

Writing a dissertation for PG Data Science

Helen Costello
Academic Skills Centre

Sessions Outline

- Session Outline:
- Research questions, structure of the introduction and outline of the aims in the introduction
- Further help available – where to go?

Session Aims

- Identify and analyse the process of writing a dissertation and compare this to the process you will experience
- Evaluate the importance of narrowing research questions for focus and direction
- Consider the structure of an introduction as a way to set up your dissertation
- Consider the importance of outlining your aims clearly and providing signposting in introductions

This course...

- **IS** about the general principles of dissertation writing
- **IT IS NOT** about answering questions specific to your discipline or topic – speak to your supervisor for help with those issues

Pairwork – 3 minutes

- Ask your partner:

What is challenging about writing a dissertation as opposed to other forms of writing?

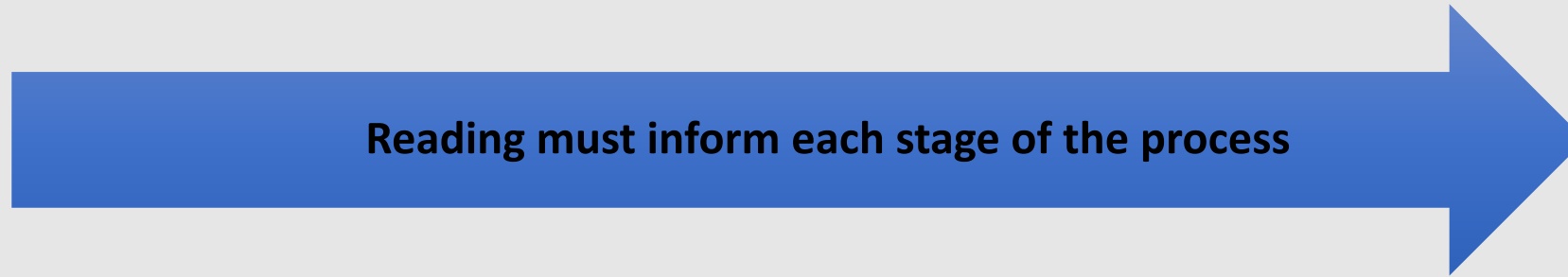
Challenges

- It's your own piece of research
- You need to establish your own research question (s)
- The extended nature of a dissertation can be challenging-
you need to remain focused and maintain continuity among
all chapters

All research projects have an aim

- Which starts with a focused research question....





Pairwork – 3 minutes

- Ask your partner:

Where are you in the reading process?

