fact the exact demands that these protesters are making? Why are they making them? And then, on a smaller level, what precisely is it about the police that we might think of as positive, and is negative, and how do you, how different funding models shape that, and how my other funding models shape it differently. Oh and then you have to think, I think, what would my utopia look like? I mean, in the best possible society, what would police look like then? And you have to do some kind of comparison between where we are now, and that distance to an ideal version, and then you kind of have to run different imaginary paths I guess. And you have to think quite carefully, quite cautiously, about the way different attempts to make things better, you have to think about the consequences. The causal chains, kind of is what I mean I guess.”

Participant K

Here, Participant K outlines a CT process around the issues raised in the Black Lives Matter protests, and this process is slow, requires the raising of questions alongside imaginative answering, all the while with reflective reference to one's own values. Thus, this exemplifies how these codes come together to characterise the ways in which CT is an experiential process.

Reading

In describing their experiences of reading, their evaluative processes during reading, and their perceptions of the relationship between reading and CT, participants described both overall ways that both fiction and nonfiction experiences were similar, and also identified some specific characteristics of their fiction and nonfiction reading that were distinct. The following categories were derived from the answers given on these areas, firstly focusing on the overall reading experience and then addressing fiction and nonfiction individually.

Reading shifts modes of thinking

Participants expressed various ways they experienced reading causing change in their thinking processes and ways of thinking, characterised by the codes *Expanding or shifting mental capacities* and *Prompting further thought and action*. Reading was experienced both as changing the ways participants thought during and immediately after reading, and also in the longer term. It was described as altering the subject matter participants thought about, and their experiences of thinking.

When describing *expanding or shifting mental capacities* there was a general sense across participants that reading helped them to develop cognitively, with commonly held views that “It helped the brain evolve” (Participant A) and “allows

you to broaden your mind” (Participant C). Furthermore, participants described adopting structures of thinking from what they read. Sometimes this could be a very explicit adopting of ways of thinking gleaned from the content of their reading, for instance several participants articulated adopting an investigative mindset based on reading mysteries, or more broadly as Participant K explains:

“Anyway, so I really liked the problem-solving, engineering, mindset that was portrayed in the novel. And so I found myself adopting it I guess. Trying to think like those characters. So maybe I was learning a way of thinking critically, by having it modelled in those characters. Like it was quite aspirational. But yeah, what I’m getting at, is that for me I think everything I read gets under my skin. It colours my thinking. Whatever it is.” Participant K

This is a very conscious, active way of adopting a way of thinking from one’s reading. In other descriptions, participants articulate a more subtle and unconscious change to their thinking resulting from their reading experiences. These adoptions of a mode of thinking driven by the reading experience are a type of *expanding or shifting mental capacities* that can be very wide-ranging:

“And I think finally, it would give you a sense of navigating, which, again, is not necessarily appraising, analysing what you read, but it’s more mapping a path through something. I think there’s a lot to be said about reading books where you have to interpret what the author is saying, to get to the final conclusion. Gaps are the nice thing about some of those books, if you can compare it to say like a darkened room. Some you start reading it, you can see 10% of the picture, then when you might get three quarters of the way through and you can still only see 60% of the picture. And then you get to the end of the book and you might only see 75 and the rest is your interpretation – what happens next?” Participant C

Here, Participant C describes reading contributing to the way we navigate through our thinking, as *expanding or shifting mental capacities* in terms of mapping and path-finding.

In addition to altering ways of thinking, reading was also associated with more concrete *prompting further thought and action*. Participants described many ways that their reading prompted alterations in their thinking around different issues and experiences. Reading that related to their personal experiences was described as changing their perspectives on those experiences: “reading something that reflects your own life experiences can also make you make you question and think about those” (Participant B). Furthermore, most participants described the subject matter that had been read as persisting, and continuing to occupy their thoughts after reading “I was thinking about it the whole time I was cooking dinner, the whole time I was eating, and you know, in the shower, before going to sleep.” (Participant J). Reading was described as changing specific opinions, as well as driving curiosity to learn more on certain topics:

“It made me think critically about the power of celebrity and it's place in a political campaign. And to the point that possibly I've reversed my view completely, because I always used to say, ‘oh, why is it that we put someone on the telly and just because they've said it, it gives it more value’ - it gives it more credibility, because of whoever said it. But now I can see that unless you use those big names and get them on side, to make a huge difference in political terms. [...] So I completely sort of changed my cynical views on the role of that. And the other thing that I thought long and hard about was how he maintained such a high level of motivation over a very long time, because obviously there would have been so many setbacks, so many knocks, so many different factions saying things should be done in a

different way. How hard it is to just know that what you're doing is correct, and stick with that, and not be swayed by other strong personalities. So, yeah, it made me think very long and hard.”

Participant E

In this example, Participant E articulates both concrete changes to opinions related to the subject matter of the reading, and also further wider thought prompted by reflection with relation to the figure described in the text. Indeed, most participants expressed a desire to read more on a topic that had been shaped by prior reading experience (e.g. Participant D for factual information relating to a nonfiction text, Participant F for the philosophical theory that underpinned a fictional text). For some participants, what they had read was unsatisfactory, and this created a desire for better reading on the newly interesting topic (e.g. Participant K seeking a novel portraying AI characters with more nuance, Participant L seeking books on design with a wider perspective).

As part of *prompting further thought and action*, participants also described how the changes reading brought about in their thinking led to concrete changes in behaviour:

“I think probably the story that impacted me the most was this story about the woman's grief around her child. And I think that, um, I found myself thinking about it. We're in a society right now, especially with lockdown and masks, we can't necessarily see everyone's emotions fully. And we're not maybe paying as much attention to people because we're kind of like ‘get away, step back’. So um, definitely it worked. Like for the couple of days afterwards I was very aware of people around me, and kind of the emotions they were feeding. And there was actually a, we lost a colleague a few days ago, he had a heart attack at 41 at work, and people's grief and they're

processing the grief - that made me think of the story in a different way. The colleagues around me that either knew him very well, or were responding. So I think, the kind of little messages of things people do that you don't pick up on. So I think from that, yes, I think that particular story probably out of all of them, had that kind of impact. So I was like, that's something I'll think of in the future.”

Participant N

Participant N here describes both an alteration to thinking about grief, and a change in relation to work colleagues. For many other participants, changes in behaviour resulting from reading were very direct, for example taking supplements (Participant A) or altering a workstation (Participant L) as a result of nonfiction reading. When it came to fiction reading, reading was described as prompting reflection into how participants might behave or what actions they might take, for example what one might have done had they lived in Nazi Germany (Participant G) or how brave one might be in a conflict (Participant E). Therefore *prompting further thought and action* encompasses both literal action taking and also reflective action modelling.

Taken together, the ways in which participants described their reading *expanding or shifting mental capacities* and *prompting further thought and action* illustrate very substantial and formative changes brought about as *reading shifts modes of thinking*.

Reading material contains critical thinking prompts

One of the key drivers participants identified for the power of reading to change their thinking lay within the content of their reading material. This was characterised in the codes of *Reading as widening exposure*, which included the sub-code of *Authors as sources*, and *Reading prompts CT circuitously*. Participants identified

prompts for thought-change in the breadth of the subject matter of their reading, as well as the breadth of authors. Furthermore, they identified subtler and less explicit ways in which reading could prompt CT and changes in thought.

Participants expressed a strong belief in the power of *reading as widening exposure* to world views and broadening the mind, by breaking through the barriers of habit and forcing the reader out of their comfort zone:

“I guess there’s always a risk of getting into a bubble, you know, like, reading always the same type of things, the same subject matter from the same perspectives. And if you did that, that would definitely influence your thinking habits, in the sense of narrowing them down I think. And conversely, I think if you read broadly – so on many different subjects from many different points of view – then that’s going to broaden out your thinking too.” Participant J

Often participants felt that books recommended by others, such they would not be within their own normal reading preferences, had an amplified effect. Furthermore, reading a very familiar author was viewed as having drawbacks, both in terms of limiting exposure to new ideas and also in terms of the authors’ written style becoming overly familiar and formulaic (e.g. Participants B and G identifying the formulas in mystery writing). When asked to recommend reading that may improve CT, many answered with a call to “read in abundance” (Participant D), rather than giving a specific text. For many participants, this *reading as widening exposure* to thinking outside of their general patterns or habits was perceived as desirable and enjoyable. However, there was a converse sense in which participants expressed a desire to read material that did not violate their expectations and conformed to their existing beliefs and ways of thinking:

“I have held for many years the view that the suffragettes was a negative movement. It didn’t help women get the vote. That is not the commonly held view. But I read something in the British history book I referred to earlier about the suffragettes, and I wasn’t critical of it, I was very pleased to read an opinion which is close to my own.”

Participant A

Here Participant A illustrates an example of reading to confirm, not to widen, one’s views. However, despite this tendency to select authors who fitted into participants’ values, participants nonetheless noted a broadening of perspectives:

“it'd be very hard to always be picking up a book that met all your expectations, that would be something that always prescribed to the exact same beliefs as yourself. I think most people are going to think differently to you about something. I guess if you were like an extremist, and all you're really reading is books about terrorism, maybe you're not really thinking, critically, you're just gonna be reaffirming your own belief system. But chances are if you're using different authors, you're going to hit something which then makes you question something, even if it's something small. So I think it'd be very hard to not think critically, I suppose.” Participant N

Participants also expressed a personal relationship with some authors they read often; a nonfiction example being Participant N’s longstanding reading of Monty Don; a fiction example being Participant G’s lifelong reading of Terry Pratchett. In these cases, it was not the variety of authorship, but rather the deep understanding of a single author’s viewpoint that was experienced as *reading as widening exposure*, through that intimacy. Thus, an overall need for breadth of exposure through wide reading of many authors was complemented by a need for depth in reading a single author.

As part of this concern for *authors as sources* within *reading as widening exposure*, participants made evaluations and assessments, thus thinking critically about

them. In the case of nonfiction, authors were primarily assessed for their credibility and authority (for instance, Participant N questioning the credibility of a heterosexual female author writing a biography of an LGBTQ figure, or Participant A's evaluation of author's qualifications), while when it came to fiction the role of the author's own opinions was perceived as less overt, and therefore offered more complexity as a prompt for CT:

“When you're writing fiction I think you have to be able to have a bit of distance. An author will have their own values and biases which will probably come out in in their work, but we have to think about to what extent do we take things in that book as an author's genuine views, for example, like if they use racist language is that a reflection on the author as a person, rather than a character they're creating? I guess it's about having a little distance between a story and a real world impact. But then to deny that stories do have a real world impact isn't good either. People read them in the real world.”

Participant B

Thus *authors as sources* acted as a prompt for CT, and readers' own reading habits in terms of their breadth and depth were also in and of themselves prompts for self-reflection and CT.

Furthermore, it was notable that for many participants the reading experience could prompt CT in a round-about, subtle manner; *reading prompts CT circuitously*. In the case of fiction, participants often described having made a choice of reading for pleasure and relaxation, but then finding themselves thinking critically despite their motivation being not to have to think:

“Like I'm reading a murder mystery, I keep on thinking along the lines of the clues, and do I have an idea of who it might be. So trying

to work that out. And then sometimes I'm not, because I want to be surprised at the end. I don't want to have guessed. I want that element of surprise, and enjoyment. And I'm reading those sort of books to relax, perhaps, rather than to challenge to my brain, or think critically. But there's always sometimes things in books where you think 'oh my gosh, I never would have thought of it in that way', or when you sort of make maybe like a historical connection." Participant B

The experience of CT when prompted by fictional material in this way was not perceived as disrupting the pleasure or relaxation of reading, but rather was experienced as part of it: "it's very much kind of like a game" (Participant G). Thus for these participants fiction could prompt CT via the pleasure of the reading experience, with no need for participants to have entered into that experience looking to think critically. In the case of nonfiction, there was a sense that the tone could cast the text as a source of amusement, similar to the fiction reading for relaxation. In this way, the tone of particular kinds of nonfiction text could also prompt CT in this circuitous manner:

"I was also just reading a book by A.A. Gill who's a journalist, and it's a bit tongue-in-cheek. How much it's tongue-in-cheek I'm not quite sure. It's a study of the English, I quite like those sort of social histories about British people and our weird quirks. There's some things in it I found quite funny, because he's very critical of the Cotswolds, which is where my family are from. That sort of thing is quite interesting because it's someone writing who to a certain extent has shared experiences, but also taking quite a critical, quite a tongue-in-cheek, quite a ranty, prejudiced view. It reminds me that we all have these prejudices. He goes on a rant about the National Trust – and who can hate the National Trust? – but then maybe he makes some valid points about land ownership." Participant B

The lack of direct confrontation in reading was seen as a key element of this circuitous CT prompting, suggesting the emotional valence of the reading experience is different in an important way:

“So by reading more, I think that’s a good way to build up discomfort, I mean build up resistance, like tolerance of it. Maybe it’s a safe way of doing it, and non-aggressive way. Because like, in the pub if somebody says something you disagree with you might want to avoid an aggressive loud argument, which would be extremely uncomfortable. You don’t want confrontation, normally. But if you’re reading a book and you disagree with the author’s argument, well that’s not nearly as uncomfortable, isn’t it.” Participant J

Therefore participants expressed a sense of *reading prompts CT circuitously*, with that circuitry bypassing defences, prior assumptions, discomfort and negative emotion, offering challenge in a “more sideways way” (Participant N).

Although there were nonfiction cases, *reading prompts CT circuitously* was most overtly discussed by participants in terms of fiction, where the indirect nature was especially valued:

“And I think the Discworld is a really good vehicle to deliver those observations in. Because if it’s set in the real world, it will probably be a lot more controversial. I think the humour is a key part of it as well - and you wouldn't be able to make it as funny. And you also wouldn't have the joy of reading it and being like, ‘oh, okay, that's what he's paralleling here’, because it'd be too obviously delivered. It's got all that layer of subtlety where, you know, he's parodying something in the real world, but you work out what it is as you read [...] I think if something makes you laugh, you're a bit more open to it. Rather than if I was reading this and I was someone who maybe didn't have those values, I think I'd be less inclined to get defensive

reading Pratchett, than reading a book set in the real world that clearly attacks the things I hold close to my heart.” Participant G

This sense that fiction reading could perhaps prompt CT by bypassing defences was echoed by other participants. The indirectness of the CT prompt in the fictional material was emphasised; as Participant G articulated in describing Pratchett’s works, the portrayal of ‘real world’ issues was described as being veiled and transposed into unreal fictional worlds. This transposition added a layer of complexity for thinking about the issues as they needed decoding, but also seemed to reduce participants’ resistance. The unique ways that fiction could indirectly prompt CT were explored further by participants, in suggestions that the emotional aspects of fiction could prompt a stronger response than cold fact (e.g. Participant M suggesting a dystopian novel on climate change may be more effective than the factual reporting), and suggestions that fiction may make complex subject matter more accessible (e.g. Participant J suggesting *Animal Farm* may promote more digestible thinking prompts than a politics textbook).

Finally, different kinds of texts were also identified by participants as having meta-prompts for CT, in the sense that the quality of the writing was in and of itself something to be thought critically about. In the case of nonfiction, the quality of the writing was analysed with reference to specific criteria (e.g. Participant C stating set markers of quality in academic writing). In the case of fiction, the artfulness and literary style of the text was something to be analysed, with an emphasis on the author’s choices in utilising a particular style (for example Participant N seeking to interpret a highly abstract short story, or Participant M critiquing the suitability of a written style

for a young audience). Overall, participants found many ways in which *reading material contains CT prompts*.

Fiction provides understanding of the real world

Although there were subtle differences in participants' experiences of fiction and nonfiction as detailed in the prior categories, there were also ways in which fiction reading was described that were entirely divergent from descriptions of nonfiction, and these were encapsulated in the codes of *Illuminating other life experiences* and *Source of social or historical information* and its sub-code of *issues around veracity prompt CT*. Fiction reading was experienced as providing a unique means of accessing alternate life experiences, and also providing special access to understandings of different social circumstances in the present or in history. The relationship of fiction to fact when it came to this social and historical portrayal was identified as a unique type of prompt for CT.

Participants expressed many ways in which fiction had a unique role in *illuminating other life experiences*, obtaining insights into the lives of others that they would not otherwise have direct access to: "I'll never know what it's like to be a gay man in Afghanistan, that will never be an experience I have, but it doesn't mean I couldn't read about it, and try and think about what might that be like." (Participant B). The way fiction did not simply describe another possible lived experience, but delivered access to that experience, was articulated as being what gave fiction this unique power: "I'd say if you read fiction, as you go along with character development in fictional works, critically you might see human behaviour as the sum of all its parts, because you've witnessed that when you read fiction." (Participant D). The process

that fiction takes the reader through, extending over time through another lived experience, was contrasted by participants against the isolated time-slice more typically presented in nonfiction:

“Because when you see news stories, it's always the aftermath. It's always a revision of what's happened and it's never as it's happening. Whereas a lot of her stories they're talking about it as if it's happening now, it's in the present it's current. And they also really look at the – so for example Room really looks at, you know, the experiences that the captives have, and it is told through the eyes of a five year old child and it is very good at kind of showing like, how children don't understand, and how children can be, like it's normal for them if they've grown up in a life of captivity, you know, so that when they finally become free, it's more of a – you get more of a human story, you get more of the kind of the struggle behind it. And the worries that both the child and the parents have both when they're being captive and when they're being released and all this kind of stuff that you don't necessarily get from an interview with the victim afterwards, or you don't get that from the news reporter standing outside the house giving you the facts of what has happened in their words. And so it definitely gives you more of it, I think a sympathetic view towards the victim while it's happening, instead of the aftermath.” (Participant M)

Thus the power of fiction in *illuminating other life experiences* was not only associated with the descriptive facets of fiction, but deeply integral to the experiential process of reading fiction across the time-course of the narrative.

Fiction was also identified as a *source of social or historical information*, presented in a relatable manner:

“certainly books set in the north of England, books about the time of the miners’ strike, or going back to the 1930s, the Durham miners,

you get fiction books around that time and I can relate to it, and I can draw and identify with the intensity of the feeling. So there are social implications in that. So fiction can frequently bring out, and indeed seeks to educate, with social views and opinions.” (Participant A)

This is echoed by participants descriptions of their engagement with and learning about many different historical or contemporary social contexts, and their descriptions of their CT about that learning (e.g. the reflections of: Participant C on post WWII Russia in *A Gentleman in Moscow*; Participant G on police culture from the novels of Ian Rankin; Participant M on Emma Donahue’s portrayal of the Irish typhoid epidemic in contrast with COVID-19 today). In all of these cases, the role of fiction as a *source of social or historical information* was identified as different to the ways that nonfiction could convey this information, in terms of richness, accessibility, and relatability of the information when presented through fiction. Participants described the immediacy of the fictional reading experience as the primary driver of its power.

Often the *illuminating other life experiences*, and being a *source of social or historical information* were connected, and furthermore participants also integrated them into their own self-reflections:

“there’s one of these books by Isabel Colegate where there’s a politician, he commits suicide when he finds out his wife has had an affair. I’m sure that does happen now, but it seems quite shocking, because he didn’t want that loss of face in public, for people to know his wife had had an affair with a younger man. Would that be such an issue now? I don’t know, necessarily, but it seemed quite shocking. I think it’s that sort of thing when it makes me really challenge, just because I don’t think less of that man just because his wife had an affair. But to that person, in that time, it meant something. It meant a lot more. I think it’s good to be challenged in those ways, and think

about our own preconceptions. Just because I make assumptions about something, or would view something in a certain way, other people wouldn't – and I think literature always really brings that home. It's a story, but it's probably come from a real-life parallel." Participant B

Here, Participant B describes gaining a window into a very different life experience, striving to understand the social context that shaped that experience, and also comparing and contrasting that with self-reflection on their own reaction. Fiction reading seemed uniquely associated for participants with this interconnectedness of social and historical information, alongside insight into other lived experiences, and this prompting of personal and self-reflective CT. As such, fictional material was identified as offering diverse avenues for developing deep understandings.

Furthermore, participants reflected on the legitimacy of what they learned from fiction and thought critically about the complexity of its relationship to factual information; *issues around veracity prompt CT*. The issue of veracity and authenticity was raised in terms of the life experiences fiction could provide access to:

"And what I think is a really interesting question, is about the legitimacy of an author's work. So should authors only write about things that they have direct experience of on some level? Is it legitimate if I were to write a story about an Aboriginal Australian? Could I put myself in their shoes and write a novel? Would that be inappropriate? Are stories about, yes just telling stories, but what role do they play in people's understanding of the world?" Participant B

And similar issues were raised in terms of the way fiction might present 'real world' events or figures (e.g. Participant F's discomfort at a fictional portrayal of Avicenna felt to be inaccurate, Participant G's view on Irish history being simplified in fiction). What is noteworthy about these reflections is that demonstrate a further way in which

participants thought critically about their fiction reading, by questioning the relationship of the fictional account to reality, and further by contemplating the relationship between fictional and nonfictional veracity.

Overall, participants expressed ways that *fiction provides understanding of the real world*, which were obtained in a richer, more experientially immediate and time-extended fashion than nonfictional information about the world. Furthermore, these understandings were inherently associated with self-reflection and with complex issues around veracity that drove further CT.

Nonfiction provides direct information

When it came to talking about their nonfiction reading specifically, participants expressed a unique direct access to information, as encompassed in the codes *Reading to gain information* and *Explicit opinions and questions directly prompt CT*.

Participants described their choice of nonfiction reading as a means of obtaining information they felt they needed in order to be able to think critically. They also experienced a directness in their nonfiction reading, with very clearly stated viewpoints and explicit questions being raised and addressed in nonfictional texts, that they identified as a direct prompt for CT.

The motivation for engaging in nonfiction reading was primarily *reading to gain information*. The information presented in nonfiction texts was felt to be both useful, and enjoyable to read, and there was also a strong sense of empowerment as a result; “It’s about educating yourself. You can have a background, go a crap school and that, but you can read” (Participant L). This empowerment was articulated by participants identifying knowledge building and expansion as the key benefit of nonfiction reading,

and furthermore this knowledge building was seen as the foundation upon which to think critically:

“I don’t think you can realistically think critically about a topic, if you aren’t well-informed about that topic. So the more you read about the topic, the more you build your knowledgebase, and the more you then have to work with. So in that sense the more reading you do, the more knowledgeable you become, the more the basis you have for your critical thinking.” Participant J

Therefore while participants were motivated in *reading to gain information*, the information was not in and of itself the goal, rather that gain then enabled CT engagement.

The means by which nonfiction delivered information was experienced by participants as key, as *explicit opinions and questions directly prompt CT*. Participants identified the way that nonfiction author’s opinions are typically explicitly stated as a prompt for directly thinking critically about them. Participants often highlighted how the author’s pre-existing beliefs are typically more explicit in the case of nonfiction than in fiction, and felt this was a key difference in how the different types of reading material could prompt CT in different ways (e.g. Participant G’s set of expectations around a historian as an author). Participants also identified the way that some nonfiction authors share the sources from which their opinions have been derived as a tool for evaluating those opinions (e.g. Participant A noting the value of reference lists in academic books). Finally, the way nonfiction posed questions directly to the reader was also identified as a key way that *explicit opinions and questions directly prompt CT*:

“I think I might recommend that one. *Sapiens*. It makes you question a lot of things, the history of man, our place in the universe, our belief systems, the things that we believe in, are they real? Are they tangible? No, they're not. Why do we believe when we can't even see them? Why do we believe in them so much? Yeah. It's a fascinating read, because it covers lots of different things. I don't know if I came out with any conclusions. But yeah, very, very good read.” (Participant E)

Here, Participant E recommends *Sapiens* by Yuval Noah Harare as a book to promote CT, specifically due its questioning content (Participant F made the same recommendation). This may seem contrary to *reading to gain information*, as here participants are reading for questions rather than answers. However, the ways participants described their experiences of *explicit opinions and questions directly prompt CT* encompassed a clarity of information on the authors and their views, including the clarity of questions where answers may not be known, and thus the informational content of the nonfiction reading entailed these more meta-information gains.

Whether in terms of information that can be used to build knowledge and create a secure base for CT, or information about authors' positions, and explicit questioning, participants emphasised how *nonfiction provides direct information* and valued that clarity.

Narrative transportation

A single category emerged from participants descriptions of their experiences of being transported, and their thoughts on how transportation arises for them, when reading.

Narrative Transportation integrates CT into reading

Prior to conducting these interviews, I thought of NT entirely in terms of the experience of reading. However, in participants' responses I found that they experienced transportation into a text as an integration of their thinking and evaluating into their reading experience. This was encompassed by the codes *Narrative transportation is active engagement* and *Narrative transportation is connection-making*. Thus, the experience of NT was both an experience of CT and of reading.

Firstly, *narrative transportation is active engagement* was characterised as a highly active experience, not as passive. For instance, Participant C identified times of day "when I suppose I can really be creative, my mind's creative" as the best times to read, in order to be transported. Similarly, Participant N described finding it hard to sleep if reading a very immersive text before bed due that active engagement. This active transportation was also often linked with problem-solving:

"That will be the absorption, time-consuming, time goes through very quickly. If I'm looking at and solving - I'm part reading a book on chess which is chess problems - how do you get out of this one, how do you get check mate, that one I can absorb and the time just flows by." Participant A

Indeed, thinking critically about the contents of a text was not seen as breaking NT, but instead supporting it:

"I think it's intrinsic. I think you're doing that during the process of reading. Otherwise, I don't think you would feel as much when reading. You wouldn't be so invested in the book if you weren't making these judgments and thought processes while you read. You do it at the time that you are reading." Participant D

Thus, *narrative transportation is active engagement* was articulated as part of the pleasure of the reading experience, and a driving force for maintaining attention and immersion into the reading, with participants describing how thinking and evaluating “gathered pace” (Participant E) during reading and helped “get into more depth” (Participant J).

A specific active experience of CT in reading was identified in *narrative transportation is connection-making*, with participants describing a plethora of associations and ways of integrating different information and experiences during reading. Connection making took place within a text, as participants described integrating what they gained from reading pieces of the text so as to make the reading experience “connect up as a whole” (Participant J). Many participants went further, building connections beyond the material of the text itself (e.g. Participant K imagining additional characters in the story line of a novel). A prevalent form of connection making lay in the experience of being transported into or identifying with a character, where connections were made between the character and the participants’ self-reflection upon their own life:

“So, obviously very different to my life, but also there’s things I’m quite interested in and can relate to, about women’s careers, having partners and marriage and that sort of thing - being in my early 30s these things are quite relevant, and things that my friends and I are kind of a bit conscious of. So, yeah, there was some sense of sort of identifying with her role as like: ‘Oh, I’m now married and people might think it’s weird I still have a career’ - and ok it wouldn’t be quite the same now, but they’re still the sort of things I’m aware of. There’s some sort of sense of maybe not quite identifying, but

definitely things that prompted me to think about my experiences.”

Participant B

This way of connecting one’s own experiences and concerns to a character, and then using that connection to drive reflective CT, was a common experience especially in fiction reading. Another prevalent form of connection-making was between the setting of the book and the participants’ own physical location or past experience with similar geographic locations. This location connection-making was most prevalent in fiction reading (e.g. Participant F connecting the portrayal of Paris to own experiences; Participant C connecting Wordsworth to own life in the Lake District). This connection of geographic experience was part of a wider trend of connecting the experiences from reading into one’s own life experiences, which was the case in both fiction and nonfiction reading:

“Like, with books on tech, that’s maybe the best example. When they described how decisions got made, and how something got developed the way it did, I can relate – because that’s the field I work in too, I’ve been in those rooms, I’ve had those conversations – I can imagine how it took place. I guess I can imagine being in the room.”

Participant L

Thus, *narrative transportation is connection-making* encompassed connections with self-reflection, with prior experiences, and with practical knowledge. Connections were also made to imagined future scenarios:

“I’ll make comparisons to things I’ve read before. And I’ll be trying to look at the, judging on the relevancy – maybe to the more practical gardening I’ll be doing in the future, and whether that’s something that would influence how I would act in that situation. So I’ll be making judgments.” Participant N

Taking the scope of *narrative transportation is connection-making* beyond association with memories, into a far more active and wide-ranging realm. Finally, the connection-making experience of NT was also described in a multi-layered manner, both as the conscious connecting of different pieces of information, and as the deeper connection-making taking place in one's thinking process:

It's interesting because I almost think reading is like the real time example of almost how your own brain connects, because then I would, you know, get a quarter of the way through and I think 'oh well this reiterates findings from another study' or 'oh, well, that seems to contrast what we've read about before'. Participant D

This emphasises how *narrative transportation is connection-making* integrates reading and CT experiences such that they cannot be disentangled.

Participants experiences of NT in their reading were not separable from their experiences of evaluating and making judgements during their reading, nor from their broader thinking experiences. As such, for participants *narrative transportation is active engagement* with a problem-solving, concentrated thinking process throughout reading. Furthermore, participants articulated wide-ranging ways that *narrative transportation is connection-making*, suggesting a constant integration and association between what is read, participants' own experiences and memories, and further imaginative connections to the future and the self. Taken together, these experiences of NT can be characterised in the category *narrative transportation integrates CT into reading*.

6.4. Discussion

6.4.1. Themes

Three themes emerged across the analysis of the data, and these will be presented here to bring out the overarching findings from the analysis.

In talking about CT, participants characterised it both in terms of a procedure, and in terms of the experience of the CT process. This is seen in the distinction between the categories *Critical thinking is an experiential process* and *Critical thinking is procedural not personal* (see Table 45). It is important to note that these are not mutually exclusive, such that CT is either a procedure or an experienced process; it can be both. In the first two themes I have drawn from the categories, I maintain the distinction between these two ways of characterising CT; in the final theme I collapse this distinction into a view of CT as described by participants overall.

Fiction gives experiential CT understanding

There is a striking similarity in the way that fiction reading was described as providing understandings of the real world through an immediate, richly experienced and time-extended process, and the way CT was characterised as an experiential process. Bringing these characterisations together allows for a theme to emerge that indicates the unique relationship between the experience of fiction reading and that of CT.

Firstly, the slow and necessarily time-extended nature of the CT process is mirrored in the slow and progressive way that fiction reading delivered understandings. While reading of all forms is inherently an activity that takes place over time, participants identified the way that fiction expanded other lived experiences and

narratives across a longer time span such that they could be intimately and gradually experienced, in contrast to the post-facto mode of reporting events more common in nonfiction. Thus, reading nonfiction may take a long and slow process, but the content is likely to be a more clearly defined time-slice; conversely, reading fiction not only in and of itself entails a time-extended experience, the content contains its own time-course that must be gradually traversed. This can also be seen in participants' experiences of CT, which were characterised by drawn-out, ever-shifting experiences of thought, reflection, and change. Thus, both fiction reading and CT experiences can be thought of as journeys, with gradual narrative developments, which must be experienced as such and cannot be reduced to simpler time-slices.

The experiential processes of both fiction reading and CT were inherently reflective for participants, with the reflection integral to both experiences and not separable from them. Again extending over the whole journey of fiction/CT, reflective insights were gradual and building, described by participants as cumulating into personally profound shifts of perspective and understanding.

The imaginative inference-making of CT was described analogously to the imaginative engagement with fictional lived experiences and contexts. Just as while thinking critically participants inferred beyond the limits of the information at hand, in reading fiction imaginative extrapolation was key to the vigour and primacy of the insights gained. In both fiction reading and CT, imaginative filling-in was essential to the experience, and to the understandings that it could yield. Here, the special role of fiction in *prompting CT circuitously* (see Table 45) can also add insight; by triggering imaginative inference making as an enjoyable process, fiction was also described as

uniquely able to bypass defences and biases and thus shift understandings even in ossified domains.

The understandings developed through fiction reading were complex, driven by the unique issues of veracity entailed in fiction, where truth was construed as more multi-layered than a direct mirroring of reality. Similarly, concerns for *authors as sources* (see Table 45) were more complex when it came to fiction, as reliability was not construed merely in terms of credentials but rather a rich, deep understanding of the author's position in relation to their fictional subject matter. Furthermore, when making connections between fictional material and themselves, the ways fiction could prompt reflection and change were similarly convoluted and indirect. Thus, CT was experienced in an ongoing process seeking fictional truths, which was distinctive from the goal-oriented procedure of CT in nonfiction that will be characterised in the following theme.

The time-extended, reflective, and imaginative aspects of the fiction reading and CT experiences were all interwoven, such that they can be seen as one path leading towards new understandings. By traversing the journey along that path, fiction reading and CT give experiential understanding.

Nonfiction builds informed CT procedure

Just as there were parallels between fiction reading and the experience of CT, so there was alignment between nonfiction reading and the procedure of CT. When discussed as a procedure (rather than as an experience) CT was described as highly goal-oriented and impersonal, and was characterised by its straightforwardness. The way participants described this very clear and direct procedure of thinking is very close

the way they described the directness of information and opinion as presented in nonfiction. Indeed, the reliability of nonfiction sources, and the reliability of CT as a procedure, were both attributed to their concision and explicitness.

In both nonfiction reading and procedural CT approaches, there was a strong emphasis on filtering information such that only kernels most relevant to one's goal could be extracted and considered in isolation (see Table 45). Both nonfiction reading and CT procedure were thus detail-oriented, reductive activities. Furthermore, when brought together nonfiction reading and the procedure of CT form a cycle of information usage: CT procedure is used to isolate relevant information from a text; this information is then integrated into one's knowledgebase; subsequently one is further empowered by that strengthened knowledgebase to fine-tune their future applications of CT procedure.

In the procedure of CT, participants articulated a desire to set aside their emotions and conduct their thinking objectively to as great an extent as possible. In line with this, the clarity of authors' positions was valued in nonfiction as it offered the possibility for such separation to be inferred between opinion and fact. The value participants placed on the transparency of the set-by-step logic of the CT process is furthermore echoed in their valuing of the transparency in the way arguments are typically presented in works of nonfiction. Likewise, the role of explicit questions posed in works of nonfiction as prompts for CT also melds with the process of CT, as a direct starting point for that process.

Participants furthermore characterised the changes in their thinking and behaviour resulting from nonfiction reading and their CT around it in very literal and

direct terms, for example taking a dietary supplement on the basis of recommendations read. This is in sharp contrast to the way fiction prompted different imagined behaviours, rather than literal action. This further emphasises the very explicit and goal-directed nature of the change sought.

Nonfiction reading and the process of CT were for participants characterised by their divisibility into discrete components, and explicit directness. Reading and thinking formed a cyclical procedure of knowledge and technique building up from nonfiction reading to an informed CT approach.

Reading as personal reimagining

Firstly, the ways in which participants described reading material as containing prompts for CT, and the ways in which they described NT as active engagement with a text (see Table 45), can be brought together to elucidate the ways reading prompted thought and change. The active nature of NT was described as a focused, creative mode of attention and thought. This highly active state was identified as a driving force for maintaining focus and generating the pleasure in reading, which built up over time and enabled deep engagement. This can be interpreted as the driving force through which the ways participants' described reading as mind-broadening could come into play. Participants expressed many ways that their reading material could prompt changes in their thinking, by challenging their habitual thought patterns and taking them beyond their comfort zone. The active, creative mode of NT was described in similar transformative and boundary-breaking ways. Furthermore, the circuitous way that reading material could deliver these prompts to CT, such that they bypassed defences and triggered thinking in subtle ways, was described by participants as

supporting rather than breaking their NT engagement such that instead of being jarred out of their transportation by surprising or disconfirmatory ideas, they were in fact drawn deeper into their immersion precisely through the active engagement these CT prompts generated.

Secondly, the ways in which reading was described as shifting modes of thinking can be brought together with the ways that NT was characterised as connection-making. Participants described profound changes in their self-reflections and behaviours as a result of reading, both in the immediate and long term. These changes pertained to explicit beliefs and opinions, as well as to implicit patterns and ways of thinking. The ways in which participants described reading as triggering these changes were highly analogous to the ways in which participants described NT as a connection-making experience, as the connections being made were also both immediate and wide-ranging, explicit and implicit. Furthermore, participants described adopting structures of thought from what they read into their own ways of thinking, and adopting mindsets from their reading, such that they changed the ways they navigated through thought processes. The ways in which participants described connection-making between their reading and themselves in the NT experience suggest that NT was the pathway through which participants' connected to these different structures of thought.

The changes promoted by reading were often self-reflective, with participants describing reading experiences prompting reflections on themselves as they are, as well as reimagining how they might be in other scenarios, or in the future; again, this is strongly echoed in the ways participants described connection-making NT as a reflective process of integrating what was read within themselves, especially in the case

of connecting with a character and integrating that character's experiences with one's own. Both involved future imagination, as reading triggered different ways of conceptualising futures, and NT connected possible futures to participants' existing experiences. Reading was described as prompting both direct changes in behaviour, as well as more indirect changes in the ways participants modelled behaving reflectively, and NT connection-making was likewise multi-layered with both conscious connections and more implicit connections involved. NT as connection-making was inherently a critical process, involving evaluation and judgement making throughout, and thereby integrated CT into reading. Thus, bringing together the experiences of reading triggering shifts in modes of thinking, and NT as connection-making, reading and CT can be integrated into process of integrative self-change.

Additionally, participants identified personal characteristics as the grounds from which CT could take place (both as a procedure, and in its experience). Firstly, participants' EO was described as the underpinning characteristic of their CT, but also as something that could be shifted and changed. Participants articulated changes in EO taking place over their lifetimes as changing their CT approaches, and also cyclically their CT changing their EO, and this could be different within different domains. In the descriptions given of ways that reading shifted participants' modes of thinking, there is ample suggestion of reading shifting EO, especially when it came to the self-reflective shifts that participants described from their reading experiences. Furthermore, the circuitous means by which reading could prompt CT was suggestive of a route for bypassing certain EO (i.e. absolutist positions) that might otherwise have limited CT engagement. Likewise, the openness and curiosity that participants identified as crucial

personal characteristics necessary for CT are mirrored in the way participants felt circuitous CT prompting could play a special role in prompting CT from reading; these indirect ways reading could trigger CT were seen as a way of bypassing defences and thus creating openness where otherwise there may have no window into CT. Therefore reading as a means of shifting modes of thinking can be seen as a means of altering personal characteristics towards CT, and also as a means of circumnavigating personal characteristics that may not be conducive to CT.

Taken in total, these findings suggest that NT is an actively engaged connection-making means of integrating CT into the reading experience, and acts as a driver for reflection and ultimately for self-change. Reading can be a process of personal reimagining, in which connections are actively made with one's pre-existing personal characteristics, and modes of thinking, such that these can be shifted and transmuted. These changes then feed into the ways in which we think critically.

6.4.2. Relation to literature review

Critical thinking

Participants described CT both as a procedure, and as an experiential process, from which I have derived different categories and themes to capture this distinction. This is an interesting finding when compared to ways CT is described in the literature, where it is typically a unified construct rather than being divided in this way. A dual-processing model (Metzger, 2007; Stanovich & Stanovich, 2010) could be seen as making such a distinction in ways of approaching CT, but in these models the two forms (type 1 autonomous and rapid; type 2 conscious and slow) form a clear dichotomy; by contrast, the experience and procedure of CT are overlapping and not

mutually exclusive. Therefore this way of conceptualising different aspects of CT appears to be novel. Furthermore, in the current CT literature the emphasis is very much on the procedure of CT (as seen in R. H. Ennis' list-based definitions, for example). Even authors who extend the conception of CT beyond obviously procedural concerns, do not necessarily discuss the experience of undertaking CT; for example Paul (1987) argues for the importance of imagining other's ways of thinking, but does not consider how that imaginative process is experienced by the critical thinker. Barbara Thayer-Bacon's (1992) emphasis on empathy and embodiment perhaps comes closest to an experiential conceptualisation of CT, but nonetheless she does not explore all the different elements that come into play in a time-extended experiential journey through CT. Further research on the experience of CT could therefore expand current conceptions of the construct.

Distinguishing between CT dispositions and abilities is a common approach in prior CT research (e.g. R. H. Ennis, 2015; Facione, 1990). The findings from this study support this division to some extent, as the category *CT is a matter of personal characteristics* implies traits that are conducive to CT, which could include the kinds of dispositions typically stipulated, in contrast to *CT is procedural not personal* and *CT is an experimental process* which describe the application of CT. However, the way that Ennis, Facione and others characterise CT dispositions and abilities is far more simplistic than the ways in which participants in this study described the personal characteristics, procedures, and experiences of CT. For example, being open-minded and being informed are dispositions that appear in both Ennis and Facione's lists, and were also identified by participants in this study as important grounding for CT.

However, the relationship between both of these dispositions and the ability to engage in CT is essentially treated as uni-directional by Ennis and Facione, with the former considered to be prerequisites to the latter. Yet participants in this study described a far more inter-mingled and cyclical process. The theme *nonfiction builds informed procedure* entails a mutually building process between gaining a stronger knowledge base and the development of a robust CT procedure, which are inherently interrelated. The theme *reading as personal reimagining* characterises personal attributes such as openness as feeding into reading and CT processes, and being actively and imaginatively reshaped through them. Therefore while it is possible to draw out attributes and abilities from the findings of this study, placing them in dichotomised lists would not be appropriate.

By contrast, there was strong alignment between the findings of this study and the prior CT literature when it came to the primacy of reflection. Being able to reflect upon one's pre-existing beliefs and biases, and upon one's CT approaches, is held to be crucial across the CT literature (e.g. Bailin et al., 1999; R. H. Ennis, 2015; Kuhn, 1991; Thayer-Bacon, 2000). Indeed, this is predicated in the origins of the CT movement in Dewey's (1910) conception of 'reflective thinking'. Therefore the ways that participants in my study described a need for reflection throughout their characterisations of CT is in keeping with the prior literature. Likewise, openness is a common feature of most CT conceptions (e.g. R. H. Ennis, 2015; Facione, 1990; Kuhn, 1991; Siegel, 1997), and therefore the need for openness as a personal characteristic conducive to CT expressed by participants in this study is also unsurprising. Interestingly, participants in this study often went beyond openness and

suggested non-conformism as a more radical predicate for CT, in keeping with bell hooks aims in critical thinking instruction as leading to “radical openness” (hooks, 1994, p. 202).

Narrative transportation

NT is typically construed as an experience of leaving the real world behind and entering fully into an immersive experience of a text (e.g. Gerrig, 1993; Green & Brock, 2013). While authors such as Green and Brock do characterise NT as an active process, they suggest it is not an effortful one, and certainly not consciously controlled. While participants in this study did experience NT as an immersion into a text, they did not describe this as an exit out of the real world or away from their own lives, rather the code *narrative transportation is connection-making* (see Table 45) highlights how the real world and participants’ memories were interconnected with the text as part of the NT experience. Furthermore, the deeply reflective characterisation of this connection-making suggests a consciously active process, with participants aware and autonomous in their connection-making. Therefore NT as described by the participants of this study is not an experience of disconnection, but in fact the very opposite.

An even more striking difference can be seen in the ways that participants construed the relationship between NT and CT in this study, in contrast to some of the prior literature. It has been suggested that NT and CT are mutually exclusive, and that NT in fact inhibits CT (Busselle & Bilandzic, 2008). By contrast, participants in my study articulated an inherent integration of CT into the experience of NT, as characterised by the category *narrative transportation integrates CT into reading* (see

Table 45). Therefore the findings of this study are more in line with those that do support an integration of CT into the NT experience (e.g. Hoeken & Fikkers, 2014). However, studies addressing NT and CT typically consider only CT around the topic of the narrative (e.g. Mazzocco et al., 2010). In this study, participants described a connection-making evaluative experience as part of their NT that often went beyond the direct topic of the text, with reflective and future-imagining facets as well as integration into participants' ways of thinking. Therefore the findings of this study support the view that CT and NT are interconnected and not mutually exclusive, and furthermore broaden the scope of CT within the NT experience.

Epistemological orientation

The prior literature on EO considers this to be a crucial factor for CT, and specifically construes evaluativism as the orientation most conducive to, or even necessary for, thinking critically (e.g. Kuhn, 1991; Siegel, 1997). Multiplism and absolutism are typically argued to be barriers to CT, as the former implies relativism and thus a rejection of the possibility of evaluating the quality of different views (Siegel, 1997), and the latter implies a limitation in engagement with differing views (Kuhn et al., 2000). However, participants in this study expressed pathways to CT from each orientation, as seen in the code *EO as a personal approach to CT* (see Table 45). It was not the case that only evaluativist orientation could be the grounds for CT, but rather each orientation offered a different grounding for different streams of thinking critically.

Furthermore, EO is often mapped to a developmental trajectory, typically cumulating in a stable evaluativist EO in adulthood (Hofer & Pintrich, 1997). Though

some adults have been found to be absolutist or multiplist, these form a minority in prior studies (Kuhn et al., 2000). This study offers a more complex picture of EO in adults, as participants expressed far more mixed and complex orientations than could be easily placed into a single EO category. Furthermore, they expressed experiences of change in EO that took place in their adult lives, as a result of their reading and CT experiences, as seen in the reciprocal changes to thinking characterised in the theme *reading as personal reimagining*. Thus the findings of this study suggest EO is a more fluid basis for CT. Although Kuhn et al. (2000) did find different EO for different domains (e.g. multiplism for judgements of value or taste, alongside evaluativism for judgements of fact), they nonetheless suggest broadly stable groupings of EO by domain. Participants in this study described a more subtle way that they could deploy different orientations to ground their CT within different facets of any given issue. Therefore the findings from this study complicate the notion of fixed EO, and integrate all orientations into CT on all issues.

Impacts of fiction reading

Based upon my review of the literature, I considered the way fiction gives access to different perspectives (e.g. Hakemulder, 2008) and engages empathy and ToM (e.g. Mar et al., 2006) as a key way in which may build CT, given the emphasis placed on the need to engage in other's perspectives for CT (e.g. Paul, 1987; Thayer-Bacon, 1992). This role played by fiction was certainly present in participants' descriptions of their reading experiences (e.g. Participant B's comment that even if we cannot know what it is like to be a gay man in Afghanistan we can learn something about that experience through reading). The need to engage with different perspectives

for CT was also frequently discussed by participants (e.g. Participant D describing a need to understand the perspective of people with anti-vaccine positions). These were connected in the category *reading shifts modes of thinking*, particularly through the code *reading as widening exposure* (see Table 45); participants emphasised the need for wide reading in order to garner exposure to different perspectives, as a means of building CT. However, this was only one channel which participant's identified as linking reading and CT, and it was not perhaps as dominant as I had anticipated. Instead of this being a primary connection, it was instead one thread in a far more complex interweaving of reading and CT. The connections participants made within their reading experience were far broader than only the connection to authors or figures in the text, and thus the engagement with other perspectives was part of a wider interconnective process. Subsequently the changes wrought on participants modes of thinking were also broader and more complex, as described in the theme *reading as personal reimagining*. Therefore the findings of this study support the link between reading and CT via engagement with different perspectives, but do place greater weight on this connection in comparison with all of the other ways participants integrated their reading and CT.

Another way in which I argued from the literature that reading could impact CT was through the training of counterfactual thinking. Counterfactual reasoning is the ability to imagine alternative scenarios, both in the past and future, with recourse to our episodic memories (Byrne, 2016), and as such I have argued it is an important factor in CT. Furthermore, studies point to reading fiction as an avenue for increasing our counterfactual reasoning capacity (Black et al., 2018), and based upon this I

reasoned that the alternative scenarios presented during reading (particularly in fiction) could build our capacity for imagining counterfactuals and thus increase our CT capacity. This line of reasoning is supported by the findings of this study, although as with modelling different viewpoints counterfactual thinking was not given a dominant weighting by participants describing their reading and CT experiences. Descriptions of counterfactual thinking primarily fell under the code *inference making and imagining* (see Table 45), in which participants expressed a need to imaginatively extrapolate in their CT, so as to consider wider possibilities. This was closely associated with fiction reading, as the time extended journey through a fictional narrative provided deep engagement with alternative situations so that they could be absorbed and understood. As such, counterfactual thinking was a component of the theme *fiction gives experiential understanding*. Furthermore, counterfactual thinking can also be seen in the theme *reading as personal reimagining*, as counterfactuals are both something inferred from scenarios read, and reflectively applied to oneself through future imagining placing oneself in such scenarios (as was frequently described by participants). Thus, counterfactual reasoning formed an element across themes, and was associated strongly though not exclusively with fiction reading.

Finally, prior literature has suggested the power of fiction may lie at least in part in its ability to bypass our defences (Bal & Veltkamp, 2013; Djikic et al., 2009a). Based upon this, it could be supposed that fiction reading may help increase CT by exposing us to wider viewpoints through this bypassing of our pre-existing defensive reactions or confirmation biases (Nickerson, 1998). This supposition was supported by the findings of this study, and indeed given far greater emphasis than I had anticipated. This was the

basis of the code *reading prompts CT circuitously*, within the category *reading material contains CT prompts*; as such, this ability of pleasurable reading to circumnavigate pre-existing beliefs and barriers was inherently tied into CT in participants' descriptions. It should be noted, however, that participants did not only experience this circuitous influence when reading fiction, but also volunteered nonfiction examples that had this same property. However, in the case of fiction this circuitousness was more prevalent, and more wide ranging, suggesting that while all reading may have some ability to bypass our defences in fiction this is amplified.