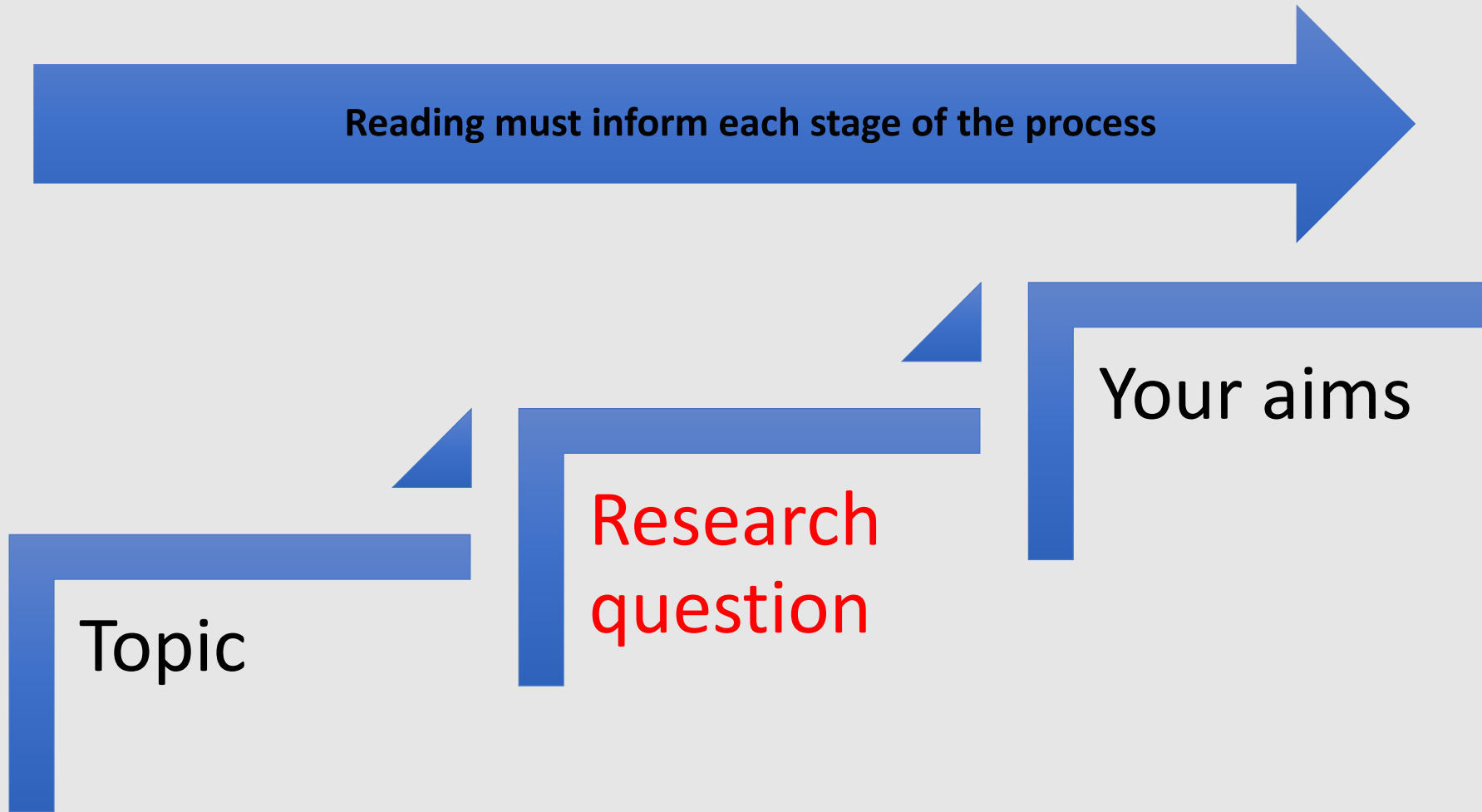
What are you doing with all the information? What is your notetaking process?

Narrowing from Topic to Research Question

- Know what you are studying and why....!
- Topic: I am studying _____
- Question: because I want to find out how/why _____
- Rationale: in order to understand how/why/whether _____

Booth (1995)

Try this with your topic...



Research questions need to be:

1. Relevant
2. Manageable in terms of research and in terms of your own academic abilities
3. Substantial
4. Clear and simple
5. Interesting

Your research question needs to be focused

Your scope can be narrowed to:

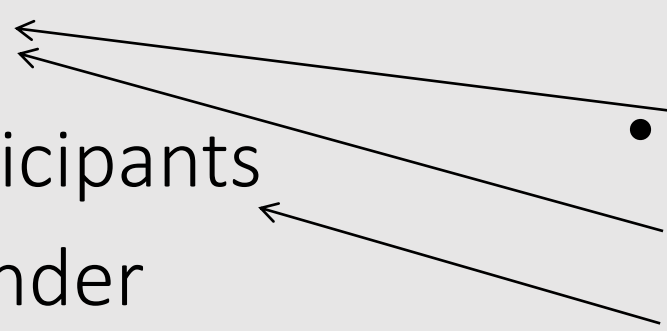
- Data Science methods
- Key concepts
- Types of participants
- Key theory under consideration
- Can you think of any other classifications that you can use to narrow your question?

Good Research Questions:

- Help us to understand what further needs to be understood/solved.
- Explore areas which are unknown and insights are useful and practical for the field

O'Leary (2018)

What's wrong with this question? What is it missing?

- Ways to narrow:
 - Data Science methods
 - Key concepts
 - Types of participants
 - Key theory under consideration
- What is the role of genes in cancer?
 - Is the **xxx gene** implicated in **cancer of the lung in non-smokers?**
 - Adapted from:
<http://www.biology.cam.ac.uk/undergrads/nst/bbs/dissertations>
- 
- The diagram consists of three arrows pointing from the specific question on the right to the general categories on the left. One arrow points from 'xxx gene' to 'Key concepts'. Another arrow points from 'cancer of the lung' to 'Types of participants'. A third arrow points from 'non-smokers' to 'Key theory under consideration'.