6.4.3. Limitations and future directions

This study took place during the COVID-19 pandemic, and was impacted by restrictions arising from emergency public health measures. My initial aim was to conduct in-person focus group interviews with book clubs and reading groups. Meeting groups in person was not possible in the spring/summer of 2020, and therefore I initially revised my plans to be through online methods. However, recruiting reading groups and book clubs online proved unfeasible; interested participants who responded to calls for participation found that their usual reading groups were unable to continue online due to increased care responsibilities, or technological limitations. Subsequently, I revised my plans further and made the decision to conduct one-to-one interviews instead. This approach was more sensitive to participants' needs, as they were able to volunteer their time individually as and when was convenient for them. As a result, the type of data I gathered differed significantly from my original aims. However, while different, I do not believe the individual interview approach has been in any way inferior to conducting focus groups.

I found that the participants who took part were enthusiastic and voluble, and speaking to them individually allowed them to express themselves in full. Conducting focus group interviews on the topics of CT and reading would be an interesting avenue to pursue in further research, as group dynamics may reveal additional insights.

It is important to note that the sample of participants recruited for this study is in no way representative of, or generalisable to, a general population. Participants were self-identified readers, and thus the perspectives of non-readers are not captured. Furthermore, participants responded to online calls for participation, thus were users of online platforms (Twitter, Reddit), or had subscribed to the research newsletter, and thus they were interested in taking part in academic research; these characteristics distinguish them from other kinds of readers who may have differing perspectives. The findings presented must therefore be taken as a reflection of this specific sample. Future research with different participant samples could add diversity and nuance to the characterisation of the reading and CT relationship described here.

Participants were explicitly asked to talk about reading and CT, and thus prompted to think of these two factors together. In the information sheet, the research project's aims in understanding the relationship between these two factors was clearly outlined. This means that participants were not spontaneously describing a relationship between their reading and CT experiences, but rather were prompted to do so. It is therefore not possible to draw any conclusions on participants' naïve experiences of any such relationship. Furthermore, I asked participants questions about fiction and nonfiction reading that prompted comparisons, and thus the distinctions drawn between the two types of reading were also not spontaneous. It should also be noted

that the lines drawn between fiction and nonfiction, and the ways they were separated out in the coding, categories, and themes drawn from the data are not absolute. Participants identified general trends between the two types of reading they experienced, and were not applying any universal rule or hard distinction between them. Thus the findings pertaining to fiction and nonfiction ought to be understood as broad patterns, not clear delineations. Further studies that discuss reading and CT without asking for any comparison of fiction and nonfiction may find no such division spontaneously made, and studies that do not expressly elicit consideration of the reading and CT relationship may not find such a relationship naturally described. Finally, reflection was a key feature in the data and played a role across codes and categories; it should be noted that in asking participants to reflect on their personal experiences of reading and CT, I may have turned their attention to more reflective modes of thinking and thus the nature of the questioning may have increased the prevalence of answers pertaining to reflection. However, for the purposes of my research aims I felt it appropriate to make these factors of interest explicit, and thus I interpret these results as framed by that prompted salience.

Lastly, it is notable that the subject matter participants used to provide examples and articulate their views around CT, and their reading, were often connected to current affairs taking place at the time of the interviews: the COVID-19 pandemic; the global environmental crisis; the Black Lives Matter movement; transgender rights and wider gender issues. I did not code for these topics, but they are present throughout the categories derived from the data. This highlights the importance of the ways we read and think critically about complex current issues. Analysis of the data with respect

to these issues fell outside of the scope of this study, but the data could be re-deployed in future to answer research questions based upon these topics.

6.5. Conclusion

6.5.1. Answering the research questions

In developing themes from the codes and categories derived from the data, some answers can be provided to the research questions that drove this study. These answers do not address the questions definitively or in full, but rather offer some initial insights and potential avenues for further research.

- a. In what ways do readers consciously experience an influence of what they read on the way they think critically?

Participants expressed a direct relationship between their reading and CT, which was active and consciously experienced. Participants had parallel experiences of moving through the fictional narrative and the CT journey, traversing these experiences of reading and CT led to shifts in ways of thinking and understanding both the world around them and themselves. By contrast, participants approached both nonfiction reading and CT from a goal-oriented, compartmentalised procedure that enabled knowledge-building and explicit verifiability, in a cyclical manner. Thus, fiction and nonfiction reading were both consciously tied to CT for participants, though in a different manner.

- b. In what ways might readers' epistemological orientations relate to their reading and CT experiences?

At the outset of this research project, I expected to find differences in CT and reading approaches between participants with different EO. However, the data gathered in these interviews with readers suggests that the interplay of reading and CT forms a unifying means of shifting one's orientation within various domains, and as such cuts through initial differences in EO. Particular emphasis was placed on the circuitous means through which reading could prompt CT and thus trigger such changes, as this seemed to circumnavigate defences and cut across all forms of EO. Furthermore, the reflective and transformative nature of CT and reading described by participants suggests a fluidity in EO, with reading and CT experiences steering changes in such foundational ways of thinking over time. EO was described as a grounding from which reading and CT experiences take place. However, reading and CT can move that grounding, and thus form a reciprocal process of self-change.

- c. How do readers' experiences of narrative transportation relate to the ways in which they think critically about a text?

At the outset of research project, I conceptualised NT as intimately tied into the experience of reading, but as distinct from that of CT. I was interested in whether engaging in CT would in fact be detrimental to NT, and would jar readers out of their flowing, immersive narrative experiences. However, the experiences of NT described by participants were contrary to these expectations, in that their NT descriptions were highly active, evaluative, connection-making, thinking processes. NT was characterised as inseparable from CT within the reading experience. Furthermore, the results portray a wider emerging picture of the ways that reading and CT wrought personal changes

within participants, and thus NT can be seen as the experiential mode within which those personal reimaginings and shifts in thinking took place.

6.5.2. Summary

This study found reading and CT to be intimately interwoven. Their relationship can be viewed through the streams of fiction and nonfiction that offer differing routes for enmeshing reading with CT, but can also be viewed in totality with overall relationships arising between reading and CT. Fiction reading was associated with the experiential process of CT, as a time-extended development of understandings. Nonfiction reading was associated with the procedure of CT, as a building of informed procedural application. Reading overall was found to be a transformative process in which NT actively and imaginatively brought CT to bear on the reading material, oneself, and wider thinking such that shifts and changes were continuously brought about. Thus, this study supports the overall argument of this research project in pointing to an important interplay of reading and CT.

7. Study four: A diary study of daily reading and critical thinking experiences

This study sought to address the ways in which the experience and evaluation of reading, and CT experiences and approaches, may be related on a day-by-day basis. The study built upon the prior quantitative studies by integrating the reading and CT relationship identified in those studies, so as to explore ways participants experienced this relationship. Furthermore, it built upon the analysis of study three by incorporating coding approaches derived from its findings, and as such offers an extension from this study rather than merely an addition to it. As this study is the final part of the mixed methods sequence forming this research project, it also sought alternative ways of understanding reading and CT so as to interrogate the validity of the previous findings. Box 9 summarises the study.

Box 9: Study four summary

This study aimed to investigate readers' daily CT and reading experiences. Self-selecting participants who self-identified as readers of either fiction or nonfiction ($N = 21$) completed a daily reading and CT diary in online format. A subset of this data was analysed ($N = 19$). A progressive coding method was applied, moving from coding each entry to coding entire diary days. This process yielded 7 codes and 3 categories. The findings showed commonality in fiction and nonfiction reading in different associations with CT, but also distinguished fiction from nonfiction through imaginative connections to CT versus more reality-grounded approaches.

Participants were asked to keep a daily diary describing these experiences for a two week period. As the way we are transported by reading can vary not only across individuals' general propensity for transportation, but also within individuals' different reading experiences at different times (Mazzocco et al., 2010), in this study the way that each participant felt transported by each day's reading was captured. Furthermore, as evaluations are made throughout the reading process (Özyürek & Trabasso, 1997), participants were asked to describe their evaluative thinking about what they read. Finally, the approach taken to, and the experience of, thinking critically was also recorded daily, to capture different instances on different subject areas across two weeks, giving room for diverse cases of CT. Participants could describe CT about reading, for example reading a news article and then thinking critically about it, but they could also describe CT independent of reading; the way in which participants described and experienced reading and CT were of interest, not the way the subject matter of reading could be thought critically about. This is an important distinction, as prior research on reading and CT is heavily focused on CT about the content of what was read (e.g. Hakemulder, 2000). In this study, it was up to participants to decide which reading and CT experiences to describe. Furthermore, this study provides a different temporal framing to study three, in which participants described reading and CT in broad terms; by asking participants to provide a daily account, a more short-term perspective is accounted for which is likely to yield different reading and thinking experiences as foregrounded for participants (Read et al., 1999). As such, this study addressed the specific daily experiences of participants, and gave space for these experiences to be described fully in as much detail as participants chose, to provide a rich characterisation of each reading and CT instance. Thus, daily entries offered a

combination of reading and CT experiences, and differences in how these were characterised across different days could be assessed. This enabled an analysis of whether there could be any distinctive ways in which reading and CT were described together, with the aim of answering the following research questions:

- a. How do day-to-day experiences of reading and thinking critically vary?
- b. How are everyday experiences of reading fiction or nonfiction related to CT approaches?

This study was pre-registered alongside the interview study; original registration is available here: <https://osf.io/3wjpr>; amendments resulting from the COVID-19 pandemic: <https://osf.io/ncp4d>.

7.1. Method

7.1.1. Participant recruitment

Participants were self-selecting, responding to calls for participation on social media (Twitter and Reddit), and via the study email newsletter. The calls for participation asked for readers or bookworms, reading fiction and/or nonfiction, and a £30 National Book Token was offered as a reward for each participant who completed the two week diary. The respondents were asked to complete a registration form in which they confirmed they were 18 years of age or older, self-identified as primarily fiction or nonfiction readers, and gave their email address; they were then added to a pool of potential participants. A target of 20 participants was aimed for in total, with an equal balance of fiction and nonfiction readers. The figure of 20 was selected as being likely to offer enough data to reach saturation, and being affordable in terms of the offered voucher rewards. From the pool, 10 fiction and 10 nonfiction readers were

drawn at random, and invited to participate in the study. Five participants drawn from the pool and invited to take part did not respond to the invitation to complete the diary, and these were replaced with others randomly drawn from the pool; however, one of these commenced and completed the diary later, yielding an additional participant to the original target. In total, 21 participants completed the study, comprised of 10 nonfiction readers and 11 fiction readers. Data collection commenced on 18/05/2020 and ended on 25/06/2020.

7.1.2. Diaries

Each participant drawn from the pool was given a unique 4 letter participant ID, sent a link to an information sheet and consent form, and upon consenting was asked for basic demographic information (highest obtained educational qualification; gender; age). The participants' ID was associated to their email address for the purposes of tracking their study completion and sending reward vouchers. Each day, participants were sent a link by email to the daily form, and asked to input their ID before answering the daily diary questions. As email addresses were deleted upon study completion, participants' responses were anonymous from then on. The daily diary form is available in Appendix E.

Each day participants were asked to think about the main thing they read, and their main experience of thinking critically that day (i.e. one example of reading, and one example of CT). They were then asked to complete the following information about these: what text they read; how they experienced transportation into the text; how they experienced evaluating the text; the topic of their CT; their prior opinion

and any change to them on this topic; their CT process; the strengths and weaknesses of their CT.

The same daily diary form was sent to participants by email each evening for the total period of 14 days. Participants could miss up to 2 days and still be considered as having completed the study. Upon completion, participants were sent an email thanking them for their participation and providing their book voucher.

7.1.3. Analysis

The study yielded 274 days of entries from 21 participants. This would present an extremely high volume of text for coding, and therefore a subset of this data was used. As the key concern in this study was the relationship between reading and CT, only those entries where both elements had been completed in the diary were kept; this resulted in 83 entries from 19 participants (this subset of data will be referred to as subset A).

In order to analyse the data in detail with deep reading of each entry, I chose to commence coding with a narrower subset of the data (to be referred to as subset B). As diversity of experiences was an important concern in this study, subset B was composed of only those participants who had at least 2 reading and CT days completed (i.e. captured at least 2 different experiences). From these participants, up to 3 entries were selected; the first and last entry where both CT and reading were complete, and the middle entry with both complete. This yielded 41 entries from 14 participants.

By subdividing the data in this manner, I was able to undertake analysis in two stages. In stage one, I used only subset B and coded this data. In stage two, I sought to apply the codes I had derived from subset B to subset A, moving from a smaller set of

the data to a wider dataset. This enabled a balance between a thorough and focused coding process with a smaller volume of text, and a wider validation of the coding derived from that highly detail-oriented process with a larger slice of diary data. Data analysis was conducted using Microsoft Excel version 2002 to display diary texts in spreadsheet format.

In stage one, using subset B, I undertook coding in stages (described in Table 46). First, I coded the responses to each prompt in the diary form (i.e. entries made in each free text box of the form): transportation into reading; evaluation of reading; topic and prompt of CT; process of CT; strengths and weaknesses of CT. The responses to each prompt will be termed ‘entries’ from hereon in. I commenced with in-vivo coding (Saldaña, 2016), extracting segments of the participants’ language that seemed most salient and representative as reflections of the topic. I chose to begin with in-vivo coding to stay true to the participants’ voice, and to keep the first stage in the coding process as a simple reduction of the data rather than a transformation of it. This enabled me to move from full answers to each prompt, to a set of codes that encapsulated the participants’ responses, still at the level of individual entries. Each answer given by a participant often had multiple in-vivo codes, for example “related to my interests / plan and predict” were two in-vivo codes applied to participant DHKX’s description of their transportation into the text. These codes will be referred to as the ‘in-vivo codes’.

In order to narrow down the volume of codes, I chose to concept code (Saldaña, 2016) the in-vivo codes, applying a single concept code to the multiple possible in-vivo codes derived from each entry. I chose concept coding as a progression

from the in-vivo codes as it permitted me to begin to identify commonalities and patterns in the data, and offered a broad enough approach to capture the different ways participants' interpreted and responded to the diary prompts. As the previous reader interview study found that transportation and evaluation were deeply intermingled, I grouped these two together into one field for coding. Thus, there would be a concept code for each of these diary topics where they had been completed: transportation and evaluation of reading; topic and prompt of CT; process of CT; strengths and weaknesses of CT. These codes were unique to each entry, as at this stage my aim was to encapsulate the participants' specific responses to each prompt. These codes will be referred to as 'entry codes' (see Table 46).

Next, I sought to find overarching codes that could capture each day's entry codes and translate them into a code for the whole day (to be termed 'day codes'), bringing together all of a day's entries on reading and CT. This was done so as to bring together the ways in which participants experienced reading and thinking critically, seeking to find a code that could fuse these for each day, as a means of addressing the research question on the relationship between daily reading and CT. Furthermore, this amalgamation of reading and CT on a daily basis would enable me to seek similarities and differences in how the interplay of reading and CT were characterised on different occasions. All of the entry codes from each day were thus concept-coded together, and this resulted in 41 day codes, each associated with a single day's diary entry (see Table 46).

These 41 day codes were then reviewed for similarities and duplication, seeking to reduce them down into a smaller set of codes (final codes) that could be applicable

to several days of diary entries. This process was done using printed paper cards for each day code, which could be moved and grouped and annotated in different ways until a meaningful clustering emerged. The 41 day codes were grouped together into 7 clusters, from which I derived 7 final codes. Table 46 gives a summary of the different codes at each stage of the process.

Table 46: Summary of code types

Code	Description
In-vivo code	Direct extracts from the participants' own language, encapsulating the meaning of an entry.
Entry code	A concept-code summarising the key meaning from the in-vivo codes; one per entry.
Day code	A concept-code summarising the whole days' entry codes, each unique to one day.
Final code	An overall code derived from grouping the day-codes into blocks of similar meanings, applied to each day of the diary.

These 7 final codes were then applied to subset B, such that each day was re-read and coded with one of these codes, verifying if the way that the individual day codes had been subsumed into these final codes in fact reflected the data. In this process all of the individual daily diary entries were coded a final code without any need to amend or add to them. Finally, these 7 final codes were used for the second stage of analysis in which all of subset A was coded (including re-coding subset B as a part of subset A). I therefore read each day of responses in subset A, and sought to apply one

of the final codes, or to identify days that these did not fit and required either an additional code to be created or one of the final codes to be amended. In this process the 7 final codes derived from subset B were found to be adequate and sufficient for coding all of subset A. Furthermore, no discrepancies between my coding of subset B in the first round, and recoding these days within subset A, were found.

Finally, having established this final set of 7 codes, I then grouped them into categories, and revisited the data in subset A, reading the responses in each category to gain a sense of how well the different responses fit together, until I was satisfied that there was sufficient relation between categories and data. This resulted in 3 categories.

7.2. Participants

The 19 participants in subset A included 11 female and 8 male, of which 4 did not have a degree as their highest educational qualification, 6 had an undergraduate degree, and 9 had a postgraduate degree. The mean participant age was 39.8, median 34, with range from 18 to 76. Participants in subset B were composed of 9 female and 5 male; 3 with no degree, 3 with a bachelor's degree, 8 with a postgraduate degree. The mean age in set B was 42.14, median 37.5, with the same range as subset A. 10 participants in subset A self-identified as primarily fiction readers, and 9 as nonfiction readers. In subset B there were 8 fiction and 6 nonfiction readers.

7.3. Results

The results will be presented by category, with each code described within the category framing. Table 47 shows an overview of the categories and codes, with a brief description for each code.

Table 47: Summary of categories and codes

Category	Code	Brief description
Personal projection	Systematic goal visioning	Future and goal-oriented imagining and planning
	Reflective perspective taking	Reflective and interpersonal imagining of alternate vantage points
Seeking and challenging	Seeking answers in information	Seeking resolution to issues and challenges through information
	Imaginative & challenging evaluation	Imaginatively manipulating, challenging, and evaluating information
Perspective balancing	Open observation	Gaining insights and observations as a result of openness
	Situational perspective taking	Grasping and situating different contextualised perspectives
	Experiential balancing	Weighing up and balancing different possibilities in light of own experiences

Personal projection

This category captures ways that participants mentally projected themselves in different time frames, recalling and reliving the past and imaginatively modelling the future. Under this category, the codes *reflective perspective taking* and *systematic goal visioning* cover this way of placing ourselves into different reference frames in our minds.

Systematic goal visioning

This code was applied to days where participants expressed a future-oriented reading and CT experience. This was typically bound to some form of goal, which was imaginatively envisaged and a path towards it was visioned. This could be seen in participants' descriptions of their reading; for instance in DHKX's reading of a Wordsworth biography, the poets' goals are envisaged and a future prediction is made:

“This sees his poetry taking more political and urban discussions, equally in becoming politically radical. This I predict formulate his cantankerism and anger in later years.” (DHKX)

This kind of future visioning associated with goals was also often the topic of participants' CT, for example in NHNL's CT about whether or not to play football:

“I considered what my coach would think about me playing with sore muscles and I considered how I would feel if I just got up and completed the training session... In previous training and gym sessions, I have sometimes been hesitant to go when suffering from post-exercise soreness. However when I just push through and go, I find that it helps to ease the muscle soreness once my blood is back pumping around my body during exercise... All of these considerations led me to believe that even though I was sore and

tired, I would actually feel better if I just got training over and done with.” (NHNL)

Here we can see NHNL including past memories in order to model future consequences, such as the coach’s reaction and the experience of muscle soreness, in CT aimed at determining whether playing football would ultimately align with the goal of feeling better. Often, participants’ engaged in reading and CT with the aim of enabling a goal, and thus their reading became building material for their visioning.

Although other future-oriented considerations appeared in various entries made by participants, this code captures an imaginative re-positioning of one’s frame of reference with relation to some goal or direction. It is notable that this code was derived from days were days when participants logged an item of nonfiction reading, and in the wider application of the code no fiction reading days were associated with it. The types of nonfiction items logged were very diverse, including news, nonfiction books, biographies, and others, suggesting a broad base of material was used to build imagined trajectories towards participants’ goals.

Reflective perspective taking

Another way of mentally positioning ourselves in different timeframes is not as goal-directed as in the previous code, but rather is more interpersonal and self-reflective, and this is captured in the code *reflective perspective taking*.

Participants gave descriptions of reflecting and positioning themselves in different perspectives to aid their reflection, sometimes this could be their own perspective but at another time (for example TVWN recalling what it was like to walk past John Betjeman’s house while reading his poetry, placing themselves in their past

viewpoint), or it could be the perspective of others (for instance EDHX thinking critically about how their family could perceive them in an online interaction). Often, the *reflective perspective taking* took place while thinking critically about the subject matter of the participants' reading, for instance:

“I tried to modify my pre-existing thoughts about this subject by just reading and avoiding unnecessary criticism of the upper elite. I understood once I did this that the elite are also suffering but in their own way in relation to their problems. A war is an issue for all, whether you are poor or rich, since it will come effect your life and change it for the next years to come. By having this rigid mindset in which I think the rich shouldn't feel bad about not being entertained, I have also understood that they want some normality rather than entertainment.” (GNYO)

Here, GNYO describes thinking critically about a reading experience in which they disliked the characters, and actively seeks to reflect on their own and the characters' perspectives. Often, these reflections were also future-oriented. For example upon reading Flaubert's *Madame Bovary*, TYVQ draws a lesson for future perspective taking:

“I think quite a lot about efficiency and what pursuits are worth my time, but the book made me think for a moment about stopping to smell the roses. enjoying the moments as they come and not worrying about whether what I'm doing is useful.” (TYVQ)

Thus participants' described reflectively moving across different imagined perspectives, their own and others', past, present and future.

This code was derived from days on which participants read fiction, and furthermore when applied to subset A there was one day with a nonfiction reading (a

memoir) coded *reflective perspective taking*. The types of fiction read included novels, poetry, and short stories. This suggests an association with a self-reflective shifting of one's perspective, and reading material that is fictional.

Seeking and challenging

Within this category, information was at the forefront as it was sought, challenged, evaluated, and imaginatively manipulated. The category was comprised of two codes, *seeking answers in information* and *imaginative & challenging evaluation*, which together build up a picture of participants' approaches to information in their reading and CT.

Seeking answers in information

Under this code, participants expressed some question, issue, or challenge they sought to resolve, and they sought its resolution through information. In some cases their information seeking resolved their questions (for example NHNL seeking information on how to find a good supervisor), but in others the questions were left open (for example BCMU seeking to understand whether opening schools was safe under the COVID-19 pandemic). Thus this code can encapsulate an ongoing, or completed, seeking process.

Very often, reading and CT were intertwined in this code. For instance, EDSR read and thought critically about brain teasers, thus forming questions about the best solutions to the puzzles, and then pursuing information about them. Participant SZHW read an article on how to find work online, and further pursuing this question thought critically about it, undertaking further information seeking:

“The first step was to research the individual possibilities, which requirements and skills one must have to carry out the work. In the second step, I examined various platforms from fiverr.com to freelancers to indeed.com and found that not only the number of people who provided their knowledge was very large, but also the individual platforms set conditions in the form of tests for example.”
(SZHW)

Thus we can see how this code often captures a cyclical process moving from reading, to CT, to reading, enmeshing the experience of both.

Given the nature of the information sought to resolve participants’ questions and issues, it is perhaps unsurprising that this code was derived from nonfiction reading days, and in the wider coding process all days with this code applied had nonfiction texts logged. Nonfiction items read were news and magazine articles, nonfiction books, and research articles. These sources both prompted, and resolved, the seeking of information to address questions and challenges.

Imaginative & challenging evaluation

This code encapsulates participants’ descriptions of holding information in mind and imaginatively manipulating it, challenging it, and evaluating it. Often, participants described actively collecting information in mind from their reading and/or CT experiences, such as GNYO who during reading a novel describes:

“I gathered all the information I had about her previously and added in the new information. Once I have done this, I also started to think about my own ideas around this and contrast them with the ideas of the character.” (GNYO)

This process was typically described as vivid in the imagination, with information being utilised to build up an imagined image in an evaluative manner:

“I initially thought this story was set a long time ago but there were frequent references to modern-day life going on outside the confines of the Abbey - cars, ambulances and the like - while at the same time mentioning time travel so I am still unsure whether the monks were living among us or ghosts of a past age.” (TTQN)

In this example, TTQN is picturing the setting while evaluating it, in order to understand the narrative and evaluate the meaning. References to other information contained within the text are also evaluated.

Often, this imaginative evaluation was described entailing challenge, either to oneself, or challenging of the information itself. This is exemplified in participant JGSB’s experience of reading a news article, and then thinking critically about that article:

“I was completely immersed. Yes, I could imagine myself in the many meetings and news conferences [...] Although I very much enjoy reading about President Trump's many mishaps, I think that the media (in the main) has a significant bias against him, and cherry pick anecdotes that present him in the worst light whilst failing to report any positive stories [...] These thoughts go through my head as I read these kind of article. To make the article interesting and smooth to read, a lot of the information is unattributed and in my view questionable [...] Whilst reading the article, I did at the same time bear in mind my own interactions with challenging individuals, somewhat similar to Trump, and the difficulties in getting your views heard and understood [...] Judging Trump fairly is challenging because he is such an unprecedented individual and president.”
(JGSB)

This shows how imagining and challenging intermingle to evaluate reading both during, and also afterwards when thinking critically about the read text.

Imaginative & challenging evaluation was equally represented in fiction and nonfiction days (7 of each from subset A), a balance in keeping the way the code was derived in subset B (4 fiction, 5 nonfiction). A broad range of reading materials were included, such as novels, magazines and news articles, nonfiction books, memoirs, etc. This suggests no specific type of reading material was associated with this code.

Perspective balancing

This category captures the way participants observed and took in different perspectives, situated themselves in them, and balanced their experiences within those different perspectives. As such, it entails openness, engagement with viewpoints, and weighing up varied possible positions. It is comprised of the codes *open observation*, *situational perspective taking*, and *experiential balancing*. Akin to *Personal projection*, this category entails an alteration of one's frame of reference, however these categories can be differentiated as *Personal projection* is temporally oriented, while *Perspective balancing* is subjectively oriented.

Open observation

Participants described a kind of observation that came to them from their reading and/or CT experiences, often very directly in an unbidden manner, as a result of being open to it. This did not tend to arise from an active, seeking orientation, but rather a more passive and absorbing state. For instance, during reading EDSR describes being a "content observer" to the "strange imagination" of the book. This experience of observation was also described as an CT experience, for example participant TYVQ described being prompted to think critically about the nature of happiness by going on a walk:

“I realised in a moment how much conscious control over our mood and state of mind we all have. Just by thinking about it, you can go from dosing off and feeling very slow and lethargic to being energised and happy in just a second.” (TYVQ)

For TYVQ, this observation contradicted a previous neuroscientific attitude to happiness, and it seems openness was required for this arise. Similarly, TVWN describes an observation on the value of music and other arts arising from a family member learning an instrument, prompting an attitude shift. Neither of these attitude changes were actively pursued, but arose from a state of openness.

In other cases, the observation could be more reflective of one’s own thinking, or layered so that both the subject matter is being openly observed as well as one’s response to it. This was the case for HCAQ when reading a sequel novel:

“There is a whole new twist - she has her own internal monologue now and I am seeing how she got to where she is. It's interesting, but also slightly ruins the original book... the author is leading me to a conclusion about her, which is slightly obvious and I think spoils things a little. I think my critical thinking is actually on the authors process in leading me, rather than on the character herself.” (HCAQ)

HCAQ is open to the characters’ development, and observes the author’s technique, and also observes their own thinking about it. Thus, *open observation* could apply to varied domains and levels of awareness.

This code was derived from days when fiction reading was logged. Although the CT subject matter within this code was very diverse, not necessarily linked to the reading, all the days coded with *open observation* contained fiction reading, and all were novels.

Situational perspective taking

In this code, participants described coming to understandings of different perspectives rooted in different situations. The situation typically drove and shaped the perspective taking, as a means of furthering an understanding of that situation. For example, the COVID-19 pandemic situation promoted participants to read about, and thinking critically about, different perspectives in response to it. For instance, SGLJ read and thought about the perspectives of those unemployed, while SZHW read and thought about different points of view on the balance between public good and restrictions. These were both experiences of engaging with and imagining alternative perspectives, rooted in the concrete situation of the pandemic.

In other instances, participants situated the perspectives of authors or figures who prompted their CT, and indeed their own perspectives. Both TTQN and TVWN read and thought about issues of race, in a concretely contextualised manner:

“I suspected that the author, being a black woman writer herself, had probably already made this connection and perhaps hoped her readers would too. I thought it possible that the author felt gender was an easier topic to write about, that it would be more readily accepted on the West, resulting in a wider readership, and hopefully in readers making links and better understanding these arguments when they see them applied in the fight against racism.” (TTQN)

Thus TTQN evaluated the author’s perspective by situating and contextualising her arguments. Furthermore, TTQN situates her own perspective as a “white woman” responding to a video on looting during the Black Lives Matter protests that prompted her CT. Similarly, TVWN connected the personal perspective of an author to the

situation, and further linked this to their own experiences in thinking critically about it:

“Whilst I knew America is divided, the article, written from a personal perspective made me think about how it must be to live with that. To live in one of the richest and most diverse countries in the world yet spend your day trying not to be the next viral video. [...] A strength is that I have experienced bigotry all my life but this is also my weakness and objectivity is impossible. Although, were such things are concerned how can there ever be objectivity?” (TVWN)

In this case TVWN further reflects on the strengths and weaknesses of this situated perspective taking in CT.

This is not the only code in which perspective taking is key; *reflective perspective taking* also entails understanding different perspectives, and is thus similar to *situational perspective taking*. However, the two codes can be distinguished by the different ways that perspective taking is described within them. In *situational perspective taking*, the focus is a specific situation or issue, and it is this that drives engagement with different perspectives on that situation; the perspectives in question are inherently grounded and contextualised. By contrast, *reflective perspective taking* does not necessarily stem from a specific situation, but rather the driving focus is a more interpersonal and self-reflective imagining of different views across different timeframes.

Perhaps because of the highly contextualised grounding that often centred on current affairs, all of the days that were coded as *situational perspective taking* were days that participants described a nonfiction reading item. Types of nonfiction included

news and magazine articles, nonfiction books, blogs, and others, forming a very diverse reading base for this code.

Experiential balancing

This code describes the ways in which participants sought balance between their own experiences and what they read, and thought critically about. It captures reading and thinking experiences that involve weighing up or comparing different views or possibilities, seeking to rank or arrive at some order to different options. Crucially, there is a combination and intermingling between experience and information, and balance is sought.

In reading, the balancing often concerned different views expressed in the text, but could also be more reflective of the balance of the text itself, for example while reading *Little Women* UKGL considered the balance of the characters: “the author seems to have created four totally different in their styles, outlooks etc and isn't that the same way the Spice Girls were created”. Such balancing could be on many different topics and could take place in different ways. Balancing was done when weighing up different possible views, but also often took place when weighing up one’s own thought processes and responses, for instance in another example:

“On the one hand, I'd been lucky in that the bulk of my money had been refunded and I will only lose £50 but on the other hand it's £50. If someone was to say to me 'go outside and throw £50 into the bin' I'd probably call them a few rude names and not do to but part of me thinks this company and all of it's employees and maybe a lot of others who have just lost their bookings will be suffering a lot more so should I be that bothered over a relatively small amount... Being able

to appreciate that sometimes you can lose something but you can also still count your blessings in the end.” (UKGL)

Here, UKGL is balancing their different responses, reflecting on their own experience and also imagining the experiences of others, and thus coming to an overall judgement on how to weight their own reactions.

Often, this experiential balancing was prompted by reading, as participants drew upon their personal experiences and weighed them up against the views presented in the text they read. For example, JGSB describes using their experience working in a university context to evaluate an article about student needs, balancing the different approaches, towards a “Road to Damascus’ moment of accepting that Universities can, and will, and perhaps, should change”. Some of these experiential balancings were far more personal and subjective in nature, for example MTPA balancing Kate Bush albums in order from preference from own experience while reading an article, or JGSB weighing up their experience of disliking fish in light of an article on the topic. Thus the domain within which balance was sought could vary widely.

This code was derived from a fairly even mixture of fiction and nonfiction reading days in subset B (2 fiction and 3 nonfiction). However, when applying the *experiential balancing* code to the wider data of subset A, there were 3 fiction days and 14 nonfiction days coded. This suggests that though not exclusive, there is an association between this kind of balancing and weighting with one’s experiences and nonfictional reading. The fiction read consisted entirely of novels, and the nonfiction was a diverse mixture of news and magazine articles, blogs, nonfiction essays, among others.

7.4. Discussion

7.4.1. Relation to the literature

The content forming the *personal projection* category is related to the way “mental time travel” is conceptualised in the cognitive sciences (Corballis, 2009); a way of drawing from memories and imaginatively modelling different scenarios in different times. It is also related to how counterfactual reasoning in a wider sense is conceptualised (Byrne, 2016). It is interesting that this category contains a code more associated with nonfiction, and another associated with fiction (see Table 47). The former, *systematic goal visioning*, is very much in line with framings of “mental time travel” as a means of imaginatively modelling the future and mapping out different possible actions to mentally test their alignment with our goals (Corballis, 2009). The latter, *reflective perspective taking*, is perhaps more representative of wider counterfactual thinking capacities concerned with different possibilities, and their resonances emotionally (McMullen, 1997). McMullen (1997) proposes different modes of counterfactual imagining, suggesting that we can enter into counterfactuals in a simulation, or as an evaluation. The former involves being entirely immersed in the counterfactual scenario, while the latter is a more detached way of reasoning through the alternative possibilities. Thus, these two codes are in keeping with these modes of thinking counterfactually, and furthermore they draw out fiction and nonfiction alignment with the two modes (fiction associated with simulation, nonfiction with evaluation). In coding the daily reading thinking entries together, this category of *personal projection* demonstrates one form of interplay between what is read and how we think. Corballis (2009) points out that our ability to share our episodic memories with one another through language permits us to share the material from which we can

construct our counterfactual imaginings. Furthermore, fictional material is also incorporated into our episodic memory (Zunshine, 2006), and therefore also builds up the store of content which can then be re-deployed counterfactually. Therefore both fiction and nonfiction may be feeding out counterfactual reasoning capacities, in two different modes.

The category *seeking and challenging* contains two codes that represent different approaches to finding and thinking about information in a text (see Table 47). Firstly, *seeking answers in information* was a very goal-directed and purposeful enmeshing of reading and CT, associated with nonfiction. Secondly, *imaginative and challenging evaluation* described ways of mentally manipulating information to interrogate it and oneself, drawing upon both fiction and nonfiction material. To some extent, these two codes align with the distinction made between “informational reading” and “experiential reading” (Hampson Lundh et al., 2018). *Seeking answers in information* is in line with how goal-oriented reading is characterised in reader response theory (Rosenblatt, 1982), and how informational reading is focused on within IL (Hampson Lundh et al., 2018). However, while *imaginative and challenging evaluation* shares many of the ways the reading experience can be characterised as “aesthetic” or “experiential” (Hampson Lundh et al., 2018; Rosenblatt, 1982), it is notable that there is far more active evaluation and thinking also contained in this code. In this way, the code is in keeping with prior research showing evaluations taking place throughout the reading process (Özyürek & Trabasso, 1997). However, it also takes reading into the realm of CT beyond the direct evaluation of what is read, as participants described imaginatively evaluating information from varied different texts

long after they had been read. Therefore this category both corroborates ways of differentiating different modes of reading, but also extends them.

The category of *perspective balancing* (see Table 47) contains three codes that all fit with prior research on both CT and reading, such that this category offers a means of bringing together different CT and reading elements and understanding their interplay as a whole. Firstly, the code *open observation* described an open stance that participants experienced as enabling them to change their opinions and engage with new perspectives. The way openness was described was as a non-judgemental stance, being an observer rather than an active seeker. As such, it corroborates calls for openness in CT (e.g. R. H. Ennis, 2015; Facione, 1990), but furthermore resonates with more empathy-driven CT approaches as a way of letting in very different points of view (Thayer-Bacon, 1992), and as a stance where nothing is pre-empted or closed off it is perhaps a good characterisation of bell hooks' "radical openness" (1994, p. 202). It also captures the way in which reading can be such an open state, without defences (Djikić et al., 2009a). Indeed, the fact that *open observation* coded fiction reading days supports this view of fiction as uniquely enabling of a fuller form of openness. Next, the code *situational perspective taking* characterises a specific way of engaging with different points of view, grounding them in their situational contexts. Reading within this code was deeply integrated into the wider background of the text, and as aligns with views on the value of reading for real-world insights (Hakemulder, 2000; Harlan, 2019). It is interesting that while many authors defining CT argue for a need to engage with different perspectives (Paul, 1987), the need to engage with points of view with a necessary contextual grounding is not argued for; this could offer an

interesting avenue for further CT research. Finally, *experiential balancing* describes a kind of weighting and ranking of personal experiences and opinions that is both normative, and deeply subjective. Participants experiences within this code articulate a way that CT prompts these kinds of shifts in attitude and belief (Dewey, 1910), alongside ways in which reading material can drive such change (Djikic & Oatley, 2014). In combination, the category of *perspective balancing* seems to take several of the puzzle pieces of reading and CT delineated by prior research and fit them together.

7.4.2. Limitations

I coded the reading diaries alone as a single coder, and thus inter-coder reliability was not measured. The coding of the data represents my own interpretations only, and cannot be generalised.

In coding the diary entries, the full day's diary was visible to me including the details of what had been read. This means that coding was not done independently of the fiction/nonfiction classification of different diary days, but rather what had been read was part of the coding process. Subsequently, the division of codes into fiction and nonfiction is not unbiased, and it cannot be concluded that *reflective perspective taking* was more likely to fit a fiction day than a nonfiction day, for example. Rather, the code *reflective perspective taking* was derived from fiction days as part of the development of this code, and it reflects a particular way that way participants described their fiction reading in relation to their CT. As there was no change in the balance of fiction and nonfiction represented in each code when coding subset A (without any focus on the reading type during this phase of coding), some validity is given to the differences in reading type interpreted in each code.

Participants were given a set diary form each day, which contained prompts on pre-defined topics. Although text boxes were in an open format allowing participants to write what they chose to, and as much as they chose, this was nonetheless not an entirely freeform diary. This means that participants may have had experiences to write about which did not fit the given prompts, and may have therefore omitted them. The diary data therefore cannot be viewed as a complete representation of participants' daily reading and CT experiences. Similarly, the diary form asked participants to select one item of reading, and one case of CT, to write about each day; it did not seek a complete record of all daily reading and CT. As participants' were asked to select the 'main' instances of reading and CT, they may have favoured certain kinds of experiences. For example, participants may have made judgements on importance or quality of what to write about, and thus the data may be reflective of only a particular type of reading and CT experience. Subsequently, the findings from this study cannot be generalised to all reading and CT experiences.

Participants were recruited online, and completed the study online. This means that they were a sample of internet users, and likely are not representative of a wider population. Furthermore, participants self-identified as readers, and were likely to be motivated by a book voucher as a reward, which further narrows their characteristics. As such, no conclusions can be made about a wider population based upon this study.

Of the total 274 days of entries from 21 participants that were collected, only 88 entries from 19 Participants were analysed for this study. Thus, a large amount of data remains unutilised, and this could be used in future research. Due to the substantial

amount of text in the complete dataset, this would potentially be suitable for text mining approaches such as sentiment analysis.

7.5. Conclusion

7.5.1. Answering the research questions

- c. How do day-to-day experiences of reading and thinking critically vary?

The 3 categories derived from the data analysis process give an insight into different ways that reading and CT were described (see Table 47). *Personal projection* captures reading and CT experiences that entailed an imagined shift into other frames of reference, with participants projecting themselves into different viewpoints across different time frames. *Seeking and challenging* characterises participants' reading and CT to address questions and evaluate and challenge information. *Perspective balancing* incorporates participants' ways of being open to different possibilities and using context and experience to make judgements and weight various options and perspectives.

Based on the differences between these categories, and the further distinctions that can be made between the codes contained within them, it is possible to answer this research question with a range of characteristics of reading and CT experiences derived from a range of day-to-day examples. The findings from this study illustrate some meaningful differences between how reading and CT together are experienced in diverse ways.

- b. How are everyday experiences of reading fiction or nonfiction related to CT approaches?

Codes associated with fiction were *reflective perspective taking* and *open observation*; these codes capture a movement into different vantage points, facilitated by openness, through which imagining and observing different possibilities and

perspectives can provide novel insights. Codes associated with nonfiction were *systematic goal visioning*, *seeking answers in information* and *situational perspective taking*; these codes comprise goal-oriented approaches, seeking resolution to challenges, and seeking evaluation of different perspectives in a highly contextualised, situated manner. Finally, codes with a mixture of fiction and nonfiction were *imaginative & challenging evaluation* and *experiential balancing*; these codes incorporate the evaluation of information in a balancing, weighing manner, with recourse to personal experiences and imagination. These differences offer an answer to this research question, with fiction and nonfiction having some distinctive codes associated with them, while also coming together in others.

One way to interpret the differences between these codes is in terms of their grounding to concrete reality, versus a more speculative character. Both *reflective perspective taking* and *open observation*, codes associated with fiction, were fluid with respect to the way that different points of view, imagined or real, were modelled and engaged with. By contrast, the codes associated with nonfiction were grounded in concrete questions and situations. The fiction-linked code *reflective perspective taking* can be contrasted with the nonfiction-linked code *situational perspective taking* to highlight this distinction. In the former, participants moved between different points of view across different frame frames, with a personally-oriented reflective approach. In the latter, participants understood different perspectives through their contextualised grounding into very specific and concrete situations. The nonfiction-linked code *systematic goal visioning* can also be contrasted with *reflective perspective taking*, as these two codes form the category of *Personal projection* and in a sense offer two sides

to the category. In *systematic goal visioning*, the mental repositioning of oneself into another time is done in order to vision set goals, and thus is a very concrete and reality-oriented activity. In *reflective perspective taking* the mental projecting is far more exploratory and meandering, more linked to one's own reflections than to any external goals or needs. Finally, *seeking answers in information* (nonfiction-based) and *open observation* (fiction-based) also offer a contrast; in the former, answers come from a concrete search, in the latter they arise without searching through a non-directed state of openness. These comparisons align well with the definitions of fiction and nonfiction adopted for this research project; the codes derived from fiction reading days are concerned with make-believe, while the codes derived from nonfiction reading days have an inner core of reality. The codes derived from a mixture of fiction and nonfiction days (*imaginative & challenging evaluation* and *experiential balancing*), perhaps offer some insight into how reality and make-believe can be intermixed in reading and CT experiences.

However, despite some interesting distinctions between some codes, all 3 categories were derived from a mixture of codes associated with fiction or nonfiction (see Table 47). Therefore at the category level, there is no clear distinction between fiction and nonfiction reading and the way reading and CT experiences can be classified. Rather, the categories offer some different ways that reading and CT described together can be experienced differently. This implies that differences between days where fiction or nonfiction were the main reading item, and how reading and CT were experienced differently across these days, operate at a smaller scale than the overarching differences between overall ways of experiencing reading

and CT. At a higher level of description, the interesting differences do not stem from reading material type, but are broader ways of engaging with reading and CT.

7.5.2. Summary

This study found meaningful differences in experiences of reading and thinking critically on a day-to-day basis, which can be grouped into codes and categories derived from a range of examples. When combined, participants' characterisations of their reading and CT could be grouped under 7 codes, each indicating a different way that these were undertaken and experienced, and furthermore these codes could be grouped into 3 categories (see Table 47). These categories suggest that participants undertake *Personal projection* in their reading and their CT, imaginatively moving into different frames of reference as they reflect and work towards their goals. Participants were also engaged in *seeking and challenging* while reading and thinking critically, to find desired answers and undertake evaluations. Finally, participants described different ways of *perspective balancing*, weighing up options and possibilities from different angles, either stemming from openness, with situational contextualisation, or with reference to their own experiences. Furthermore, these different experiences of reading and CT were connected to reading fiction or nonfiction that was described as the main reading each day. However, the main categorisation process in this study merged these together, seeking their commonalities rather than their differences. Indeed, this enables the overall findings of this study to be seen as a counterpoint to the previous studies, and as such these findings can be used as a comparator for gauging the validity of the separate fiction and nonfiction findings (Shadish et al., 2002). By identifying commonalities in how reading may relate to CT,

we can better articulate differences in fiction and nonfiction. Within this study, these differences can be found at the level of different codes, but not at the level of different categories; differences in fiction and nonfiction at a fine-grained level of analysis were subsumed together into larger over-arching ways in which reading and CT could be characterised. On the level of codes, the differences between the ways participants described their reading CT on days when they logged a fiction or a nonfiction text can be characterised through their grounding in reality versus an engaging in make-believe. Nonfiction-linked codes were more concrete, contextualised, and reality based; fiction-based codes were more fluid, reflective, and openly exploratory. Therefore this study contributes both insight into possible differences in how fiction and nonfiction reading may be associated with CT, and also gives insight into the ways that both can be interconnected in broader ways of experiencing reading and CT.

8. Discussion

This research project investigated the influence of reading fiction, as differentiated from nonfiction, upon CT. The research encompassed exploration and hypothesis testing; normative and experiential approaches; quantitative and qualitative methods. Four studies were conducted: one, an observational questionnaire study testing for a correlation between print exposure and factors of CT; two, a quantitative reading log study testing experimental manipulation and assessing a causal direction between reading and CT; three, an interview study seeking to gain insights into readers' experiences of fiction and nonfiction reading and thinking critically; four, a reading diary study seeking ways in which reading and CT may differ on a day to day basis. Firstly, a summary of the most important results from the research project will be given. Following this, a more in-depth summary of the findings from across all four studies will be given. This discussion chapter will then expand upon and contextualise these findings.

The overall finding of this research project as a whole, that fiction and nonfiction have distinctive relationships with CT, is an important contribution to research on the impacts of reading, as well as broader IL and CT research. Within this, the relationship between fiction reading and CT has been elucidated with new insights garnered; fiction was found to correlate with a greater disposition to think critically, and with less absolutist and more evaluativist epistemological orientation; fiction reading had an association with greater CT change, and with improvement in CT, over a two week period; fiction was connected with experiential and imaginative CT approaches, as well as with prompting CT in circuitous and complex ways. All of these

findings present evidence to support increased fiction provision, and provide a grounding for future research. The finding with perhaps the greatest practical value yielded by this research is that of the efficacy of assigning fiction reading to nonfiction readers as an intervention for improving CT (study two). This demonstrates the power of exposure to fiction for those who may not otherwise engage in fiction reading, and opens an array of future research possibilities for investigating this effect further.

Study one found that fiction print exposure was associated with increased CTD, and greater likelihood of having an evaluativist EO rather than absolutist. Conversely, nonfiction exposure was associated with a greater likelihood of being an absolutist than any other EO. No relationship between nonfiction exposure and CTD was found. In this study, NT had both a direct relationship with CTD, and mediated the reading effects upon it. Study two supported the correlations established in study one, as participants' baseline CT score was found to be associated with increased reading engagement in total, and in terms of fiction, though not nonfiction. Study two also found that the time participants spent reading in total in between two CT tests was predictive both of change in CT score, and of that change being positive. When comparing fiction and nonfiction, only fiction was associated with increased change in CT scores. Addressing improvement in score, reading fiction as measured in entries logged also had a significant relationship. Conversely, larger numbers of nonfiction entries were predictive of negative or stable CT score, but when time was used as the predictor variable, increased nonfiction reading was associated with having improved CT score. For a differentiation of CT change and improvement, refer to Box 7. Thus, fiction and nonfiction reading had a more complex relationship with changes in CT

over a two week period. Additionally, assigning participants readings as an experimental manipulation revealed that when nonfiction readers were assigned fiction, their CT score increased, and was more likely to be positive. However, contrary to study one, in study two NT had no mediation effect. Exploring this further, it was found that reading more fiction and nonfiction at a peak of one's NT did significantly impact change in CT. Additional exploratory analyses also showed that spending more time per entry of reading, both for fiction and nonfiction, was associated with having improved CT score, and for fiction that also had an association with increased CT change.

Study three found that reading and CT were deeply interconnected in participants' experiences. NT into a text involved active evaluative thinking, thus constituting a bridge between reading and CT. Through this, reading was found to be a deeply transformative process, with transported evaluative thinking leading to reflective changes in oneself, as well as to changes to actions and beliefs. Furthermore, distinctions were found between the ways participants experienced fiction and nonfiction reading in relation to CT. Fiction reading was connected to the experiential facets of CT, as a time-extended, developing and unfolding conscious experience. Nonfiction reading was associated with procedural approaches to CT, as a means of building knowledge and expertise to refine CT skill. Study four aligned with study three in finding both commonalities in reading and CT, and distinctions in the way reading and CT could be associated with fiction and nonfiction. Codes connected to fiction reading days were more exploratory, reflective, and fluid, similar to the experiential emphasis found in study three. Codes from nonfiction days were more

grounded in context, and rooted in reality, in keeping with the way nonfiction was used to build knowledge in study three. However, in study four it was the blend of fiction and nonfiction reading together that produced the overall categories. These categories reflected day-to-day differences in how reading and CT together were experienced. Participants described projecting themselves into their reading and CT, moving between different mental frames of reference. They also interrogated information in imaginative and evaluative ways. Throughout reading and thinking critically, participants also balanced their perspectives, contextualised them to concrete situations, and reached shifts in their beliefs from a position of openness.

For the exploratory and interpretive aims of this discussion, the qualitative findings offer broader ways of approaching the topics of reading and CT than the quantitative hypotheses would. Therefore for the purposes of structuring the discussion into a coherent interplay of all four studies, I have opted to use two of the themes that emerged in study three as the section headings for discussing the key findings pertaining to fiction and to nonfiction. Considerably more time is spent discussing the first, fiction-related, theme; this is due to the focus on fiction in this research project, with nonfiction utilised as a comparator rather than as a variable of primary interest in its own right. The third theme from study three pertained to reading and CT overall, and thus relates to the wider context within which this research is situated; this will not be treated separately, but will be present within the discussion of fiction and nonfiction. Following on from this discussion of fiction and nonfiction findings, nuance in the delineation of reading materials will be considered.

8.1. Fiction gives experiential critical thinking understanding

8.1.1. Time

The findings from the four studies taken in conjunction demonstrated ways that participants described their fiction reading experiences as providing deep, actively experienced, understandings. An important aspect of this was the way that fiction was experienced over an ongoing timespan across which a gradual and intimate comprehension of lived experiences could develop, i.e. a moment-by-moment flow of reading and thinking. This aligns with previous research that finds the temporality of reading experiences to be central in their impacts (Braun & Cupchik, 2001; Kuiken et al., 2004). Indeed, Kuiken and colleagues argue that the sequential process of reading may be “the crucible for self-modifying feelings” (Kuiken et al., 2004, p. 179). Participants actively contrasted this sequential fictional flow with the reporting mode of nonfiction which (though also taking time to read) truncated events into post-facto description. As a participant described it:

“Because when you see news stories, it's always the aftermath. It's always a revision of what's happened and it's never as it's happening. Whereas a lot of her [Emma Donoghue] stories they're talking about it as if it's happening now, it's in the present it's current” (Participant M, study three)

This draws a line between the moment-by-moment experience of events in fiction, and the nonfictional emphasis on closure of events. This is also aligned with Braun and Cupchik's (2001) findings of a distinction between ‘close’ and ‘far’ reading orientation, where the former is highly absorbed into the passage of narrative, and the later observes events from outside of their flow. Participants connected this experience of fictional reading with an associated experience of CT, in which there was also this form of

drawn-out time extension, permitting continuous shifts of thought, rather than a more closed final evaluation. The findings of this research thus suggest that the gradually developing, time-extended nature of fiction reading, in which every step in the narrative is experienced and evaluated in and of itself, promotes a different kind of CT that is more expansive and open to flux, than the nonfiction reading and evaluating experience. It may be that both reading and CT can take a 'close' or 'far' mode, and fiction in particular drives close engagement, through which the close mode of CT is also amplified.

This argument, that the expansive experience of fiction reading that entails a gradual journey through the narrative's time may shift modes of CT, may appear to be contradicted by some of the findings in study two, however. Findings differed when the number of entries logged by participants, or when the time logged, were taken as the predictor variables. Time spent reading fiction, and entries logged of fiction, were associated with increased change in CT. However, only fiction logged in entries, not in time, had a significant association with having an improved score at t2 in contrast with t1. This may imply that time spent reading fiction does drive some change in CT, but not necessarily an improvement. If we interpret the number of entries logged of fiction reading as the number of sessions a participant spent reading, then this would suggest that engaging in more fiction reading frequency is important for CT, both for driving and change and improvement. Thus, this would appear misaligned with an explanation of fiction's efficacy rooted in the time-course of reading. However, this misalignment evaporates upon closer attention to the way time features in the *fiction gives experiential CT understanding* theme. Participants were not talking about the

time they spent reading fiction, rather they were describing the passage through time entailed in the nature of fictional narrative. One might spend the same amount of minutes reading fiction and nonfiction, but the former was experienced by participants in study three as valuable in engaging them in the time flow of the narrative, while the latter was viewed in a more punctuated manner. Thus, whether reading for five minutes or five hours, one experiences a move through time within the fictional text that may not always be present in nonfiction. Indeed, reading speed has been used as a measure of immersion into the flow of a story, with the same length of text read faster or slower depending on how transported the reader is (Cupchik & Laszlo, 1994; Miall, 2008), thus shorter time spent may not necessarily indicate less read, but rather a stronger grip of the narrative flow upon the reader. This may be one explanation for fiction time in study two not predicting CT improvement. Each entry of fiction logged in study two represents an instance in which the participant stepped into the time-flow of the fictional narrative, and perhaps might have experienced the kind of gradual developing evaluative change described by participants in study three. Each dip into the fictional time stream may thus be important. Conversely, Cupchik and Laszlo (1994) argue slower reading is indicative of greater emotional and cognitive connection making during reading, and Miall (2008) suggests slower reading may result from foregrounded sections of text being particularly salient and personally relevant to readers. Exploratory analyses in study two did also find that spending more time per each entry made was associated with CT change and improvement; it may therefore be that each dip into fictional experiencing of the narrative flow is important, and also that taking time to experience each such engagement fully, relative to one's own level of transportation, emphasises its impacts.

The way that fiction reading entails a shift into different timeframes was also evident in the *reflective perspective taking* code from study four. This code was derived from days when participants read fiction, with the exception of one nonfiction day which was a memoir and thus perhaps entailed a similar narrative mode to fiction texts (Freeman, 2003). Under this code, participants described a very reflective move into different perspectives across different times, often evaluatively linking their reading to their own past experiences or to projections of the future. This is closely aligned with the way previous research has found the engagement with narrative sequences to cause shifting patterns of self-change (Kuiken et al., 2004). This finding expands the potential time associated link between the fiction reading and CT, also implying a connection between the content of fiction reading and the kinds of evaluations made with recourse to memory, which also entail mental shifts across time. In both ways, fiction reading is associated with ways of thinking that are fluid, connection-making, and continuously changing.

As study one found fiction readers to be more likely to have an evaluativist EO than absolutist, but no difference in their likelihood of being multiplist in contrast with either other orientation, this finding also fits with this characterisation of the fiction reading and CT relationship. Evaluation was intricately linked to the experience of moving through fictional narrative time, and this was actively contrasted by participants in study three from an absolute nonfictional position. This experience of the fictional time course was open to multiple perspectives. Likewise, in study four multiple perspectives and timeframes were taken, and used as grounds for evaluations, in reading and thinking critically. Thus, the experience of fiction reading may be associated with

both evaluativist and multiplist experiences of thinking critically. This is further commensurate with the study three code *EO as a personal approach to CT*, which found that participants could ground their CT in different EOs at different times, thus diverging from a view of EO as fixed, with only evaluativism associated with CT (Kuhn et al., 2000). With relation to study one, this suggests that rather than interpreting these findings as reflective of participants EO at all times, the measure of EO used (McGinnis, 2016) may instead indicate participants' most common or most comfortable EO, but with others likely to come into play in different instances. Taken together, these studies build a picture of an association between entering into the time course of fictional narratives, and a flexible experiential mode of CT.

Two avenues emerge, drawn from the previously described ways of engaging in narrative time, for explaining these findings with relation to how reading fiction may be different to nonfiction; the counterfactual thinking entailed in moving across different temporal and perspectival frames of reference; the imaginative and engaged nature of moving into these different perspectives.

8.1.2. Counterfactual reasoning

Findings from the four studies provide some support for a pathway as part of the causal link between fiction reading and CT via counterfactual reasoning. Most notably, participants in study three actively described ways reading fiction required and promoted counterfactual thinking, and within the code *inference making and imagining* also emphasised the importance of counterfactual thinking in CT, neatly described by one participant as: “you kind of have to run different imaginary paths” (Participant K, study three). In study four, this was also seen in the *personal projection*

category, and as discussed in the study four chapter this can be related to “mental time travel” as a particular form of counterfactual imagining (Corballis, 2009).

In study one, nonfiction readers’ likelihood of being absolutist rather than multiplist, in contrast with fiction readers’ having no differences in multiplism but only a higher likelihood of evaluativism versus absolutism, may also fit here. It may be that counterfactual reasoning is well aligned with an evaluativist EO, and may also align with a multiplist EO, as thinking counterfactually requires engaging with a multiplicity of perspectives and possibilities. By contrast, absolutist EO may be least conducive to counterfactual reasoning, as it implies a rejection of alternative possibilities in favour of seeking a single solution. However, at the time of writing no studies linking EO and counterfactual reasoning were found, thus this remains a space for future research to explore further.

Finally, in study four, I discussed how the two codes that made up the *personal projection* category (*reflective perspective taking* associated with fiction reading days; *systematic goal visioning* associated with nonfiction reading days) could be argued to align with McMullen’s (1997) proposed different modes of counterfactual reasoning as either simulation, or as evaluation, respectively. This suggests that fiction is not exclusively associated with counterfactual reasoning, as nonfiction has an association also, but rather that there are associations to different modes of counterfactual reasoning between the two forms of reading. Fiction, as a simulation activity (Oatley, 2011), enables a more personal, fluid, and emotional counterfactual imagining. This may therefore promote increases in these ways of engaging in CT.

The literature review suggested that reading fiction may drive increases to counterfactual reasoning (Black et al., 2018), and also suggested counterfactual reasoning is important for CT (Byrne, 2016; Hoeken et al., 2020). The engagement with different possible scenarios, perspectives, explanations, and views entailed in thinking counterfactually (McMullen, 1997), is furthermore highly analogous to calls for openness and the ability to engage with different perspectives in CT (e.g. Paul, 1987). This pathway, from fiction increasing counterfactual reasoning capacity, to this increase then furthering CT, was part of the original argument for the proposed association between fiction reading and CT to be tested through this research project. The findings across the four studies support it.

8.1.3. Imaginative capacities

In addition to counterfactual reasoning, a second proposed pathway by which fiction reading may change and increase CT capacities is also supported by the findings of this research; the imaginative and engaged nature of the fiction reading experience. In study four, the code *reflective perspective taking*, which was associated with fiction reading days, was characterised by imaginative descriptions of participants engaging with a multitude of perspectives. This contained imagining the experiences and views of others, highly analogously to the ways in which reading fiction is often argued to drive improvements to ToM (e.g. Kidd & Castano, 2013; Zunshine, 2006). What is of particular note about this finding from study four, is that the code was derived from the combination of participants' descriptions of their reading and CT in each daily entry. Thus, this code summarises a relatedness in participants' fiction reading and CT experiences in imagining the alternate perspectives. Further qualitative findings support

the role of imagination and engagement in CT. In study three, the imaginative facets of CT were strongly emphasised. Under the category *CT is an experimental process*, participants described a mode of thinking critically that entailed fully imaginatively immersing into issues and using that imagining to make inferences and to reflect. Similarly, the study four code *imaginative and challenging evaluation* contained descriptions of highly imaginative approaches to CT, with evaluations made using visual, simulative extrapolations from the texts read and problems thought about. This was very close to Braun and Cupchik's (2001, p. 96) list of experiential and textual properties: corporeality; sensory perception; spatiality; temporality; agency. Participants descriptions in study four, and in study three, included imaginative simulation across all dimensions in the Braun and Cupchik list, but furthermore these were not only elicited in questions on reading, but also those on CT. Thus, prior arguments ways in which fiction reading may train imaginative simulation (e.g. Oatley, 2011; Zunshine, 2006) are supported in the findings of the qualitative studies, as are arguments for imagination as a facet of CT (e.g. Bailin, 1987).

One of the ways imagination may form a link between reading, especially fiction, and CT is in the filling-in that participants described as essential to the reading experience in both study three and four. Fiction reading offered material which participants used to create broader, richer scenarios in their mind's eye. This active process of adding to what is contained in fictional texts while reading has also previously been argued to be central to the power fiction has, notably by David Lodge (2002) who argues the lacunae in fictional narratives offer the opportunity to exercise imaginative processes we then also utilise in building the stories of our own lives. For

participants, this imaginative extrapolative process was intimately tied to NT. In study three it can be seen under the category *narrative transportation integrates CT into reading*, within which NT was characterised as an active process that is as evaluative as it is imaginative, assimilating CT fully into the reading experience. In study four participants' responses to diary questions on reading evaluations and NT were grouped so as to reflect this interconnectedness, and again the interplay of imagination and evaluation is seen across the various codes, but with particular concentration in those derived from fiction reading days (*reflective perspective taking* and *open observation*). These codes entail imagination both in shifting into the possible positions of others, but also in envisaging alternative futures and histories, and in engaging in imaginative journeys across the unfolding timespan of fictional narratives and the CT experience. These qualitative findings offer some explanation for the causal relationships found in study two, as they interconnect imagination with evaluation.

8.1.4. Openness, veracity, and bypassing defences

Both counterfactual reasoning, and imagination, were deployed by participants in their reading and critical thought, and these could sometimes be prompted explicitly or indirectly by the contents of the texts they read. Examining the ways that fiction texts in particular could prompt CT in unique ways yielded further insight into the relationship between fiction and critical thought.

One of the most interesting ways in which evaluations were made by participants about the content of fiction reading was surrounding issues of veracity. In study three, participants articulated ways in which fiction uniquely challenged them to

think critically about its truthfulness, in concrete ways that nonfiction could not. For example, Participant B notes:

“An author will have their own values and biases which will probably come out in in their work, but we have to think about to what extent do we take things in that book as an author’s genuine views, for example, like if they use racist language is that a reflection on the author as a person, rather than a character they’re creating?”

(Participant B, study three)

This illustrates the way a fictional author’s invitation to make-believe complicates interpretations about that author’s other intentions. This can perhaps be a prompt for more complex evaluation, requiring a high level of imaginative modelling of the author’s possible intentions. The study four category *perspective balancing* also finds participants describing similar ways of evaluative thinking about authors within their fictional reading experiences. Furthermore, inside of this study four category, the code *open observation* was associated with fiction reading days. This code characterised a state of radical openness, which permitted participants to fully engage in the imagination of authors, and through this complete openness reach new perspectives and beliefs which often dramatically differed from their original views, in keeping with calls for radical openness in CT (hooks, 1994).

In the openness entailed in fiction reading and some experiences of CT, we can see a potential link to the study one characterisation of the EO as an interwoven space where evaluativist and multiplist perspectives can coincide. Furthermore, this could also be one possible explanation for the way that assigned fiction reading impacted nonfiction readers in study two; perhaps the openness facilitated by fiction, and the kinds of evaluations it can prompt, were particularly transformative for readers for

whom these were unusual experiences. This connection between the unusuality of fictional experiences, their complex relationship to reality, and an arising openness, fits with Bruner's argument that literary texts are concerned with the real world, but they "render that world newly strange, rescue it from obviousness." (Bruner, 1986, p. 24). He therefore argues that literature 'subjunctivises', making the world less fixed and less banal, and as a consequence, it makes us more open to intuition. This argument has also been put forward by many other researchers investigating the impacts of fiction reading (Mar et al., 2008; Miall, 2001; Oatley, 2011). The findings from this research project therefore support this pathway from the strangeness of fiction, to openness, and extend that openness into CT.

The efficacy of assigning fiction reading to nonfiction readers shown in study two is further interesting, as less openness to texts that violate one's habits could have been expected, and it could have been hypothesised that reading texts incongruous with one's preferences would be arduous and unpleasant, and thus may in fact negatively impact performance on the second round of the test by demotivating and fatiguing these participants. However, for participants who considered themselves to be nonfiction readers, the assigned fiction reading was associated with CT improvement. In the literature review, one key theme in the arguments for the power of fiction was in the safe, or 'offline', mode of fiction reading, which allows it to bypass defences (Djikic et al., 2009a), without the sense obligation carried by nonfiction (Bal & Veltkamp, 2013). This could be an explanation for why fiction reading was not demotivating, but it may also offer some explanation as to how fiction reading could drive positive CT change. As study one found fiction associated with greater CTD, but

not nonfiction, and as a comfort with different views is a part of the disposition to think critically as measured by the scale used (Sosu, 2013), this further lends credence to a connection between fiction reading and greater capacity to comfortably engage with different views. The value of fictional safety was strongly echoed by participants in study three, perhaps most notably by Participant G's articulation of how Terry Pratchett's fantasy novels may promote CT on real-world issues:

“if I was reading this and I was someone who maybe didn't have those values, I think I'd be less inclined to get defensive reading Pratchett, than reading a book set in the real world that clearly attacks the things I hold close to my heart.” (Participant G, study three)

However, it must be noted here that the power of reading to bypass defences was not only identified by participants within fiction. Interestingly, there were also ways in which all reading was described as having some safety in study three, for instance:

“in the pub if somebody says something you disagree with you might want to avoid an aggressive loud argument, which would be extremely uncomfortable [...] But if you're reading a book and you disagree with the author's argument, well that's not nearly as uncomfortable” (Participant J, study three)

This is perhaps a simpler version of safety than that offered specifically by fiction, but nonetheless this could be one avenue through which all types of reading can prompt personal change, as it can be safer to experiment with deeply held beliefs and stances without an adversary in the flesh. In study four, several codes capture ways in which participants could privately think through and evaluate their own positions with relation to what they had read, for instance in *experiential balancing* by weighing up their own experiences against alternative perspectives, which could be a gradual and

oscillating process without the need for immediate commitments to any position. In a sense, the stability of text offers an entirely open-ended timeframe for response, with scope to think critically and engage with it over and over, for as long as desired, which is inherently less pressurised than a ‘real world’ interaction.

However, there are also indications that fiction reading has stronger, more unique associations with the bypassing of defences, which go beyond the safety of text on a given subject in comparison with a real-world interaction on that subject. The study four code *open observation* described a state of openness that transcended a single domain, but was a more general frame of mind one could enter into, resulting in very wide ranging acceptance of diverse possibilities. It may therefore be the case that spending time in the safe, open mode of reading fiction allows for that un-defensive openness to extend beyond the boundaries of the text at hand. Indeed, as Djikic and Oatley (2014) have also found, fiction can bring about non-directed changes, perhaps by prompting this type of openness that is uniquely conducive to flux in how we think about and make evaluations on topics we have had prior beliefs about that would otherwise have remained fixed. This prompting of change was also deeply reflective in participants descriptions across study three and four. In this way, reading and CT together intermingled in personal reflections, suggestive of an avenue from reading towards highly reflective conceptions of CT (R. H. Ennis, 2015; Ward, 2006). Changes to perspectives and beliefs, and ultimately to ones’ self, were often very intimate and personal, and yielded profound shifts for participants. This chimes with arguments for the power of fiction in prompting non-directed changes: “we open ourselves up to greater possibilities for who we may become.” (Mar et al., 2008, p.

133). Linking this back to the argument that it is the temporal, sequential unfolding of a narrative that provides its power, Gray Hardcastle (2003) identifies narratives as an essential part of our development of a sense of self, as she argues stories enable the development of a framework that allows for temporal ordering and the anticipation of future possibilities. In this way, it may be that the fluid unfolding of the fictional narrative, throughout which reflective connections are made, provides an open vantage point not only upon the kinds of perspectives we could hold, but also upon our selves as we extend into an unknown future, making the possibility of change more real.

8.1.5. Narrative transportation

Finally, from the literature review NT emerged as the likely mechanism through which all of these simulative, imaginative ways of engaging in fiction reading would take place. However, the findings from these four studies complicate this assumption. While study one did find that NT mediated the impact of fiction reading upon CTD, it had no such mediation effect on EO. Study two found no mediation effect by NT at all, although upon more detailed exploration spending more time and making more entries of fiction reading at a peak of one's NT was associated with increased CT change, and peak NT entries (but not time) with likelihood of improvement in CT. It does, therefore, seem that NT plays a role in the relationship between fiction reading and CT, but this is not as simple as if being more transported always led to increased impact from fiction. Rather, being more transported may be part of the pathway from fiction to CT in some more specific ways.

The finding that more entries of reading made at a peak NT level was associated with improvement CT, while time spent reading fiction at a peak NT level did drive

change in CT but not necessarily improvement, aligns with the previously discussed overall relationship between fiction reading entries and time. It may be that each instance of entering fully and being completely transported by a fictional narrative confers a benefit, regardless of how long one spends in that state. Given the lack of mediation effect in study two, it also seems likely that NT does not influence the relationship between fiction and CT in a linear way (i.e. it is not the case that as NT increases this relationship is strengthened), but rather it may be that very high NT is distinctive from low or no NT, and thus it is reading one's personal peak NT level that is important. This is commensurate with the ways in which participants in study three described their NT, as participants either described a disconnection with text when NT was absent, or described a very rich NT experience, but no findings emerged for any experience of gradual or relative increases in NT being meaningful. This contrasts with the study one mediation effect upon fiction and CTD, which suggests that when it comes being disposed to think critically, increases in NT may influence the impact of exposure to fiction. These findings taken together suggest a complex interplay between reading fiction, NT, and different facets of CT.

Based upon the literature review, it was assumed that NT is a departure away from the real world, and as such being highly transported would not involve CT during the transportation (Gerrig, 1993). Thus, the proposed pathway from fiction reading to CT was one where NT was centred entirely on the fiction reading, and formed the basis for kinds of simulation training fiction is thought to provide (Oatley, 2011; Zunshine, 2006), and the outcome of this training would then yield improvements to factors that could feed into CT. However, study three found NT and

CT to be enmeshed. As evaluations have been found to take place throughout fictional reading (Özyürek & Trabasso, 1997), this supports the study three finding that NT as an active, evaluative experience. Therefore NT is not best thought of as a departure away from reality (Gerrig, 1993), but as a process that integrates the time-extended, imaginative, engaged experience of reading fiction directly with CT. Therefore a clean step-by-step causal explanation from fiction, through NT, to CT is not warranted in either the quantitative or qualitative findings, but instead a more complex and entangled relationship emerges between these factors.

8.1.6. Fiction and critical thinking connection limitations and summary

The essential limitation in all of the above arguments for the connection between fiction reading and CT, is that these proposed links do not apply exclusively to fiction. It is not the case that nonfiction reading always takes a post-facto perspective on events, in contrast with a solely fictional mode of moving through the time course of events. Many nonfiction texts describe events in an unfolding, flowing manner. However, there is perhaps a sense of closure entailed in the nonfictional association with reality, or inviting belief, in contrast with the fictional openness in inviting make-believe. It could be argued that even a flowing nonfictional narrative is leading to some conclusions, and that the conclusions form an end-point from which the narrative journey is ultimately to be cast. For example, reading Jon Krakauer's *Into Thin Air* takes us through a temporally unfolding narrative of events leading to a catastrophe on mount Everest, but taking this narrative journey ultimately leads to some conclusions about why this disaster took place, and thus there is ultimately a nonfictional post-facto accounting of events. It is the latter that grounds this text in reality, and provides its

nonfictionality. By contrast, Michelle Paver's *Thin Air* is analogous to Krakauer's similarly named work in Himalayan location and themes of the vicissitudes of mountaineering, however as a fictional horror story it is only the experience of the gradually developing fearful narrative that the novel invites one to make-believe, there is no concluding terminal vantage point. In this way, there may be a difference between fictional and nonfictional works even when they both share similar narrative time-extended descriptions of events. However, there will certainly be counterexamples, and a liminal space will always exist between these categories. Therefore the relationship argued for between fiction and CT is perhaps best thought of as an argument for fictionality and CT; i.e. what has been discussed here is the potential influence of reading fictional narratives inviting make-believe, which may sometimes also be found in nonfiction texts. These issues will be discussed further in a later section of this discussion.

In summary, fiction was found to provide experiential, time-extended, imaginative, reflective and connection-making experiences that entailed CT. These findings support views of fiction reading as a simulation activity that yields changes in approaches, perspectives, and oneself, and furthermore extend these proposed changes into CT. The causal pathway from fiction to CT may include increased counterfactual thinking and imaginative capacities, and it may be the temporality of fiction and its complex relationship to reality that makes it particularly suited to drive these changes.

8.2. Nonfiction builds informed critical thinking procedure

A simplified portrayal of the distinction between fiction and nonfiction related to CT, would be that fiction is associated with reading and thinking in a way that

emphasises continuous experience, while nonfiction is associated reading and thinking focused on obtaining an end. This fits neatly with reader response theory's distinction between "aesthetic" and "efferent" reading (Rosenblatt, 1982), and Hampson Lundh et al.'s (2018) analogous "informational reading" versus "experiential reading" approach distinction. The study three theme *nonfiction builds informed CT procedure* fits this distinction, and can be used to group findings from across all four studies such as to highlight this contrast with the previous theme of *fiction gives experiential CT understanding*. However, this dichotomy between experiential versus procedural is not absolute, and there is more nuance in the ways nonfiction reading was found to relate to CT across the four studies. Firstly, the ways in which findings pertaining to nonfiction reading align with this dichotomy will be outlined. This will be followed by a discussion of the ways in which these findings extend beyond such distinctions.

Study one found that nonfiction readers were more likely to be absolutist than to fall into either other EO category. This suggests an association between nonfiction reading and a view that most questions and issues have true or false solutions, which are complete. This finding therefore aligns with an association to efferent/informational reading, as a means of seeking such true answers. This is seen in the study four code *seeking answers in information*, in which participants articulate very clear ideas of what they wished to obtain from particular instances of nonfiction reading and CT. Furthermore, in study three the reliability of nonfiction sources, and the reliability of CT as a procedure, were both attributed to their concision and explicitness, also aligning with a view of a demarcated procedure focused on a clear end point.

However, the boundaries around nonfiction reading and CT were not as rigid upon closer examination. The study four code *seeking answers in information* furthermore contained an enmeshed reading and thinking, as participants often drew questions from their reading about which to think critically, and also used reading as a means of answering questions from their CT. This was also seen in study three, under the categories *nonfiction provides direct information*, and *critical thinking is procedural not personal*, as within both nonfiction reading and the procedure of CT participants moved back and forth between seeking relevant information, and utilising that information to further their thinking. The way participants described building their knowledgebase by extracting information from their nonfiction reading was characterised by experiences of empowerment in their CT, which in turn enabled more confident information seeking. Furthermore, not only did participants describe building knowledge, they also described this highly directed and extractive reading for information as a means of improving their CT procedures, as seen in the study three theme *nonfiction builds informed CT procedure*. There is an interesting parallel here between the ways in which fiction reading is argued to train social ability (e.g. Zunshine, 2006), and this way in which participants described their nonfiction reading training their applications of very procedural CT approaches. In the case of fiction reading, de Nooy (2001) argues that people derive scripts for social action from narratives. It may be that people also derive scripts for performing step-by-step CT procedures of the kind described in study three as *critical thinking is procedural not personal*, from nonfiction reading.

Subsequently, it would be reasonable to expect improvements in CT resulting from increased nonfiction reading. Indeed, in study two spending more time reading nonfiction was associated with greater likelihood of having an improvement in CT score (although there was no association between nonfiction reading and the amount of change in CT score). Based upon this finding, the supposition that nonfiction reading improves CT is supported. However, when measured in entries made by participants, rather than in time logged, nonfiction in fact had a negative association with improvement in CT; i.e. logging more entries of nonfiction made it more likely to have a decreased or unchanged CT score. This apparently paradoxical finding contrasts with the previously discussed way in which fiction was effective in terms of more entries, but not necessarily reading time. It seems that when it comes to nonfiction, more instances of reading are not conducive to CT, but spending more time is. Exploratory analyses supported this by finding that more time spent on each nonfiction entry was associated with higher likelihood of improved CT (though again, no impact was found on CT change). This therefore suggests that the way in which nonfiction may build knowledge and CT procedure is not a simple question of gathering information from nonfiction texts, as this would imply a strategy of gathering as much as possible from as many texts as possible would be successful. Instead, a longer engagement with each nonfiction text is needed, suggesting that this extraction of information is not a simple matter of the more the better, but perhaps the deeper the better.

Another finding from study two appears to complicate, but perhaps suggests a direction for explaining, the other nonfiction findings from the study. Exploratory

analysis found that reading nonfiction at a peak of one's NT resulted in an increase in CT score change, but was not associated with CT improvement. This contrasts the previously discussed findings where the outcome of CT improvement was influenced by nonfiction, while change in CT score was not. For the latter, it seems that being transported into the nonfiction text is important. However, highly transported nonfiction reading prompts CT change, without always improving CT. Furthermore, NT had no mediation effects for nonfiction in either study one or study two, which implies that the way NT interacts with nonfiction reading is not a simple magnifying relationship. Perhaps reading nonfiction at a peak of one's NT is a different type of reading experience, prompting a different CT approach. It seems that some ways of reading nonfiction can lead to increasing changes in CT, and other ways of reading nonfiction can lead to improvements in CT, and these are distinctive processes. Concretely, highly transported nonfiction reading was the only way nonfiction showed the same effect upon CT change as fiction, suggesting a commonality with fiction reading brought about through high levels of transportation.

These differences in how nonfiction can be read, and how different nonfiction reading experiences may related to differences in CT, is reflected in the qualitative findings. In addition to the previously discussed findings that align with efferent/informational understandings of reading nonfiction (Hampson Lundh et al., 2018; Rosenblatt, 1982), considering nonfiction reading with NT opens a more experiential way of understanding nonfiction reading. The study four code *imaginative & challenging evaluation* incorporated both fiction and nonfiction reading days, and it

was notable than with both types of reading NT took place and formed a grounds for thinking critically, as exemplified by participant JGSB:

“I was completely immersed. Yes, I could imagine myself in the many meetings and news conferences. [...] Although I very much enjoy reading about President Trump's many mishaps, I think that the media (in the main) has a significant bias against him, and cherry pick anecdotes that present him in the worst light whilst failing to report any positive stories [...] These thoughts go through my head as I read these kind of article.” (JGSB, study four)

Here we see an interplay of imaginatively entering into the scenario of the text, while evaluating the issues. Furthermore, the code *experiential balancing* included both fiction and nonfiction days, and entailed participants stepping into the perspectives of others in order to then critically weigh up and order their responses, thus combining a ToM type of imagining more commonly associated with fiction (e.g. Zunshine, 2006), with a very goal-oriented informational reading approach. Here then, the boundaries between experiential and procedural reading and thinking are blurred, and NT is connected with nonfiction. These findings would suggest that NT interacts with nonfiction reading experiences, and while not increasing the likelihood of improvement in CT through this interaction, it may nonetheless yield to changes in CT. The literature review found some association between NT and nonfiction reading in previous research, with NT cast as a driver of engagement with nonfiction texts and the issues they discuss (Borum Chattoo & Feldman, 2017), and these findings offer support for this association. These descriptions of NT in more imaginative nonfiction reading given by participants were very distinctive to the procedural capacity building already discussed. These may therefore suggest two different pathways from reading

nonfiction; one leading to increased CT change through heightened transported imagining, the other leading to CT improvement through building knowledge and procedures.

This research project was focused on the role of fiction reading in changing and potentially improving ways in which people think critically. Nonfiction reading was not a variable of interest, but rather was selected as a comparator for fiction in order to allow fiction-specific impacts to drawn out in contrast to nonfiction. Based upon the literature review, I argued that nonfiction reading would feed into CT through the building of a knowledge base, but that fiction would have a stronger association with CT as it could be connected through a broader array of pathways (including social and imaginative capacities). The findings of the four studies support this initial argument to the extent that the proposed pathway from nonfiction to CT improvement is associated with knowledge and procedure building. However, the possibility of a second pathway from nonfiction to CT change through imagination and NT suggests a way nonfiction reading can share some characteristics and effects with fiction reading. The extent to which different modes of experiencing nonfiction may be associated with different ways of changing CT, and how similar or distinctive these may be from fictional experiences and effects, is an area that would be suitable for future research.

8.3. Nuance in reading materials

The findings of this research project demonstrate some broad differences in how different forms of reading can be related to different ways of thinking. However, there are also many further ways of sub-dividing these materials, liminal spaces between such divisions, and interconnections between them that create nuance worth exploring.

8.3.1. Diversity of reading

The non-directed and continuous ways in which reading was connected to changes in participants' thinking in the qualitative studies suggests that different cases of reading can have very different impacts. It is not the case that, for example, reading *Loneliness Is in Your Blood* by Cadwell Turnbull will always result in change in CT. Different connection-making will likely take place during reading the story for different participants, and with different effects arising. However, study two did find a significant trend between assigning such short stories to nonfiction readers, and change as well as improvement in their CT scores. Indeed, when it came to likelihood of having an improved CT score, with reading time controlled for, assigned fiction to nonfiction readers had the highest effect size in this research project. It is interesting that the assignment of reading was only significantly impactful for fiction given to nonfiction readers; no other group showed any impact of their assigned reading. A possible explanation is that reading fiction when one is not habitually accustomed to doing so is a more powerful experience. Indeed, it is the power of fiction to defamiliarise, or make things strange, that has been argued to drive its impacts (Bruner, 1986; Miall, 2008). Furthermore, it seems likely that fiction readers would already have been reading nonfiction also, and since the assigned nonfiction readings were magazine articles these were of a very common type that all readers are likely to have been encountering. By contrast, fiction short stories may be a more rarefied material. In study three, Participant N (a self-described nonfiction reader) directly commented this upon being gifted a short story collection, and described the stories as highly impactful. Therefore one key difference between the assigned fiction and nonfiction texts in study two may be in their distance from readers' usual types of texts. This is further

supported by participants in study three, who strongly associated diverse reading with impacts on CT. This was most prominent within the code *reading as widening exposure*, within which a very commonly held view was summarised by Participant J as: “I think if you read broadly – so on many different subjects from many different points of view – then that’s going to broaden out your thinking too.” This perspective aligns with prior research suggesting reading, especially fiction, can permit a widening of cultural perspectives (Hakemulder, 2001). Furthermore, participants specified the necessity to exit one’s comfort zone, and read material contrasting with one’s usual habits. This may be precisely what the assignment of the fiction short stories to nonfiction readers in study two accomplished. It would be interesting to follow this up with nonfiction texts that are less likely to be familiar to participants, rather than magazine articles that are very commonplace. This would permit a differentiation of the impact of the novelty of the type of text, from the fictional or nonfictional nature of the text.

Another way in which participants in study three suggested diverse reading may be beneficial was in reading different genres. However, in study two the assigned short stories were deliberately selected from varied genres, and it is not possible to separate out the effects of any one story, thus no genre impacts of the assigned reading can be tested. In study one, the ART-G does permit the differentiation of the fiction recognition score into genre (Mar & Rain, 2015); however, as this creates 6 additional predictor variables large sample sizes would be needed to achieve statistical power in testing them (Bujang et al., 2018). Therefore these analyses would not have yielded robust results from the current sample. Similarly, the material participants logged in

study two can also be classified by genre, however again the proliferation of predictor variables raises participant number requirements for statistical power beyond the obtained sample (Knofczynski & Mundfrom, 2007). An investigation into the role of fiction genre with regard to the effects found in study one and two, therefore, remains for future research. It should also be noted that fiction genre has been found to have different impacts in previous studies (Black et al., 2018; Fong et al., 2015). Indeed, fiction genre has been associated with different personal constructs connected to members of different ethnic groups, genders, and ages (Birdi, 2011), which are the kinds of differences fiction is often argued to overcome (Hakemulder, 2001). These are also the kinds of differences participants in study three specifically emphasised as important in the ways reading fiction could diversify thinking. These findings further indicate that future research into genre effects following from this research project would be worthwhile. However, it should also be noted that differentiating fiction genres poses major issues around defining and classifying texts into these categories, which have many grey areas and liminal spaces between them (Hider & Spiller, 2020). Indeed, each category added also adds borderlands between it and others, thus the finer grained a division to be made the less clear that division will be.

Another possible division that could be made within fiction is between literary and popular texts. Researchers investigating the effects of fiction commonly focus on the literary, and argue for its unique value. For instance, Cupchik and Leonard (2001) argue that we relate high culture to other past art, and expect it to challenge us. By contrast, they propose we approach popular culture works within the immediate timeframe only, and we expect catharsis, or emotional evocation, from them. Yet, this

differentiation in the temporality of art versus popular culture is not borne out by the findings of the qualitative studies, which found participants integrated all kinds of fictional texts into challenging shifts of perspective across timeframes. Very similarly, Oatley (2011), distinguishes “pastimes” from works of art, arguing the former follow a direct and predictable emotional trajectory so when we read them “we have run a schema of habit, not much different than putting on our clothes in the morning” (2011, p. 117). Art, for Oatley, violates the habitual and familiar, and presents us with emotional trajectories that are unfamiliar and sometimes uncomfortable. Again, the ways in which participants discussed how reading could shift and challenge their beliefs, pushing them outside of their comfort zones and habits, was not only connected to literary or artful works. This suggests a tension between the findings from this research project, and the division made between literary and popular fiction, which would be worthy of further investigation. However, as noted in the case of genre, differentiation between these two categories will likely be problematic.

8.3.2. Fiction and nonfiction

While the findings of this project highlighted differences in fiction and nonfiction reading and their relation to CT, there were also many commonalities. As was discussed in the introduction, fiction and nonfiction do not have clear boundaries between them. Therefore rather than attempting to use some definition based on necessary and sufficient conditions to clearly distinguish them, an approach of looking for features and practices associated with each kind of material was taken (S. Friend, 2008). To this end, Currie’s (1985) intentional approach was adopted, under which fiction is that which invites make-believe, while nonfiction invites belief. Study four in

particular supports this approach to dividing fiction and nonfiction, as codes derived from nonfiction reading days were strongly rooted in reality and highly contextualised, while codes associated with fiction reading days were connected with imaginative perspective shifting; in this way it is possible to see how nonfiction reading material was connected with participants' CT around beliefs, and fiction reading was linked to creative make-believe CT simulations. However, study four also shows how these can be blended together, as ultimately the categories derived from these codes were intermixed between fiction and nonfiction. Thus, the way fiction and nonfiction have been discussed throughout this thesis ought not be understood as a strict demarcation, but rather as a grouping around the characteristics of inviting make-believe or belief, with a necessary range of how tightly around these centres texts may cluster.

The common ground between fiction and nonfiction features within the study three theme *reading as personal reimagining*, within which both make-believe and belief appear. From this common ground, the distinctive associations between fiction, nonfiction, and their different relationships to CT can be better understood, as considering ways in which texts can elude these categories can also help to clarify their shared and distinctive effects. The cases that arguably complicate fiction and nonfiction classification most are those which contain both invitations to make-believe, and to believe. For example, in science fiction some authors are highly concerned with the scientific accuracy and plausibility of the technologies and events that they weave into an otherwise fictional narrative; e.g. Stephenson's (2016) novel *Seveneves* merges a fictional storyline with highly detailed accounts of physics (Elvik, 2018). The author's intentions can be mercurial, shifting from passage to passage. The reader, too, seems to

be able shift from make-belief to belief within a given text (S. Friend, 2008). Indeed, participants in study three actively described ways in which a fictional author's invitation to make-believe complicates interpretations of fictional texts relationships with veracity, and about that author's other intentions such as delivering historical or social messages. This complexity in unpicking invitations to make-believe and believe was part of what drove CT in fictional reading. Likewise a nonfiction book can contain fictional elements that invite make-belief, as shown in study three and four's participants imaginatively entering and expanding nonfiction scenarios. Many of the devices associated with fiction (such as giving insight into a figure's thoughts that would be impossible for the author to know) are often deployed in nonfiction writing (S. Friend, 2008). Thus, it may be that in texts such as these, both the effects of fiction and nonfiction reading are present, building both experiential and procedural CT approaches. For the purposes of future research, it may thus be more interesting to consider fiction and nonfiction as quantities with any given text, rather than categories to be applied to a text as a whole.

In keeping diaries of their reading, participants in study two and four included many kinds of texts, such as novels, poetry, memoirs, plays, textbooks, news articles, instructional guides, nonfiction books, etc. In study three, participants also chose varied types of texts in between the two interview sessions, and in conversation drew from many kinds of examples to illustrate their descriptions. By contrast, the ART-G used as a measure of text exposure in study one includes specifically authors of books (Mar & Rain, 2015). This firstly suggests a major limitation in using the ART-G as a measure of reading engagement. Secondly, this opens questions around which forms of texts

both within fiction and within nonfiction may have different, or more impactful, effects. Hawkins' (2012) research on voluntary reading impacts on CT in undergraduates compared books and "nonbooks" (the former including electronic book formats as well as print, and the latter being comprised of social media, personal communications, and internet reading), and found book reading to be predictive of higher CT test performance but not nonbook reading. This finding suggests a difference between the impacts of book reading and some other types of texts, and this may be more pronounced when it comes to nonfiction, as the majority of fiction logged across studies two and four was likely to be in books (novels, poems, etc.), while far more variation can be seen in nonfiction as this spans many non-book formats such as articles in print or online, blog posts, and wider opinion pieces. Indeed, non-book fiction examples are fairly rare, such as literary journals, or online presentations of poems, short stories, etc. It may be that some of the findings pertaining to nonfiction in particular, therefore, will be subject to differences between book and non-book reading. Such differences may be commensurate with the study two findings on time spent per instance of reading; one may be more likely to spend more time reading a book than an online article. There may also be differences in NT across these different mediums, as it is frequently argued that online media promote more fragmented attention and increase cognitive load (Macedo-Rouet et al., 2003), and differences have been found in the way we process electronic versus print text (Hou et al., 2017). It may be therefore by simply containing fewer online possibilities, fiction reading captured more book reading than nonfiction, and this was inherently more amenable to more focused, transported reading experiences. However, nonfiction reading could be highly transported, as seen in both the quantitative ranking on transportation in

study two, and in participants descriptions of their reading experiences in studies three and four. Conversely, NT only mediated fiction effects in study one, where only book reading was considered, implying NT was not important in nonfiction book reading. Overall, this is an area requiring further investigation, to separate out the effects of book versus other formats, and online versus other mediums.

As the previous research on the impacts of reading has been on a trajectory from investigating the outcomes of reading overall (Stanovich & Cunningham, 1992), through to disentangling fiction and nonfiction effects upon social skill (Mar et al., 2006), towards then unpicking genre impacts (Fong et al., 2015), this research project can be situated within this pathway from broad to narrow investigation. As such, it falls within the scale of work such as that conducted by Raymond Mar and his research group, which focused on fiction and nonfiction and their comparative influences on social skill, but applies this logic to the investigation of fiction and nonfiction in relation to CT. Thus, further research may pursue a finer grained approach, and explore the role of genre, and the role of different types of reading material more focused manner.

9. Conclusion

This research project aimed to address two overarching questions:

- Are there differences in fiction and nonfiction readers in their CT approaches?
- Can reading, particularly fiction, cause increases in CT?

These questions were explored through four distinct studies, two quantitative and two qualitative, comprising a sequential mixed methods approach. Through combining the findings from across the four studies, some answers can be given to these questions.

Firstly, fiction and nonfiction readers were found to differ in their CT approaches. They were found to have different underlying EOs, with fiction readers most likely to be evaluativist than absolutist, and nonfiction readers likely to be either evaluativist or absolutist rather than multiplist. These orientations were found to underpin different starting points to CT, with all orientations used to think critically, but with differences in context and approach. Furthermore, fiction readers were found to be more disposed towards CT, and higher engagement with fiction (but not nonfiction) was associated with higher baseline CT test scores; these results imply differences in CT between fiction and nonfiction readers. Additionally, the experience of reading fiction was associated with an experiential, ongoing process of CT, which was deeply reflective and entailed openness in shifting one's frame of reference across different perspectives, resulting in personal change. Nonfiction reading, by contrast, was more associated with a bounded, goal-oriented information seeking approach that could build knowledge and procedure for CT. However, nonfiction reading could also take a mode closer to that of fiction, with highly imaginative and transported

reading experiences associated with more experiential aspects of CT, suggesting that the distinction between fiction readers and nonfiction readers is often blurred. Instead of fiction versus nonfiction, fictionality and nonfictionality within texts may be more indicative of distinctive CT effects. Subsequently fiction and nonfiction readers may be better differentiated on a gradient of the kinds of texts they prefer across a range rather than categorically.

Secondly, different causal relationships were found for fiction and nonfiction. When assigned short fiction and nonfiction texts to read over a two week period, nonfiction readers who were given fiction had a significant improvement in their CT test scores. This suggests that reading fiction when it is not one's normal habit to do so can cause increases in CT. Furthermore, reading outside of one's habits was affirmed by participants as a means of increasing CT, and the complexity of fictional texts' relationship to truth, and to their authors' beliefs, was seen as a driver for CT. Fiction's power to bypass our defences, and prompt CT circuitously, was also identified as a way it could promote CT. When comparing amounts of fiction and nonfiction read over a two week period, different connections to changes in, and improvement to, CT were found. Change in CT was predicted by more instances of fiction reading, and by more time spent reading fiction; spending more time per item of fiction was also associated with CT change. Therefore fiction reading does cause change in CT in this context. Improvement in CT was predicted by more instances, though not more time, of fiction reading; more time spent per fiction item was however predictive of CT improvement. Thus, engaging in fiction reading more times, and spending longer in each engagement, was found to cause CT improvement. This is also seen in the way

participants linked both diversity of experiences, and the depth of engagement, as important in fiction reading. The way reading fiction immerses the reader into the timeframe of the fictional narrative may drive the power of reading fiction to change CT, as by stepping into the narrative time course more shifts in mental perspective may be possible, simulating the experiential shifting and fluidity of CT.

Nonfiction reading was also found to have a causal relationship with CT. More instances of nonfiction reading were predictive of one's CT score remaining the same or decreasing, but more time spent reading nonfiction predicted CT improvement, while more time per item of nonfiction was also predictive of improvement in CT. Therefore reading nonfiction for longer periods of time can cause improvement to CT, while larger numbers of very short nonfiction reading sessions may in fact be detrimental. Participants articulated ways in which slow detail-focus is essential in procedural, logic-based CT approaches, and in informational nonfiction reading, which may explain why longer reading times are required for nonfiction to be beneficial. It is interesting that CT change was not influenced by nonfiction, while it was by fiction. Fiction, but not nonfiction, therefore, can cause flux in CT, and this may be non-directed, and open-ended. However, when taking NT into account, reading at a peak of one's NT in the case of nonfiction also caused change in CT, suggesting that this non-directed promotion of change is not exclusively a feature of fiction, and can occur in nonfiction also. This may be a feature of highly transported reading states, which participants associated with features of fictionality, but these features can also be present in some nonfiction texts.

These findings show that differences in fiction and nonfiction reading were associated with differences in CT. Reading fiction was shown to cause change in, and improvement to CT, in different ways to nonfiction. In these ways, the findings from the four studies supported the arguments that originated this research project, derived from literature on how fiction can improve social and cognitive capacities. However, findings across the four studies also challenged some of the originating assumptions that were part of the research design, and prompted change in how some of the original arguments were cast.

Firstly, an important framing for this research project lay in the way CT was conceptualised. Synthesising the way content evaluation is discussed in IL, with wider conceptions of CT in the philosophy, education, and psychology literature, CT was construed as follows:

CT as discussed throughout this document entails both trait and state manifestations of how one evaluates information, incorporating logical reasoning about its consistency and possible bias, with an openness and sensitivity to points of view, an ability to imaginatively simulate other viewpoints empathetically, as well as an internal reflective sensitivity to one's own views and reasoning processes and epistemological beliefs, and broader creative and imaginative capacity to engage with and model alternative possibilities, actively and effortfully aimed towards a creative outcome that could be the generation of new information, behaviour change, or shift in belief. (Thesis section 2.1.)

The findings from the four studies support this view of CT. Overarching (trait) manifestations can be seen in the factors of CT measured in study one, and in the ways participants in study three described their CT overall. Case-by-case (state) expressions of CT can be seen in CT test responses in study two, the diary entries of study four,

and also in the specific examples participants gave in study three. Logical reasoning was present in the ways participants described engaging in CT in both qualitative studies, and so was an experiential, open, perspective-taking way of thinking critically. Imaginatively engaging with different perspectives, and different possibilities, was also a key feature of the qualitative CT findings. Fluidity was perhaps underestimated in this original conception, as the flux of experientially traversing through CT as described by participants was not captured here. Furthermore as epistemological beliefs were found to be highly fluid, and this was linked to reflection and personal change. Overall, this way of understanding CT, and content evaluation in IL, held up throughout the research process. This conception of content evaluation and CT, with an added emphasis on fluidity, is therefore valuable for researchers and practitioners seeking to understand and promote CT.

In the original research design, NT was conceived as a mechanism via which reading could impact CT, and as such was treated as a distinct factor separate from reading and CT. In designing the quantitative studies, this resulted in NT being construed as a covariate that may play a mediation role. In study one this was found to be the case for the association between fiction and CTD; however no other mediation effects by NT were found in either quantitative study. As NT also had direct effects in study one, and reading fiction and nonfiction at a peak of one's personal NT level was impactful in study two, this suggested that NT was not simply an additional step in a possible path from reading to CT, but rather could play direct role in reading, and in CT. In study three, the way participants described NT contradicted a view of it as a distinct and separable factor. Instead, NT integrated CT into the reading experience,

and was a highly active link-making process. This shifted the way NT had been framed in the original research design, and this change in conceptualisation was applied in the coding process of study four, with the experience of evaluation in reading and of NT coded together, yielding insights into how transported reading and CT changed from day to day. This approach that combines NT with reading and CT, may be fruitful for further conceptualisations of, and research into, the relationship between reading and CT.