Reading must inform each stage of the process

Topic

Research
question

Your
aims

Dissertation Structure (IMRD)

- A data-based (empirical) dissertation *may* contain:
 - Abstract
 - Introduction
 - Literature Review
 - Methodology
 - Results
 - Discussion
 - Conclusion
 - Reference List

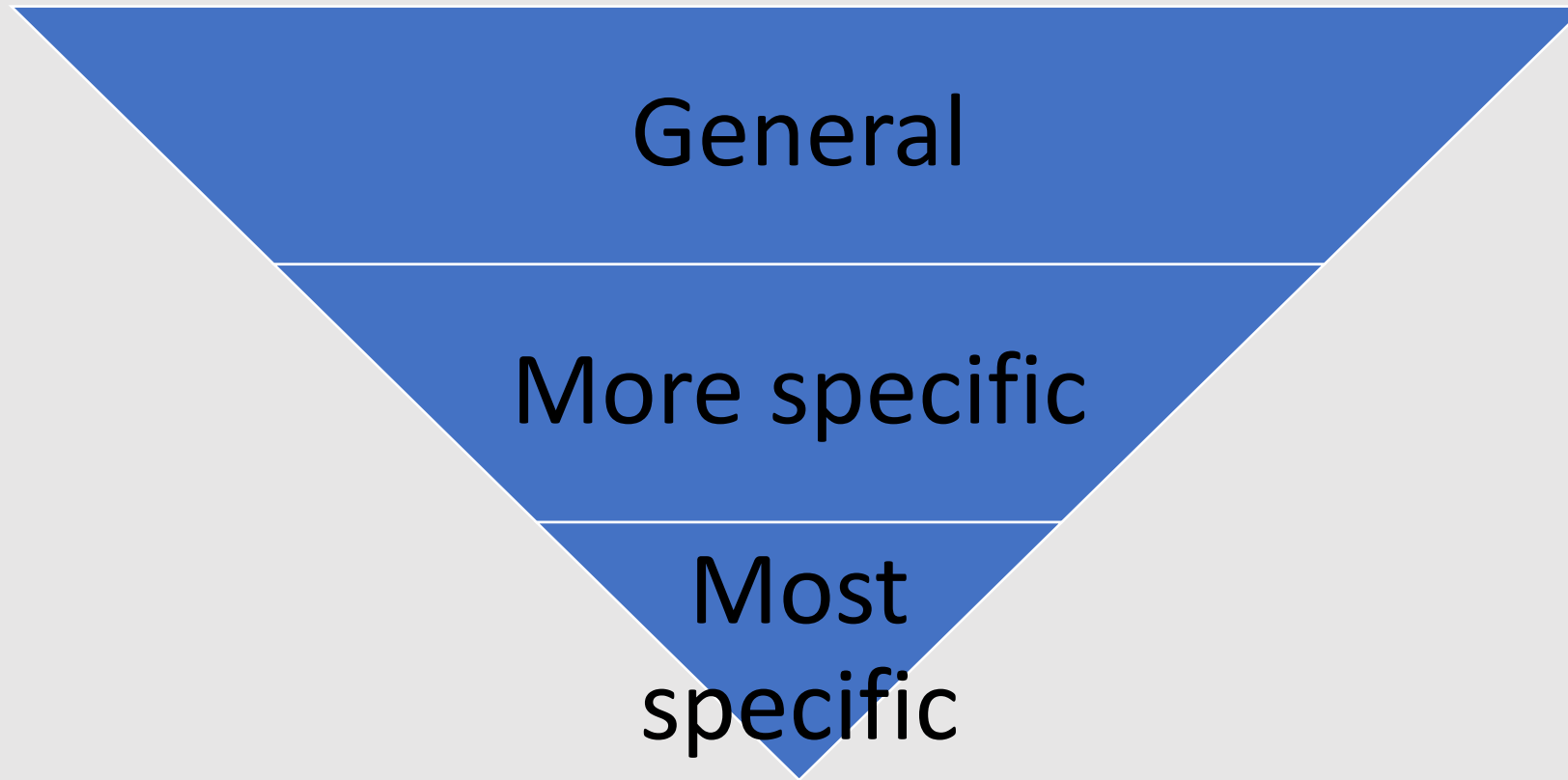
This means that:

Some dissertations will have:

- A combined Introduction and Literature Review
- A combined Discussion/conclusion chapter

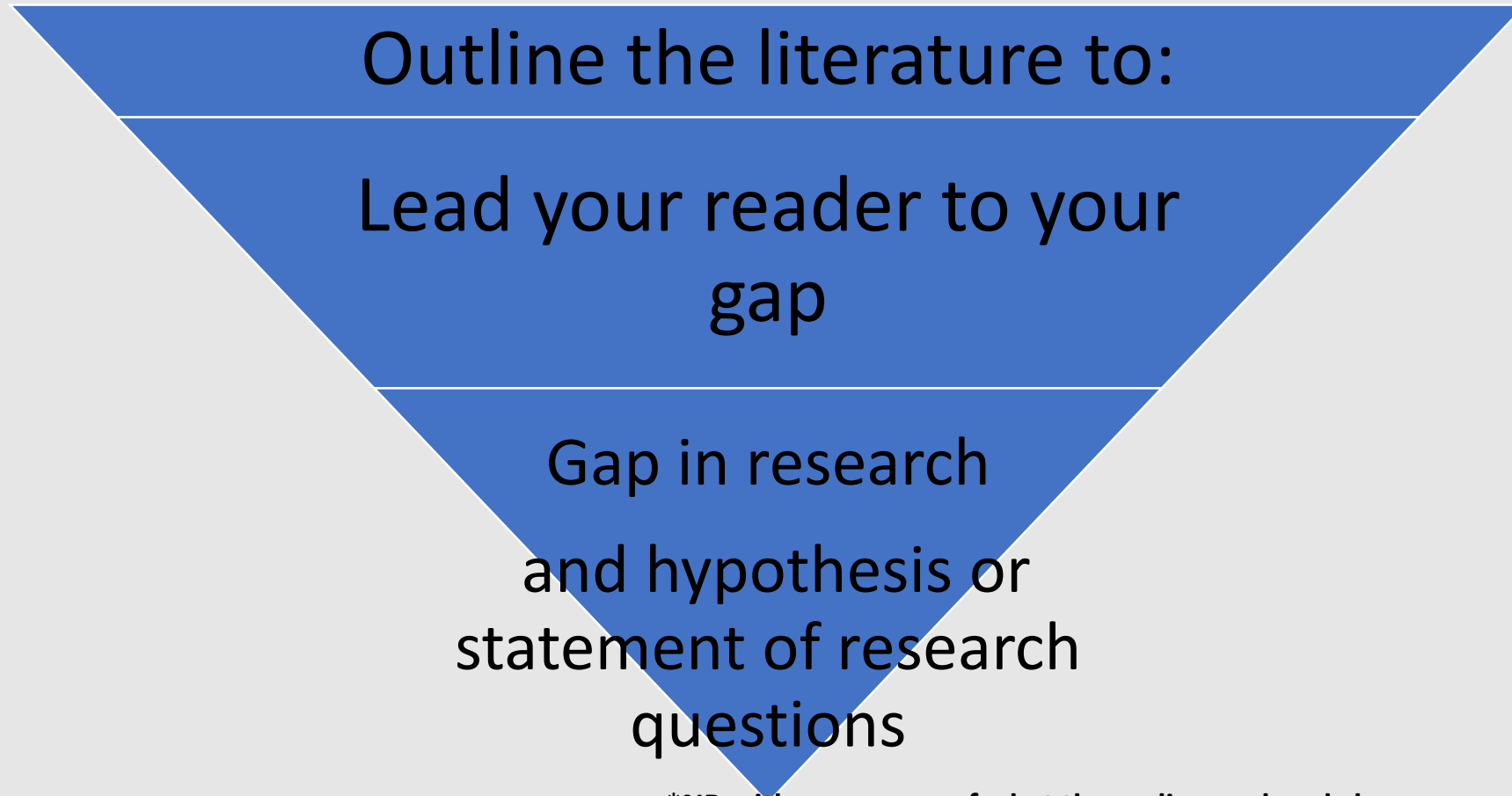
Speak to your supervisor about what is appropriate for your dissertation