As part of the argument that formed the basis of this research project, I made a case for Bruner's (1986) narrative and paradigmatic reasoning modes to not be viewed as fully separate and entirely compartmentalised. Rather, I proposed that given the highly reflective, empathetic, and creative conceptualisations of CT identified through the literature review (e.g. Bailin, 1987; Paul, 1987; Thayer-Bacon, 2000), paradigmatic reasoning must also tap into features of what Bruner would consider narrative reasoning. These results have supported this supposition, and furthermore have shown how narrative reasoning also draws from paradigmatic reasoning, as NT bridges reading and CT. Such a breakdown in these boundaries is supported by the findings of this research. However, this original argument proposed a single direction of influence from narrative to paradigmatic. A further breaking of boundaries was in fact found, exceeding beyond the originally proposed ways in which fiction reading experiences could feed into CT. The complex issues of veracity in fiction suggest an informational, belief-oriented paradigmatic thinking entering into the narrative or aesthetic realm. The way NT was characterised as highly evaluative also implies a flow of paradigmatic reasoning into narrative experience. Furthermore, fictionality, which prompts make-

believe of narratives in an aesthetic experiential way, can also be present in texts that set out to invite belief and informational reading with paradigmatic thinking. Thus, some of the effects found for fiction were also present in nonfiction. Overall, these boundaries and dichotomies are not entirely shattered by the findings of this research, as distinctive ways of thinking critically, and different associations with fiction and nonfiction reading were shown. This thesis does not, therefore, refute Bruner's modes of thinking, or Rosenblatt's reader response theory. Rather it demonstrates interconnections and avenues of influence between them, implying wider interplay between make-belief and belief. Therefore, considering the relationship between reading and CT in the ways that have been described here, it may be wiser to think of dichotomies such as fiction/nonfiction, experiential/procedural CT, as poles around which certain experiences and approaches may be clustered, but which can influence each other.

9.1. Contributions and implications

9.1.1. For theory

This research has demonstrated a causal link between reading fiction, distinct from the impact of nonfiction, and CT. These findings contribute to the body research empirically studying the effects of fiction reading as an individual difference (Stanovich & Cunningham, 1992). Furthermore, fictional texts given to habitual nonfiction readers have been found to be an efficacious intervention for improving CT. This finding furthers those of previous studies utilising fiction as an intervention (D. R. Johnson, 2012; Kidd & Castano, 2013). These are novel findings that further understanding both of the impacts of fiction reading as addressed in psychology and literary studies, and of influences upon CT of interest across wider disciplines.

This research contributes to the understanding of several constructs and furthers the way they can be conceptualised across fields. Firstly, findings have enriched a conception of CT, and concretely its relationship with IL within LIS. This research was conducted within the context of Information Studies, and specifically IL, as the subject focus. Though CT was adopted as the key construct of interest, this encompassed content evaluation within IL, which I have argued is the richest and strongest sense of IL (Hollis, 2019). As such, the findings of this research are relevant to IL researchers. At the outset of this research project, I aligned with Hampson-Lundh and colleagues' argument that "Information Literacy research would have much to gain from reintroducing an empirical focus on reading activities" (Hampson Lundh et al., 2018, p. 1043), and this research project was in part a response to that call. However, the findings from this project further emphasise the need for reading activities to be considered beyond an informational focus in IL research. The relationship between fiction reading and IL warrants further investigation. Furthermore, the way CT was characterised in Box 1 derived from the literature review contributes a useful conception bridging LIS and wider disciplines. Secondly, the empirical findings of study three and four, providing readers' experiences of CT, offer support for the conception derived from the literature review while furnishing it with further nuance. As CT has been increasingly emphasised in modern definitions of IL (CILIP, 2018), the way CT has been described in this research offers an important contribution to how it can be understood in this field.

Additionally, based upon this research more nuanced understandings of EO and NT are offered. The findings of study one and three demonstrate an interplay between

traditional EO categories (Kuhn et al., 2000), finding these to be inter-related and deployed by individuals in different ways in different cases as a basis for their CT. This challenges a view that evaluativism is a superior, or developmentally final, orientation and only this orientation can be conducive to CT (Hofer & Pintrich, 1997; Kuhn et al., 2000). Researchers working with the concept of EO would therefore benefit from the findings of this research in shaping their approaches. In terms of NT, this research does not offer a single clear finding on its relationship to CT, but complicates previous assumptions about NT impairing CT (e.g. Busselle & Bilandzic, 2008). As participants experienced NT as integrating evaluative thinking into their reading experiences, and as higher NT was associated with higher CTD, this research provides some evidence for a positive relationship between the two. However, hypothesised mediation by NT in study two was not found, thus this is not a simple relationship. This complexity is worth further investigation, and may be useful for researchers working towards deeper understandings of NT.

9.1.2. For practice

CT is argued to be necessary for, and foundational to, democratic engagement (Facione, 1990a; Frímannsson, 2016; Gainer, 2012; Siegel, 1997); likewise IL (CILIP, 2018; Hall, 2010; Pawley, 2003; Reece, 2005; UNESCO, 2003). This is emphasised by current concerns surrounding “fake news” (Batchelor, 2017; Machete & Turpin, 2020). Given this importance, any practical means identified as an avenue for improving CT ought to be valued and promoted. As this research has identified ways in which reading fiction can yield changes and improvements to CT, it presents

evidence for the value of fiction reading as an activity that can be promoted in a range of settings.

Instructors and practitioners seeking to improve CT in library contexts could utilise fiction. Specifically, the way fiction has been shown to be impactful without instruction, and the non-directed nature of its power to drive CT change, implies that an informational reading approach applied to the specific subject content of fiction in IL instruction (e.g. Harlan, 2019), misses the potential of aesthetic fiction reading in impacting CT. Rather than using fiction in IL instruction with a focus on pre-set expectations of what students should extract from the fictional text, a more open approach to assigning fiction as an exploratory device inviting reflection and CT is recommended (e.g. Prinsloo, 2018). Furthermore, this also implies that while fiction can be utilised to promote IL around specific subjects, e.g. ecology (Via Rivera, 2019), sex and gender (Lips, 1990); it could also be used more broadly without this content-focus, but with emphasis on the experiential journey narratives enable and the very individual impacts these have. This presents a novel direction for the use of fiction in IL instruction, suggesting a more open-ended engagement with fictional material without specific subject-focus as a new way of including reading material either in-class or as assigned reading. Furthermore, using fictional materials to promote CT in this manner does not require instructional techniques which LIS practitioners outside academic settings can find off-putting (Harding, 2008). Thus all practitioners could find ways of promoting fiction engagement, and through this promote CT, without any requirement for active instruction on *how* to approach the fictional text nor *what* should be gained from it.

The fact that fiction has been shown to drive change and improvement in CT, and that CT is vital for social justice and democracy, implies that exposure to fiction can be seen as a social good. Furthermore, the individual and non-directed effects of fiction suggest that ongoing encounters with fictional texts offer an avenue for continuous personal change. Exposure to diverse texts, which contravene habits, and which can bypass defences, may also be important for CT. Specifically contravening the reading habits of nonfiction readers by assigning them fiction was found to be efficacious for promoting CT; in educational settings, targeting students in subjects where fiction is unlikely to be a regular feature, and assigning these students some fiction reading, would therefore be recommended. Unlike existing uses of fiction in these subjects where it is sometimes used to teach content (e.g. accounting: Crumbley & Smith, 2000), a non-directed assignment of fiction reading may offer wider avenues for CT development. Beyond educational contexts, enabling people to access a diverse range of fiction texts across their lifetimes can therefore be argued to be important for CT across society. The public library is perhaps the obvious avenue for providing access to fiction reading materials. This research project was conducted against a background of ongoing cuts to UK public library funding (Flood, 2020), this implies a trend in a negative direction for access to fiction, and thus access to its benefits for CT. Public library provision of fiction has been argued to be key in promoting equality across society by providing access to fiction on social justice and diversity themes (Chapman & Birdi, 2008). This research takes this argument further, as not only may the themes of fiction be important, but the experience of reading fiction independent of content has been shown to be beneficial. Furthermore, as public libraries provide access to a range of nonfiction materials including books that may provide a longer

reading engagement than online nonfiction texts, the findings pertaining to nonfiction also imply that public library nonfiction provision may have benefits to CT. Thus, public libraries offer access to all members of society to resources that can promote CT, and should be valued as such.

Finally, fiction can be seen as an element in a wider spectrum of art forms which may share some of its effects. Arts and humanities are frequently argued to be beneficial to CT (e.g. Dumitru, 2019), and this research presents evidence for such arguments in the case of fiction. However, at the time of writing a trend in arts and humanities education being cast as “poor value” is accelerating (Phillips, 2019), with resulting funding cuts (C. Smith, 2021). While this research was not on value of arts and humanities education, nonetheless the role fiction plays in shaping CT can be seen as a part of a wider range of ways in which exposure to arts may benefit CT, and thus forms evidence in the case that reducing arts and humanities exposure is detrimental. Furthermore, the argument for low economic value derived from arts and humanities (Phillips, 2019) is set in a wider context of economic and labour market changes which are highly likely to see dramatic shifts in the kinds of skills required in work (Suzman, 2021), that in fact are largely in an opposite direction implying greater need for humanistic skills. Progress in automation by AI, as part of this picture, has significant bottlenecks in creative and social intelligence (Frey & Osborne, 2017), thus signifying that CT related skills will be less susceptible to replacement by these technologies. This further emphasises the value of CT not only within the current context, but as likely to remain important in the foreseeable future. Thus, any means of building CT ability

ought to be promoted, and this research provides evidence for how engagement with fiction as framed in the wider arts and humanities can be one such driver of CT.

9.2. Future directions

The design of some parts of this research project was impacted by the COVID-19 pandemic and ensuing restrictions (these impacts are described in Appendix F). Subsequently, future research could address some of the limitations that arose due to this crisis, such as including in-person conversations.

Future research could utilise the methods deployed in the four studies presented here, but with a focus on identifying genre differences. For instance, study one and two could be replicated on a much larger scale with a sufficiently large sample size to permit for a breakdown of reading measured into genre variables while retaining statistical power. Similarly, interviews and reading diaries could be designed to capture participants' experiences of different genres. Such further work would fit into the pattern of research in prior studies on the impacts of individual differences in reading, which have previously focused on social and cognitive skill, but not CT. These studies commenced with comparison of reading against other activities such as television watching (Stanovich, 1993), and then moved into comparisons of fiction and nonfiction (Mar et al., 2006), and further narrowed into genre comparisons (Fong et al., 2013). As such, this research establishes differences in fiction and nonfiction impacts on CT, and further focus into potential genre differences would be a logical next step.

Further studies into different pathways between fiction reading and CT could obtain deeper detail and explanation for the relationships identified in this research project. This research project was a first step in investigating the potential relationship

between reading fiction and CT, based upon arguments for how the factors previously connected to fiction (i.e. empathy, ToM, counterfactual thinking) could feed into CT. As such, the findings of this research validate the original proposed relationship, but more investigation remains to test each proposed causal pathway. Thus, studies measuring reading, CT, and each of the proposed intermediary variables would be valuable. Structural equation modelling (SEM) seems likely to be an appropriate method, enabling the tracing of different causal pathways (Bullock et al., 1994). Additionally, as has been the case in the prior research on the impacts of reading (Panero et al., 2016, 2017), replication studies will be essential.

The finding that assigning fiction reading to nonfiction readers can increase CT, both in terms of change and improvement in test score, could be explored further. The effects of different types of reading materials remain to be tested, for example a nonfiction book versus a fiction book may be different to assigning short stories or articles. Furthermore, the impact of fictionality, in contrast to the impact of incongruity, could be further tested by comparing the assignment of nonfiction reading that is very different to nonfiction readers' normal habits (for instance science articles to someone who generally does not read them), to the assignment of fiction reading. Genre effects are a further area for investigation with different assigned text groups. Further qualitative studies are also recommended to interrogate the experiences of readers when encountering texts that diverge from their usual reading habits, and thus clarify and further explore how and why such reading is impactful.

As has been discussed, the boundary between fiction and nonfiction is not absolute, and these may be best thought of on a continuum, or as cluster concepts

(Gaut, 2000), than as delimited categories. As such, further research into the impacts of fictionality as a set of features, rather than fiction as a category, would be valuable. Comparisons between texts that mix invitations to belief and make-belief, in contrast with more paradigmatic cases at either end of the spectrum, may yield interesting nuance. Further exploration into readers experiences of fictionality and nonfictionality within different kinds of texts would also enable further understanding of some of the differences suggested in this research.

Further research with different participant samples would also be valuable. This research project focused in on readers, but further comparisons between readers and non-readers may offer other directions. Differentiating different types of readers beyond simply those who read more fiction or nonfiction could also provide more detailed insights into how CT differences may be associated with such individual differences in reading. Reaching out to participants offline, and to more diverse participant groups, would also enable greater generalisability of findings.

Finally, this research project suggests further research in LIS, particularly with focus on IL, looking at fiction reading would be a worthwhile direction. For instance, previous research has identified different fiction reader constructs in public library users (Birdi, 2011), and these types could form groups for comparison in more detailed research differentiating the effects of different fiction reading upon different readers. Further work comparing informational and aesthetic reading experiences would also be beneficial, and could shed further light on the ways that fiction and nonfiction reading may differ and relate. Research in educational contexts could also further explore differences in fiction impacts with and without instruction. This research project

demonstrates that fiction reading is a complex activity, entailing thinking mingled with imagination and transportation, conducive to deep personal change and critical thought. It implies LIS researchers whose interests are broader than the facilitation of information searches, who are interested in how we experience and think critically about all forms of information we find, would benefit from giving fiction their attention.

9.3. Chapter summary

This section presents a simplified summary of each prior chapter.

- **Literature review:** prior literature on IL and CT reveals commonalities that can be brought together into an overarching CT conception including: logic and bias detection; openness to different points of view; empathy; self-reflection; counterfactual imagination; making judgements informing action and/or belief. Previous research on the effects of fiction reading has found associations with several of these facets of CT. IL research that considers fiction frames this within instructional settings, and focuses on topic content of fictional texts, which is commensurate with other fiction and CT research in broader fields. However, no previous research has empirically explored the impact of fiction reading upon CT.
- **Methodology:** A sequential mixed methods approach was adopted for this research project, guided by a scientific realist paradigm. Validity was taken as contextual, and ensured by maintaining vigilance to threats to validity in each context, as well as incorporating polyphony, dialogue,

and openness to transformation as much as possible. The four studies comprising this research project were designed to complement one another such that they could be brought together with a results point of interface to yield insights that were greater than the sum of their parts.

- **Study one:** This observational survey study identified correlations between fiction and nonfiction exposure, EO, and CTD. Fiction, but not nonfiction, has a significant positive association with CTD. Fiction and nonfiction were also associated with different EO categories. This study therefore validated the proposed direction for the research project overall.
- **Study two:** Utilising a reading log with pre- and post- CT testing, and including assigned readings as an experimental manipulation, this study tested a causal relationship between fiction reading and CT. Reading more fiction was associated with greater change in CT, but nonfiction was not. When comparing the improved and no improvement CT groups, fiction entries were predictive of improvement but time was not, while conversely nonfiction time was associated with being in the improved group but more nonfiction entries had the inverse effect. Assigning fiction reading to readers who normally identify as nonfiction readers was found to improve CT.
- **Study three:** Interviews with readers were conducted to capture their experiences of reading and thinking critically. Reading fiction was connected with the experiential process of CT, in a moment-by-moment development of understanding. Nonfiction reading was linked

to CT procedure, as a cyclical building of knowledge and technique. NT was experienced by readers as integrative of CT into the reading experience.

- **Study four:** A diary of readers' experiences reading and thinking critically each day was used to explore how these differed on a day-to-day basis. This revealed nonfiction linked to more reality-oriented and contextualised reading and thinking experiences, while fiction was linked to more fluid and reflective experiences. However, the study overall found an interplay across fiction and nonfiction which together formed combined patterns in reading and thinking.
- **Discussion:** bringing together the findings across the four studies, differences in fiction and nonfiction reading and CT were identified, and the liminal space between these groupings was highlighted as an area for further investigation. Fiction reading was found to provide expansive, time-extended experiences leading to the deployment of counterfactual reasoning and imagination. The unique issues of veracity raised by fiction, alongside the safety it offers to facilitate openness, were found to be important drivers of its influence on CT. Nonfiction reading was connected with more explicit, contextualised CT procedures. Traditional boundaries demarkating different modes of thinking were challenged, revealing an interplay between aesthetic and informational reading, and thought.

Appendix A: Study One

Measures

ART-G

This author recognition test presented a list of 200 names, with participants asked to check each name they recognize to be an author (Mar & Rain, 2015).

Below you will see a list of names. Some of the people in the list are popular writers (of books) and some are not. You are to read the names and put a check mark next to the names of those individuals who you know to be writers. Do not guess, but only check those who you know to be writers. Remember, some of the names are people who are not writers, so guessing can easily be detected.

Aimee Emery; Eric Schlosser; John Saul; Alastair Reynolds; Erma Bombeck; John Searle; Albert Camus; Ernst Mayr; John Steinbeck; Alice Munro; Faith Popcorn; John Updike; Alice Sebold; Fern Michaels; Jonathan Kellerman; Alister Yussen; Frances Gresham; Jonathan T. Cortes; Amir D. Aczel; Frank Bluth; José Saramago; Amy Tan; Frank Herbert; Joseph Heller; Anne McCaffrey; Frank Killarney; Joseph LeDoux; Antonio Damasio; Franklin D. Manis; Joy Fielding; Arthur C. Clarke; Gabriel Garcia Marquez; Jude Deveraux; Audrey Niffenegger; Gary Baron; Judith Krantz; Barry Z. Posner; George R. R. Martin; Julia London; Bertrand Russell; Geraldine Dickson; K. Warner Sexton; Bob Woodward; Greg Bear; Karen Marie Moning; Carl Daniels; Gregory Maguire; Ken Follett; Carol Shields; Harlan Coben; Kenneth H. Blanchard; Catherine Anderson; Harold Gallivan; Kim Harrison; Cathy Reichs; Hilda Blyth; Larry Niven; Charlaine Harris; Hugh Liben; Lauren Amsel; Christopher Moore; Ian Rankin; Lisa Kleypas; Chuck Palahniuk; Iris Johansen; Lynn H. Larson; Clive Cussler; Italo Calvino; M. D. Johnson Spencer; Dale Blass; J. D. Sallinger; M. Scott Peck; Daniel Goleman; Jack Canfield; Maeve Binchy; Danielle Steele; Jack Higgins; Margaret Weis; David Passman; Jackie Collins; Margaritia Barrera; Dean Koontz; Jacqueline Carey; Marianne Williamson; Deepak Chopra; James Mendelson; Mark Sorenson; Denise Cuneo; James Patterson; Martin Faulkner; Diana Gabaldon; Jayne Ann Krentz; Mary Higgins Clark; Diana Palmer; Jean Baudrillard; Matt Ridley; Diane Ackerman; Jean Vanier; Meg Cabot; Diane Corter; Jeffrey Deaver; Melody Beattie; Dick Francis; Jeffrey Gray; Michael Connelly; Donna Leon; Jim Butcher; Michael Jecks; Douglas Adams; Jim Collins; Michael Moore; Douglas Coupland; Jo Davis; Michel Foucault;

Douglas Rushkoff; Jodi Picoult; Milan Kundera; Edward Condry; John Grisham; Miriam Schaie; Elliot Bever; John Irving; Morton Lytton; Emily Giffin; John LeCarré; Naomi Klein; Eric Adamson; John Maynard Smith; Naomi Wolf; Napoleon Hill; Sidney Sheldon; Neil Gaiman; Sinclair Ross; Nelson DeMille; Sophie Kinsella; Nicholas Sparks; Stephen C. Lundin; Noam Chomsky; Stephen Hawking; Nora Roberts; Stephen J. Gould; Norman Mailer; Stephen R. Covey; Oliver Sacks; Steve Yorel; Orson Scott Card; Sue Grafton; Oscar Asmitia; Susan Sontag; P. D. James; Terry Brooks; Patricia Cornwell; Terry Goodkind; Paulo Coelho; Terry Pratchett; Peter F. Drucker; Thomas Borko; Peter S. Pande; Thomas Kuhn; Philip C. McGraw; Thomas Mann; Philip K. Dick; Timothy Findley; Philippa Gregory; Tom Robbins; Pierre Berton; Toni Morrison; Piers Anthony; Umberto Eco; R. A. Salvatore; Ursula K. Le Guin; Ray Bradbury; Vince Flynn; Reed Inness; W. G. Sebald; Reuben Beauchamp; W. O. Mitchell; Richard Dawkins; W. Patrick Dawson; Richard Plath; Wally Lamb; Robert A. Heinlein; William Faulkner; Robert B. Parker; William Gibson; Robert D. Kaplan; Yann Martel; Robert Donahue; Yukio Mishima; Robert Fulghum; Robert Irons; Robert Jordan; Robert Ludlum; Robert Saul; Robert T. Kiyosaki; Rohinton Mistry; Roland Barthes; Sandra Brown; Scott Parson; Sherrilyn Kenyon.

IRI-F

The Fantasy subscale of the Interpersonal Reactivity Index (IRI-F) was used as a measure of *NT* (Davis, 1983).

The items were rated by participants on a 5 point scale from strongly disagree to strongly agree. Items marked with an asterisk were reverse coded.

1 I daydream and fantasize, with some regularity, about things that might happen to me.

2 I really get involved with the feelings of the characters in a novel.

3 I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.*

4 Becoming extremely involved in a good book or movie is somewhat rare for me.*

5 After seeing a play or movie, I have felt as though I were one of the

characters.

6 When I watch a good movie, I can very easily put myself in the place of a leading character

7 When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.

JCI

The Justifying Conclusions Inventory (JCI) was used as a measure of *EO* (McGinnis, 2016). This is a Likert scale with 1-6 agreement rating.

1. Most debatable issues have one correct conclusion, even though it seems that they do not
2. Evidence must be evaluated to determine if it is really relevant to the issue under consideration
3. Usually there is not a person who is right and one who is wrong. They each have their own truths so their conclusions are equally valid
4. Most debatable issues have one right conclusion
5. Reasonable thinking about debatable issues requires that we are willing to evaluate our own reasoning processes relevant to the issue
6. If I have one conclusion and someone else holds a different conclusion, both conclusions are equally accurate
7. For most debatable issues, there is a right position and a wrong position
8. Understanding debatable issues is an ongoing process requiring the evaluation of new evidence.
9. Because drawing conclusions is based on our own personal perceptions, it is not possible to determine if one conclusion is better than others.
10. Debatable issues are associated with one correct conclusion, even though I may not know what that conclusion is.
11. Conclusions are compelling if they are supported by solid reasoning about the most relevant evidence.

12. Deciding among conclusions is unnecessary because everyone has a right to their opinion.
13. Each person has his or her own truth and beliefs, so there can be more than one acceptable conclusion to a debatable issue
14. Drawing a conclusion about a debatable issue involves reasoning about why one conclusion is well supported and others are not.
15. Disagreements about debatable issues are really unnecessary because every person's opinion should be viewed as accurate.
16. Most issues are not truly debatable: there is one correct answer.
17. Determining the best conclusion requires the critical evaluation of all of the relevant evidence
18. Because people have a right to their opinion, there is more than one conclusion that is correct.
19. Each person sees the issues in his or her own unique way, so there is often more than one acceptable conclusion to a debatable issue.
20. Drawing a solid conclusion about a debatable issue involves determining which evidence provides the best support for that conclusion.
21. If someone else has come to a conclusion that differs from my conclusion, and he or she can justify their conclusion, then we are both right.
22. The best conclusion reflects the most compelling understanding of an issue given the relevant evidence.
23. There is one correct conclusion to issues, and other conclusions are wrong

CTDS

The Critical Thinking Disposition Scale (CTDS) was used as a measure of *CTD* (Sosu, 2013).

- 1 I usually try to think about the bigger picture during a discussion
- 2 I often use new ideas to shape (modify) the way I do things
- 3 I use more than one source to find out information for myself
- 4 I am often on the lookout for new ideas

- 5 I sometimes find a good argument that challenges some of my firmly held beliefs
- 6 It's important to understand other people's viewpoint on an issue
- 7 It is important to justify the choices I make
- 8 I often re-evaluate my experiences so that I can learn from them
- 9 I usually check the credibility of the source of information before making judgements
- 10 I usually think about the wider implications of a decision before taking action
- 11 I often think about my actions to see whether I could improve them

CTD Model Selection

In order to develop a model to test with CTD as the outcome variable, relationships between this and the measured demographic variables (age, education, and gender) were tested to identify which of these ought to be included. Tests of homogeneity of variance, and normality of distribution, are listed in a following appendix under the heading: Differences across CTD levels.

Participant ages across CTD levels were broken down as follows: Low M 39.41, SD 17.03; Medium M 32.66, SD 12.37; High M 35.34, SD 12.37. Given the equality of variances, and the robust sample size, an ANOVA was used and showed there was no significant difference in ages across the CTD levels [$F(2, 318) = 2.94, p = .0544$]. Age was therefore not considered a relevant predictor variable for CTD.

Due to low sample numbers in some categories, Fisher's exact test was carried out and found a significant relationship CTD level and education level ($p = .002$); education was therefore relevant.

Figure 11 shows a breakdown of educational level within each CTD category. Additionally, including only male and female participants ($n = 311$), Fisher's exact test

showed a significant relationship between CTD level and gender ($p = .024$), thus gender was also relevant. Figure 12 shows the breakdown of each CTD level by gender.

Figure 11: Bar chart of CTD levels by educational level

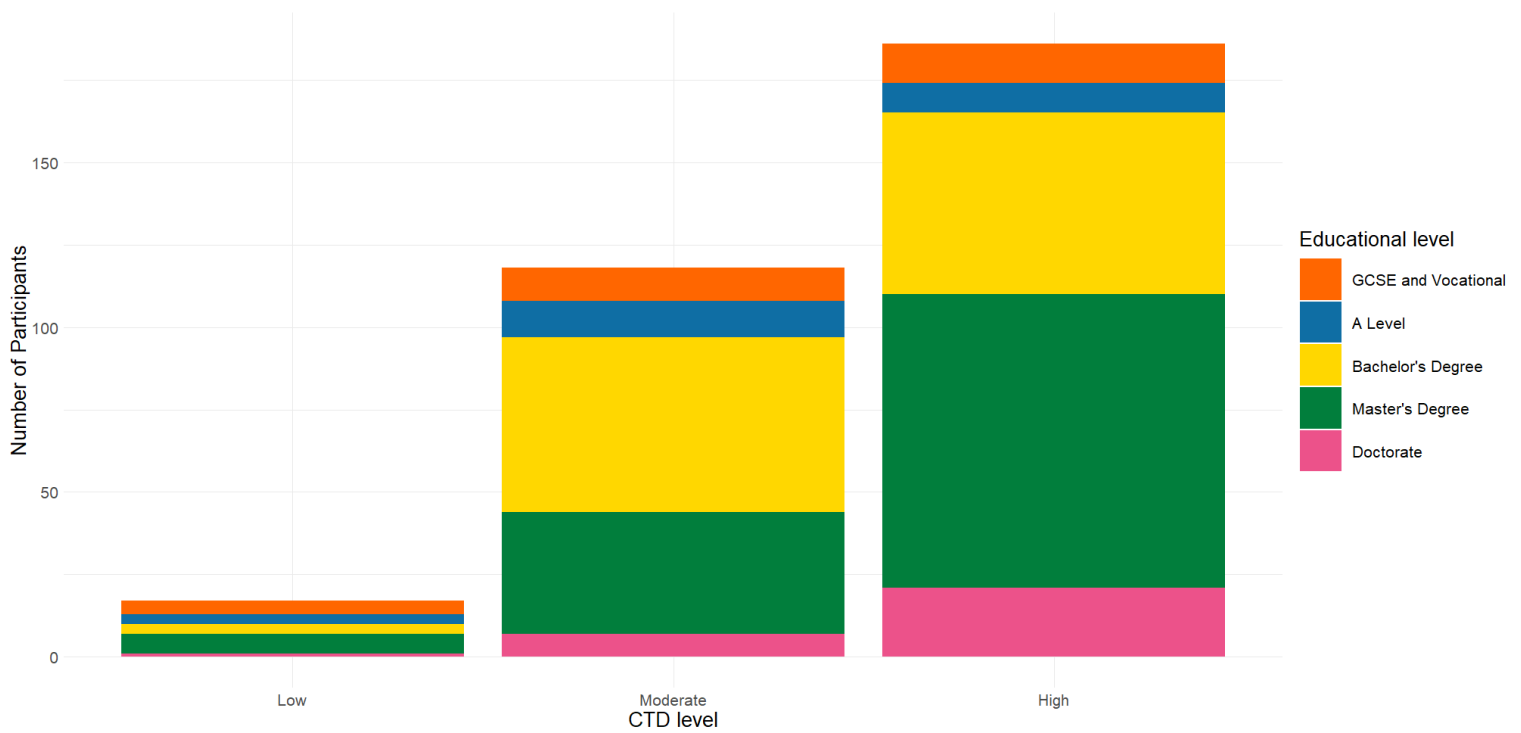
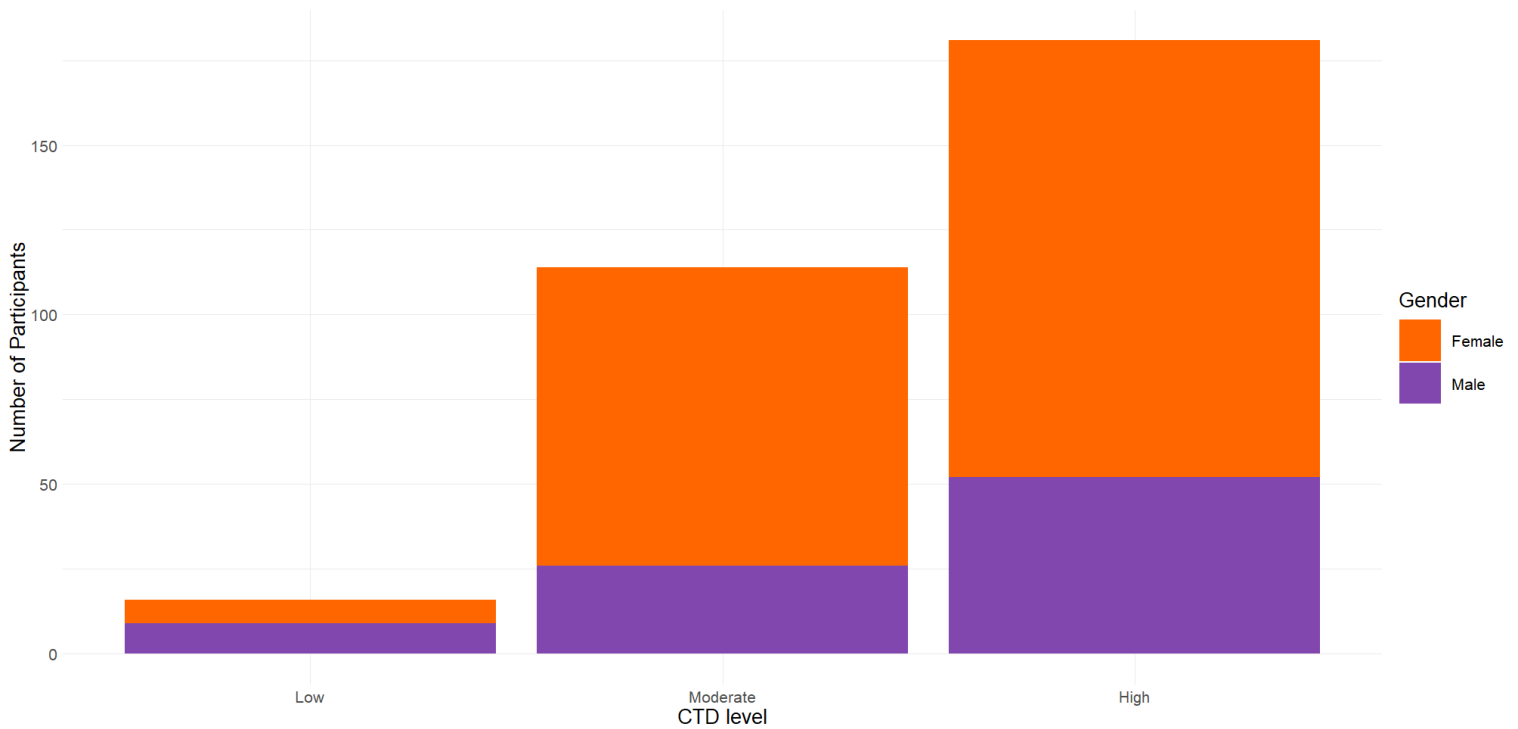
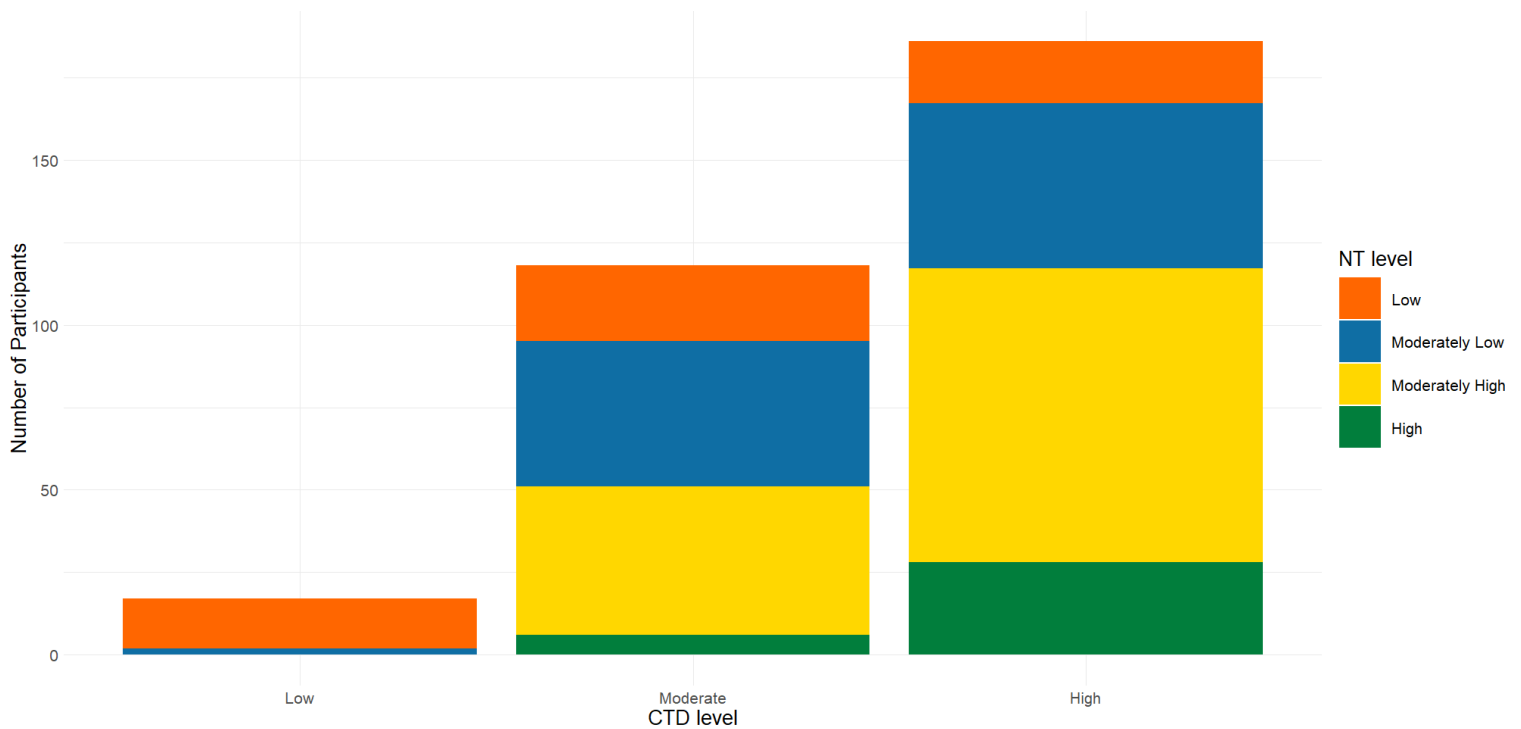


Figure 12: Bar chart of CTD levels by gender



Next, the direct relationships between the predictor variables of interest (ART-G total, fiction and nonfiction scores; NT level) and CTD level were assessed. Fisher's exact test found a significant relationship between CTD levels and NT level ($p = <.001$), supporting the relevance of NT. Figure 13 shows a breakdown of NT level within each CTD level.

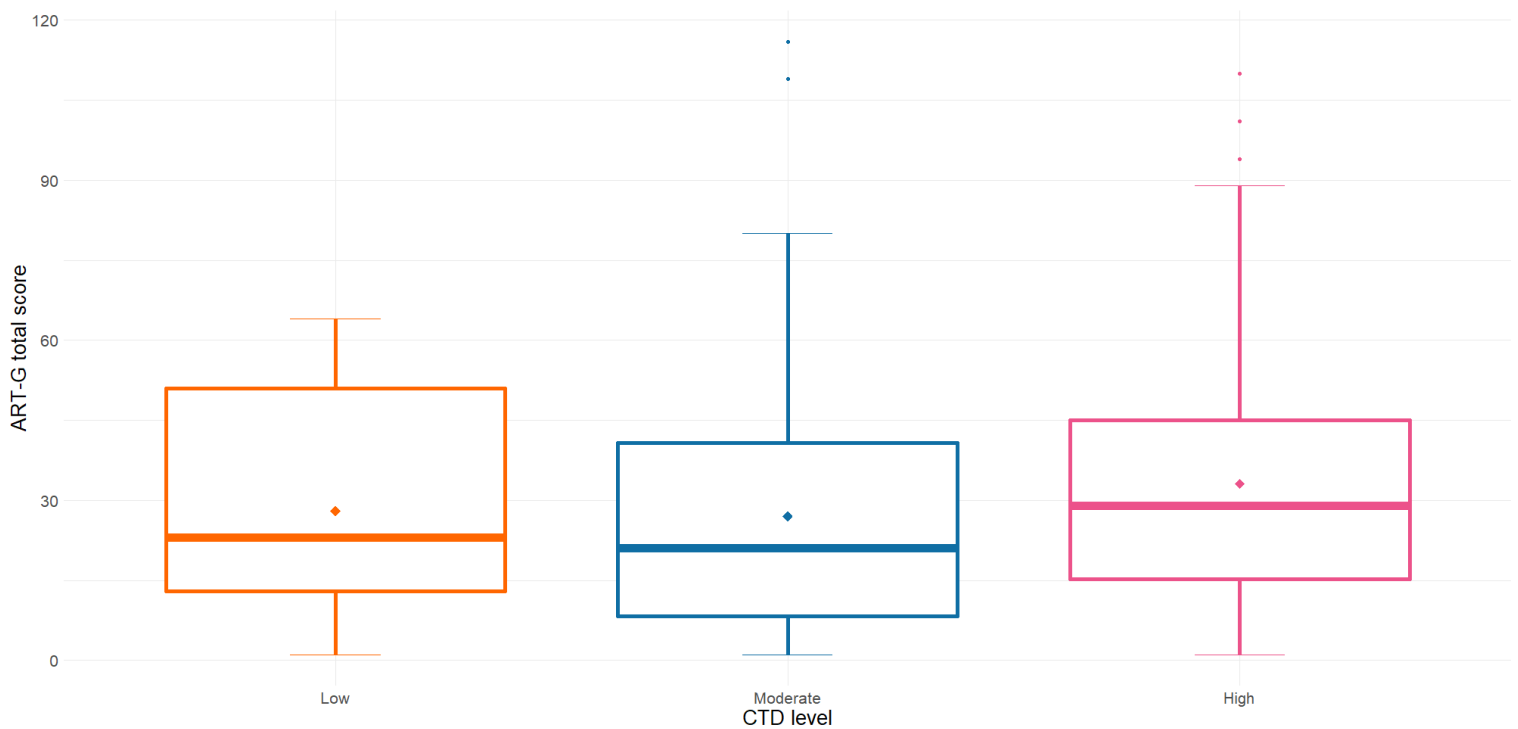
Figure 13: Bar chart of CTD levels by NT level



A direct relationship between ART-G total score and CTD level was tested. Given the equality of variances, and the robust sample size, an ANOVA was used and showed there was no significant difference in ART-G total score [$F(2, 318) = 2.80, p = .062$] between participants in different CTD level groups. This suggests that total

ART-G scores are not a relevant predictor of CTD level. Figure 14 shows the distribution of ART-G total scores within each CTD level.

Figure 14: Box plot of ART-G total score by CTD levels

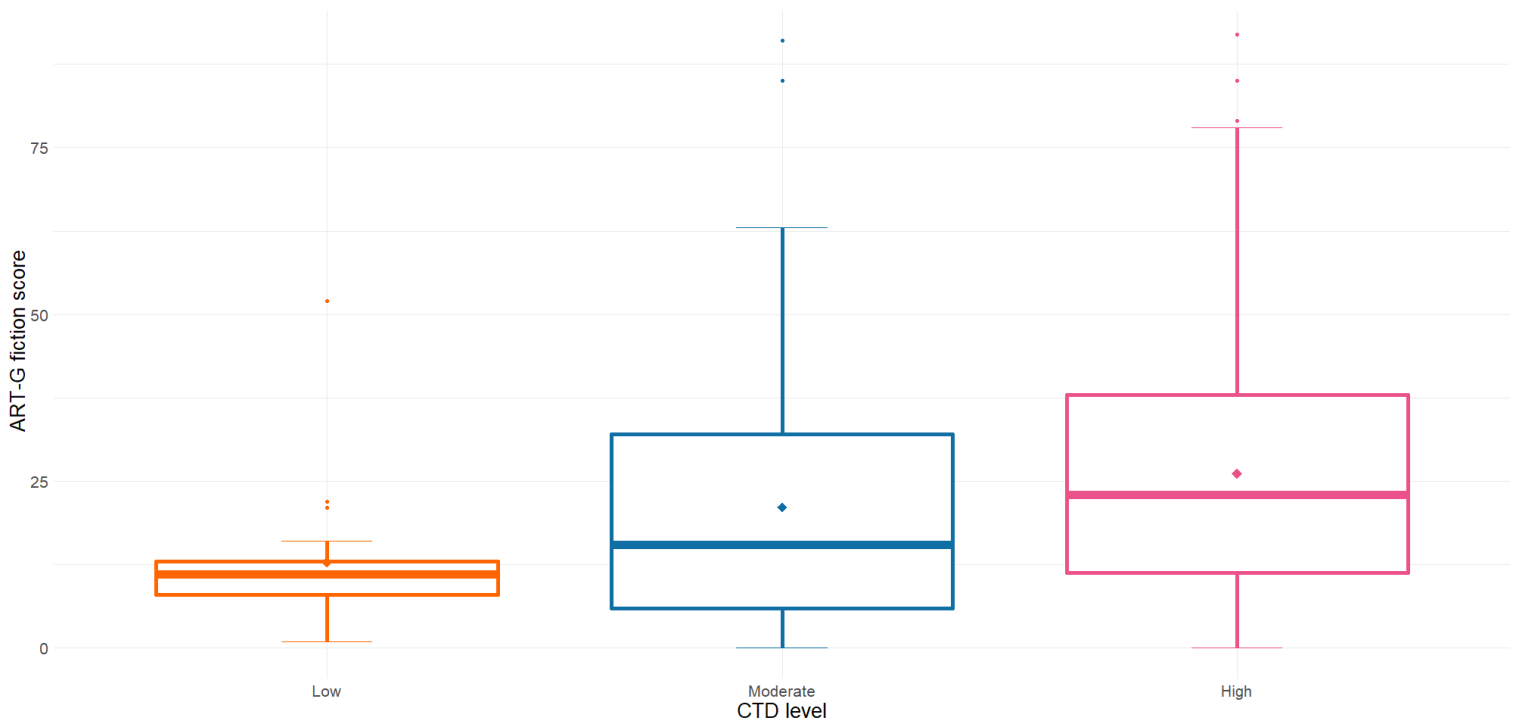


Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

In the case of ART-G fiction scores, variance across CTD levels was unequal. An ANOVA would therefore not be appropriate. Subsequently, a Kruskal-Wallis rank sum test was conducted, which showed a significant difference in median fiction scores between participants in different CTD level groups [$X^2(2, 318) = 15.85, p < .001$]. Pairwise comparisons using a Wilcoxon rank sum test with continuity correction showed the only significant difference to be between low ($M = 12.88, SD = 11.56, Mdn = 11$) and high ($M = 26.11, SD = 18.32, Mdn = 23$) ($p = .003$); moderate ($M = 21.11, SD = 18.99, Mdn = 15.5$) and high ($p = .005$) CTD level groups, with no significant difference for low and moderate ($p = .159$) CTD groups. This suggests

fiction score is a relevant predictor variable for CTD. Figure 15 shows the distribution of ART-G fiction scores across the different CTD levels.

Figure 15: Box plot of ART-G fiction score by CTD levels

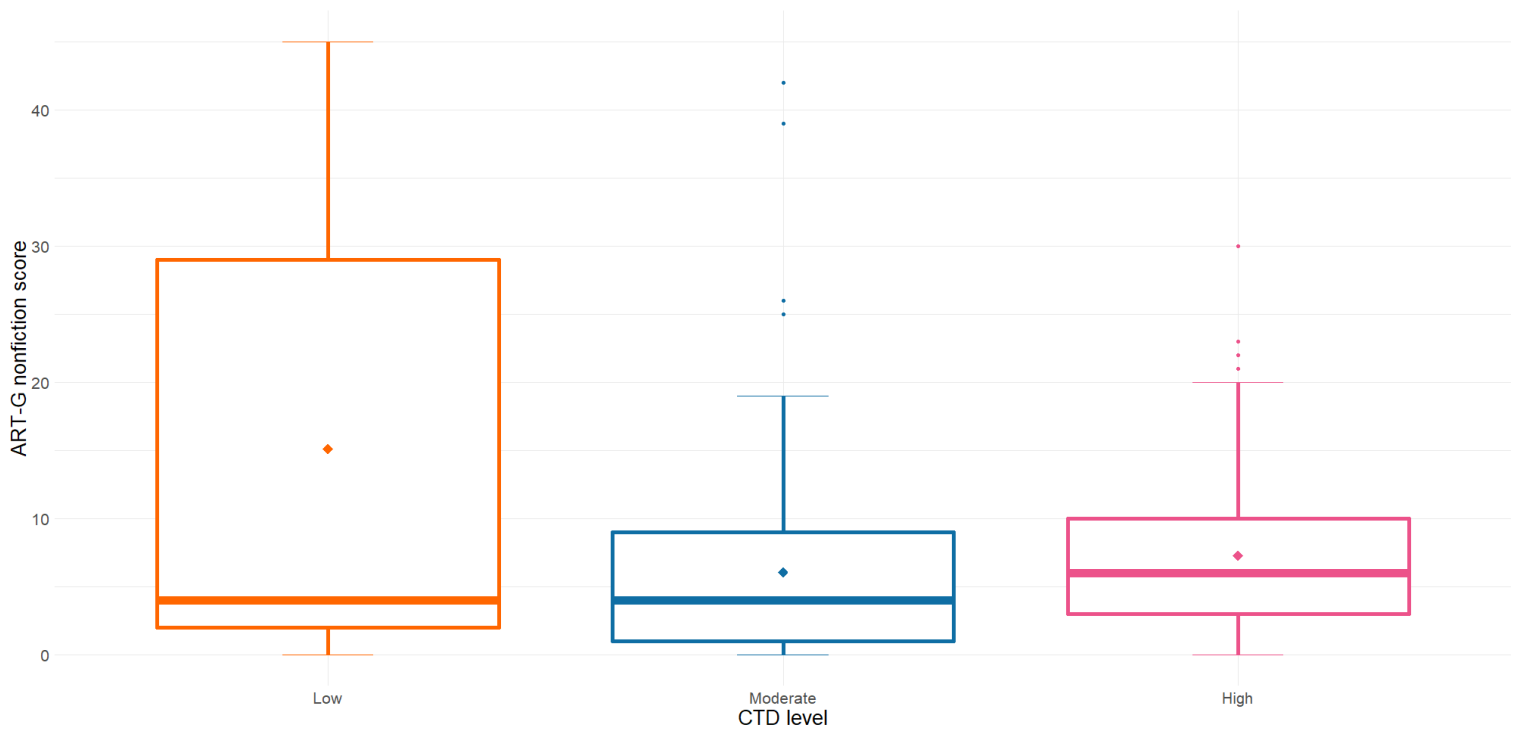


Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

ART-G *nonfiction* scores also had differing variance across CTD level. An ANOVA would therefore not be appropriate. Subsequently, a Kruskal-Wallis rank sum test was conducted, which showed a significant difference in median nonfiction scores between participants in different CTD level groups [$X^2(2, 318) = 9.67, p = .008$]. Pairwise comparisons using a Wilcoxon rank sum test with continuity correction showed the only significant difference to be between moderate ($M = 6.08, SD = 6.9, Mdn = 4$) and high ($M = 7.3, SD = 5.43, Mdn = 6$) CTD level groups ($p = .006$), with no significant difference for low ($M = 15.12, SD = 17.23, Mdn = 4$) and moderate ($p = .399$) or high ($p = .764$) CTD levels. This suggests nonfiction score is a

relevant predictor variable for CTD. Figure 16 shows the distribution of ART-G nonfiction scores within different CTD levels.

Figure 16: Box plot of ART-G nonfiction score by CTD levels



Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

As the ART-G total score is equal to the combined fiction and nonfiction scores with a subtracted penalty for foil checking (up to a maximum of -2, as any greater number of foils checked resulted in participant exclusion), and as foils are included in the ART-G to discourage guessing and remove participants who deploy a guessing strategy, the penalty for foil checking does not constitute a variable of interest in this study. The ART-G total score does not therefore provide essential additional information to the fiction and nonfiction scores. As the main focus of this study was to compare the influence of fiction and nonfiction engagement, which cumulatively give

an indication of the total influence of print exposure, fiction and nonfiction were included as predictor variables for further analyses and total scores were not.

Based on these individual analyses, the variables of interest in predicting CTD level were education, gender, NT, fiction ART-G score and nonfiction ART-G score. As CTD is an ordinal variable, a cumulative link model was used to test which predictor variables best explain the variance in CTD scores. Models including different predictors were tested, to determine which variables ought to be included for best model fit. As 10 participants were neither male nor female, and a category of $n = 10$ is too small for comparison, these participants were excluded from model comparisons. This means that model comparisons were made with a subset of the full data where gender was binary as male and female ($n = 311$).

Model permutations with those variables that were shown to have a relationship to CTD in the individual analyses were compared by AIC values. Table 48 shows variables included in each model with AIC values.

Table 48: Comparison of single level models without interactions

Model	Variables	AIC
clm1	CTD level ~ Fiction + Nonfiction + Gender + Education + NT level	472.52
clm2	CTD level ~ Fiction + Nonfiction + Education + NT level	477.21
clm3	CTD level ~ Fiction + Nonfiction + Gender + NT level	485.19
clm4	CTD level ~ Fiction + Nonfiction + NT level	488.46
clm5	CTD level ~ Fiction + Nonfiction + Gender + Education	504.50
clm6	CTD level ~ Fiction + Nonfiction + Gender	510.67
clm7	CTD level ~ Fiction + Nonfiction + Education	503.72
clm8	CTD level ~ Fiction + Nonfiction	509.68

The model with the lowest AIC is clm1. This was therefore selected as the best single level model without interactions.

Next, interactions were introduced and models with interactions compared. The demographic variables, and NT, were included as interactions with fiction and nonfiction ART-G scores. This was done as fiction and nonfiction scores are conceptually the predictor variables of interest, and other variables were considered relevant only in conjunction with these. Furthermore, NT level is of particular interest in relation to fiction versus nonfiction reading, and therefore was also included as an interaction with fiction or nonfiction alone. Thus, the model permutations tested were not intended as an exhaustive list of every possible variable combination, but rather constrained to those models that represented the aims of the study. Model

permutations were compared by AIC. Table 49 shows variables included in each model with AIC values.

Table 49: NT Interaction model comparison

Model	Variables	AIC
clmi1	CTD level ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Gender + Fiction * Gender + Nonfiction * Education + Fiction * Education	451.62
clmi2	CTD level ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Education + Fiction * Education	450.78
clmi3	CTD level ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Gender + Fiction * Gender	462.88
clmi4	CTD level ~ Nonfiction * Gender + Fiction * Gender + Nonfiction * Education + Fiction * Education	501.71
clmi5	CTD level ~ Fiction * NT level + Nonfiction + Nonfiction * Education + Fiction * Education	467.38

clmi6	CTD level ~ Fiction + Nonfiction * NT level + Nonfiction * Education + Fiction * Education	466.77
clmi7	CTD level ~ Fiction * NT level + Nonfiction + Nonfiction * Gender + Fiction * Gender	474.14
clmi8	CTD level ~ Fiction + Nonfiction * NT level + Nonfiction * Gender + Fiction * Gender	481.90
clmi9	CTD level ~ Fiction * NT level + Nonfiction + Gender	475.63
clmi10	CTD level ~ Fiction + Nonfiction * NT level + Gender	479.35
clmi11	CTD level ~ Fiction * NT level + Nonfiction + Education	469.68
clmi12	CTD level ~ Fiction + Nonfiction * NT level +	472.60

	Education	
clmi13	CTD level ~ Fiction * NT level + Nonfiction + Education + Gender	465.20
clmi14	CTD level ~ Fiction + Nonfiction * NT level + Education + Gender	470.09

Based upon this comparison, clmi2 is has the lowest AIC and therefore represents the best fit. This model includes NT interacting with fiction and nonfiction, and educational level interacting with fiction and nonfiction.

Next, a multilevel model was tested. Education level was selected as the clustering variable, as conceptually it offers the most relevant means of stratifying participants. Furthermore, it is subjectively determined (in contrast to NT which is based on a subjective Likert scale rating), making it more suited. Gender was not used as a clustering variable as two levels would be insufficient. Indeed, even the 5 levels of education falls beneath the minimum typically recommended for multilevel models (Bryan & Jenkins, 2016). However, while this may render conclusions about educational level difference less reliable, the individual-level predictors ought not be impacted, and therefore such a model would still be relevant for the purposes of this study. Different model permutations with education as the random intercept were therefore compared. See Table 50 for model versions and AIC scores.

Table 50: Random intercept model comparison

Model	Variables	AIC
clmm1	CTD level ~ fiction * NT level + nonfiction * NT level + (1 Education)	461.15
clmm2	CTD level ~ fiction + nonfiction + NT level + (1 Education)	484.10
clmm3	CTD level ~ fiction * NT level + nonfiction * NT level + Gender + (1 Education)	460.14
clmm4	CTD level ~ fiction * NT level + nonfiction * NT level + fiction * Gender + nonfiction * Gender + (1 Education)	461.62
clmm5	CTD level ~ fiction + nonfiction + NT level + Gender + (1 Education)	479.79
clmm6	CTD level ~ fiction * NT level + nonfiction + Gender + (1 Education)	471.96

clmm7	CTD level ~ fiction + nonfiction * NT level + Gender + (1 Education)	476.33
clmm8	CTD level ~ fiction * NT level + nonfiction (1 Education)	476.08
clmm9	CTD level ~ fiction + nonfiction * NT level (1 Education)	478.40

Based upon AIC, clmm3 displayed the best fit. This model includes NT interaction with fiction and nonfiction scores, and gender without interaction.

From these three types of model, the best versions were then compared: clm1, clmi2 and clmm3.

Table 51: Final model comparisons

Model	Variables	AIC
clm1	CTD level ~ fiction + nonfiction + Gender + Education + NT level	472.52
clmi2	CTD level ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Education + Fiction * Education	450.78
clmm3	CTD level ~ fiction * NT level + nonfiction * NT level + Gender + (1 Education)	460.14

An ANOVA comparing clm1 and clmi2 found a significant difference between the models score ($p < .001$), as did comparing clm1 with clmm3 ($p < 0.001$), and clmi2 and clmm3 ($p = .001$). This suggests the models are significantly different from one another. Furthermore, comparing each model against an empty null model using a comparison of Nagelkerke's Pseudo R^2 values, found clm1 to improve fit by 0.25, clmi2 by 0.39, and clmm3 by 0.31. This implies that clmi2 is the model that best fits the data, i.e. the model with interactions, but without random intercept, is preferred.

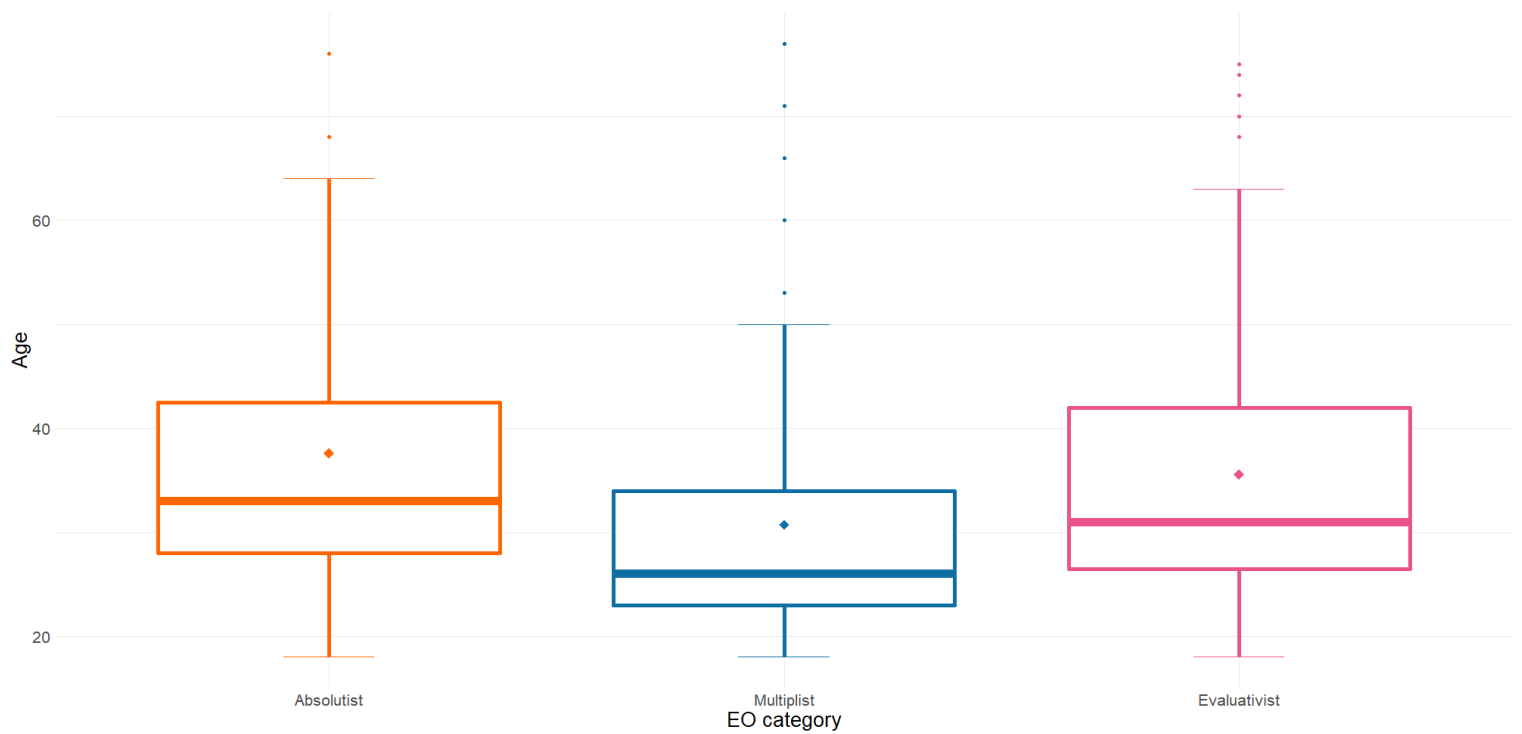
EO Model Selection

As was done for CTD as the outcome variable, the different measured variables were first directly tested for association with EO in order to identify which would be relevant in a full model. The following analyses include only the 311 participants who completed the IRI-F scale. Additionally, where gender is tested as a variable a subset is used of those who identified as either male or female only ($n = 302$). Tests of equality

of variance, and normality of distribution, are in the following appendix section on Differences across EO level.

First, significant differences in age across EO levels were sought. Given the equality of variances, and the robust sample size, an ANOVA was used and showed there was a significant difference in ages [$F(2, 308) = 6.51, p = .002$] between participants in different EO groups. A Tukey multiple comparison of means found significant difference in age between absolutists ($M = 37.59, SD = 13.05$) and multiplists ($M = 30.73, SD = 12.16$) ($p = .003$), and multiplists and evaluativists ($M = 35.55, SD = 12.32$) ($p = .01$), with age decreasing on average by 6.86 (95% CI 1.95, 11.77) years and increasing by 4.82 (95% CI 0.96, 8.67) years respectively across these groups. Figure 17 shows the distribution of age across EO categories. Age was taken as a relevant variable.

Figure 17: Box plot of Age by EO category

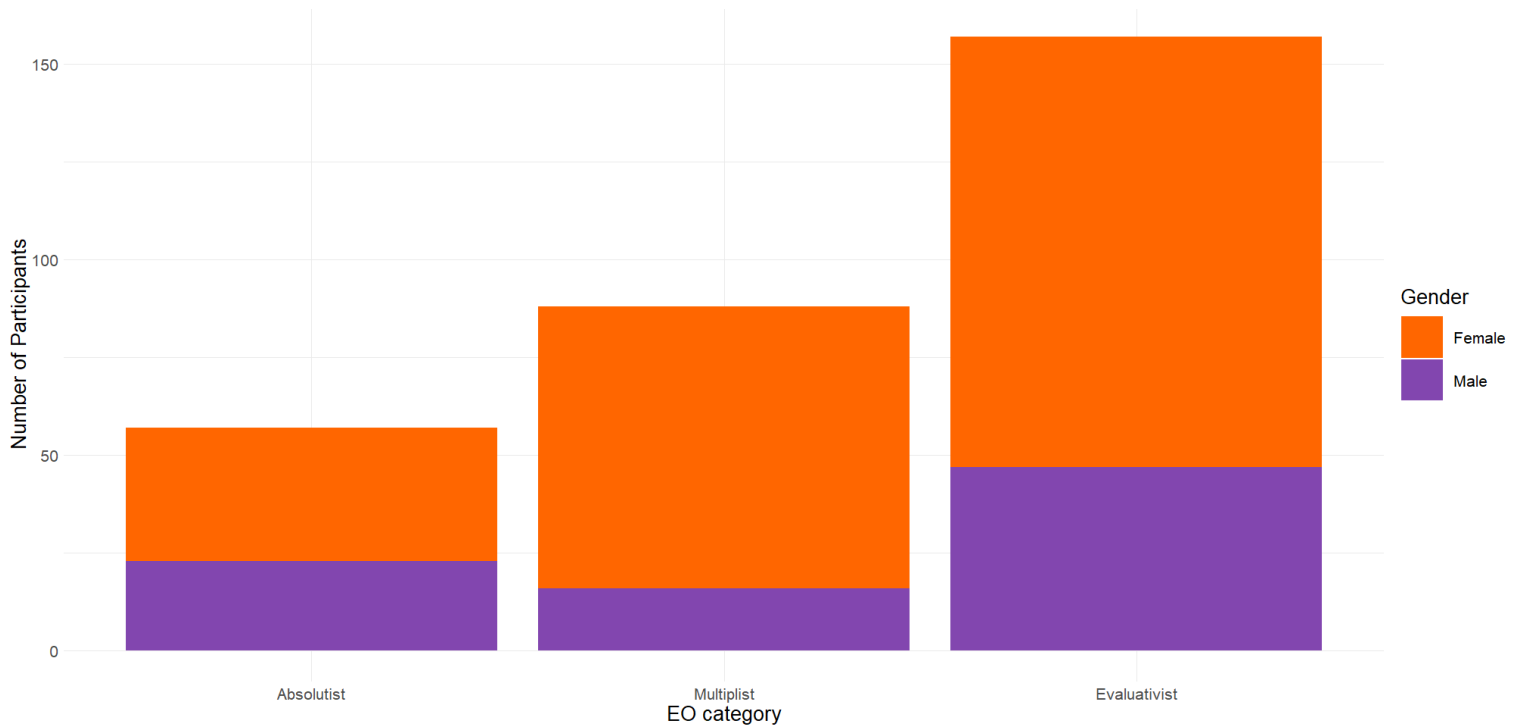


Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

Next, gender differences across EO categories were assessed with Fisher's exact test, which found a significant difference ($p = .012$), suggesting that gender is a relevant variable. Figure 18 shows the breakdown of gender within each EO category. Finally, Fisher's exact test found no significant difference between EO categories by education level ($p = .065$), and therefore education was not included when modelling EO.

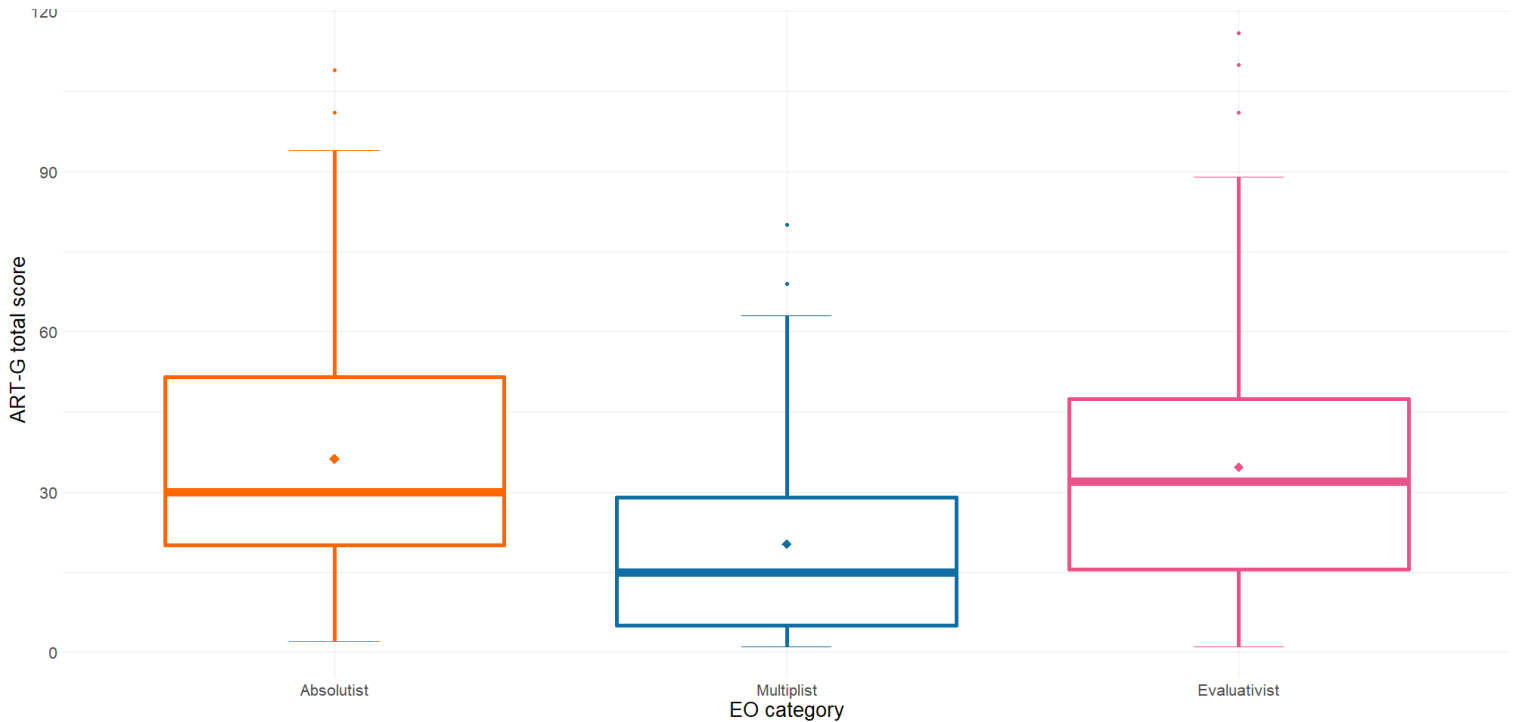
The predictor variables of interest were similarly individually tested for relevance to EO. Fisher's exact test showed no significant difference between EO categories by NT level ($p = .373$), suggesting NT is not a relevant predictor of EO.

Figure 18: Bar chart of EO categories by gender



Next, ART-G total scores were compared across EO levels. Given the equality of variance, an ANOVA was used and showed there was a significant difference in ART-G total score [$F(2, 308) = 14.74, p < .001$] between participants in different EO groups. A Tukey multiple comparison of means found significant differences in ART-G total score between absolutists ($M = 36.27, SD = 24.9$) and multiplists ($M = 20.27, SD = 18.36$) ($p < .001$), and multiplists and evaluativists ($M = 34.67, SD = 22.43$) ($p < .001$), with ART-G total score decreasing on average by 16 (95% CI 7.36, 24.64) items and increasing by 14 (95% CI 7.62, 21.19) items respectively across these groups. ART-G total scores were therefore a relevant predictor. Figure 19 shows the distribution of ART-G total scores across EO categories.

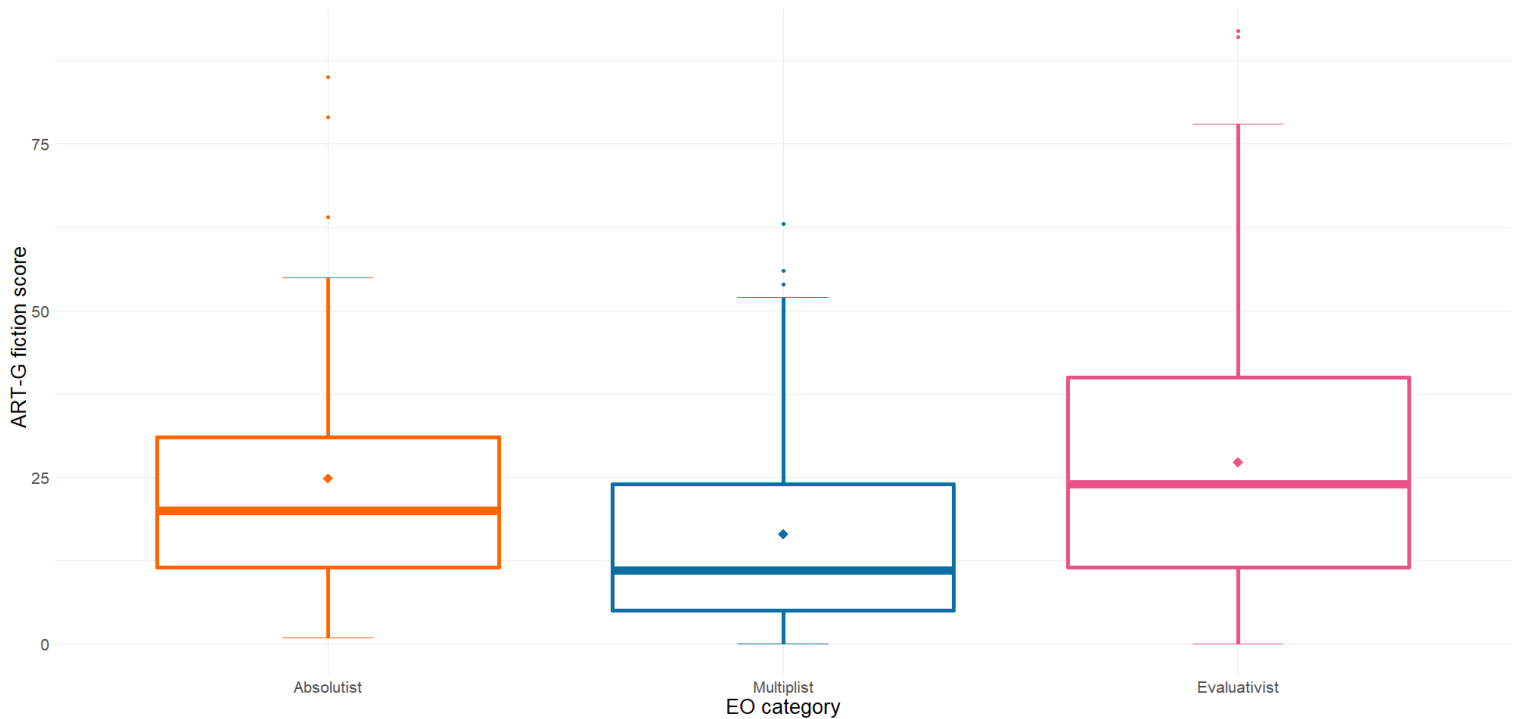
Figure 19: Box plot of ART-G total score by EO categories



Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

Next, fiction scores were compared across EO levels. Given the equality of variance, an ANOVA was used and showed a significant difference in ART-G fiction score [$F(2, 308) = 10.5, p < .001$] between participants in different EO groups. A Tukey multiple comparison of means found significant difference in ART-G fiction score between absolutists ($M = 24.85, SD = 20.08$) and multiplists ($M = 16.47, SD = 15.49$) ($p = .017$), and multiplists and evaluativists ($M = 27.3, SD = 18.56$) ($p < .001$), with ART-G fiction score decreasing on average by 8 (95% CI 1.24, 15.51) items and increasing by 11 (95% CI 5.23, 16.43) items respectively across these groups. This suggests fiction score is a relevant predictor variable for EO. Figure 20 shows the distribution of fiction scores across the EO categories.

Figure 20: Box plot of fiction score by EO categories

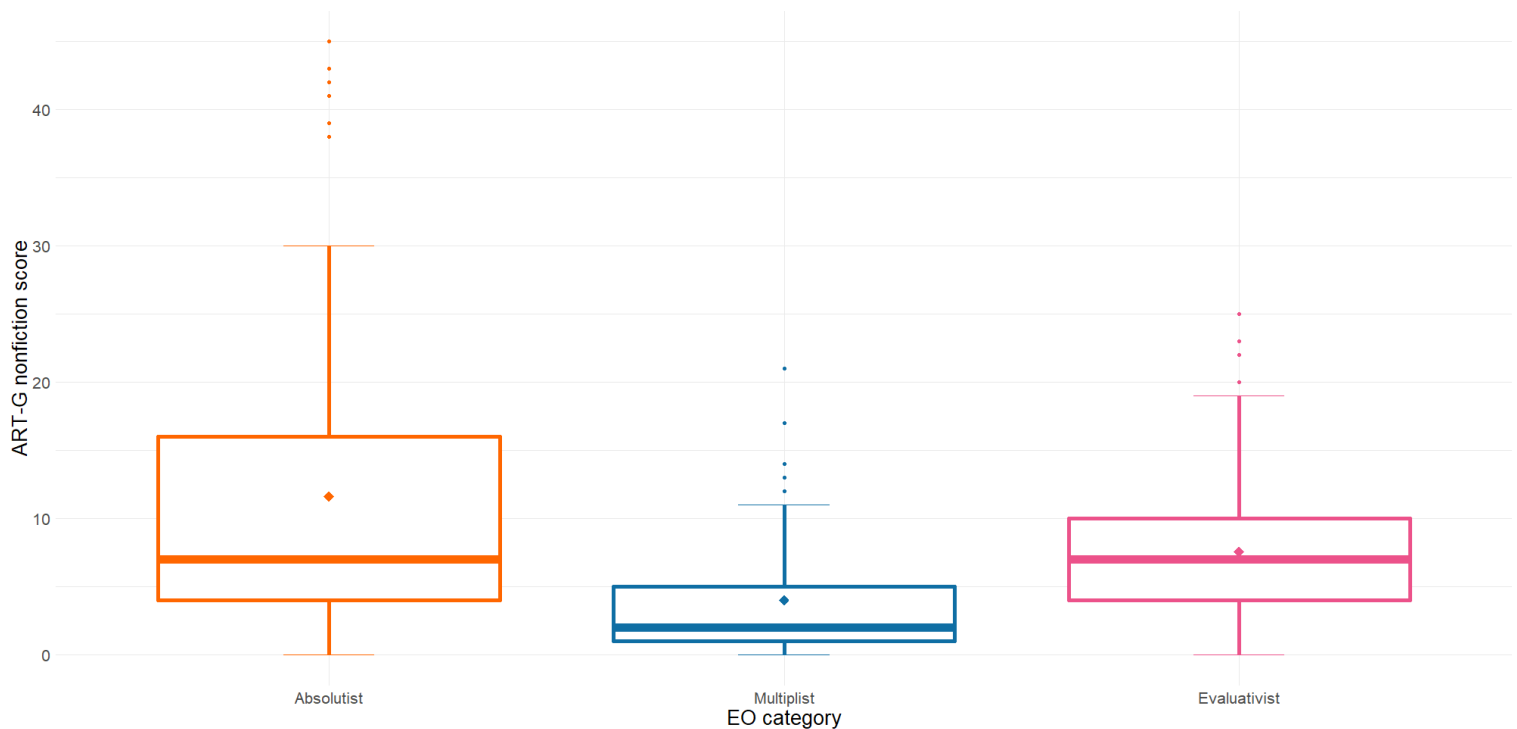


Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

Finally, nonfiction scores were tested, however these had differing variance across EO levels. An ANOVA would therefore not be appropriate. Subsequently, a Kruskal-Wallis rank sum test was conducted, which showed a significant difference in median nonfiction scores between participants in different EO groups [$X^2(2, 308) = 42.71, p < .001$]. Pairwise comparisons using a Wilcoxon rank sum test with continuity correction showed the significant differences to be between absolutists ($M = 11.64, SD = 12.16, Mdn = 7$) and multiplists ($M = 4.01, SD = 4.38, Mdn = 2$) ($p < .001$); multiplists and evaluativists ($M = 7.59, SD = 5.2, Mdn = 7$) ($p < .001$), with no significant difference for multiplist and evaluativist ($p = .33$) EO categories. This

suggests nonfiction score is a relevant predictor variable for EO. Figure 21 shows the distribution of nonfiction scores across EO categories.

Figure 21: Box plot of ART-G nonfiction score by EO categories



Note. Diamond symbol = mean; horizontal bar = median; top to bottom of box shows interquartile range; whiskers = smallest and largest value 1.5 times the interquartile range; small points = outliers

Therefore the relevant predictor variables for model comparisons with the outcome variable of EO were age, gender, ART-G total score, ART-G fiction score, and ART-G nonfiction score. NT was further added, as although there was no difference in NT across the EO categories this variable has theoretical interest within this study. As EO levels are not ordered (e.g. absolutist is simply a different category from multiplist and is determined by scoring highly on scale items for absolutism rather than by a lower score on an overall scale), they were treated as a series of dichotomous binary outcomes for modelling. Therefore, absolutists and multiplists were compared separately, as were absolutists and evaluativists, and multiplists and evaluativists.

For the purposes of this analysis, the reference category was always selected to be prior in the theoretical trajectory that moves from absolutism, through multiplism, and into evaluativism as the end point (Kuhn et al., 2000). Therefore absolutist was the reference category, aside from comparing multiplists and evaluativists where multiplist was the reference category. The sample sizes for the different comparisons were: absolutists and evaluativists $n = 214$; absolutists and multiplists $n = 145$; multiplists and evaluativists $n = 245$. These are all exclusive of 10 participants who did not identify their gender as either male or female, and thus constitute too small a category for comparison.

Generalised linear models were tested, to determine which variables ought to be included for best model fit. EO categories of multiplists versus evaluativists were used for model testing, as these combined to the largest sample size for the most robust model comparisons.

First, a full model was tested including all relevant variables:

Model 1: EO level \sim ART-G total + Fiction + Nonfiction + NT + Gender + Age

This model had an AIC value of 299.8.

As previously discussed, the ART-G total score does not add information of interest beyond what is provided by the fiction and nonfiction scores, and therefore a model without it was tested:

Model 2: EO level \sim Fiction + Nonfiction + NT + Gender + Age

This model had an AIC value of 299.91.

Rao's analysis of deviance was used to compare model 1 and 2, and found no significant difference ($p = 0.14$). This suggests that while excluding the ART-G total score does slightly lower the AIC, it does not in fact make a significant difference to model fit. Either model could therefore be considered equally well fitted. As conceptually ART-G total score does not contribute to the variables of interest, the lower AIC and more parsimonious model without it was therefore preferred and the total score was not included in further model testing,

Further model variations were compared using AIC and Nagelkerke's Pseudo R^2 values, see Table 52 for comparisons.

Table 52: Comparison of EO Multiplist vs Evaluativist GLM models without interactions

Model	Variables	AIC	Pseudo R ²
glm1	EO level ~ Fiction + Nonfiction + Gender + Age + NT level	299.91	0.19
glm2	EO level ~ Fiction + Nonfiction + Gender + NT level	298.62	0.18
glm3	EO level ~ Fiction + Nonfiction + Age + NT level	300.64	0.17
glm4	EO level ~ Fiction + Nonfiction + NT level	299.81	0.17
glm5	EO level ~ Fiction + Nonfiction + Gender + Age	295.34	0.18
glm6	EO level ~ Fiction + Nonfiction + Gender	293.76	0.18
glm7	EO level ~ Fiction + Nonfiction + Age	295.3	0.17
glm8	EO level ~ Fiction + Nonfiction	294.16	0.17

Based on these comparisons, glm6 has the lowest AIC, suggesting best model fit. However, glm1 has the highest Pseudo R² value, suggesting it explains more of the variance in EO. Rao's analysis of deviance showed no significant difference between the two models ($p = .763$). Subsequently, AIC value was favoured, as well as favouring

the more parsimonious model, and therefore glm6 was selected as the best single level model without interactions.

Next, a model with interactions was tested. Table 53 shows a comparison of interaction model permutations.

Table 53: Comparison of EO Multiplist vs Evaluativist interaction models

Model	Variables	AIC	Pseudo R ²
glmi1	EO ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Gender + Fiction * Gender + Nonfiction * Age + Fiction * Age	308.68	24
glmi2	EO ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Age + Fiction * Age	305.49	23
glmi3	EO ~ Fiction * NT level + Nonfiction * NT level + Nonfiction * Gender + Fiction * Gender	304.02	23
glmi4	EO ~ Nonfiction * Gender + Fiction * Gender + Nonfiction * Age + Fiction * Age	300.63	19
glmi5	EO ~ Fiction * NT level + Nonfiction +	301.47	22

		Nonfiction * Age +		
		Fiction * Age		
glmi6	EO ~ Fiction +		304.04	21
		Nonfiction * NT level +		
		Nonfiction * Age +		
		Fiction * Age		
glmi7	EO ~ Fiction * NT level +		300.31	22
		Nonfiction +		
		Nonfiction * Gender +		
		Fiction * Gender		
glmi8	EO ~ Fiction +		302.3	21
		Nonfiction * NT level +		
		Nonfiction * Gender +		
		Fiction * Gender		
glmi9	EO ~ Fiction * NT level +		296.89	22
		Nonfiction +		
		Gender		
glmi10	EO ~ Fiction +		299.43	21
		Nonfiction * NT level +		
		Gender		
glmi11	EO ~ Fiction * NT level +		298.73	21
		Nonfiction +		
		Age		

glmi12	EO ~ Fiction + Nonfiction * NT level + Age	301.43	20
glmi13	EO ~ Fiction * NT level + Nonfiction + Age + Gender	298.54	22
glmi14	EO ~ Fiction + Nonfiction * NT level + Age + Gender	300.77	21

Based upon this comparison, glmi9 is the preferred model with interactions as it has the lowest AIC, although glmi1- 3 do have slightly higher Pseudo R² values. AIC was favoured as a better indicator of model suitability, as well as preferring the more parsimonious model, and therefore glmi9 was selected.

Next, a multilevel model was considered. As was the case with CTD, only educational level was suitable as a clustering variable, and even this was a low number of categories (Bryan & Jenkins, 2016). To determine whether this would be suitable to proceed with, an empty model was compared with a model with educational level as the random intercept. The empty model had an AIC of 321.94, while the random intercept model had an AIC of 320.7. The difference in log-likelihoods between the models was 1.64; using this test statistic with 1df and deriving the *p* value from a chi squared table, no statistically significant difference was found [$X^2(1) = 3.28, p = .070$]. This implies that the addition of education as a random intercept does not make a

significant difference from the empty model, and therefore this was not adopted.

Multilevel models were subsequently not assessed.

Based upon these model fit comparisons, the best model without interactions was glm6, while the best model with interactions was glmi9. Rao's analysis of deviance showed no significant difference between these models ($p = .182$). The AIC value and Pseudo R^2 offer conflicting information, as the former indicates glm6 has the better fit while the latter indicates that glmi9 explains more of the variance in EO. Ultimately, given the lack of a statistically significant difference between the models, either would be suitable. Conceptually, glmi9 was deemed more relevant as NT is a variable of interest within the broader aims of this study. Therefore this model was adopted.

Assumption checking

Differences across CTD levels

Levene's test for homogeneity of variance found no significant difference in the variance of age across CTD levels ($F(2, 318) = 1.93, p = .146$). Shapiro-Wilk normality tests showed that ages at all three CTD levels were not normally distributed (low: $W = 0.89, p = .044$; moderate: $W = 0.84, p < .001$; high: $W = 0.90, p < .001$).

Levene's test for homogeneity of variance found no significant difference in the variance of ART-G total scores across CTD levels ($F(2, 318) = 0.14, p = .873$). Shapiro-Wilk normality tests showed that ART-G total scores at two out of three CTD levels were not normally distributed (low: $W = 0.92, p = .135$; moderate: $W = 0.89, p < .001$; high: $W = 0.93, p < .001$).

Levene's test for homogeneity of variance found a significant difference in the variance of ART-G fiction scores across CTD levels ($F(2, 318) = 3.56, p = .03$).

Shapiro-Wilk normality tests showed that ART-G fiction scores at all CTD levels were not normally distributed (low: $W = 0.72, p < .001$; moderate: $W = 0.88, p < .001$; high: $W = 0.93, p < .001$).

Levene's test for homogeneity of variance found a significant difference in the variance of ART-G nonfiction scores across CTD levels ($F(2, 318) = 22.97, p < .001$). Shapiro-Wilk normality tests showed that ART-G nonfiction scores at all CTD levels were not normally distributed (low: $W = 0.80, p = .002$; moderate: $W = 0.74, p < .001$; high: $W = 0.92, p < .001$).

Differences across EO level

In terms of age, Levene's test for homogeneity of variance found no significant difference in the variance of age between EO categories [$F(2, 308) = 0.77, p = .462$]. Shapiro-Wilk normality tests showed that ages in all three EO groups were not normally distributed (absolutist: $W = 0.88, p < .001$; multiplist: $W = 0.78, p < .001$; evaluativist: $W = 0.91, p < .001$).

Levene's test for homogeneity of variance found no significant difference in the variance of ART-G total scores between EO categories ($F(2, 308) = 2.66, p = .072$). Shapiro-Wilk normality tests showed that ART-G total scores in all three EO groups were not normally distributed (absolutist: $W = 0.92, p < .001$; multiplist: $W = 0.87, p < .001$; evaluativist: $W = 0.94, p < .001$).

Levene's test for homogeneity of variance found no significant difference in the variance of ART-G fiction scores between EO categories ($F(2, 308) = 2.1, p = .124$). Shapiro-Wilk normality tests showed that ART-G fiction scores in all three EO groups

were not normally distributed (absolutist: $W = 0.87, p < .001$; multiplist: $W = 0.87, p < .001$; evaluativist: $W = 0.94, p < .001$).

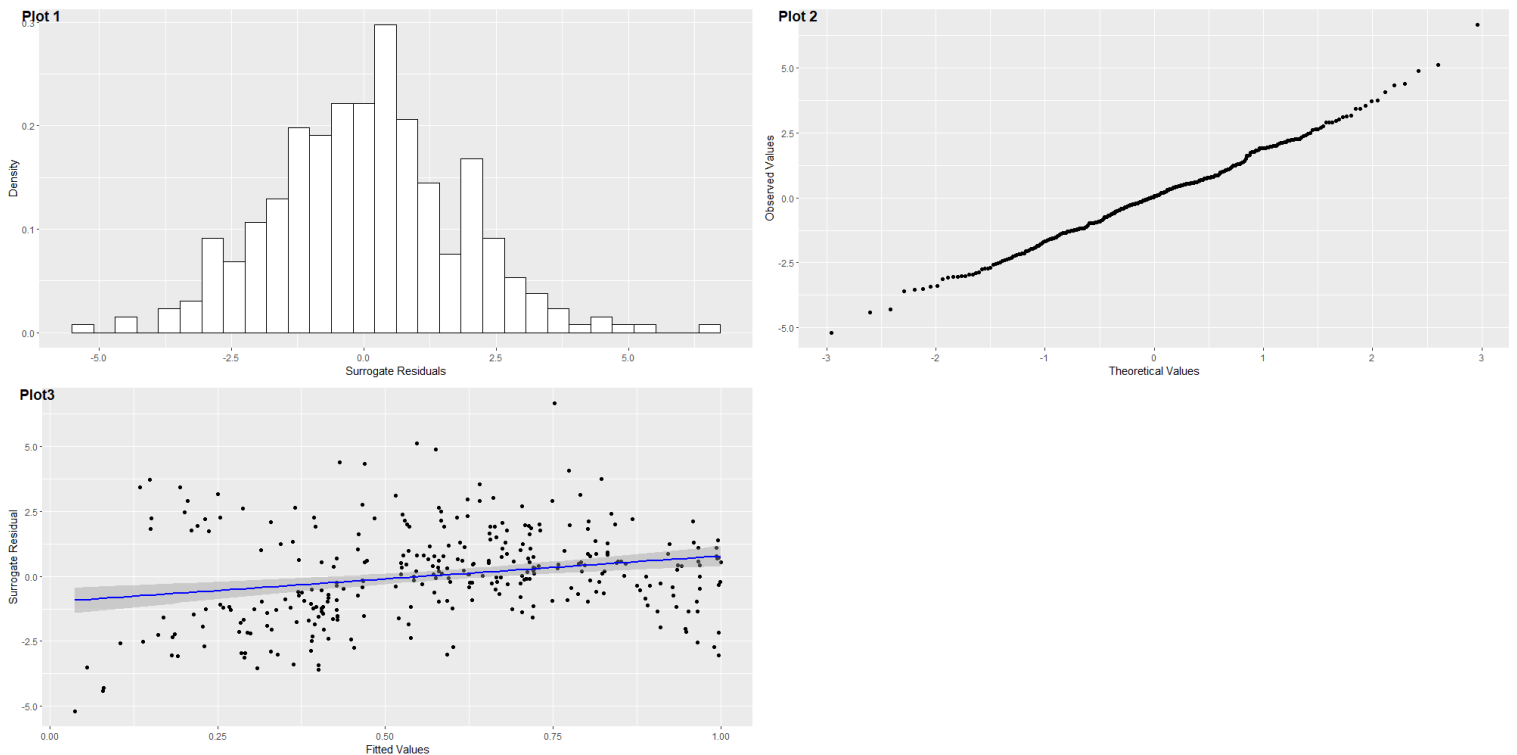
Levene's test for homogeneity of variance found a significant difference in the variance of ART-G nonfiction scores between EO categories ($F(2, 308) = 17.24, p < .001$). Shapiro-Wilk normality tests showed that ART-G nonfiction scores in all three EO groups were not normally distributed (absolutist: $W = 0.76, p < .001$; multiplist: $W = 0.80, p < .001$; evaluativist: $W = 0.94, p < .001$).

Model accuracy testing

Accuracy testing for the cumulative link model described in Table 3

As residual diagnostics are not straightforward in cumulative link models, a surrogate residuals approach was taken, using the sure package in R (Greenwell et al., 2018). Plot 1 shows a histogram of the surrogate residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of surrogate residuals against predicted values. These plots suggest the model conforms to the assumption of normality. However, it must be noted that surrogate residual testing is not without drawbacks (Lorenzo-Arribas, 2019), and as such these findings must be interpreted with caution.

Figure 22: Plots for cumulative link model described in Table 3



Accuracy testing for the binary glm model described in Table 4

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 425, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 21.25 (5%) cases would have standardised residuals outside of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 4.25 cases were expected to have standardised residuals outside ± 2.5 ; however only 3 cases were found to have a standardised residual outside ± 2 and none outside ± 2.5 . A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.91, p = .45$). Finally, the linearity of logit was tested, using interactions between the numerical predictors and the log; the interaction for

nonfiction was significant in the model, and therefore the assumption of linearity of logit was not met. Therefore a robust bootstrapped (100,000 replications) version of the logistic regression was conducted, which found values close to the original model. The distance from the original coefficient value to the bootstrapped mean was: constant = -0.13; nonfiction = 0.01. Overall, this suggests the model fits the data and can be generalised.

Accuracy testing for the binary glm model described in Table 5

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 214, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 10.70 (5%) cases would have standardised residuals outside of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 2.14 cases were expected to have standardised residuals outside ± 2.5 ; however no cases were found to have a standardised residual outside ± 2 . A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.81$, $p = .154$). Finally, the linearity of logit was tested, using interactions between the numerical predictors and the log; the interaction for nonfiction was significant in the model, and therefore the assumption of linearity of logit was not met. Therefore a robust bootstrapped (100,000 replications) version of the logistic regression was conducted, which found values close to the original model. The distance from the original coefficient value to the bootstrapped mean was: constant = -0.07; nonfiction = -0.003. Overall, this suggests the model fits the data and can be generalised.

Accuracy testing for the binary glm model described in Table 6

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 145, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 7.25 (5%) cases would have standardised residuals outside of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.45 cases were expected to have standardised residuals outside ± 2.5 ; however only one case had a standardised residual outside ± 2 , and did not exceed ± 2.5 . A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 2.28$, $p = .082$). Finally, the linearity of logit was tested, using interactions between the numerical predictors and the log; no interaction was significant in the model, and therefore the assumption of linearity of logit was met. Overall, this suggests the model fits the data and can be generalised.

Appendix B: Study Two

Measures

ICTET-A

Table 54: ICTET-A questions

Number	Question
1	<p>The main purpose of this text is...</p> <p>(Here you are trying to state as accurately as possible the author's purpose for writing the article. What was the author trying to accomplish?)</p>
2	<p>The most important information in this text is...</p> <p>(You want to identify the key information the author used, or presupposed, in the article to support his/her main arguments. Here you are looking for facts, experiences, data the author is using to support her/his conclusions).</p>
3	<p>The main conclusion(s) in this text is/are...</p> <p>(You want to identify the most important conclusions that the author comes to and presents in the article)</p>
4	<p>The main idea(s) we need to understand in order to understand this text is/are...</p> <p>(To identify these concepts, ask yourself: What are the most important ideas that you would have to understand in order to understand the author's line of reasoning?)</p>
5	<p>Here is a short explanation of what the author means by this/these concept(s)...</p> <p>(This refers to the concepts you answered the previous question with)</p>
6	<p>The main assumption(s) underlying the author's thinking is/are...</p> <p>(Ask yourself: What is the author taking for granted (that might be questioned). The assumptions are generalizations that the author does not think s/he has to defend in the context of writing the article, and</p>

they are usually unstated. This is where the author's thinking logically begins).

7 The main implications of this line of reasoning is/are...

(What consequences are likely to follow if people take the author's line of reasoning seriously? Here you are to follow out the logical implications of the author's position. You should include implications that the author states, if you believe them to be logical, but you should do your best thinking to determine what you think the implications are.)

The scoring rubric for the ICTET-A:

Detailed scoring rubric for Form A

Points	Level	Criteria
0-2	Unacceptable-Unskilled	<ul style="list-style-type: none"> The answer is inaccurate or unclear. The student has either fundamentally misunderstood what the author is saying or the answer is so unclear as to communicate no distinct meaning. No critical thinking skills are displayed by the student.
3-4	Poor-Marginally skilled	<ul style="list-style-type: none"> The answer, though partially accurate and minimally clear, is significantly inaccurate or misleading. The answer contains at least one important misunderstanding. At this level, the student is unlikely to be taking ownership of the author's ideas, even at a beginning level. The student's work marginally adheres to relevant intellectual standards.
5-6	Mixed-Partially skilled	<ul style="list-style-type: none"> The answer is of mixed quality, or inconsistent, showing some beginning insight, but also some important inaccuracies or significant vagueness. The analysis includes either some significant misunderstanding or insufficient elaboration or explanation (making it difficult to assess the depth of understanding). The student's work inconsistently adheres to relevant intellectual standards.
7-8	Commendable-Skilled	<ul style="list-style-type: none"> The answer is not only fundamentally accurate, it also contains some important insight(s) and is relatively complete (though not as detailed as excellent work). The answer contains sufficient details to show some depth of understanding, though may contain minor problems. The work primarily adheres to relevant intellectual standards.
9-10	Excellent-Highly skilled	<ul style="list-style-type: none"> The answer is accurate, insightful, clearly and precisely stated, and well-exemplified. The student appears to understand the author's thinking in depth. The student provides sufficient details to show deep understanding of the author's ideas. The work adheres to relevant intellectual standards.

This rubric makes reference to “intellectual standards” which can be accessed from the Foundation for Critical Thinking, furthermore example answers with grading are also available (The Foundation for Critical Thinking, 2019).

Reading log form

Reading Log

Welcome to your reading log. For the purposes of this study, please only log fiction or nonfiction texts that you have read. This means items such as articles, blog posts, short stories, poems, opinion pieces, books (sections, chapters, a few pages) etc. It is up to you if you would like to include audio versions as reading or not.

Please do not log communications. This means no emails, text messages, letters, etc.

Please also do not log social media posts, such as tweets, Facebook updates, Instagram captions etc. unless you consider them to be substantial enough to constitute a blog post or article.

Please also do not log items that you did not read thoroughly, or that you do not consider to be substantial reading. This means that skim reading, or briefly reading a small segment, can be omitted. It is up to you to gauge what you consider to be substantial reading.

What you choose to log is at your own discretion. You are not expected to log each and every text you read. There is a daily maximum of 15 items, and no minimum.

Please do not input any personal identifying or sensitive information into the fields provided; any such information will be deleted.

-

Basic information about you

Please input the unique Participant ID you were given in the invitation email for this study. Please double-check it is typed correctly:

Please input the date for which you are logging your reading in DD/MM format

(for example for the 3rd of February, put 03/02)

Did you read anything today?

Yes / No

-

Please complete the following questions for each item you read today:

Please input the Author's name:

Please input the Title of what you read:

What type of text was this? (For example, a news article, a novel, a blog post etc.)

Approximately how long did you spend reading this item? (in minutes)

What format did you read this in? (i.e. electronic such as on a computer, e-reader etc., or paper based print formats)

Electronic / Print

How transported or immersed did you feel while reading?

1 2 3 4 5

Not at all transported or immersed / Completely transported or immersed.

Did you read any more items that you'd like to log today?

Yes, I'd like to log more items. / No, this was the last item I'd like to log today.

Assumption checking

Descriptive statistics

The following section is presented as an appendix as it contains too much detail for the main study chapter. Here, successive sets of statistics are presented in a repetitive pattern to fully describe the data and justify the use of selected statistical models as presented in the study chapter.

ICTET-A scores

Levene's test for homogeneity of variance was used to assess whether scores with text and A at t1 and at t2 had differing variance. There was no difference in the variance of scores between t1 and t2 ($F(1,119) = 0.79, p = .375$). Levene's test showed a significant difference for text B scores between t1 and t2 ($F(1,119) = 3.95, p = .049$). Combining text and A and B scores, Levene's test showed a significant difference in variance between t1 and t2 ($F(1,240) = 4.53, p = .034$).

A Shapiro-Wilk normality test found CT scores at t1 to be normally distributed in women ($W = 0.99, p = .871$), men ($W = 0.98, p = .635$), neither gender ($W = 0.91, p = .529$); likewise for t2 in women ($W = 0.97, p = .183$), men ($W = 0.97, p = .293$), neither gender ($W = 0.90, p = .417$). Levene's test for homogeneity of variance found no significant difference in the variance of ICTET-A scores at t1 between genders ($F(2,118) = 1.08, p = .343$), but there was a significant difference at t2 ($F(2,118) = 4.17, p = .018$).

A Shapiro-Wilk normality test found CT scores at t1 to be normally distributed in participants with no degree ($W = 0.98, p = .568$), UG degrees ($W = 0.97, p = .59$), PG degrees ($W = 0.95, p = .070$); likewise for t2 in participants with no degree ($W = 0.97, p = .181$), UG degrees ($W = 0.97, p = .681$), PG degrees ($W = 0.99, p = .960$). Levene's test for homogeneity of variance found no significant difference in the variance of ICTET-A scores across education levels at t1 ($F(2,118) = 1.15, p = .32$), nor at t2 ($F(2,118) = 1.62, p = .203$).

Change in CT scores

A Shapiro-Wilk normality test found CT change to be normally distributed in women ($W = 0.98, p = .855$), not normally distributed in men ($W = 0.95, p = .029$), normally distributed in those of neither gender ($W = 0.88, p = .326$). Levene's test for homogeneity of variance found no significant difference in the variance of change scores across genders ($F(2,118) = 1.08, p = .343$).

A Shapiro-Wilk normality test found CT change not to be normally distributed in participants with no degree ($W = 0.95, p = .029$), normally distributed in those with UG degrees ($W = 0.93, p = .096$), and normally distributed in those with PG degrees

($W = 0.98, p = .639$). Levene's test for homogeneity of variance found no significant difference in the variance of change scores across *educational level* ($F(2,118) = 0.77, p = .465$).

Reading engagement

A Shapiro-Wilk normality test found total engagement not to be normally distributed ($W = 0.93, p < .001$), likewise for fiction engagement ($W = 0.95, p < .001$), and nonfiction engagement ($W = 0.77, p < .001$). A Shapiro-Wilk normality test fiction engagement was not normally distributed in fiction ($W = 0.94, p = .004$), or nonfiction ($W = 0.95, p = .017$), readers. Nonfiction engagement was also not normally distributed in fiction ($W = 0.58, p < .001$), or nonfiction ($W = 0.90, p < .001$), readers. Levene's test for homogeneity of variance found no significant difference in the variance of fiction engagement scores between participants who identified as mainly fiction readers or nonfiction readers ($F(1,119) = 1.29, p = .257$), but there was a significant difference in variance of nonfiction engagement scores between reader types ($F(1,119) = 36.41, p < .001$).

A Shapiro-Wilk normality test found total reading engagement to not be normally distributed in women ($W = 0.90, p < .001$), men ($W = 0.95, p = .047$), nor in those of neither gender ($W = 0.78, p = .05$). Fiction engagement was found to not be normally distributed in women ($W = 0.94, p = .003$), men ($W = 0.96, p = .052$), but normally distributed in those of neither gender ($W = 0.98, p = .96$). Nonfiction engagement was found to not be normally distributed in women ($W = 0.68, p < .001$), men ($W = 0.80, p < .001$), but normally distributed in those of neither gender ($W = 0.95, p = .754$). Levene's test for homogeneity of variance found no significant

difference in the variance of total engagement scores between genders ($F(2,118) = 0.45, p = .638$), nor for fiction engagement ($F(2,118) = 0.15, p = .865$), nor nonfiction engagement ($F(3,117) = 2.01, p = .139$).

A Shapiro-Wilk normality test found total reading engagement not to be normally distributed in participants with no degree ($W = 0.94, p = .001$), nor in those with UG degrees ($W = 0.84, p < .001$), and normally distributed in those with PG degrees ($W = 0.92, p = .062$). Levene's test for homogeneity of variance found no significant difference in the variance of total engagement scores between education levels ($F(2,118) = 1.03, p = .362$), nor for fiction engagement ($F(2,118) = 1.13, p = .325$), nor nonfiction engagement ($F(2,118) = 0.52, p = .599$).

Reading logged

A Shapiro-Wilk normality test found total reading entries not to be normally distributed ($W = 0.94, p < .001$), likewise for total time ($W = 0.82, p < .001$), fiction entries ($W = 0.92, p < .001$), fiction time ($W = 0.69, p < .001$), nonfiction entries ($W = 0.89, p < .001$), and nonfiction time ($W = 0.75, p < .001$). Transportation scores were also not normally distributed ($W = 0.94, p < .001$).

Levene's test for homogeneity of variance found no significant difference in the variance of total entries logged between participants in different experimental groups ($F(5,115) = 0.57, p = .724$), nor for total time ($F(5,115) = 0.60, p = .7$). A significant difference in variance in fiction entries logged was found between experimental groups ($F(5,115) = 3.85, p = .003$), and for fiction time ($F(5,115) = 2.31, p = .049$). There was no significant difference in variance of nonfiction entries logged between

experimental groups ($F(5,115) = 2.29, p = .05$), but nonfiction time variance was significantly different ($F(5,115) = 2.96, p = .015$).

Levene's test for homogeneity of variance found no significant difference in the variance of *average transportation* score between different experimental groups ($F(5,115) = 0.92, p = .47$).

Levene's test for homogeneity of variance found no significant difference in the variance of total entries logged between genders ($F(3,117) = 0.90, p = .443$), nor for total time ($F(3,117) = 0.32, p = .811$), fiction entries ($F(3,117) = 1.14, p = .335$), fiction time ($F(3,117) = 0.50, p = .683$), nonfiction entries ($F(3,117) = 0.53, p = .665$), nonfiction time ($F(3,117) = 0.57, p = .636$).

Levene's test for homogeneity of variance found no significant difference in the variance of average transportation scores between genders ($F(3,117) = 1.46, p = .228$).

Levene's test for homogeneity of variance found no significant difference in the variance of total entries logged between educational level ($F(2,118) = 2.56, p = .082$), nor for total time ($F(2,118) = 1.14, p = .324$), fiction entries ($F(2,118) = 0.34, p = .715$), fiction time ($F(2,118) = 0.14, p = .867$). There was a significant difference in variance between educational levels for nonfiction entries ($F(2,118) = 4.06, p = .02$), but not for nonfiction time ($F(2,118) = 2.50, p = .087$).

Levene's test for homogeneity of variance found no significant difference in the variance of average transportation scores between educational levels ($F(2,118) = 0.83, p = .437$).

Inferential statistics

CT change across experimental groups

A Shapiro-Wilk normality test found CT score change to be normally distributed within each experimental group (Cn: $W = 0.95$, $p = .247$; Cf: $W = 0.96$, $p = .385$; Ff: $W = 0.97$, $p = .767$; Fn: $W = 0.95$, $p = .360$; Nf: $W = 0.92$, $p = .142$; Nn: $W = 0.97$, $p = .856$). Levene's test for homogeneity of variance found no significant difference in the variance of CT change scores across different experimental groups ($F(5,115) = 1.26$, $p = .288$).

Improved versus no improvement groups

A Shapiro-Wilk normality test found that age was not normally distributed in either group (No improvement: $W = 0.91$, $p < .001$; improved: $W = 0.93$, $p = .001$). Levene's test for homogeneity of variance found no significant difference in age variance between the groups ($F(1,119) = 1.33$, $p = .251$).

Shapiro-Wilk normality tests found neither reading in entries per group (no improvement: $W = 0.91$, $p < .001$; improved: $W = 0.93$, $p = .002$) nor in time (no improvement: $W = 0.86$, $p < .001$; improved: $W = 0.87$, $p < .001$) to be normally distributed. Levene's test for homogeneity of variance was used to assess whether total entries logged, and total time, had differing variance for the improved and no improvement groups; no difference in the variance of total entries between the two groups was found ($F(1,119) = 0.43$, $p = .515$); total time was significantly different in variance across groups ($F(1,119) = 10.58$, $p = .0015$).

Shapiro-Wilk normality tests found fiction entries made to not be normally distributed in the no improvement group ($W = 0.85$, $p < .001$), but normally

distributed in the improved group ($W = 0.96, p < .074$). In the case of fiction reading time, neither group was normally distributed (no improvement: $W = 0.54, p < .001$; improved: $W = 0.78, p < .001$). Levene's test for homogeneity of variance was used to assess whether fiction entries logged, and fiction reading time, had differing variance for the improved and *no improvement* groups; no difference in the variance of fiction entries between the two groups was found ($F(1,119) = 1.01, p = .318$); fiction reading time was significantly different in variance across groups ($F(1,119) = 10.44, p = .002$).

Shapiro-Wilk normality tests found neither nonfiction entries (no improvement: $W = 0.91, p < .001$; improved: $W = 0.87, p < .001$) nor nonfiction time (no improvement: $W = 0.86, p < .001$; improved: $W = 0.72, p < .001$) to be normally distributed within the two groups. Levene's test for homogeneity of variance found no difference in the variance of nonfiction entries between the two groups ($F(1,119) = 1.20, p = .276$); nonfiction reading time also was not significantly different in variance across groups ($F(1,119) = 1.59, p = .21$).

Exploratory analyses

Peak and minimum NT

Shapiro-Wilk normality tests found peak and minimum fiction entries logged not to be normally distributed (peak: $W = 0.85, p < .001$; min: $W = 0.67, p < .001$). Neither were nonfiction entries (peak: $W = 78, p < .001$; min: $W = 70, < .001$).

Shapiro-Wilk normality tests showed neither fiction nor nonfiction entries with a minimum NT level to be normally distributed in either the improved or no improvement groups (nonfiction imp: $W = 0.71, p < .001$; nonfiction no imp: $W = 0.79, p < .001$; fiction imp: $W = 0.66, p < .001$; fiction no imp: $W = 0.73, p < .001$).

001). Levene's test for homogeneity of variance found a significant difference in the variance of nonfiction entries at minimum NT level between the improved and no improvement CT groups ($F(1, 104) = 4.16, p = .044$), however there was no difference in variance for fiction entries at minimum NT level ($F(1,93) = 1.05, p = .307$).

Shapiro-Wilk normality tests showed neither fiction nor nonfiction time (in half-hours) with a minimum NT level to be normally distributed in either the improved or no improvement groups (nonfiction imp: $W = 0.43, p < .001$; nonfiction no imp: $W = 0.77, p < .001$; fiction imp: $W = 0.54, p < .001$; fiction no imp: $W = 0.84, p < .001$). Levene's test for homogeneity of variance found no significant difference in the variance of nonfiction time at minimum NT level between the improved and no improvement CT groups ($F(1,104) = 3.74, p = .056$), and no difference in variance for fiction entries at minimum NT level ($F(1,93) = 2.42, p = .123$).

Shapiro-Wilk normality tests showed neither fiction nor nonfiction entries with a peak NT level to be normally distributed in either the improved or no improvement groups (nonfiction imp: $W = 0.75, p < .001$; nonfiction no imp: $W = 0.80, p < .001$; fiction imp: $W = 0.89, p < .001$; fiction no imp: $W = 0.75, p < .001$). Levene's test for homogeneity of variance found no significant difference in the variance of nonfiction entries at peak NT level between the improved and no improvement CT groups ($F(1,104) = 0.24, p = .623$), likewise for fiction entries at peak NT level ($F(1,93) = 2.38, p = .126$).

Shapiro-Wilk normality tests showed neither fiction nor nonfiction time (in half-hours) with a peak NT level to be normally distributed in either the improved or no improvement groups (nonfiction imp: $W = 0.54$, $p < .001$; nonfiction no imp: $W = 0.70$, $p < .001$; fiction imp: $W = 0.77$, $p < .001$; fiction no imp: $W = 0.64$, $p < .001$). Levene's test for homogeneity of variance found no significant difference in the variance of nonfiction time at peak NT level between the improved and no improvement CT groups ($F(1,104) = 1.65$, $p = .202$), however for fiction time at peak NT level there was a significant difference in variance ($F(1,93) = 5.89$, $p = .017$).

Mediation analysis results

Mediation by narrative transportation on entries read

Separate regressions were used to compare the effect of total, fiction and nonfiction reading amount (in entries) on CT change, with and without NT as a mediator. Model-based causal mediation analysis was performed (using the 'mediation' package for R), with nonparametric bootstrapped (1000 simulations) confidence level estimation using the percentile method, to calculate the average causal mediation effect (Tingley et al., 2014).

A linear regression model testing for a relationship between total entries read and CT change was found to be significant as compared to an empty model ($F(1, 119) = 4.90$, $p = .029$), with an R^2 of .04. Participants predicted CT change score was equal to -2.79 (score change) + 0.2 (total entries); participants CT change score increased on average by 0.2 for every fiction entry logged ($p = .029$). However, a linear regression model testing for a relationship between total entries made and average transportation rating was found not to be significantly different to an empty model ($F(1, 119) = 1.72$,

$p = .193$), with an R^2 of .01. Furthermore, a multiple regression model was used to test the relationship between both total entries made and average transportation rating, and CT score change; this model significantly improved fit compared to an empty model ($F(1, 119) = 3.21, p = .044$); however, only total entries were a significant predictor ($p < .029$), while average transportation rating was not significant ($p = .222$).

Although the multiple regression model did not indicate a significant effect of average transportation rating on CT change score, a mediation analysis was nonetheless conducted to provide a more robust test of the relationship between total entries logged and average transportation, upon CT change score. The average causal mediation effect was 0.01 [95% CI -0.01, 0.05], and was not significant ($p = .318$). The average direct effect was 0.19 [95% CI 0.01, 0.41], and this was significant ($p = .038$), as was the total effect of 0.20 [95% CI 0.02, 0.42] ($p = .018$). The proportion of the effect of total entries upon CT change score that was mediated by average transportation score was 0.07 [95% CI -0.09, 0.59], and this was not significant ($p = .328$). The effect of total entries logged on change in CT scores from t1 to t2 was not mediated via average transportation scores.

A linear regression model found that nonfiction entries made had no direct effect on CT change (i.e. the model was not significantly different from an empty model) [$F(1, 119) = 0.81, p = .37, R^2 = .007$]. Likewise, nonfiction entries made had no effect on mean transportation scores [$F(1, 119) = 0.25, p = .62, R^2 = .002$]. Thus, a mediation analysis is not suitable.

A linear regression model testing for a relationship between fiction entries made and CT change was found to be significant as compared to an empty model ($F(1, 119)$

= 25.58, $p < .001$), with an R^2 of .18. Participants predicted CT change score was equal to -4.18 (score change) + 0.69 (fiction entries); participants CT change score increased on average by 0.69 for every fiction entry logged ($p < .001$). Additionally, a linear regression model testing for a relationship between fiction entries made and mean transportation rating was found to be significant as compared to an empty model ($F(1, 119) = 7.99, p = .006$), with an R^2 of .06. Participants predicted mean transportation score was equal to 3.40 (mean transportation rating) + 0.03 (fiction entries); participants mean transportation rating increased on average by 0.03 for every fiction entry logged ($p = .006$). A multiple regression model was used to test the relationship between both fiction entries made and mean mediation rating, and CT score change. A significant regression equation was found as compared to an empty model ($F(2, 118) = 13.243, p = <.001$), with an R^2 of .18. Participants predicted CT change score is equal to -5.49 (CT score change) + 0.67 (fiction entries) + 0.39 (mean transportation rating); participants change in CT score increased on average by 0.67 for every fiction entry logged and 0.39 for every increase in mean transportation rating. However, only fiction entries were a significant predictor ($p < .001$), while mean transportation rating was not significant ($p = .749$).

Although the multiple regression model did not indicate a significant effect of mean transportation rating on CT change score, a mediation analysis was nonetheless conducted to provide a more robust test of the relationship between fiction entries logged and mean transportation, upon CT change score. The average causal mediation effect was 0.01 [95% CI $-0.04, 0.08$], and was not significant ($p = .65$). The average direct effect was 0.67 [95% CI $0.42, 0.95$], and this was significant ($p < .001$), as was the

total effect of 0.69 [95% CI 0.45, 0.95] ($p < .001$). The proportion of the effect of fiction entries upon CT change score that was mediated by mean transportation score was 0.02 [95% CI -0.07, 0.12], and this was not significant ($p = .65$). The effect of fiction entries logged on change in CT scores from t1 to t2 was not mediated via mean transportation scores.

Mediation by narrative transportation on reading time

Separate regressions were used to compare the effect of total, fiction and nonfiction reading time (in half-hours) on CT change, with and without NT as a mediator. Model-based causal mediation analysis was performed (using the 'mediation' package for R), with nonparametric bootstrapped (1000 simulations) confidence level estimation using the percentile method, to calculate the average causal mediation effect (Tingley et al, 2014).

A linear regression model testing for a relationship between total reading time and CT change was found to be significant as compared to an empty model ($F(1, 119) = 19.78, p < .001$), with an R^2 of .14. Participants predicted CT change score was equal to -3.35 (score change) + 0.16 (total reading time); participants CT change score increased on average by 0.16 for every half-hour of reading logged ($p < .001$).

However, a linear regression model testing for a relationship between total half-hours read and mean transportation rating was found not to be significant as compared to an empty model ($F(1, 119) = 3.57, p = .061$), with an R^2 of .03. A multiple regression model was used to test the relationship between both total reading time and mean mediation rating, and CT score change. The regression including both variables did significantly improve model fit compared to an empty model ($F(2, 118) = 10.20, p <$

.001). However, only total reading time was a significant predictor of CT change in the model ($p < .001$), while mean transportation rating was not significant ($p = .414$).

Although the multiple regression model did not indicate a significant effect of mean transportation rating on CT change score, a mediation analysis was nonetheless conducted to provide a more robust test of the relationship between total reading time and mean transportation, upon CT change score. The average causal mediation effect was 0.005 [95% CI -0.004, 0.02], and was not significant ($p = .32$). The average direct effect was 0.016 [95% CI 0.08, 0.26], and this was significant ($p < .001$), as was the total effect of 0.16 [95% CI 0.09, 0.26] ($p < .001$). The proportion of the effect of total reading time upon CT change score that was mediated by mean transportation score was 0.03 [95% CI -0.03, 0.17], and this was not significant ($p = .32$). The effect of total reading time on change in CT scores from t1 to t2 was not mediated via mean transportation scores.

A linear regression model found that nonfiction reading time had no direct effect on CT change (i.e. the model was not significantly different to an empty model) [(F(1, 119) = 3.17, $p = .078$), $R^2 = .03$]. Likewise, nonfiction reading had no effect on mean transportation scores [(F(1, 119) = 0.10, $p = .751$), $R^2 = .001$]. Thus, a mediation analysis is not suitable.

A linear regression model testing for a relationship between fiction reading time and CT change was found to be significant as compared to an empty model (F(1, 119) = 18.19, $p < .001$), with an R^2 of .13. Participants predicted *CT change* score was equal to -1.85 (score change) + 0.21 (fiction reading time); participants CT change score increased on average by 0.21 for every fiction minute logged ($p < .001$).

Additionally, a linear regression model testing for a relationship between fiction minutes read and mean transportation rating was found to be significant as compared to an empty model ($F(1, 119) = 5.12, p = .025$), with an R^2 of .04. Participants predicted mean transportation score was equal to 3.5 (mean transportation rating) + $.008$ (fiction half-hours); participants mean transportation rating increased on average by 0.008 for every fiction minute logged. A multiple regression model was used to test the relationship between both fiction reading time and mean transportation rating, and CT score change. A significant regression equation was found as compared to an empty model ($F(2, 118) = 9.30, p = <.001$), with an R^2 of .14. Participants predicted CT change score was equal to -4.88 (CT score change) + 0.20 (fiction half-hours) + 0.87 (mean transportation rating); participants change in CT score increased on average by 0.20 for every fiction minute logged and 0.87 for every increase in mean transportation rating. However, only fiction reading time was a significant predictor ($p < .001$), while mean transportation rating was not significant ($p = .482$).

Although the multiple regression model did not indicate a significant effect of mean transportation rating on CT change score, a mediation analysis was nonetheless conducted to provide a more robust test of the relationship between fiction reading time and mean transportation, upon CT change score. The average causal mediation effect was 0.007 [95% CI $-0.01, 0.03$], and was not significant ($p = .36$). The average direct effect was 0.20 [95% CI $0.12, 0.38$], and this was significant ($p < .001$), as was the total effect of 0.21 [95% CI $0.12, 0.39$] ($p < .001$). The proportion of the effect of fiction reading time upon CT change score that was mediated by mean transportation score was 0.03 [95% CI $-0.04, 0.14$], and this was not significant ($p = .36$). The effect

of fiction reading time on change in CT scores from t1 to t2 was not mediated via mean transportation scores.

Model testing

Baseline CT scores

Accuracy testing for regression model 2 described in Table 14

First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 5 cases were found to have standardised residuals outside ± 2 . Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; one case (107) had a standardised residual of -3.34. The model therefore conforms to predictions as an accurate model. Cook's distance was calculated for the 5 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (2+1)/121 = 0.02$. Taking the 3x average leverage as a cut-off (Stevens, 2002), the maximum was 0.06 and none of the cases with large standardised residuals exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(2+1)/121] = 1.06$$

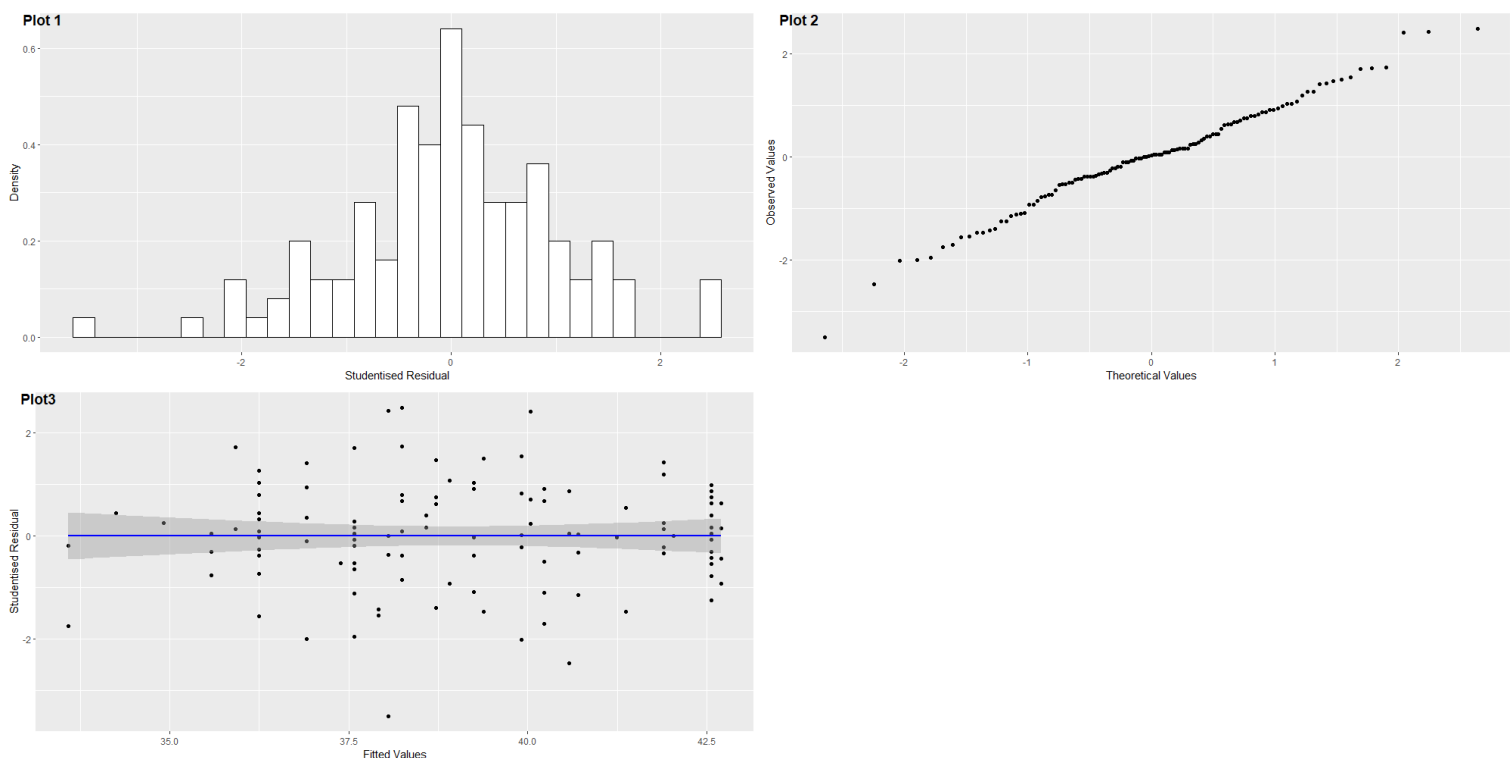
$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(2+1)/121] = 0.94$$

All of the cases with large standardised residuals fell beneath the bottom limit of these covariance ratio boundaries. However, as all of these cases had low Cook's

distance values (< 0.1), this is not a major concern. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.95$, $p = .772$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Educational level, Tolerance = .89, VIF = 1.12; Reading engagement, Tolerance = .89, VIF = 1.12), with the mean VIF = 1.22. This suggests that regression 2 met the relevant assumptions and this model can be taken to be accurate for the sample.

Standardised residuals from regression 2 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggest the model conforms to the assumption of normality.

Figure 23: Plots for regression model described in Table 14



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = -0.02$, ND - UG = 0.00, ND - PG = -0.05, reading engagement $b = 0.01$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Accuracy testing for regression model 5 described in Table 16

Regression 5 was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 7 cases were found to have standardised residuals outside ± 2 . Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; one case (107) had a standardised residual of -3.36; all other cases rounded to ± 2 . The model therefore conforms to predictions as an accurate model. Cook's distance was calculated for the 7 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (3+1)/121 = 0.03$. Finally, covariance ratios were checked; the cut-offs were set as:

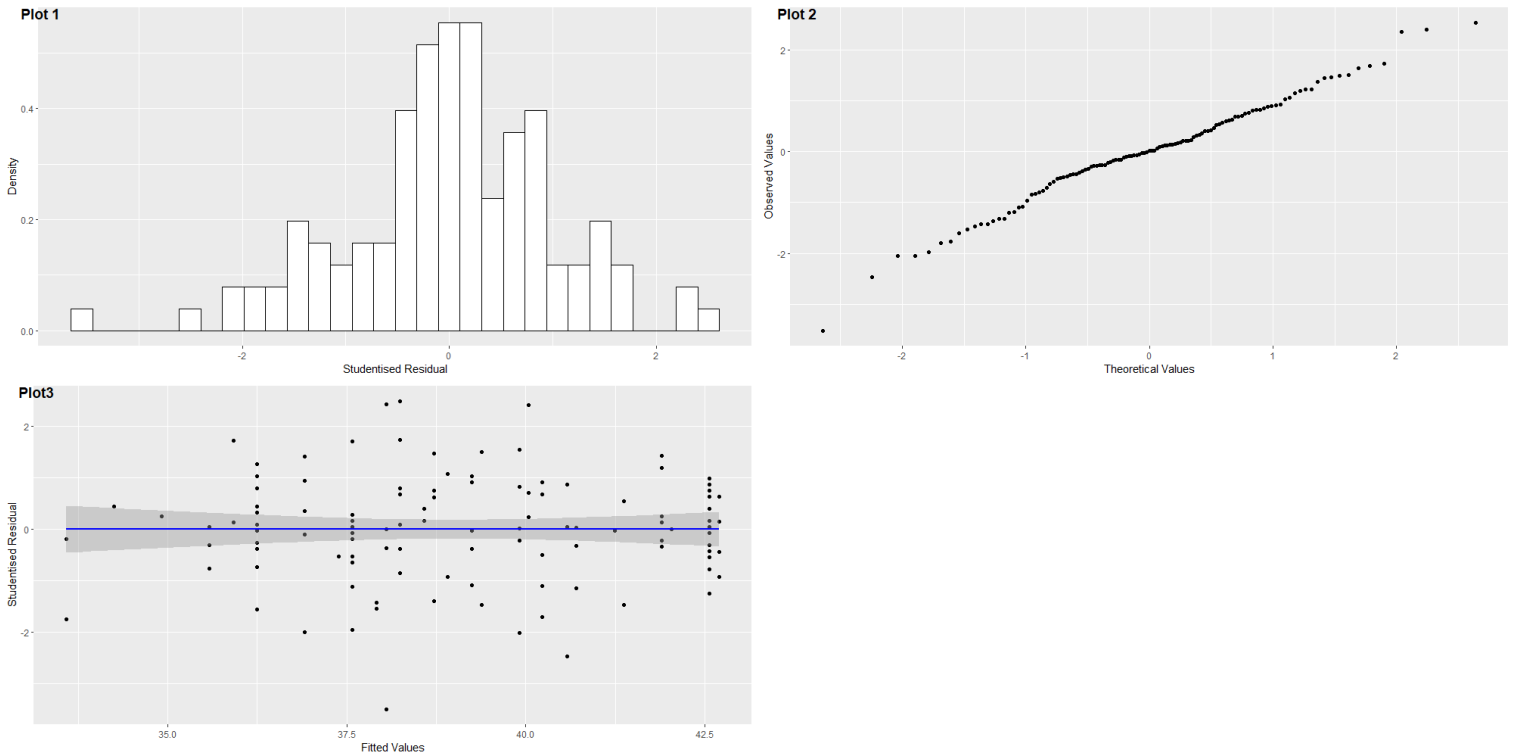
$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(3+1)/121] = 1.09$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(3+1)/121] = 0.10$$

None of the 7 cases with large standardised residuals fell outside of these covariance ratio boundaries. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.94$, $p = .698$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Educational level, Tolerance = .88, VIF = 1.14; Nonfiction engagement, Tolerance = .83, VIF = 1.20, Fiction engagement, Tolerance = .75, VIF = 1.34), with the mean VIF = 1.23. This suggests that regression 5 met the relevant assumptions and this model can be taken to be accurate for the sample.

Standardised residuals from regression 5 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggest the model conforms to the assumption of normality.

Figure 24: Plots for regression model described in Table 16



CT change from t1 to t2

Accuracy testing for regression model 1 described in Table 17

Regression 1 was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 5 cases were found to have standardised residuals outside ± 2 .

Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 3 cases had a standardised residual in excess of this (case 1 = 2.70, case 31 = 2.94, case 80 = 3.24). Cook's distance was calculated for the 5 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was

assessed; average leverage for the model is $(k+1)/n : (1+1)/121 = 0.02$. Taking the 3x average leverage as a cut-off (Stevens, 2002), the maximum was 0.06; none of the cases with large standardised residuals exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

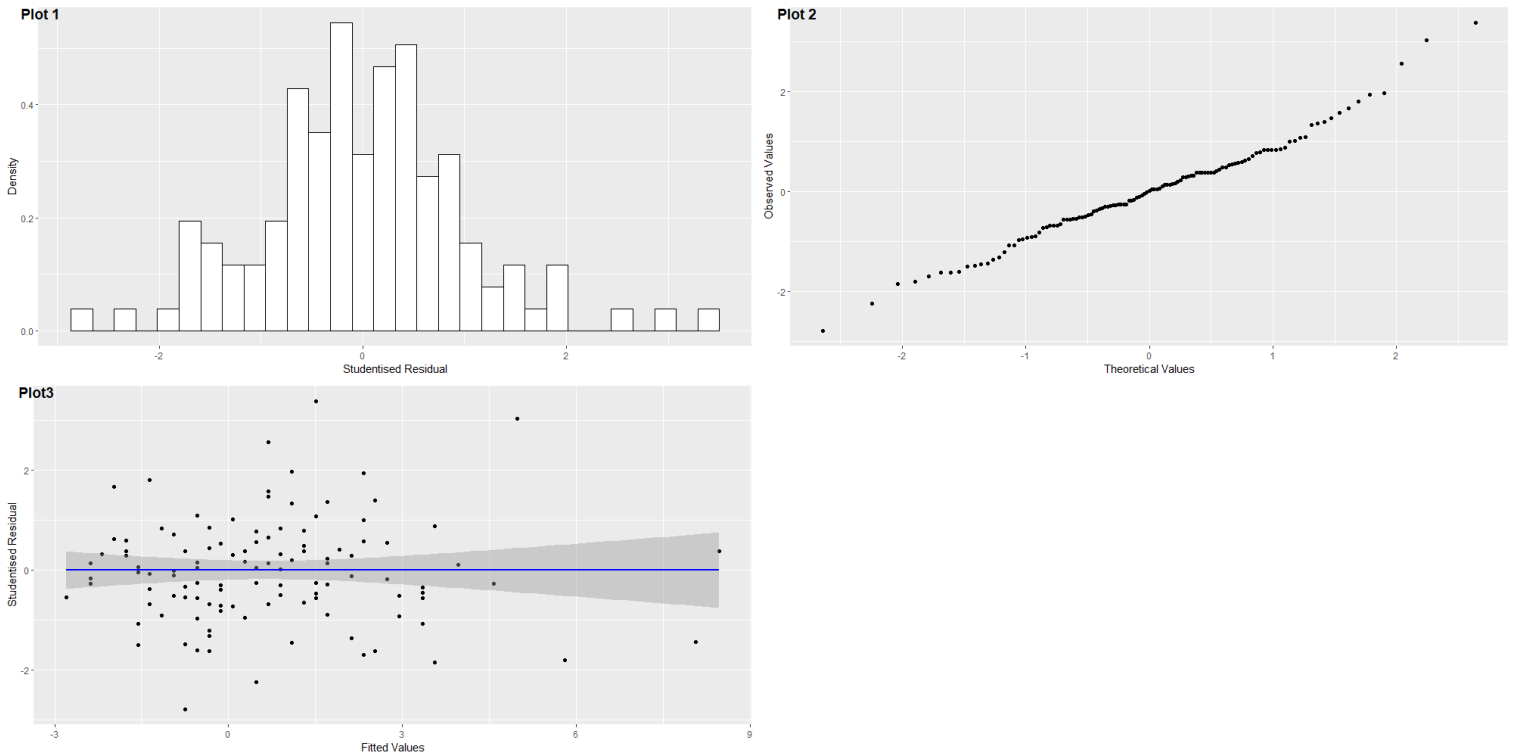
$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(1+1)/121] = 1.06$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(1+1)/121] = 0.94$$

All of the cases but one (case 34) with large standardised residuals fell beneath the bottom limit of these covariance ratio boundaries. However, as all of these cases had low Cook's distance values (< 0.2), this is not a major concern. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model (D-W = 1.90, $p = .572$). This suggests that regression 1 met the relevant assumptions and this model can be taken to be accurate for the sample.

Standardised residuals from regression 1 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggests the model conforms to the assumption of normality.

Figure 25: Plots for regression model described in Table 17



Accuracy testing for regression model 3 described in Table 18

Regression 3 was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 4 cases were found to have standardised residuals outside ± 2 . Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 4 cases had a standardised residual exceeding this limit (case 1 = -3.06, case 31 = 3.58, case 47 = 2.78, case 80 = 3.75); this means that 3% of cases fell outside standardised residual values of ± 2.5 . Cook's distance was calculated for the 4 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model.

Next, leverage was assessed; average leverage for the model is $(k+1)/n : (2+1)/121 = 0.02$. Taking the $3x$ cut-off limit (Stevens, 2002), the maximum was 0.06; no cases exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(2+1)/121] = 1.06$$

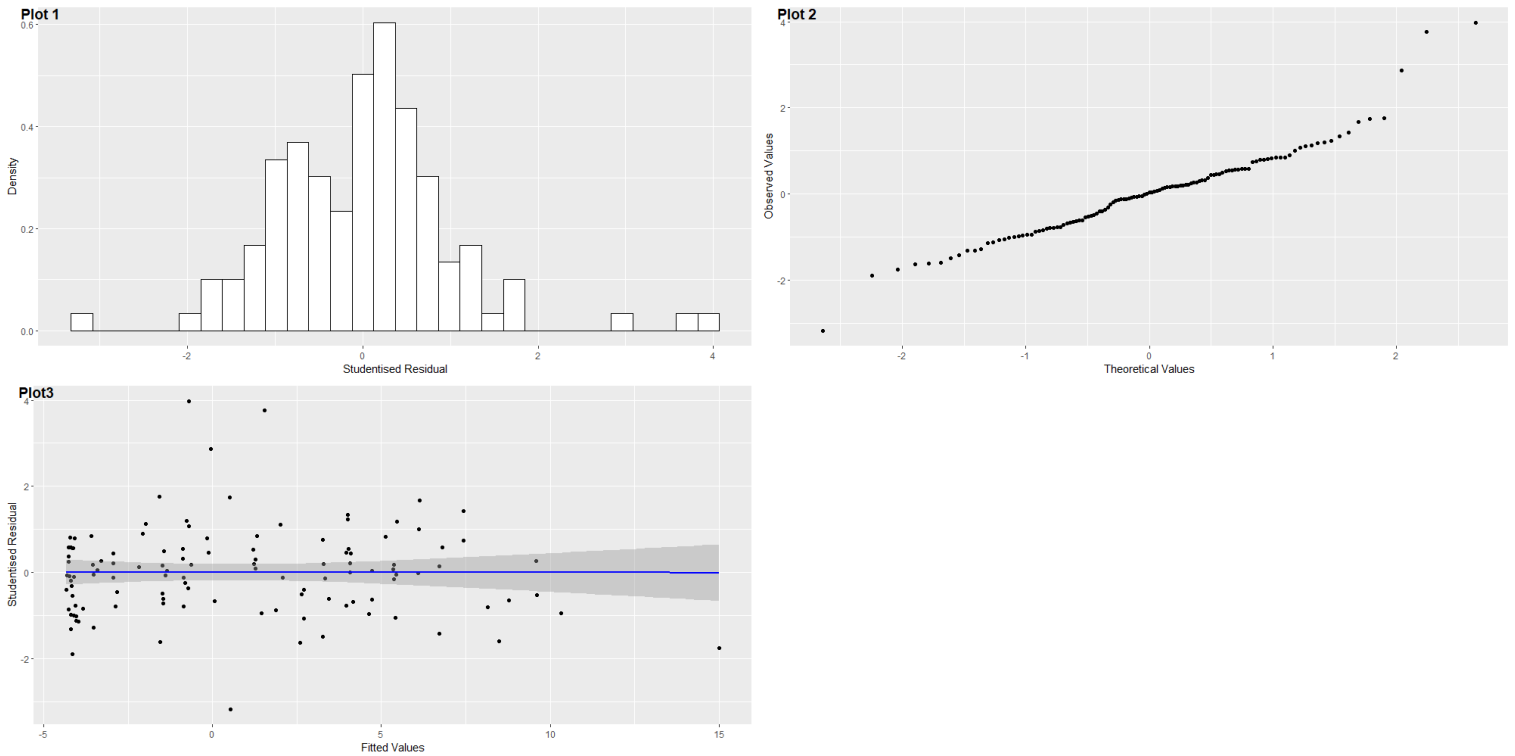
$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(2+1)/121] = 0.94$$

All of the 4 cases with large standardised residuals fell beneath the lower limit of these covariance ratio boundaries. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.95, p = .772$).

Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Nonfiction entries, Tolerance = .95, VIF = 1.05, Fiction entries, Tolerance = .95, VIF = 1.05), with the mean VIF = 1.05. This suggests that regression 3 met the relevant assumptions and this model can be taken to be accurate for the sample.

Standardised residuals from regression 3 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggest the model conforms to the assumption of normality.

Figure 26: Plots for regression model described in Table 18



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = -0.11$, nonfiction entries $b = 0.005$, fiction entries $b = 0.012$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Accuracy testing for regression model 1 described in Table 19

Regression 1 was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 6 cases were found to have standardised residuals outside ± 2 .

Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 2 cases had a standardised residual in excess of this (case 1 = -2.74, case 80 = 3.36). Cook's distance was calculated for the 6 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (1+1)/121 = 0.02$. Taking the 3x average leverage as a cut-off (Stevens, 2002), the maximum was 0.06; none of the cases with large standardised residuals exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

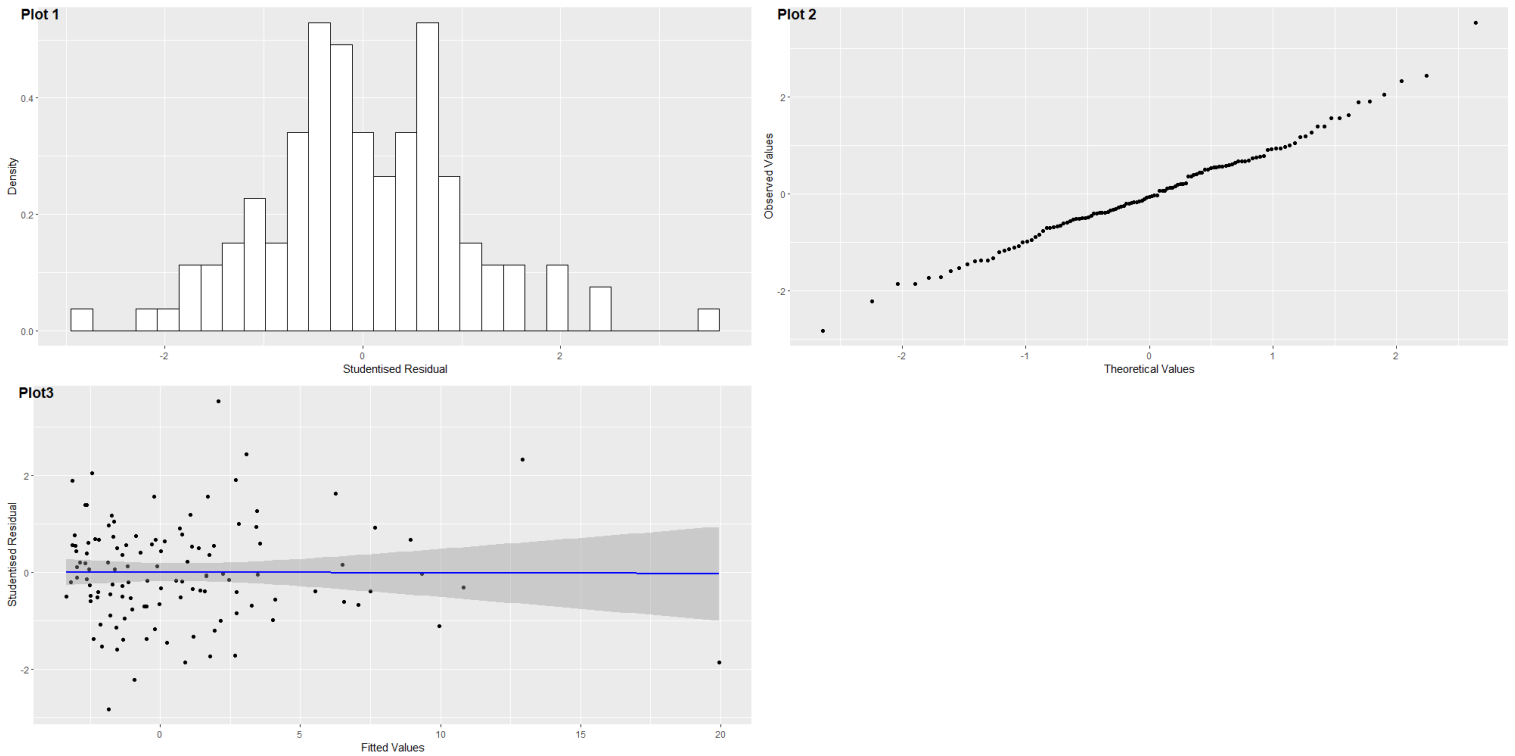
$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(1+1)/121] = 1.06$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(1+1)/121] = 0.94$$

3 of the cases with large standardised residuals fell beneath the bottom limit of these covariance ratio boundaries (case 1, 47, 80). However, as all of these cases had low Cook's distance values (< 0.05), this is not a major concern. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model (D-W = 1.82, $p = .312$).

Standardised residuals from regression 1 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggests the model conforms to the assumption of normality.

Figure 27: Plots for regression model described in Table 19



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = -0.01$, total time $b = 0.004$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Accuracy testing for regression model 3 described in Table 20

Regression 3 was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 6 cases were found to have standardised residuals outside ± 2 .

Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 2 cases exceeded this limit. Cook's distance was calculated for the 6 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (2+1)/121 = 0.02$. Taking 3x (0.06) average leverage as a cut-off (Stevens, 2002), two cases with large standardised residuals exceeded this; case 31 had a leverage of 0.18, and case 95 had a leverage of 0.29. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(2+1)/121] = 1.06$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(2+1)/121] = 0.94$$

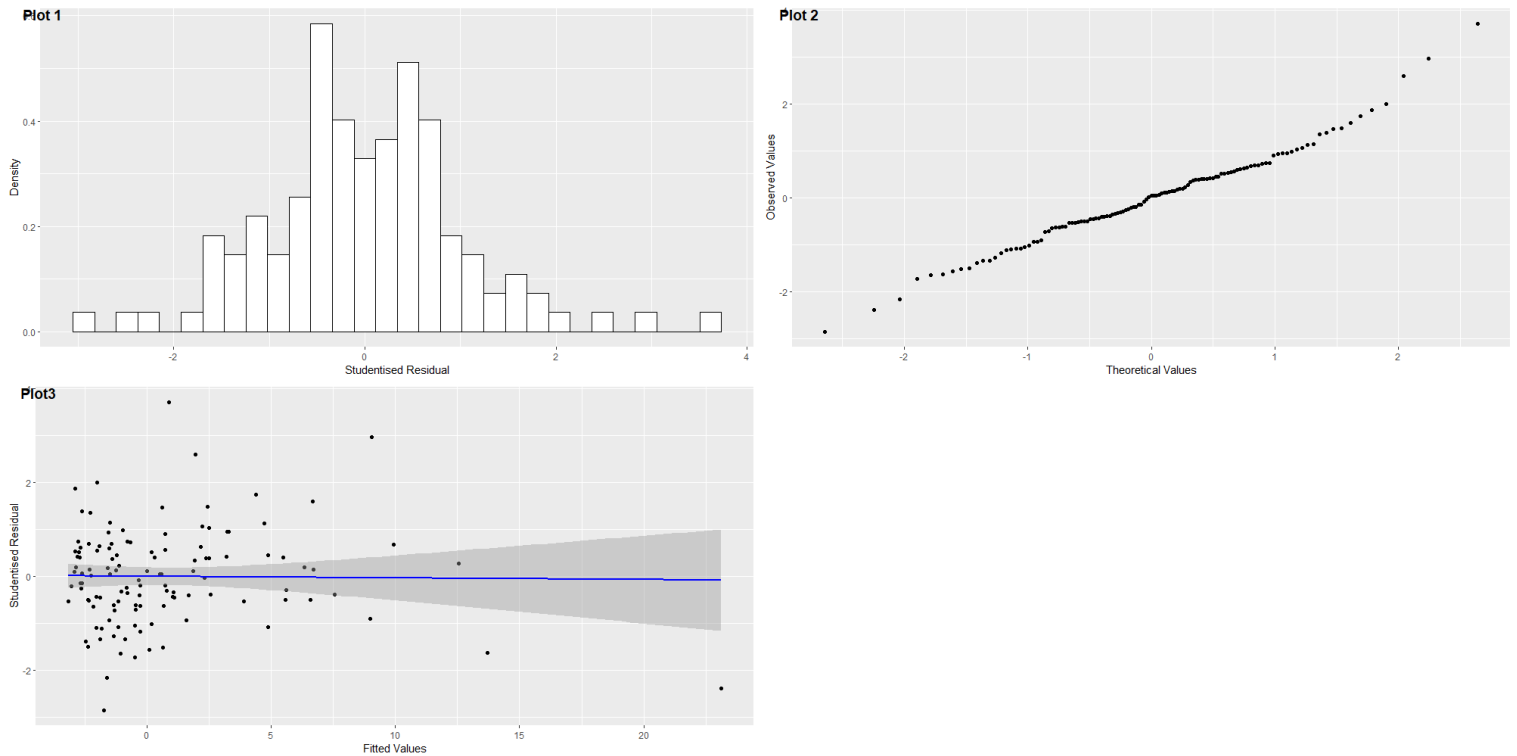
4 cases with large standardised residuals fell beneath these covariance ratio boundaries (case 1, 34, 47,80), however these cases had low Cook's distances (< 0.7) which suggests they were not concerning. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.79, p = .202$).

Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Nonfiction time, Tolerance = 0.10, VIF = 1.00, Fiction time, Tolerance = .10, VIF = 1.00), with the mean VIF = 1.00.

Standardised residuals from regression 3 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of

Standardised residuals against predicted values. These plots suggests the model conforms to the assumption of normality.

Figure 28: Plots for regression model described in Table 20



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = 0.08$, nonfiction time $b = -0.01$, fiction time $b = 0.02$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Experimental Groups

Accuracy testing for the regression model described in Table 22

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 5 cases were found to have standardised residuals outside ± 2 .

Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 3 cases fell outside this limit (case 1, 31, 80).

Cook's distance was calculated for the 5 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n$: $(3+1)/121 = 0.03$. Taking the 3x average leverage as a cut-off (Stevens, 2002), the maximum was 0.09; one of the cases with large standardised residuals exceeded this (case 31 = 0.11). Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(3+1)/121] = 1.09$$

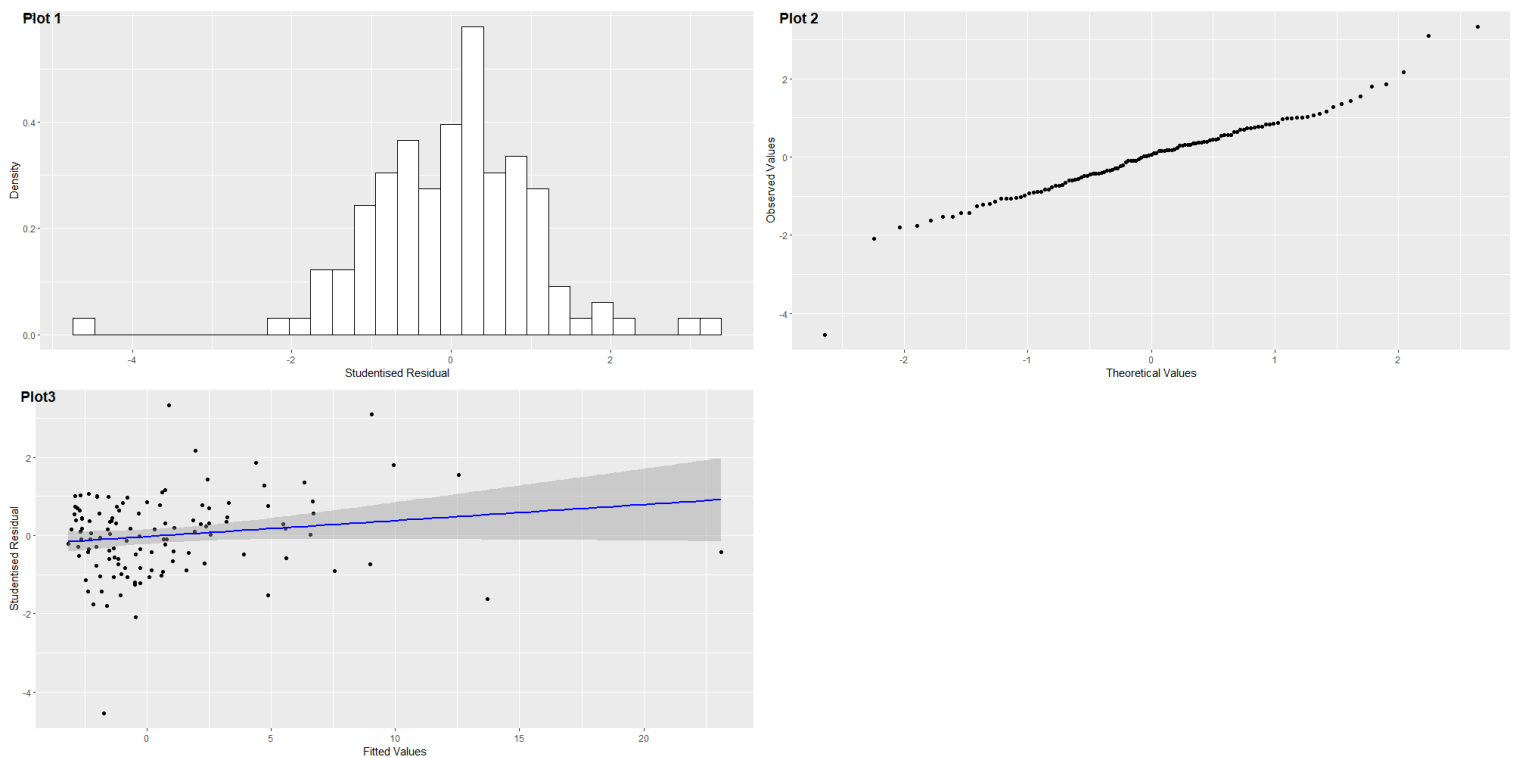
$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(3+1)/121] = 0.91$$

All 5 of the cases with large standardised residuals fell beneath these covariance ratio. However, all of these cases had very low Cook's distance values (< 0.1), and therefore are not a cause for concern. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.95$, $p = .84$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Nonfiction entries, Tolerance = .80 VIF = 1.26, Fiction entries,

Tolerance = .73, VIF = 1.37; Group, Tolerance = 0.61, VIF = 1.65), with the mean VIF = 1.65. This suggests that the regression met the relevant assumptions and this model can be taken to be accurate for the sample.

Standardised residuals from regression 1 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggest the model conforms to the assumption of normality.

Figure 29: Plots for regression model described in Table 22



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression. Table 55 shows the difference between the

original regression coefficients, and the distance to the mean of bootstrapped coefficient values.

Table 55: Comparison of original and bootstrapped regression

Variable	Original B	Distance to bootstrap M	SE
Constant	-5.59	-0.02	2.15
Nonfiction entries	0.01	0.01	0.12
Fiction entries	0.73	0.01	0.13
Group:			
Cf	-0.64	-0.07	2.15
Ff	-2.05	-0.11	2.81
Fn	8.33	-0.23	3.62
Nf	-0.95	-0.04	2.37
Nn	1.28	-0.09	1.99

This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Accuracy testing for the regression model described in Table 23

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 5 cases exceeded ± 2 . Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 2 cases exceeded 2.5. Cook's distance was calculated for the 4 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (3+1)/121 = 0.03$. Taking the 3x average leverage as a cut-off (Stevens, 2002), the maximum was 0.09; 2 of the cases with large standardised residuals exceeded this (case 31 = 0.21, case 95 = 0.30). Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(3+1)/121] = 1.09$$

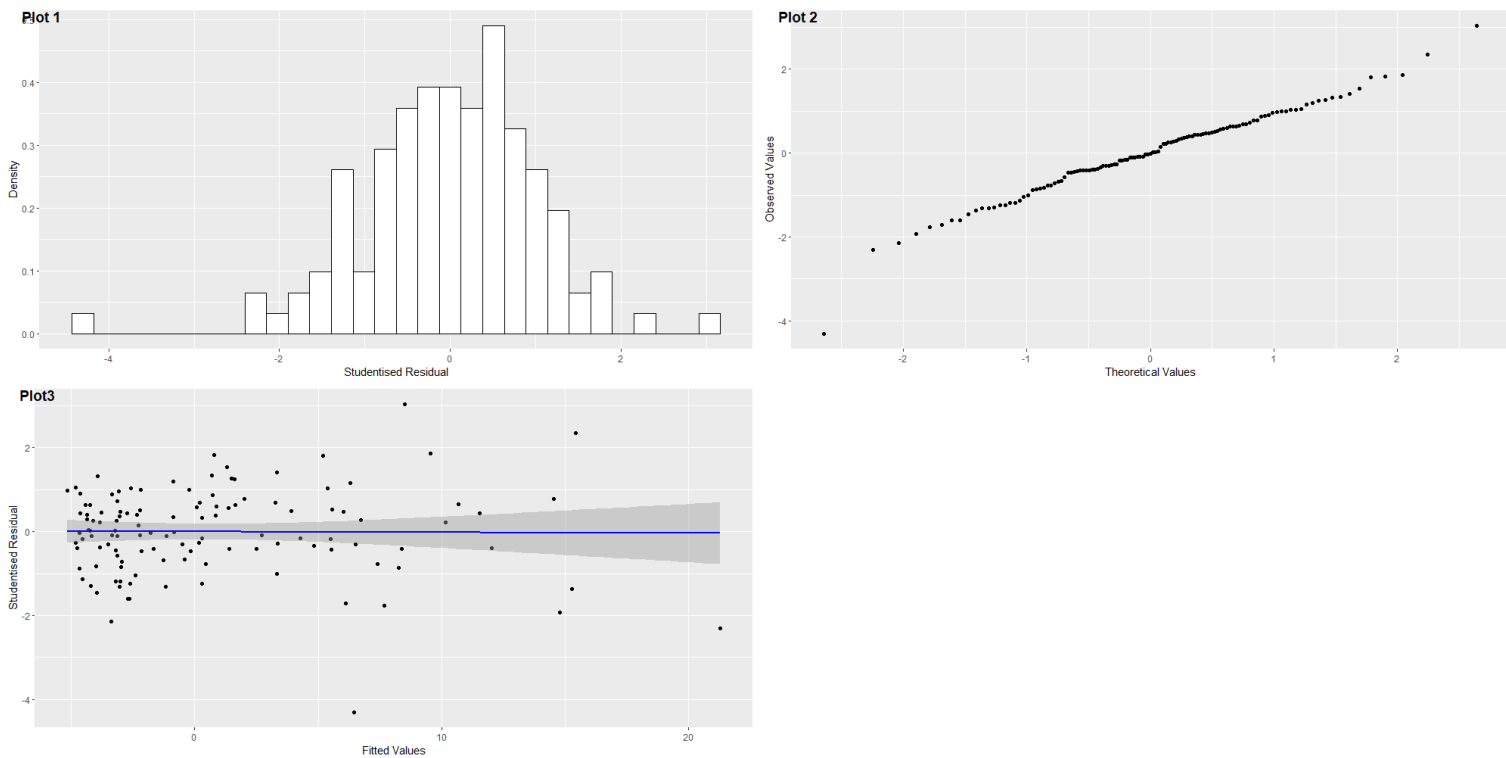
$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(3+1)/121] = 0.91$$

3 of the cases with large standardised residuals fell beneath these covariance ratio limits; however, all of these cases had very low Cook's distance values (< 0.2), and therefore are not a cause for concern. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.86, p = .44$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Nonfiction time, Tolerance = .88, VIF = 1.14, Fiction time, Tolerance = .88, VIF = 1.14; Group, Tolerance = .78, VIF = 1.28), with the mean VIF = 1.52. This suggests that the regression met the relevant assumptions and this model can be taken to be accurate for the sample.

Standardised residuals from regression 2 were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot

of the theoretical values against observed values. Plot 3 shows a scatterplot of studentised residuals against predicted values. These plots suggest the model conforms to the assumption of normality.

Figure 30: Plots for regression model described in Table 23



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression.

Table 56 shows the difference between the original regression coefficients, and the distance to the mean of bootstrapped coefficient values.

Table 56: Comparison of original and bootstrapped regression

Variable	Original B	Distance to bootstrap M	SE
Constant	-5.51	-0.04	1.83
Nonfiction time	0.08	-0.01	0.07
Fiction time	0.22	0.02	0.07
Group:			
Cf	0.77	0.04	2.14
Ff	3.04	-0.07	2.56
Fn	10.63	-0.06	3.23
Nf	0.97	-0.10	2.35
Nn	0.94	0.17	2.19

This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Distinguishing improvement in CT score from t1 to t2 from decrease

Accuracy testing for regression model 3 described in Table 25

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside

of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 1 case (28) was found to have a standardised residual of -2.34 . Cook's distance was 0.18, therefore this case had no undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (2+1)/121 = 0.02$. Taking the 3x average leverage (0.06) as a limit (Stevens, 2002), case 28 was within this with a leverage of 0.05. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(2+1)/121] = 1.06$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(2+1)/121] = 0.94$$

Case 28 fell within these limits (0.98). A Durbin-Watson test confirmed that the assumption of independent errors was met for the model (D-W = 1.86, $p = .428$).

Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (Total entries, Tolerance = .59, VIF = 1.71, Total time, Tolerance = .59, VIF = 1.71), with the mean VIF = 1.71. This suggests that the regression met the relevant assumptions and this model can be taken to be accurate for the sample. Finally, the linearity of logit was tested, using interactions between each predictor and the log; no interaction was significant in the model and therefore no main effect violated the assumption of linearity of the logit.

Accuracy testing for regression model 3 described in Table 27

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside

of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 3 cases (14, 28, 106) were found to have a standardised residual outside ± 2 , of which one case (28) fell outside ± 2.5 . None of these cases exceeded a Cook's distance of 1, therefore these cases had no undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (4+1)/121 = 0.04$. Taking the 2x average leverage as a cut-off (Hoaglin & Welsch, 1978) the maximum was 0.08, taking a 3x average leverage (Stevens, 2002) the limit was 0.12; case 14 exceeded this with a leverage of 0.20. However case 14 had a low Cook's distance of 0.25 and therefore is not highly concerning. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(4+1)/121] = 1.12$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(4+1)/121] = 0.88$$

Case 28 fell just below these limits (0.82). A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.89, p = .602$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not highly concerning for this model (Nonfiction entries, Tolerance = .46, VIF = 2.17, Fiction entries, Tolerance = .45, VIF = 2.24, Nonfiction time, Tolerance = .43, VIF = 2.31, Fiction time, Tolerance = .48, VIF = 2.10), with the mean VIF = 2.21. This suggests that the regression met the relevant assumptions and this model can be taken to be accurate for the sample. Finally, the linearity of logit was tested, using interactions between each predictor and the log; no interaction was significant in the model and therefore no main effect violated the assumption of linearity of the logit. This suggests the regression results are accurate and generalisable.

Accuracy testing for the logistic regression model described in Table 28

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 3 cases were found to have a large standardised residual. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; however no cases were found to have a standardised residual outside ± 2.5 . A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.88$, $p = .548$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not highly concerning for this model (Nonfiction items, Tolerance = .79, VIF = 1.27, Fiction items, Tolerance = .67, VIF = 1.44, Group, Tolerance = .57, VIF = 1.77), with the mean VIF = 1.65. This suggests that the regression met the relevant assumptions and this model can be taken to be accurate for the sample. Finally, the linearity of logit was tested, using interactions between each predictor and the log; no interactions were significant in the model, and therefore the assumption of linearity of logit was met. Overall, this suggests the model fits the data and can be generalised.

Accuracy testing for the logistic regression model described in Table 29

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 3 cases were found to have a large standardised residual. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; one case was found to have a standardised residual outside ± 2.5 (case 28 = 3.06). A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.89$, $p = .528$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not highly concerning for this model (Nonfiction time, Tolerance = .87, VIF = 1.17, Fiction time, Tolerance = .86, VIF = 1.17, Group, Tolerance = .75, VIF = 1.34), with the mean VIF = 1.54. This suggests that the regression met the relevant assumptions and this model can be taken to be accurate for the sample. Finally, the linearity of logit was tested, using interactions between each predictor and the log; no interactions were significant in the model; the interaction for fiction time was significant in the model, and therefore the assumption of linearity of logit was not met. Therefore a robust bootstrapped (100,000 replications) version of the logistic regression was conducted, which found values close to the original model.

Table 57 shows the distance from the original coefficient value to the bootstrapped mean. Overall, this suggests the model fits the data and can be generalised.

Table 57: Comparison of original and bootstrapped regression

Variable	Original B	Distance to bootstrap M	SE
Constant	-1.34	-0.15	0.73
Nonfiction time	0.02	0.00	0.02
Fiction time	0.07	0.01	0.04
Group:			
Cf	0.56	0.03	0.79
Ff	0.43	0.03	1.06
Fn	2.36	0.98	3.69
Nf	-0.94	-0.27	1.69
Nn	0.15	0.00	1.18

Average hours per entry made

Accuracy testing for the logistic regression model described in Table 30

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; however no cases were found to have a standardised residual outside ± 2 . A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.86$, $p = .388$). Finally, the linearity of logit was tested, using interactions between the predictor and

the log; the interaction was significant in the model, and therefore the assumption of linearity of logit was not met. Therefore a robust bootstrapped (100,000 replications) version of the logistic regression was conducted, which found values close to the original model. The distance from the original coefficient value to the bootstrapped mean was: constant = -0.03; $T t/e = 0.02$. Overall, this suggests the model fits the data and can be generalised.

Accuracy testing for the logistic regression model described in Table 31

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits. Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; however no cases were found to have a standardised residual outside ± 2 . A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.84$, $p = .386$). Finally, the linearity of logit was tested, using interactions between each predictor and the log; no interactions were significant in the model, and therefore the assumption of linearity of logit was met. Overall, this suggests the model fits the data and can be generalised.

Accuracy testing for the regression model described in Table 32

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 6 cases were found to have standardised residuals outside ± 2 .

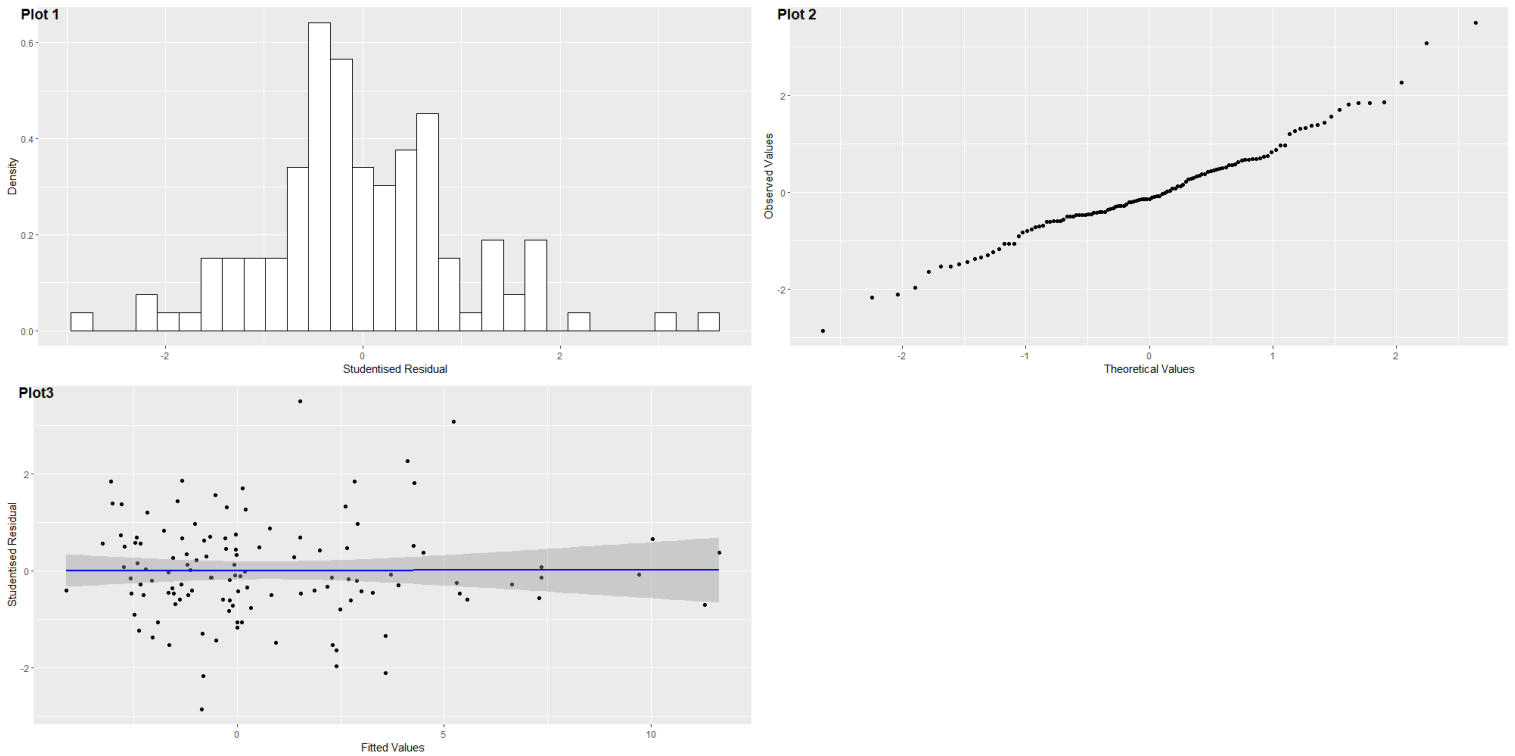
Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 3 cases exceeded this limit (case 1 = -2.78; case 31 = 2.97; case 80 = 3.34). This means that 2% of cases fell outside standardised residual values of ± 2.5 . Cook's distance was calculated for the 6 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (2+1)/121 = 0.02$. Taking the 3x (0.06) average leverage as a cut-off (Stevens, 2002), no cases exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(2+1)/121] = 1.06$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(2+1)/121] = 0.94$$

3 of the cases with large standardised residuals fell beneath the bottom limit of these boundaries (case 1 = 0.90; case 31 = 0.89; case 80 = 0.84). A Durbin-Watson test confirmed that the assumption of independent errors was met for the model (D-W = 1.80, $p = .31$). Standardised residuals from the regression were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggests the model conforms to the assumption of normality.

Figure 31: Plots for regression model described in Table 32



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = 0.02$, total half-hours per entry $b = -0.007$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Accuracy testing for the regression model described in Table 33

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 121, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 6.05 (5%) cases would have standardised residuals outside of these limits; 5 cases were found to have standardised residuals outside ± 2 .

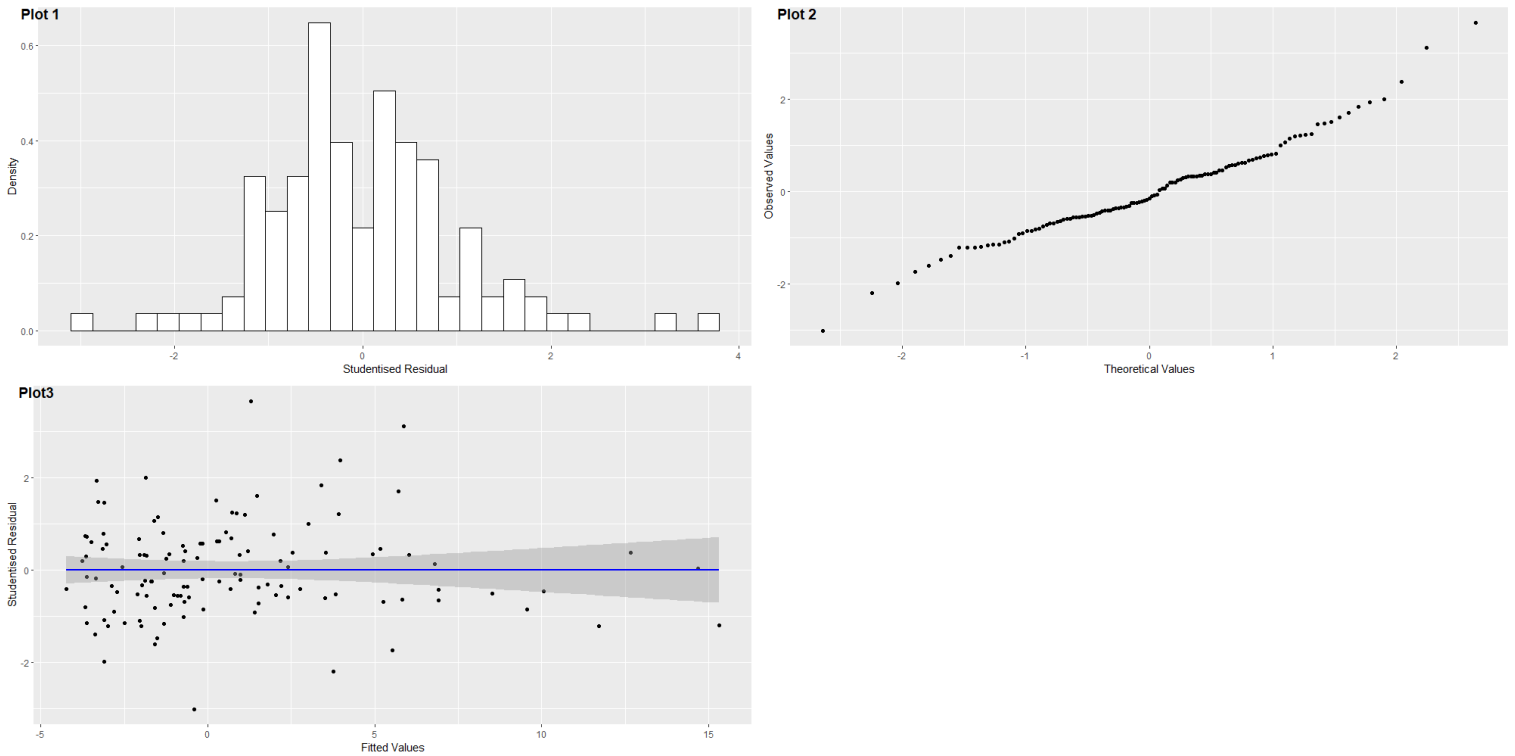
Furthermore, assuming that 99% of cases will be between ± 2.5 , 1.21 cases are expected to have standardised residuals outside ± 2.5 ; 3 cases exceeded this limit (case 1 = -2.91; case 31 = 3.00; case 80 = 3.47). This means that 2% of cases fell outside standardised residual values of ± 2.5 . Cook's distance was calculated for the 5 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (3+1)/121 = 0.03$. Taking either the or 3x (0.09) average leverage as a cut-off, no cases with large standardised residuals exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(3+1)/121] = 1.09$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(3+1)/121] = 0.91$$

3 of the cases with large standardised residuals fell beneath these boundaries (case 1 = 0.83, 31 = 0.84, 80 = 0.75). A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.79$, $p = .254$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (NF h/i, Tolerance = 0.91, VIF = 1.10, F h/i, Tolerance = .91, VIF = 1.10), with the mean VIF = 1.10. Standardised residuals from the regression were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggests the model conforms to the assumption of normality.

Figure 32: Plots for regression model described in Table 33



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = -0.05$, nonfiction $t/e b = -0.01$, fiction $t/e b = 0.06$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Narrative Transportation

Accuracy testing for the regression model described in Table 37

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 79, and the

assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 3.95 (5%) cases would have standardised residuals outside of these limits; 4 cases were found to have standardised residuals outside ± 2 .

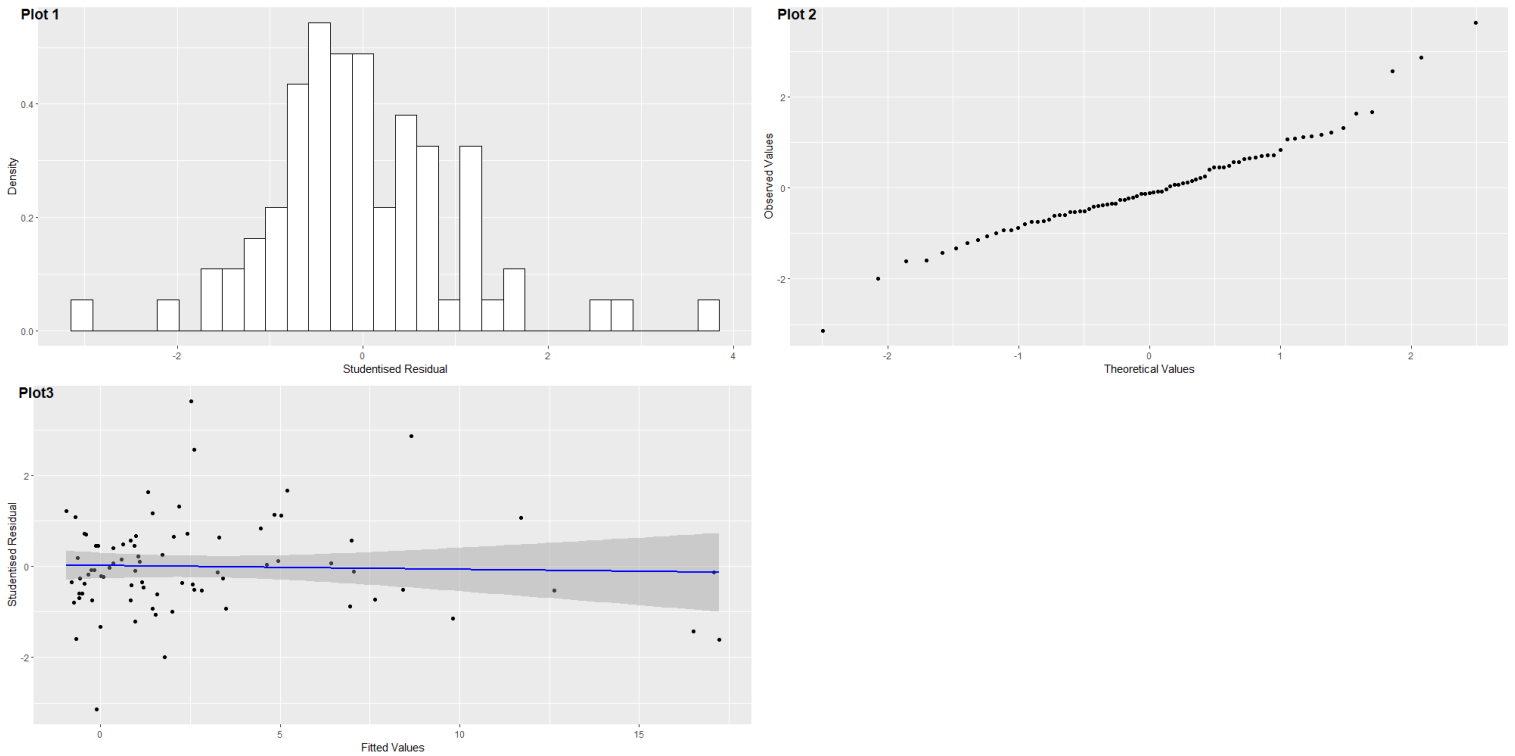
Furthermore, assuming that 99% of cases will be between ± 2.5 , 0.79 cases are expected to have standardised residuals outside ± 2.5 ; 3 cases exceeded this limit (case 1 = -2.97; case 31 = 2.74; case 80 = 3.36). Cook's distance was calculated for the 4 cases with large standardised residuals, and none were found to exceed 1; therefore none of these cases had an undue influence on the model. Next, leverage was assessed; average leverage for the model is $(k+1)/n : (2+1)/121 = 0.02$. Taking the 3x (0.06) average leverage as a cut-off (Stevens, 2002), none of the cases with large standardised residual exceeded this. Finally, covariance ratios were checked; the cut-offs were set as:

$$\text{CVR} > 1 + [3(k+1)/n] = 1 + [3(2+1)/121] = 1.06$$

$$\text{CVR} < 1 - [3(k+1)/n] = 1 - [3(2+1)/121] = 0.94$$

All of the cases with large standardised residuals fell beneath these boundaries. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model ($D-W = 1.67$ $p = .132$). Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (NF Peak Time, Tolerance = 0.97, VIF = 1.03, Fiction Peak Time, Tolerance = 0.97, VIF = 1.03), with the mean VIF = 1.03. Standardised residuals from the regression were used to test the normality of the model. Plot 1 shows a histogram of the Standardised residuals. Plot 2 shows a Q-Q plot of the theoretical values against observed values. Plot 3 shows a scatterplot of Standardised residuals against predicted values. These plots suggest the model conforms to the assumption of normality.

Figure 33: Plots for regression model described in Table 37



Due to some possible outliers in the model, a bootstrapped robust version of the regression was run (bootstrap replications = 2000) and this found coefficient values very close to the original regression (distance from mean of bootstrap samples: constant $b = -0.29$, nonfiction peak $b = 0.08$, fiction peak $b = 0.001$). This suggests that although there were some potential outliers in the data, these did not significantly impact the model as the coefficient values are very close to those derived from a bootstrapped version. This suggests the model is accurate and generalisable.

Accuracy testing for the logistic regression model described in Table 42

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 79, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was

expected that approximately 3.95 (5%) cases would have standardised residuals outside of these limits; three cases were found to have standardised residuals outside ± 2 . Furthermore, assuming that 99% of cases will be between ± 2.5 , 0.79 cases are expected to have standardised residuals outside ± 2.5 ; no cases exceeded this limit. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model (D-W = 1.68, $p = .166$). Finally, the linearity of logit was tested, using interactions between the predictor and the log; the interactions were not significant in the model, and therefore the assumption of linearity of logit was met. Overall, this suggests the model fits the data and can be generalised.

Accuracy testing for the logistic regression model described in Table 43

The regression was assessed for accuracy. First, outliers and influential cases were sought. Standardised residuals were assessed; given the sample size of 79, and the assumption that 95% of cases will have standardised residuals falling within ± 2 , it was expected that approximately 3.95 (5%) cases would have standardised residuals outside of these limits; one case was found to have standardised residuals outside ± 2 . Furthermore, assuming that 99% of cases will be between ± 2.5 , 0.79 cases are expected to have standardised residuals outside ± 2.5 ; no cases exceeded this limit. A Durbin-Watson test confirmed that the assumption of independent errors was met for the model (D-W = 1.66, $p = .152$). Finally, the linearity of logit was tested, using interactions between the predictor and the log; two interactions were significant in the model (Log F Min time $p = .017$, Log F Peak time $p = .032$) and therefore the assumption of linearity of logit was not met. Therefore a robust bootstrapped (100,000 replications) version of the logistic regression was conducted, which found values close

to the original model. The distance from the original coefficient value to the bootstrapped mean was: constant = -0.15; F Min time = 0.03, NF Min time = 0.02, F Peak time = 0.02 , NF Peak time = 0.01. Overall, this suggests the model fits the data and can be generalised.

Appendix C: Validation of ICTET-A with text B

As two texts were used with the ICTET-A, only one of which (text A) was previously validated, scores with text A and text B were assessed for consistency at t1 and at t2, to ensure that performance with either text was similar enough such that the two texts can be considered to be equivalent.

The mean score on the ICTET-A at t1 with text A ($n = 62$) was 39.31 ($SD = 9.3$), with a median of 39.5. The lowest score was 10 and the highest 60, giving a range of 50. As the maximum score available in the ICTET-A is 70, the sample mean represents 56% of the possible marks. For t1 text B ($n = 59$) the mean score was 38.75 ($SD = 8.42$), with a median of 39. The lowest score was 20 and the highest 59, giving a range of 39. The sample mean represents 55% of the possible marks. Combining texts A and B, the mean ICTET-A score at t1 was 39.03 ($SD = 8.85$), median 39, minimum 10 and maximum 60 (range = 50), with the sample mean being 56% of the total possible score. A Shapiro-Wilk normality test showed combined A and B scores at t1 to be normally distributed ($W = 0.99$, $p = .5$).

At t2, the mean ICTET-A with text A ($n = 59$) score was 38.49 ($SD = 10.73$), with a median of 38. The lowest score was 11 and the highest 64, giving a range of 53. The sample mean represents 55% of the possible marks. For t2 text B ($n = 62$) the mean score was 40.89 ($SD = 11.19$), with a median of 41. The lowest score was 13 and the highest 64, giving a range of 51. The sample mean represents 58% of the possible marks. Combining texts A and B, the mean ICTET-A score at t2 was 39.7 ($SD = 10.99$), median 40, minimum 11 and maximum 64 (range = 53), with the sample mean being 57% of the total possible score. A Shapiro-Wilk normality test

showed combined A and B scores at t2 to be normally distributed ($W = 0.99$, $p = 0.541$).

Levene's test for homogeneity of variance (centred on the median score) was used to assess whether scores with texts and A and B had differing variance. There was no difference in the variance of scores between the two texts at t1 ($F(1,119) = 0.37$, $p = .545$), nor at t2 ($F(1,119) = .21$, $p = .644$). Student's two sample t-test was therefore used to test for difference in scores between text A and B, and found no significant difference at t1 ($t(119) = 0.35$, $p = .729$), nor at t2 ($t(119) = 1.20$, $p = .232$). This suggests the two texts are equally challenging for participants as they yield similar scores.

Cronbach's alpha was used to assess the internal consistency of the ICTET-A with text A (combining t1 and t2); alpha was 0.87, with a bootstrapped (1000 samples) 95% CI of 0.80 to 0.90. Likewise for text B (combining t1 and t2), alpha was 0.88 with a bootstrapped (1000 samples) 95% CI of 0.84 to 0.91. Therefore the ICTET-A with both texts was shown to have good internal consistency.

Appendix D: Study three

Topic sheet for interview session 1

1. Could you name a few of your favourite authors/writers?
2. Are you mainly a fiction or nonfiction reader?
 - a. How would you describe your habits or preferences in terms of reading fiction versus nonfiction?
3. Transportation
 - a. When you read, do you tend to be very transported (so you lose track of your surroundings)? Please describe...
 - b. Do you put yourself inside of the text, or do you feel quite separate?
 - i. Do you step into the author's position, or put yourself in the shoes of the figures in the text?
4. What is critical thinking?
 - a. What kinds of adjectives would best describe a critical thinker?
 - b. Do you think there is single 'true' or correct answer to most issues?
 - i. Do you think there are issues on which we can all be equally right even if we disagree?
 - c. Do you think your critical thinking approaches have changed over time across your lifespan?
5. Critical Thinking when reading

- a. Can you describe any kinds of evaluations that usually go through your mind while reading?
 - i. For example, do you guess what will come next in text? Do you judge the author's intentions?
 - b. How would you characterise the similarities or difference in the way you think during reading fiction versus nonfiction?
 - i. Can you describe the evaluations you typically make when reading nonfiction?
 1. What about fiction?
6. What might reading contribute to how you think critically in everyday life?
- a. In what ways might fiction reading improve critical thinking in everyday life?
 - b. In what ways might nonfiction reading improve critical thinking in everyday life?

Topic sheet for interview session 2

1. Can you briefly describe the text you chose to read?
2. How immersive was your reading experience?
 - a. Did you identify strongly with the author or any figures in the text?
 - b. How clearly did you picture the scenes being described?

c. Did you read the text in long reading stretches? Or did you split into small bites?

i. What do you think you gain from small bites of reading? And what about long sessions?

3. Critical thinking about your text

a. Can you describe an example of something in the text that made you think critically?

i. Did you stop reading to think critically, or did you do the thinking afterwards?

ii. Did you come to any conclusions?

b. Did the book contain different beliefs and perspectives from your own?

i. How did it feel to read those?

4. Change to your thinking

a. Can you give any examples of how this text has changed the way you think about anything?

b. Is there anything from the text that you will continue to think critically about?

5. Reading choice

a. How do you think your choice of books influences your thinking habits?

- b. What do you want from the next book you are going to read?
- c. I am looking to improve my critical thinking – can you recommend me a book to read that you think would boost my critical thinking?
 - i. Why this book?

Elucidating quantitative findings

Naming favourite authors

In the reading log study, participants were asked to name their favourite authors as a measure of reading engagement, and these were broken down into fiction and nonfiction. As participants tended to name more fiction than nonfiction authors, this raised the issue of whether being asked for ‘favourites’ was inherently associated with fiction and thus prompted more fiction answers. In the interviews, participants were also asked to name their favourite authors, and spontaneously responded with both fiction and nonfiction answers. When asked if they considered themselves to be primarily fiction or nonfiction readers, participants’ answers broadly aligned with their favourite authors. However, Participant J presents a case where favourite authors were given as literary fiction figures, and only after an explicit prompt were nonfiction authors included. When asked about this, Participant J answered:

“I guess I’m used to that referring to fiction. Although now you point it out, I’m really not sure why. I mean, I have favourite authors, there’s no reason why that would have to be fiction really.”

Therefore serving as an example of a case where ‘favourite’ can indeed prompt fiction over nonfiction. There is therefore some support for the concern that the listing of

favourite authors is not an appropriate measure of fiction engagement in comparison with nonfiction engagement.

Influence of education or life stage

Though no demographic variables were found to be relevant predictors in study two, study one did suggest educational level may have some influence. However, conclusions about this could not be drawn from the study one data. To explore the role of education, and to broaden this out to wider considerations of different life stages, events, and contexts that may shape CT across one's lifespan, participants were asked to reflect on how their CT approaches may have changed over their lifetimes.

For many participants, their CT was something they felt was innate, expressing itself in their childhood and maintaining its essential character over the course of their lives: "it could be there's a genetic element" (Participant L). However, even if their CT was stable, participants often characterised their self-awareness and reflection upon it as shifting:

"I think it's probably the approach I've always had. But I'm not sure if I would have been able to name it, or recognise it, or explore it."
(Participant B)

Conversely, many participants did experience changes in their ways of thinking critically over their lifespan. For many, specific life experiences (e.g. education for Participant M, career change for Participant D) were identified as leading to changes in CT approach. Furthermore, a life-long building of knowledge was seen as a grounding for increased CT capability: "the more sort of information I get from the world, the more connections I'm able to make" (Participant F). Additionally, participants described changes to their ways of perceiving themselves as critical thinkers. Self-

confidence was described as increasing over time, in line with the experience and knowledge-building. These ways of developing over time may alter the motivations to think critically, as Participant J described:

“I think when you’re doing critical thinking because you’ve been told to, because you’re going to be graded on it, in an exam or an essay or whatever, that’s very different motivation. It’s quite a different thinking processes I think. In some sense a think it’s critical thinking to get it over and done with. You know, you submit the essay, you leave the exam hall, and immediately it ceases to matter at all. It evaporates straight out of my head. Whereas now, if I think critically about something it’s because I myself have decided it’s worthy, it’s important for me to spend time and think critically about it. So it doesn’t have an endpoint, past which I switch off, and stop thinking or caring about it.” (Participant J)

Thus there was both some stable basis, and many changing aspects, to the ways participants perceived their CT across their lives.

Connecting this back to the previous studies, the inconclusive findings of study one with regards to educational level may be explained through the fact that participants felt it was not necessarily what they had been taught in educational settings that contributed to their CT, but the different experiences they had in education and beyond it. This suggests that level of attainment is not the most important factor. Furthermore, some of the contradictory findings (study one found educational level could reverse reading effects, and reading more with a higher educational level could in fact be negatively associated with CTD), could be explained through the difference in motivation to think critically described by Participant J. It may be that too much extrinsic motivation is an inhibitor. This could also have had some influence in study

two. Although all participants took part in the reading log study voluntarily, they were also 'told to' take the CT test as part of the study, thus there was an extrinsic motivation for thinking critically within the test. Likewise the assigned readings. However, motivation was not measured in the study, and thus this remains an avenue open for future investigation.

Influence of reading time duration

The reading log study yielded interesting and mixed results pertaining to the influence of time spent reading on CT. Because of the mixed findings in terms of entries logged and time spent reading these were combined, and time spent per entry made in the reading log was used as a variable for exploratory analysis. On the basis of this, it was found that overall spending more reading time per entry had a significant positive association with CT score change. This was also true of fiction, but not of nonfiction. Reading both fiction and nonfiction and items for a longer duration made it more likely that one's CT would be improved rather than not. Therefore I asked participants how they experienced longer blocks of reading, or reading in shorter bites, and what they felt either way of reading had to offer.

Concision was seen as a mark of quality by many participants, and this was particularly true when participants described their nonfiction reading (as is well characterised by Participant A's Yorkshire adage disparaging verbose articles: "happen he talks a lot but he says nowt"). Brevity in nonfiction was described as beneficial to understanding in offering a manageable portion of information, and also in leaving some ideas suggested for further thinking without fully expounding on them (e.g. Participant D finding great impact in a short National Geographic article). These views

are in keeping with the finding that more time spent reading each nonfiction item was not associated with CT change (though it was for improvement vs. none); in the case of nonfiction reading, shorter pieces may be impactful precisely by not overloading readers, and by leaving more room for further thought.

However, participants' views on the benefits of shorter reading were not only limited to nonfiction, but some participants also felt short fiction reading had benefits (e.g. Participant N finding short stories more impactful and intense, Participant C valuing the immersion into short fast-paced thriller chapters). In the case of fiction, the value of short readings was tied to the nature of reading material:

“I guess with the longer reading blocks, because I sort of covered more ground, and because the ground didn't seem like, the different stories didn't seem to link together that well, I guess I was left feeling a little more confused. Whereas in my shorter reading segments, because I only sort of focused on a smaller bit, it actually made sense because I wasn't then trying to keep in mind everything that had happened in the reading session. So keeping it to sort of shorter, more digestible segments, maybe it works better for this one. It's interesting. It really is just disparate events just happening.” (Participant F)

Here, Participant F reflects that the nature of the book, which comprised of small and disconnected stories, lent itself to short readings. Therefore it may be that findings pertaining to fiction reading time are highly influenced by the nature of fictional texts, with some forms of fiction having a greater or lesser impact with different timescales of reading.

Many participants articulated greater benefits gained from longer reading periods, particularly as these longer reading sessions allowed greater integration:

“I think if you read just a tiny little bite of it you might take away one idea, one thing he’s saying, but then that idea isn’t going to necessarily connect up with all the other things in the book.”

(Participant J)

It may be reading in longer time frames allows for more connection-making, and by engaging that integrative process more readers strengthen their CT. Furthermore, many participants expressed a need for more time not only to read, but to reflect. Participants described a need to pause while reading in order to have “time to mull over what I read so far” (Participant K). Increased reading time per text may be beneficial for CT by permitting slowness and reflective pauses. Furthermore, participants described a need for this reflective time in between reading different texts: “I can't just dive straight into another book, I need a bit to think about everything I've just gone through” (Participant G). This implies more time reading may not be key, but rather more time to also think over and reflect upon one a text after a reading session may be more essential, particularly since reflection has been found to be such a key feature of reading and CT relationship.

Appendix E: Study four

Reading and Critical Thinking Diary Study

Welcome to your reading and critical thinking diary.

Please think about what you read today, and pick the one item you consider to be your main reading for today. This could be a novel, a news article, a poem, a blog post – it is up to you what you consider to be your main reading item.

Similarly, please pick the one main case of critical thinking that you undertook today. You can include topics that range in importance or scale, for example deciding on an item to purchase can involve critical thinking, and so can considering a philosophical issue. Please be aware that any sensitive topics you include are entirely at your own discretion.

Your reading and critical thinking do not have to be related, they can be two separate topics.

Please do not input any personal identifying information into the fields provided; any such information will be deleted.

Completing the daily form should take 10-20 minutes.

-

Information about you

Please input your unique participant ID. This was included in the email inviting you to take part.

Please input the date for which you are logging your reading and critical thinking, in DD/MM format (for example, for 14th April you would enter 14/04)

Did you read anything you'd like to describe today?

Yes / No

Did you think critically about anything you'd like to describe today?

Yes / No

-

Your one main item read today

What was the name of the Author?

What was the title?

What kind of text was this? (for example, a novel, a magazine article, a blog post, etc.)

Please describe how immersed or transported you were by what you read. For example, did you lose track of time and your surroundings, or were you distracted by your environment? Did you imagine yourself in the scenes described, or did you read the text more from a distance? You can describe your absorption into the text however you feel is relevant.

[text entry box]

In the following box, please describe the evaluative experience of reading this item. You could describe any evaluations you made about the text as you read, or any predictions or expectations you formed as you read. For example, did you consider the view point of the author, or any figures described in the text? Did you think of questions as you read? You can describe the evaluations you made during reading however you wish.

[text entry box]

-

Your one main case of critical thinking today

What topic or question did you think critically about?

What prompted you to think about it?

What (if any) were your pre-existing opinions about this topic/question?

Did your opinion about this topic/question change?

I changed my opinion in some way / I did not not change my opinion at all

In the following box, please describe the process you went through in thinking critically. Try to describe the steps you took, the different views you considered, any evidence you sought out, any changes to your own thinking, and if/how you reached a conclusion.

[text entry box]

What do you think the strengths and weaknesses of your critical thinking process on this occasion were?

[text entry box]

Appendix F: COVID-19 pandemic impacts

This research project took place in part during the COVID-19 pandemic (World Health Organisation, 2020). This resulted in changes to the planned qualitative studies. Study three had originally been intended to utilise focus groups, rather than individual interviews, with reading groups and book clubs. Each book club or reading group would have been visited twice, before and after they read their selected text. These would have been conducted in person, rather than online. Members of the reading groups would have been invited to take part in study four, and asked to keep their reading diaries in between the two focus group sessions. However, communicating with groups as the pandemic developed and restrictions came into place, it became clear that the stresses of the situation disrupted the groups' plans and rendered online group meetings impractical. The move to individual interviews permitted far greater flexibility for participants, enabling them to fit the meetings around their other commitments in a simpler, less pressurised way. Likewise, conducting the reading diaries in a timeframe of participants' choice, and decoupling the diaries from the interviews, permitted greater adaptability to participants' needs. However, these changes did yield a different format and subsequently different data than the original research design. The individualisation of the interviews permitted thorough, deep descriptions fully expounded by each participant. However, other insights may be gained by engaging with readers as a group, where differences of opinion and experience could arise and drive further discoveries. Furthermore, including a reading diary in between two phases of interview could also yield reflections on the diary keeping process itself, which would have been valuable in addition to the content of the diaries. Finally, the changes resulting from the pandemic

necessitated the use of online channels, as in-person meeting was impossible. This means that all data gathered throughout this research project was collected online, and this means only internet users were sampled. The original plans for visiting reading groups in person were made in part to capture some participants who may not be amenable to online interactions, and would thus have been more inclusive. However, the use of online methods permits inclusivity in another sense, as anyone who identified themselves as a reader could take part regardless of their geographic location or reading group membership. It would be interesting for future research to explore group dynamics, and to capture a different sample of participants.

In addition to altering the plans for the way data was collected, the COVID-19 pandemic also influenced the content of the data. Studies two, three and four all contained subject matter related to the pandemic. In studies two and four, participants logged various kinds of reading about the pandemic and related issues. In studies three and four participants also described their thinking about these issues. These may be very different to participants reading and thinking in different times, and had the studies been conducted prior to the spread of COVID-19 the data may have been very different. It seems unlikely, for instance, that all participants would have engaged in reading and thinking on the same topic so consistently, as it is rare for one topic to impact the lives of all kinds of individuals so pervasively. The data from these studies therefore offers further uses, firstly for researchers interested in reading during the pandemic, and secondly for comparisons against similar datasets from different time periods.

Reference List

- Aberbach, D. (2018, July 26). *Environmental scientists must learn the language of literature*. Times Higher Education (THE).
<https://www.timeshighereducation.com/opinion/environmental-scientists-must-learn-language-literature>
- ACRL. (2016). *Framework for information literacy for higher education* [Text]. Association of College & Research Libraries (ACRL).
<http://www.ala.org/acrl/standards/ilframework>
- Ahern, K. J. (1999). Ten tips for reflexive bracketing. *Qualitative Health Research*, 9(3), 407–411. <https://doi.org/10.1177/104973239900900309>
- ALA. (2000). *Information literacy competency standards for higher education*.
<http://www.ala.org/Template.cfm?Section=Home&template=/ContentManagement/ContentDisplay.cfm&ContentID=33553>
- Alaszewski, A. (2006). *Using diaries for social research*. SAGE Publications, Inc.
<https://doi.org/10.4135/9780857020215>
- Albitz, R. S. (2007). The what and who of information literacy and critical thinking in higher education. *Portal: Libraries and the Academy; Baltimore*, 7(1), 97–109.
<http://dx.doi.org.libproxy.ucl.ac.uk/10.1353/pla.2007.0000>
- Aloqaili, A. S. (2012). The relationship between reading comprehension and critical thinking: A theoretical study. *Journal of King Saud University - Languages and Translation*, 24(1), 35–41. <https://doi.org/10.1016/j.jksult.2011.01.001>
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

- Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, *12*, 176–176. PubMed. <https://doi.org/10.1186/s12955-014-0176-2>
- Armstrong, M. (2015). Using literature in an EFL context to teach language and culture. *The Journal of Literature in Language Teaching*, *4*(2), 7–24.
- Arslan, S., & Demirtas, Z. (2016). Social emotional learning and critical thinking disposition. *Studia Psychologica*, *58*(4), 276–285.
<https://doi.org/10.21909/sp.2016.04.723>
- Associated Press. (2008, March 24). Clarke was cleared of child sex charges. *Irish Examiner*.
- Avin, S. (2019). Exploring artificial intelligence futures. *Journal of AI Humanities*, *2*, 171–193. <https://doi.org/10.17863/CAM.35812>
- Bailin, S. (1987). Critical and creative thinking. *Informal Logic*, *9*(1), 23–30.
<https://doi.org/10.22329/il.v9i1.2656>
- Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies*, *31*(3), 285–302.
<https://doi.org/10.1080/002202799183133>
- Bal, M. P., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PLoS ONE*, *8*(1), e55341. <https://doi.org/10.1371/journal.pone.0055341>
- Baron-Cohen, S. (2001). Theory of mind and autism: A review. *International Review of Research in Mental Retardation: Autism*, *23*, 169–184.

- Batchelor, O. (2017). Getting out the truth: The role of libraries in the fight against fake news. *Reference Services Review*, *45*(2), 143–148.
<https://doi.org/10.1108/RSR-03-2017-0006>
- Bialostosky, D. (2006). Should college English be close reading? *College English*, *69*(2), 111–116. <https://doi.org/10.2307/25472195>
- Birdi, B. (2011). Investigating fiction reader characteristics using personal construct theory. *Aslib Proceedings*, *63*(2–3), 275–294.
<http://dx.doi.org/10.1108/00012531111135709>
- Black, J. E., Capps, S. C., & Barnes, J. L. (2018). Fiction, genre exposure, and moral reality. *Psychology of Aesthetics, Creativity, and the Arts*.
<https://doi.org/10.1037/aca0000116>
- Blanca, M. J., Alarcón, R., & Arnau, J. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29.4*, 552–557.
<https://doi.org/10.7334/psicothema2016.383>
- Boell, S. K., & Cecez-Kecmanovic, D. (2010). A hermeneutic approach for conducting literature reviews and literature searches. *Australian Academic & Research Libraries*, *41*(2), 129–144.
<https://doi.org/10.1080/00048623.2010.10721450>
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616.
<https://doi.org/10.1146/annurev.psych.54.101601.145030>
- Borum Chattoo, C., & Feldman, L. (2017). Storytelling for social change: Leveraging documentary and comedy for public engagement in global poverty. *Journal of Communication*, *67*(5), 678–701. <https://doi.org/10.1111/jcom.12318>

- Braun, I. K., & Cupchik, G. C. (2001). Phenomenological and quantitative analyses of absorption in literary passages. *Empirical Studies of the Arts*, 19(1), 85–109.
<https://doi.org/10.2190/W6TJ-4KKB-856F-03VU>
- Bruner, J. S. (1986). *Actual minds, possible worlds*. Harvard University Press.
- Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, 32(1), 3–22.
<https://doi.org/10.1093/esr/jcv059>
- Bujang, M. A., Sa'at, N., Sidik, T. M. I. T. A. B., & Joo, L. C. (2018). Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *The Malaysian Journal of Medical Sciences : MJMS*, 25(4), 122–130. PubMed. <https://doi.org/10.21315/mjms2018.25.4.12>
- Bulgurcuoglu, A. N. (2016). Relationship between critical thinking levels and attitudes towards reading habits among pre-service physical education teachers. *Educational Research and Reviews*, 11(8), 708–712.
<https://doi.org/10.5897/ERR2016.2713>
- Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(3), 253–267. <https://doi.org/10.1080/10705519409539977>
- Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A Model of narrative comprehension and engagement. *Communication Theory*, 18(2), 255–280. <https://doi.org/10.1111/j.1468-2885.2008.00322.x>
- Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, 67(1), 135–157. <https://doi.org/10.1146/annurev-psych-122414-033249>

- Byrnes, J. P., & Dunbar, K. N. (2014). The nature and development of critical-analytic thinking. *Educational Psychology Review*, *26*(4), 477–493.
<https://doi.org/10.1007/s10648-014-9284-0>
- Cardoso, L. M., & Silva, N. M. (2019). Communication, informational literacy and critical thinking. *European Journal of Multidisciplinary Studies*, *3*(4), 215–220.
<https://doi.org/10.26417/ejms.v3i4.215-220>
- Castro, F. G., Kellison, J. G., Boyd, S. J., & Kopak, A. (2010). A Methodology for Conducting Integrative Mixed Methods Research and Data Analyses. *Journal of Mixed Methods Research*, *4*(4), 342–360.
<https://doi.org/10.1177/1558689810382916>
- Chan, C. (2019). Using digital storytelling to facilitate critical thinking disposition in youth civic engagement: A randomized control trial. *Children and Youth Services Review*, *107*, 104522.
<https://doi.org/10.1016/j.childyouth.2019.104522>
- Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541–557. <https://doi.org/10.1037/0022-3514.54.4.541>
- Chapman, E., & Birdi, B. (2008). Fiction for all. *Public Library Journal*, *23*(1), 8–11.
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*(1), 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Cheung, K. L., ten Klooster, P. M., Smit, C., de Vries, H., & Pieterse, M. E. (2017). The impact of non-response bias due to sampling in public health studies: A

- comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health*, 17(1), 276.
<https://doi.org/10.1186/s12889-017-4189-8>
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Christensen, R. H. B. (2015). *Analysis of ordinal data with cumulative link models: Estimation with the R-package ordinal*. 1–31.
- CILIP. (2018). *CILIP definition of information literacy 2018*.
<https://infolit.org.uk/ILdefinitionCILIP2018.pdf>
- Colucciello, M. L. (1997). Critical thinking skills and dispositions of baccalaureate nursing students—A conceptual model for evaluation. *Journal of Professional Nursing*, 13(4), 236–245. [https://doi.org/10.1016/S8755-7223\(97\)80094-4](https://doi.org/10.1016/S8755-7223(97)80094-4)
- Coolican, H. (2014). *Research methods and statistics in psychology* (6th ed.). Psychology Press.
- Corballis, M. C. (2009). Mental time travel and the shaping of language. *Experimental Brain Research*, 192(3), 553–560. <https://doi.org/10.1007/s00221-008-1491-9>
- Corrall, S. (2017). Crossing the threshold: Reflective practice in information literacy development. *Journal of Information Literacy*, 11(1), 23–53.
<http://dx.doi.org/10.11645/11.1.2241>
- Critten, J. (2015). Ideology and critical self-reflection in information literacy instruction. *Communications in Information Literacy*, 9(2), 145–156.
<https://doi.org/10.15760/comminfolit.2015.9.2.191>
- Crumbley, D. L., & Smith, L. M. (2000). Using short stories to teach critical thinking and communication skills to tax students. *Accounting Education*, 9(3), 291–296.
<https://doi.org/10.1080/09639280010017248>

- Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. Basic Books.
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, 22(Spring/Summer), 1–8.
- Cupchik, G. C. (2001). Constructivist realism: An ontology that encompasses positivist and constructivist approaches to the social sciences. *Forum: Qualitative Social Research*, 2(1). <https://doi.org/10.17169/fqs-2.1.968>
- Cupchik, G. C., & Laszlo, J. (1994). The landscape of time in literary reception: Character experience and narrative action. *Cognition and Emotion*, 8(4), 297–312. <https://doi.org/10.1080/02699939408408943>
- Cupchik, G. C., & Leonard, C. (2001). High and popular culture from the viewpoints of psychology and cultural studies. In G. Steen, D. H. Schram, & E. Ibsch (Eds.), *The psychology and sociology of literature: In honor of Elrud Ibsch* (pp. 407–419). John Benjamins.
- Cupchik, G. C., Leonard, G., Axelrad, E., & Kalin, J. D. (1998). The landscape of emotion in literary encounters. *Cognition and Emotion*, 12(6), 825–847. <https://doi.org/10.1080/026999398379457>
- Currie, G. (1985). What is fiction? *The Journal of Aesthetics and Art Criticism*, 43(4), 385–392. JSTOR. <https://doi.org/10.2307/429900>
- Dal Cin, S., Zanna, M. P., & Fong, G. T. (2004). Narrative persuasion and overcoming resistance. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion*. (pp. 175–191). Lawrence Erlbaum Associates.

- Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Research Reviews*, 26(2), 83–86. [https://doi.org/10.1016/S0165-0173\(97\)00064-7](https://doi.org/10.1016/S0165-0173(97)00064-7)
- Davies, D. (2010). Learning through fictional narratives in art and science. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention: Representation in art and science* (Vol. 262, pp. 19–31). Springer.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
- de Nooy, W. (2001). Stories and social structure: A structural perspective on literature in society. In G. Steen, D. H. Schram, & E. Ibsch (Eds.), *The psychology and sociology of literature: In honor of Elrud Ibsch* (pp. 359–377). John Benjamins.
- Deitering, A.-M., & Jameson, S. (2008). Step by step through the scholarly conversation: A collaborative library/writing faculty project to embed information literacy and promote critical thinking in first year composition at Oregon State University. *College & Undergraduate Libraries*, 15(1–2), 57–79. <https://doi.org/10.1080/10691310802176830>
- Denney, N. W. (1995). Critical Thinking During the Adult Years: Has the Developmental Function Changed over the Last Four Decades? *Experimental Aging Research*, 21(2), 191–207. <https://doi.org/10.1080/03610739508254277>
- Denzin, N. K. (1970). *The research act in sociology: A theoretical introduction to sociological methods*. Butterworths.
- Dewey, J. (1910). *How we think*. D.C. Heath & Co. <https://www.gutenberg.org/files/37423/37423-h/37423-h.htm>

- Djikic, M., & Oatley, K. (2014). The art in fiction: From indirect communication to changes of the self. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 498–505. <https://doi.org/10.1037/a0037999>
- Djikic, M., Oatley, K., Zoeterman, S., & Peterson, J. B. (2009a). Defenseless against art? Impact of reading fiction on emotion in avoidantly attached individuals. *Journal of Research in Personality*, *43*(1), 14–17. <https://doi.org/10.1016/j.jrp.2008.09.003>
- Djikic, M., Oatley, K., Zoeterman, S., & Peterson, J. B. (2009b). On being moved by art: How reading fiction transforms the self. *Creativity Research Journal*, *21*(1), 24–29. <https://doi.org/10.1080/10400410802633392>
- Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PLOS ONE*, *8*(11), e81279. <https://doi.org/10.1371/journal.pone.0081279>
- Dorocak, J. R., & Purvis, S. E. C. (2004). Using fiction in courses: Why not admit it? *Law & Literature*, *16*(1), 65–91. <https://doi.org/10.1525/lal.2004.16.1.65>
- Dumitru, D. (2019). Creating meaning: The importance of arts, humanities and culture for critical thinking development. *Studies in Higher Education*, *44*(5), 870–879. <https://doi.org/10.1080/03075079.2019.1586345>
- Dunham-LaGree, C., Chibnall, D., & Becker, S. (2017, May 19). *The power of stories: Using fiction & nonfiction to develop information literacy skills*. Iowa Library Association - Association of College and Research Libraries Conference. https://ir.uiowa.edu/ilaacrl_conference/2017/spring/2

- Elmborg, J. (2006). Critical Information Literacy: Implications for Instructional Practice. *The Journal of Academic Librarianship*, 32(2), 192–199.
<https://doi.org/10.1016/j.acalib.2005.12.004>
- Elvik, S. (2018). The orbits of Seveneves: A book review with a touch of orbital dynamics. *Principium*, 20(February), 16–18.
- Ennis, P. H. (1965). *Adult book reading in the United States* (No. 105). National Opinion Research Center, University of Chicago.
<https://files.eric.ed.gov/fulltext/ED010754.pdf>
- Ennis, R. H. (2015). Critical thinking: A streamlined conception. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 31–47). Palgrave Macmillan.
<https://doi.org/10.1057/9781137378057.0005>
- Ennis, R. H. (2016). Definition: A three-dimensional analysis with bearing on key concepts. *OSSA Conference Archive*, 115.
https://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/105?utm_source=scholar.uwindsor.ca%2Fossaarchive%2FOSSA11%2Fpapersandcommentaries%2F105&utm_medium=PDF&utm_campaign=PDFCoverPages
- Ennis, R. H., & Millman, J. (1985). *The Cornell Critical Thinking Test Level Z*. Midwest Publications.
- Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test: Test manual, criteria, scoring sheet*. Midwest Publications.
- Erkaya, O. R. (2005). Benefits of using short stories in the EFL context. *Asian EFL Journal*, 8, 1–13.

- Escalas, J. E. (2004). Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising*, *33*(2), 37–48.
<https://doi.org/10.1080/00913367.2004.10639163>
- Facione, P. A. (1990a). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. <https://eric.ed.gov/?id=ED315423>
- Facione, P. A. (1990b). *The California Critical Thinking Skills Test-College Level. Technical report #2. Factors predictive of CT skills*. California Academic Press.
- Facione, P. A., & Facione, N. C. (1992). *The California Critical Thinking Dispositions Inventory*. California Academic Press.
- Fagerheim, B., & Shrode, F. (2010). Information literacy rubrics within the disciplines. *Communications in Information Literacy*, *3*(2), 158–170.
- Farley, M. J., & Elmore, P. B. (1992). The relationship of reading comprehension to critical thinking skills, cognitive ability, and vocabulary for a sample of underachieving college freshmen. *Educational and Psychological Measurement*, *52*(4), 921–931. <https://doi.org/10.1177/0013164492052004014>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.
<https://doi.org/10.3758/BRM.41.4.1149>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications, Inc.
- Fitzgerald, M. A. (2004). Making the leap from high school to college. *Knowledge Quest: Chicago*, *32*(4), 19–24.

- Flick, U. (2018). *Doing triangulation and mixed methods*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781529716634>
- Flood, A. (2020, December 4). *UK's public libraries record another year of cuts, with yet more on the way*. The Guardian.
<http://www.theguardian.com/books/2020/dec/04/uks-public-libraries-record-another-year-of-cuts-with-yet-more-on-the-way>
- Fong, K., Mullin, J. B., & Mar, R. A. (2013). What you read matters: The role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, 7(4), 370–376. <https://doi.org/10.1037/a0034084>
- Fong, K., Mullin, J. B., & Mar, R. A. (2015). How exposure to literary genres relates to attitudes toward gender roles and sexual behavior. *Psychology of Aesthetics, Creativity, and the Arts*, 9, 274–285. <https://doi.org/10.1037/a0038864>
- Freeman, M. (2003). Rethinking the fictive, reclaiming the real: Autobiography, narrative time, and the burden of truth. In G. D. Fireman, T. E. McVay, Jr., & O. J. Flanagan (Eds.), *Narrative and consciousness: Literature, psychology, and the brain* (pp. 115–128). Oxford University Press.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Friend, C. M., & Zubek, J. P. (1958). The effects of age on critical thinking ability. *Journal of Gerontology*, 13(4), 407–413.
<https://doi.org/10.1093/geronj/13.4.407>
- Friend, S. (2008). Imagining fact and fiction. In K. Stock & K. Thomson-Jones (Eds.), *New waves in aesthetics* (pp. 150–169). Palgrave Macmillan.

- Frímansson, G. H. (2016). Reasons and normativity in critical thinking. *Studier i Pædagogisk Filosofi*, 4(1), 3. <https://doi.org/10.7146/spf.v4i1.18594>
- Gainer, J. (2012). Critical Thinking: Foundational for Digital Literacies and Democracy. *Journal of Adolescent & Adult Literacy*, 56(1), 14–17. <https://doi.org/10.1002/JAAL.00096>
- Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. New York University Press; Project MUSE. muse.jhu.edu/book/22815.
- Gaut, B. (2000). "Art" as a cluster concept. In N. Carroll (Ed.), *Theories of art today* (pp. 25–44). University of Wisconsin Press.
- Gearing, R. E. (2004). Bracketing in research: A typology. *Qualitative Health Research*, 14(10), 1429–1452. <https://doi.org/10.1177/1049732304270394>
- Genç, G. (2017). The Relationship between academic achievement, reading habits and critical thinking dispositions of Turkish tertiary level EFL learners. *Educational Research Quarterly*, 41(2), 43–74.
- Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. Yale University Press: New Haven, CT. <https://doi.org/10.2307/j.ctt1dr37cb>
- Goldstein, S. (2018). *Information Literacy (re)defined*. <https://www.cilip.org.uk/page/InformationLiteracydefinition>
- Gorichanaz, T., & Latham, K. F. (2019). Contemplative aims for information. *Information Research*, 24(3). <http://informationr.net/ir/24-3/paper836.html>

- Grafstein, A. (2017). Information Literacy and Critical Thinking. In *Pathways Into Information Literacy and Communities of Practice* (pp. 3–28). Elsevier.
<https://doi.org/10.1016/B978-0-08-100673-3.00001-0>
- Gray Hardcastle, V. (2003). The development of the self. In G. D. Fireman, T. E. McVay, & O. J. Flanagan (Eds.), *Narrative and consciousness: Literature, psychology, and the brain* (pp. 37–50). Oxford University Press.
- Green, M. C., & Brock, T. C. (2013). In the mind's eye: Transportation-imagery model of narrative persuasion. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 287–313). New York : Psychology Press.
- Green, M. C., Chatham, C., & Sestir, M. A. (2012). Emotion and transportation into fact and fiction. *Scientific Study of Literature*, 2(1), 37–59.
<https://doi.org/10.1075/ssol.2.1.03gre>
- Greenwell, B., M., McCarthy, A., J., Boehmke, B., C., & Liu, D. (2018). Residuals and diagnostics for binary and ordinal regression models: An introduction to the sure package. *The R Journal*, 10(1), 381. <https://doi.org/10.32614/RJ-2018-004>
- Guba, Egon G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, 29(2), 75–91.
- Hakemulder, J. (2000). *The moral laboratory: Experiments examining the effects of reading literature on social perception and moral self-concept*. John Benjamins Publishing.
- Hakemulder, J. (2001). How to make alle menschen bruder: Literature in a multicultural and multiform society. In G. Steen, D. H. Schram, & E. Ibsch

- (Eds.), *The psychology and sociology of literature: In honor of Elrud Ibsch* (pp. 225–242). John Benjamins.
- Hakemulder, J. (2008). Imagining what could happen: Effects of taking on the role of a character on social cognition. In S. Zyngier, M. Bortolussi, A. Chesnokova, & J. Auracher (Eds.), *Directions in empirical literary studies: In honor of Willie van Peer* (pp. 139–153). John Benjamins.
- Halcomb, E. J. (2019). Mixed methods research: The issues beyond combining methods. *Journal of Advanced Nursing*, 75(3), 499–501.
<https://doi.org/10.1111/jan.13877>
- Hall, R. (2010). Public praxis: A vision for critical information literacy in public libraries. *Public Library Quarterly*, 29(2), 162–175.
<https://doi.org/10.1080/01616841003776383>
- Halpern, D. F. (2007). *Halpern Critical Thinking Assessment using everyday situations: Background and scoring standards*. Claremont McKenna College.
- Halpern, D. F. (2008). Is intelligence critical thinking? Why we need a new definition of intelligence. In P. C. Kyllonen, R. D. Roberts, & L. Stankov (Eds.), *Extending intelligence: Enhancement and new constructs* (pp. 293–310). Lawrence Erlbaum Associates.
- Hamby, A., Brinberg, D., & Daniloski, K. (2017). Reflecting on the journey: Mechanisms in narrative persuasion. *Journal of Consumer Psychology*, 27(1), 11–22. <https://doi.org/10.1016/j.jcps.2016.06.005>
- Hampson Lundh, A., Dolatkah, M., & Limberg, L. (2018). From informational reading to information literacy. *Journal of Documentation*, 74(5), 1042–1052.
<https://doi.org/10.1108/JD-11-2017-0156>

- Harding, J. (2008). Information literacy and the public library: We've talked the talk, but are we walking the walk? *Australian Library Journal*, 57(3), 274–294.
- Harlan, M. A. (2019). Fiction as information: A look at reading as information source. *International Association of School Librarianship. Selected Papers from the ... Annual Conference*, 1–10. ProQuest Central; Social Science Premium Collection. <https://search.proquest.com/scholarly-journals/fiction-as-information-look-at-reading-source/docview/2343150585/se-2?accountid=14511>
- Hartel, J. (2019, June 17). *Turn, turn, turn*. Conceptions of Library and Information Science, 10th International Conference, Ljubljana, Slovenia.
- Haven, T. L., & Van Grootel, L. (2019). Preregistering qualitative research. *Accountability in Research*, 26(3), 229–244. <https://doi.org/10.1080/08989621.2019.1580147>
- Hawkins, K. (2012). *Thinking and reading among college undergraduates: An examination of the relationship between critical thinking skills and voluntary reading* [Doctoral dissertation, University of Tennessee]. https://trace.tennessee.edu/utk_graddiss/1302
- Hider, P., & Spiller, B. (2020). Fiction genres in bookstores and libraries: A comparison of commercial and professional classifications. *Cataloging & Classification Quarterly*, 58(8), 664–682. <https://doi.org/10.1080/01639374.2020.1850590>
- Hoaglin, D. C., & Welsh, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17–22. <https://doi.org/10.1080/00031305.1978.10479237>

- Hoeken, H., & Fikkers, K. M. (2014). Issue-relevant thinking and identification as mechanisms of narrative persuasion. *Poetics*, *44*, 84–99.
<https://doi.org/10.1016/j.poetic.2014.05.001>
- Hoeken, H., Hornikx, J., & Linders, Y. (2020). The importance and use of normative criteria to manipulate argument quality. *Journal of Advertising*, *49*(2), 195–201.
<https://doi.org/10.1080/00913367.2019.1663317>
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*(1), 88–140.
<https://doi.org/10.3102/00346543067001088>
- Hollis, H. (2018). Information literacy as a measurable construct: A need for more freely available, validated, and wide ranging instruments. *Journal of Information Literacy*, *12*(2), 76–88. <https://doi.org/10.11645/12.2.2409>
- Hollis, H. (2019). Information Literacy and Critical Thinking: Different concepts, shared conceptions. *Proceedings of the Tenth International Conference on Conceptions of Library and Information Science, Information Research 24*(4), paper 1921. <http://www.informationr.net/ir/24-4/colis/colis1921.html>
- Hollis, H., Rachitskiy, M., van der Leer, L., & Elder, L. (in press). *Validation of the International Critical Thinking Essay Test Form A (ICTET-A)*.
https://discovery.ucl.ac.uk/id/eprint/10093714/1/Hollis_ICTETPrePrint.pdf
- Holmes, F. L. (1981). The fine structure of scientific creativity. *History of Science*, *19*(1), 60–70. <https://doi.org/10.1177/007327538101900105>
- hooks, bell. (1994). *Teaching to transgress: Education as the practice of freedom*.
 Routledge.

- Hou, J., Rashid, J., & Lee, K. M. (2017). Cognitive map or medium materiality? Reading on paper and screen. *Computers in Human Behavior*, *67*, 84–94. <https://doi.org/10.1016/j.chb.2016.10.014>
- Husserl, E. (1969). *Ideas: General introduction to pure phenomenology* (W. R. B. Gibson, Trans.). Allen & Unwin.
- Hyytinen, H., Holma, K., Toom, A., Shavelson, R. J., & Lindblom-Ylänne, S. (2014). The complex relationship between students' critical thinking and epistemological beliefs in the context of problem solving. *Frontline Learning Research*, *2*(5), 1–25. <https://doi.org/10.14786/flr.v2i4.124>
- IFLA. (2012). *The Moscow declaration on media and information literacy*. <https://www.ifla.org/files/assets/information-literacy/publications/moscow-declaration-on-mil-en.pdf>
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, *18*(1), 3–20. <https://doi.org/10.1177/1525822X05282260>
- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>
- Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences*, *52*(2), 150–155. <https://doi.org/10.1016/j.paid.2011.10.005>
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107. [https://doi.org/10.1016/S0160-2896\(03\)00062-X](https://doi.org/10.1016/S0160-2896(03)00062-X)

- Johnson, W., Nijenhuis, J. te, & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81–95.
<https://doi.org/10.1016/j.intell.2007.06.001>
- Khatib, M., & Alizadeh, Iman. (2012). Critical thinking skills through literary and non-literary texts in english classes. *International Journal of Linguistics*, *4*(4), 563–580. <https://doi.org/10.5296/ijl.v4i4.2928>
- Khatib, M., & Mehrgan, K. (2012). Achieving critical thinking skills through reading short stories. *Advances in Digital Multimedia*, *1*(3), 166–172.
- Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*(6156), 377–380. <https://doi.org/10.1126/science.1239918>
- Kilic, İ., Yazici, T., & Topalak, S. İ. (2017). Critical thinking disposition of music teachers. *Eurasian Journal of Educational Research*, *17*(72), 185–202.
- King, P. M., & Kitchener, K. S. (2004). Reflective Judgment: Theory and Research on the Development of Epistemic Assumptions Through Adulthood. *Educational Psychologist*, *39*(1), 5–18.
https://doi.org/10.1207/s15326985ep3901_2
- Knofczynski, G. T., & Mundfrom, D. (2007). Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement*, *68*(3), 431–442. <https://doi.org/10.1177/0013164407310131>
- Kohnen, A. M., & Saul, E. W. (2018). Information literacy in the internet age: Making space for students' intentional and incidental knowledge. *Journal of Adolescent & Adult Literacy*, *61*(6), 671–679. <https://doi.org/10.1002/jaal.734>

- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121–1134.
- Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.
- Kuhn, D. (2005). *Education for thinking*. Harvard University Press.
- Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*(3), 309–328.
[https://doi.org/10.1016/S0885-2014\(00\)00030-7](https://doi.org/10.1016/S0885-2014(00)00030-7)
- Kuiken, D., Miall, D. S., & Sikora, S. (2004). Forms of self-implication in literary reading. *Poetics Today*, *25*(2), 171–203. <https://doi.org/10.1215/03335372-25-2-171>
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.
- Largent, E. A., & Lynch, H. F. (2017). Paying Research Participants: The Outsized Influence of 'Undue Influence'. *IRB*, *39*(4), 1–9. PubMed.
- Leichner, N., Peter, J., Mayer, A.-K., & Krampen, G. (2014). Assessing information literacy programmes using information search tasks. *Journal of Information Literacy*, *8*(1), 3–20. <https://doi.org/10.11645/8.1.1870>
- Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly*, *15*(1), 37–46.
<https://doi.org/10.2307/20009693>
- Lips, H. M. (1990). Using science fiction to teach the psychology of sex and gender. *Teaching of Psychology*, *17*(3), 197–198.
https://doi.org/10.1207/s15328023top1703_17

- Lipsey, M. W., & Hurley, S. M. (2009). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (pp. 44–76). SAGE Publications, Inc.
<https://www.doi.org/10.4135/9781483348858>
- Liu, S.-Y., Lin, C.-S., & Tsai, C.-C. (2011). College students' scientific epistemological views and thinking patterns in socioscientific decision making. *Science Education, 95*(3), 497–517. <https://doi.org/10.1002/sce.20422>
- Lloyd, A. (2012). Information literacy as a socially enacted practice: Sensitising themes for an emerging perspective of people-in-practice. *Journal of Documentation, 68*(6), 772–783. <https://doi.org/10.1108/00220411211277037>
- Lodge, D. (2002). *Consciousness and the novel: Connected essays*. Secker & Warburg.
- Lorenzo-Arribas, A. (2019). *Statistical methods for the analysis of ordinal response data* [Phd, University of Southampton]. <https://eprints.soton.ac.uk/436174/>
- Macedo-Rouet, M., Rouet, J.-F., Epstein, I., & Fayard, P. (2003). Effects of online reading on popular science comprehension. *Science Communication, 25*(2), 99–128. <https://doi.org/10.1177/1075547003259209>
- Machete, P., & Turpin, M. (2020). The use of critical thinking to identify fake news: A systematic literature review. In M. Hattingh, M. Mathee, H. Smuts, I. Pappas, Y. K. Dwivedi, & M. Mäntymäki (Eds.), *Responsible Design, Implementation and Use of Information and Communication Technology* (Vol. 12067, pp. 235–246). Springer International Publishing.
- Makani-Lim, B., Agee, A., Wu, D., & Easter, M. (2014). Research in action: Using rubrics to assess information literacy skills in business education. *Journal of Business and Educational Leadership, 5*(1), 3–17. ProQuest Central.

- Mar, R. A., Djikic, M., & Oatley, K. (2008). Effects of reading on knowledge, social abilities, and selfhood: Theory and empirical studies. In S. Zyngier, M. Bortolussi, A. Chesnokova, & J. Auracher (Eds.), *Directions in empirical literary studies: In honor of Willie van Peer* (pp. 127–137). John Benjamins.
- Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality, 40*, 694–712. <https://doi.org/10.1016/j.jrp.2005.08.002>
- Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications, 34*(4), 407–428. <https://doi.org/10.1515/COMM.2009.025>
- Mar, R. A., & Rain, M. (2015). Narrative fiction and expository nonfiction differentially predict verbal ability. *Scientific Studies of Reading, 19*(6), 1–15. <https://doi.org/10.1080/10888438.2015.1069296>
- Masterson, J., & Hayes, M. (2007). Development and data for UK versions of an author and title recognition test for adults. *Journal of Research in Reading, 30*, 212–219. <https://doi.org/10.1111/j.1467-9817.2006.00320.x>
- Maxwell, J. A., & Mittapalli, K. (2010). *SAGE handbook of mixed methods in social & behavioral research* (2nd ed.). SAGE Publications, Inc. <https://doi.org/10.4135/9781506335193>
- Mazzocco, P. J., Green, M. C., Sasota, J. A., & Jones, N. W. (2010). This story is not for everyone: Transportability and narrative persuasion. *Social Psychological and*

Personality Science, 1(4), 361–368.

<https://doi.org/10.1177/1948550610376600>

McGinnis, D. (2016). Epistemological orientations and evidence evaluation in undergraduates. *Thinking Skills and Creativity*, 19, 279–289.

<https://doi.org/10.1016/j.tsc.2016.01.002>

McLeish, T. (2019). *The poetry and music of science: Comparing creativity in science and art*. Oxford University Press.

McMullen, M. N. (1997). Affective Contrast and Assimilation in Counterfactual Thinking. *Journal of Experimental Social Psychology*, 33(1), 77–100.

<https://doi.org/10.1006/jesp.1996.1313>

McMullin, S. L. (2018). The correlation between information literacy and critical thinking of college students: An exploratory study [Ph.D., University of North Texas]. In *ProQuest Dissertations and Theses* (2120950703). ProQuest Central; ProQuest Dissertations & Theses Global; Social Science Premium Collection.

<https://search.proquest.com/dissertations-theses/correlation-between-information-literacy-critical/docview/2120950703/se-2?accountid=14511>

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *The Behavioral and Brain Sciences*, 34(2), 57–74.

<https://doi.org/10.1017/S0140525X10000968>

Messick, S. (1974). The standard problem: Meaning and values in measurement and evaluation. *ETS Research Report Series*, 1974, i–37.

<https://doi.org/10.1002/j.2333-8504.1974.tb01034.x>

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the*

- American Society for Information Science and Technology*, 58(13), 2078–2091.
<https://doi.org/10.1002/asi.20672>
- Miall, D. S. (2001). An evolutionary framework for literary reading. In G. Steen, D. H. Schram, & E. Ibsch (Eds.), *The psychology and sociology of literature: In honor of Elrud Ibsch* (pp. 407–419). John Benjamins.
- Miall, D. S. (2008). Foregrounding and feeling in response to narrative. In S. Zyngier, M. Bortolussi, A. Chesnokova, & J. Auracher (Eds.), *Directions in empirical literary studies: In honor of Willie van Peer* (pp. 89–102). John Benjamins.
- Molden, D. C. (2014). Understanding Priming Effects in Social Psychology: An Overview and Integration. *Social Cognition*, 32(Supplement), 243–249.
<https://doi.org/10.1521/soco.2014.32.sup.243>
- Moon, J. A. (1999). *Reflection in learning and professional development*. Routledge.
- Moutafi, J., Furnham, A., & Tsaoasis, I. (2006). Is the relationship between intelligence and trait Neuroticism mediated by test anxiety? *Personality and Individual Differences*, 40(3), 587–597.
<https://doi.org/10.1016/j.paid.2005.08.004>
- Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 109–120.
<https://doi.org/10.1037/aca0000089>
- Nell, V. (2013). Mythic structures in narrative: The domestication of immortality. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 17–39). New York : Psychology Press.
- Nezlek, J. (2012). *Diary methods for social and personality psychology*. SAGE Publications, Inc. <https://doi.org/10.4135/9781446287903>

- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*(2), 175–220.
<https://doi.org/10.1037/1089-2680.2.2.175>
- Oatley, K. (2011). *Such stuff as dreams: The psychology of fiction*. Wiley-Blackwell.
- Office for National Statistics. (2011). *2011 Census: Key statistics for England and Wales, March 2011*.
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuskeystatisticsforenglandandwales/2012-12-11#qualifications>
- Owusu-Ansah, E. K. (2003). Information literacy and the academic library: A critical look at a concept and the controversies surrounding it. *The Journal of Academic Librarianship, 29*(4), 219–230. [https://doi.org/10.1016/S0099-1333\(03\)00040-5](https://doi.org/10.1016/S0099-1333(03)00040-5)
- Owusu-Ansah, E. K. (2005). Debating definitions of information literacy: Enough is enough! *Library Review, 54*(6), 366–374.
<https://doi.org/10.1108/00242530510605494>
- Özyürek, A., & Trabasso, T. (1997). Evaluation during the understanding of narratives. *Discourse Processes, 23*(3), 305–335.
<https://doi.org/10.1080/01638539709544995>
- Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *Journal of Personality and Social Psychology, 111*(5), e46–e54. <https://doi.org/10.1037/pspa0000064>

- Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2017). No support for the claim that literary fiction uniquely and immediately improves theory of mind: A reply to Kidd and Castano's commentary on Panero et al. (2016). *J Pers Soc Psychol*, *112*, e5–e8.
<https://doi.org/10.1037/pspa0000079>
- Pardede, P. (2019). Using fiction to promote students' critical thinking. *JET (Journal of English Teaching)*, *5*(3), 166–178. <https://doi.org/10.33541/jet.v5i3.1309>
- Paterson, S., & Gamtso, C. (2019). Excavating visual texts: Information literacy, critical thinking, and the graphic novel in the crime fiction classroom. In O. Miller & S. Grimm (Eds.), *Comics and Critical Librarianship: Reframing the Narrative in Academic Libraries* (pp. 275–289). Library Juice Press.
https://scholars.unh.edu/unhmlibrary_pub/19
- Patnaik, E. (2013). Reflexivity: Situating the researcher in qualitative research. *Humanities and Social Science Studies*, *2*(2), 98–106.
- Paul, R. (1981). Teaching critical thinking in the 'Strong' sense: A focus on self-deception, world views, and a dialectical mode of analysis. *Informal Logic*, *4*(2).
<https://doi.org/10.22329/il.v4i2.2766>
- Paul, R. (1987). Dialogical thinking: Critical thought essential to the acquisition of rational knowledge and passions. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 127–148). W. H. Freeman and Company.
- Paul, R., & Elder, L. (2005). *A guide for educators to critical thinking competency standards*.
http://www.criticalthinking.org/files/SAM_Comp%20Stand_07opt.pdf

- Paul, R., & Elder, L. (2009). Critical Thinking, Creativity, Ethical Reasoning: A Unity of Opposites. In T. Cross & D. Ambrose (Eds.), *Morality, Ethics, and Gifted Minds* (pp. 117–131). Springer US. https://doi.org/10.1007/978-0-387-89368-6_8
- Pawley, C. (2003). Information literacy: A contradictory coupling. *Library Quarterly*, 73(4), 422–452. IIs.
- Phillips, S. (2019, March 20). *Art under threat: The growing crisis in higher education*. Royal Academy of Arts. <https://www.royalacademy.org.uk/article/art-under-threat-crisis-britain-higher-education>
- Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The Role of Demographics, Engagement and Personality. *Research in Higher Education*, 46(2), 127–152. <https://doi.org/10.1007/s11162-004-1597-2>
- Prinsloo, C. (2018a). Literature as catalyst of homogenous and heterogeneous patterns of disciplinary thinking. *Thinking Skills and Creativity*, 27, 147–159. <https://doi.org/10.1016/j.tsc.2018.02.005>
- Prinsloo, C. (2018b). Students' intrinsic perspectives on the diverse functions of short stories beyond language learning. *System*, 74, 87–97. <https://doi.org/10.1016/j.system.2018.02.019>
- Pu, D., Ni, J., Song, D., Zhang, W., Wang, Y., Wu, L., Wang, X., & Wang, Y. (2019). Influence of critical thinking disposition on the learning efficiency of problem-based learning in undergraduate medical students. *BMC Medical Education*, 19(1), 1. <https://doi.org/10.1186/s12909-018-1418-5>

- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48.
<https://doi.org/10.1006/cogp.1999.0735>
- Read, D., Loewenstein, G., & Kalyanaraman, S. (1999). Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic. *Journal of Behavioral Decision Making*, *12*(4), 257–273.
[https://doi.org/10.1002/\(SICI\)1099-0771\(199912\)12:4<257::AID-BDM327>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1099-0771(199912)12:4<257::AID-BDM327>3.0.CO;2-6)
- Reece, G. J. (2005). Critical thinking and cognitive transfer: Implications for the development of online information literacy tutorials. *Research Strategies*, *20*(4), 482–493. <https://doi.org/10.1016/j.resstr.2006.12.018>
- Reed, S. L., & Stavreva, K. (2006). Layering knowledge: Information literacy as critical thinking in the literature classroom. *Pedagogy*, *6*(3), 435–452.
<https://doi.org/10.1215/15314200-2006-004>
- Richardson, D., Griffin, N., Zaki, L., Stephenson, A., Yan, J., Curry, T., Noble, R., Hogan, J., Skipper, J., & Devlin, J. (2018). *Measuring narrative engagement: The heart tells the story*. <http://dx.doi.org/10.1101/351148>
- Richmond, S. (2004). Being in others: Empathy from a psychoanalytical perspective. *European Journal of Philosophy*, *12*(2), 244–264. <https://doi-org.libproxy.ucl.ac.uk/10.1111/j.0966-8373.2004.00209.x>
- Ripley, E. B. D. (2006). A review of paying research participants: It's time to move beyond the ethical debate. *Journal of Empirical Research on Human Research Ethics*, *1*(4), 9–19. <https://doi.org/10.1525/jer.2006.1.4.9>

- Root, R. L. (2004). Variations on a theme of putting nonfiction in its place. *Pedagogy*, 4(2), 289–295. <https://doi.org/10.1215/15314200-4-2-289>
- Rosenblatt, L. M. (1982). The literary transaction: Evocation and response. *Theory Into Practice*, 21(4), 268–277. <https://doi.org/10.1080/00405848209543018>
- Rosman, T., Mayer, A.-K., & Krampen, G. (2015). Measuring psychology students' information-seeking skills in a situational judgment test format. *European Journal of Psychological Assessment*, 32(3), 220–229. <https://doi.org/10.1027/1015-5759/a000239>
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd edition.). SAGE Publications, Inc.
- SCONUL. (2015). *Perceptions of the seven pillars of information literacy: A brief review*. <https://www.sconul.ac.uk/sites/default/files/documents/Seven%20Pillars%20Review%202015.pdf>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (Donald T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63–75. <https://doi.org/10.3233/EFI-2004-22201>
- Shihab, I. A. (2011). Reading as critical thinking. *Asian Social Science*, 7(8), 209–218. <https://doi.org/10.5539/ass.v7n8p209>

- Shukri, N. A., & Mukundan, J. (2015). A review on developing critical thinking skills through literary texts. *Advances in Language and Literary Studies*, 6(2), 4–9. <https://doi.org/10.7575/aiall.v.6n.2p.4>
- Siegel, H. (1997). *Rationality redeemed? Further dialogues on an educational ideal*. Routledge.
- Smith, C. (2021, March 11). *UKRI Official Development Assistance letter 11 March 2021*. UK Research and Innovation. <https://www.ukri.org/our-work/ukri-oda-letter-11-march-2021/>
- Smith, M. C. (2000). The real-world reading practices of adults. *Journal of Literacy Research*, 32(1), 25–52. <https://doi.org/10.1080/10862960009548063>
- Smith, M. C., & Stahl, N. A. (1993). *The use of the reading diary for obtaining reading behavior data among adults* (# ED 362844; Literacy: Celebration and Challenge). Illinois Reading Council, Silver Anniversary Monographs.
- Sosu, E. M. (2013). The development and psychometric validation of a Critical Thinking Disposition Scale. *Thinking Skills and Creativity*, 9, 107–119. <https://doi.org/10.1016/j.tsc.2012.09.002>
- Stanovich, K. E. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 24, pp. 133–180). JAI. [https://doi.org/10.1016/S0065-2407\(08\)60302-X](https://doi.org/10.1016/S0065-2407(08)60302-X)
- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, 20(1), 51–68. <https://doi.org/10.3758/BF03208254>

- Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85*, 211–229.
- Stanovich, K. E., & Cunningham, A. E. (2004). Inferences from correlational data: Exploring associations with reading experience. In N. K. Duke & M. H. Mallette (Eds.), *Literacy Research Methodologies* (pp. 28–45). The Guilford Press.
- Stanovich, K. E., Cunningham, A. E., & West, R. F. (1998). Literacy experiences and the shaping of cognition. In S. G. Paris & H. M. Wellman (Eds.), *Global prospects for education: Development, culture, and schooling*. (pp. 253–288). American Psychological Association. <https://doi.org/10.1037/10294-009>
- Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 195–237). Springer.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*, 402–433.
- Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31*, 811–826.
- Stansfield, J., & Bunce, L. (2014). The relationship between empathy and reading fiction: Separate roles for cognitive and affective components. *Journal of European Psychology Students, 5*, 9–18.
- Stephenson, N. (2016). *Sevenses*. The Borough Press.

- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and Instruction, 18*(6), 513–527.
<https://doi.org/10.1016/j.learninstruc.2007.11.001>
- Subedi, D. (2016). Explanatory sequential mixed method design as the third research community of knowledge claim. *American Journal of Educational Research, 4*(7), 570–577. <https://doi.org/10.12691/education-4-7-10>
- Suzman, J. (2021). *Work: A deep history, from the stone age to the age of robots*. Penguin Press.
- Tabačková, Z. (2015). Outside the classroom thinking inside the classroom walls: Enhancing students' critical thinking through reading literary texts. *Procedia - Social and Behavioral Sciences, 186*, 726–731.
<https://doi.org/10.1016/j.sbspro.2015.04.042>
- Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience, 11*(2), 215–224.
<https://doi.org/10.1093/scan/nsv114>
- Tang, L. (2016). Exploration on cultivation of critical thinking in college intensive reading course. *English Language Teaching, 9*(3), 18–23.
<http://dx.doi.org/10.5539/elt.v9n3p18>

- Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. *Research in Higher Education, 36*(1), 23–39. <https://doi.org/10.1007/BF02207765>
- Thayer-Bacon, B. J. (1992). A feminine reconceptualization of critical thinking theory. *Journal of Thought, 27*(1/2), 4–17.
- Thayer-Bacon, B. J. (2000). *Transforming critical thinking: Thinking constructively*. Teachers College Press.
- The Foundation for Critical Thinking. (2019). *International Critical Thinking Test*. <https://www.criticalthinking.org/pages/international-critical-thinking-test/619>
- Thelwall, M., & Kousha, K. (2017). Goodreads: A social network site for book readers. *Journal of the Association for Information Science and Technology, 68*(4), 972–983. <https://doi.org/10.1002/asi.23733>
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5). <https://doi.org/10.18637/jss.v059.i05>
- Toth, A. A., Banks, G. C., Mellor, D., O'Boyle, E. H., Dickson, A., Davis, D. J., DeHaven, A., Bochantin, J., & Borns, J. (2020). Study preregistration: An evaluation of a method for transparent reporting. *Journal of Business and Psychology*. <https://doi.org/10.1007/s10869-020-09695-3>
- Tourangeau, R. (2009). Remembering what happened: Memory errors and survey reports. In Arthur A. Stone, Christine A. Bachrach, Jared B. Jobe, Howard S. Kurtzman, & Virginia S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 29–48). Lawrence Erlbaum Associates.

- Tsui, L. (2000). Effects of campus culture on students' critical thinking. *The Review of Higher Education*, 23(4), 421–441. <https://doi.org/doi:10.1353/rhe.2000.0020>
- Tung, C.-A., & Chang, S.-Y. (2009). Developing critical thinking through literature reading. *Feng Chia Journal of Humanities and Social Sciences*, 19, 287–317.
- Ulu, H. (2019). Examining the relationships between the attitudes towards reading and reading habits, metacognitive awarenesses of reading strategies and critical thinking tendencies of pre-service teachers. *International Journal of Contemporary Educational Research*, 6(1), 169–182. <https://doi.org/10.33200/ijcer.549319>
- UNESCO. (2003). *The Prague declaration: Towards an information literate society*. www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/.../PragueDeclaration.pdf
- van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2013). The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research*, 40(5), 797–817. <https://doi.org/10.1086/673383>
- Via Rivera, M. (2019, September 3). Is reading for fun teaching information literacy? Reflections on the Cambridge EcoFiction Bookgroup. *Cambridge Information Literacy Network*. <https://camiln.org/2019/09/03/is-reading-for-fun-teaching-information-literacy-reflections-on-the-cambridge-ecofiction-bookgroup/>
- Ward, D. (2006). Revisioning information literacy for lifelong meaning. *The Journal of Academic Librarianship*, 32(4), 396–402. <https://doi.org/10.1016/j.acalib.2006.03.006>

- Watson, G., & Glaser, E. M. (1980). *Watson–Glaser Critical Thinking Appraisal: Forms A and B*. PsychCorp.
- Weiler, A. (2005). Information-seeking behavior in generation Y students: Motivation, critical thinking, and learning theory. *The Journal of Academic Librarianship*, *31*(1), 46–53. <https://doi.org/10.1016/j.acalib.2004.09.009>
- Weiner, J. M. (2011). Is there a difference between critical thinking and information literacy? *Journal of Information Literacy*, *5*(2), 81–92. <https://doi.org/10.11645/5.2.1600>
- Weinstock, M., Kienhues, D., Feucht, F. C., & Ryan, M. (2017). Informed reflexivity: Enacting epistemic virtue. *Educational Psychologist*, *52*(4), 284–298. <https://doi.org/10.1080/00461520.2017.1349662>
- Whitworth, A. (2014). *Radical information literacy: Reclaiming the political heart of the IL movement*. Chandos Publishing.
- World Health Organisation. (2020, June 29). *Listings of WHO's response to COVID-19*. <https://www.who.int/news/item/29-06-2020-covidtimeline>
- Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Ohio State University Press.