Movement in introductions and literature reviews



Combined Literature Review and Introduction Chapter

Overall structure for a combined introduction and literature review



***NB with awareness of what the audience already knows**

Adapted from Weissburg and Buker (1990, p.24)

Combined Introduction and Literature Review Structure

1. General introduction to the topic (establish that this is an area of interest and/or debate; include definitions if necessary)
2. More specific statements about aspects of the problem already studied by past writers. Organised logically
3. Statement of need for more investigation (identify a gap)
4. Indication that you will occupy the gap
- 5) Signposting - A brief guide to the structure of the dissertation (“Firstly, this dissertation will discuss... Next, the issue of ... will be considered... Finally, I will turn to...”)

Stand Alone Introduction Chapter

Stand-alone Introduction structure

1. Introduce the topic/establish controversy
2. Define key concepts and limiting scope of research
3. Move toward the gap in research
4. State the gap in research
5. State the aims of your study/possibly state your hypothesis and signpost to future chapters

The research question from our sample dissertation

Can Machine Learning Accurately Predict the Success of a Start-Up?

What are the main ideas in this question?

```
graph LR; A[Can machine learning] --> B[Accurately predict]; B --> C[The success of a startup?]
```

Can
machine
learning

Accurately
predict

The success
of a
startup?

This dissertation's structure:

Table of Contents

1. INTRODUCTION	1
1.1 Motivation	1
1.2 Research Questions	3
1.3 Structure.....	4
2. LITERATURE REVIEW.....	5
2.1 Section Introduction	5
2.2 Definitions	5
2.2.1 Start-Ups	5
2.2.1 Success	6
2.3 Venture Capital	7
2.3.1 Fund Lifecycle and Venture Capital Investment Process	7
2.3.2 Benefits	8
2.4 Data Driven Investing.....	9
2.4.1 Start-Ups	10
2.4.2 Data Analysis and Machine Learning in Venture Capital	11
2.4.3 Theoretical use of Machine Learning in Venture Capital Investment Process.....	11
2.4.4 Practical use of Machine Learning in Venture Capital Investment Process.....	12
2.5 Literature Review Summary	13
3. METHODOLOGY	14
3.1 Section Introduction	14
3.2 Research Approach and Purpose.....	14
3.3 Data Set Curation	15
3.3.1 Primary Data Source - PitchBook.....	15
3.3.2 Secondary Data Source - LinkedIn	17
3.4 Analysis and Modelling	20
4. ANALYSIS, RESULTS, & DISCUSSION	22
4.1 Section Introduction	22
4.2 Exploratory Data Analysis.....	22
4.2.1 Missingness	22
4.2.2 Categorical Variable Exploration	24
4.2.3 Numerical Variable Exploration.....	27
4.3 Results and Discussion	31
4.3.1 Logistic Regression	31
4.3.2 Classification Trees.....	32
4.3.3 Model Comparison.....	39
4.3.4 Important Variables.....	40
5. CONCLUSIONS.....	47
5.1 Section Introduction	47
5.2 Reflections on Research Questions	47
5.2.1 Primary Research Question.....	47
5.2.2 Secondary Research Question.....	48
5.3 Implications and Recommendations	48
5.3.1 Implications of Findings	48
5.3.2 Recommendations to Venture Capital Firms and Venture Capitalists.....	49
5.3.3 Recommendations to Founders and Entrepreneurs	49
5.4 Limitations and Future Research Directions	50
5.4.1 Limitations.....	50
5.4.2 Future Research Directions.....	51

Activity

1. Read through part of the introduction and
 2. Identify the different sections (where relevant)
-
- Dissertation Title: *Can Machine Learning Accurately Predict the Success of a Start-Up?*

In conclusion

- Establish a clear research question and follow it through in all chapters
- Lead your reader to your aims by following a general to specific structure in your combined introduction and literature review or stand-alone introduction

Help available to you...



Academic Skills Drop-ins

**Need study or writing advice?
Come and see us!**

In-person

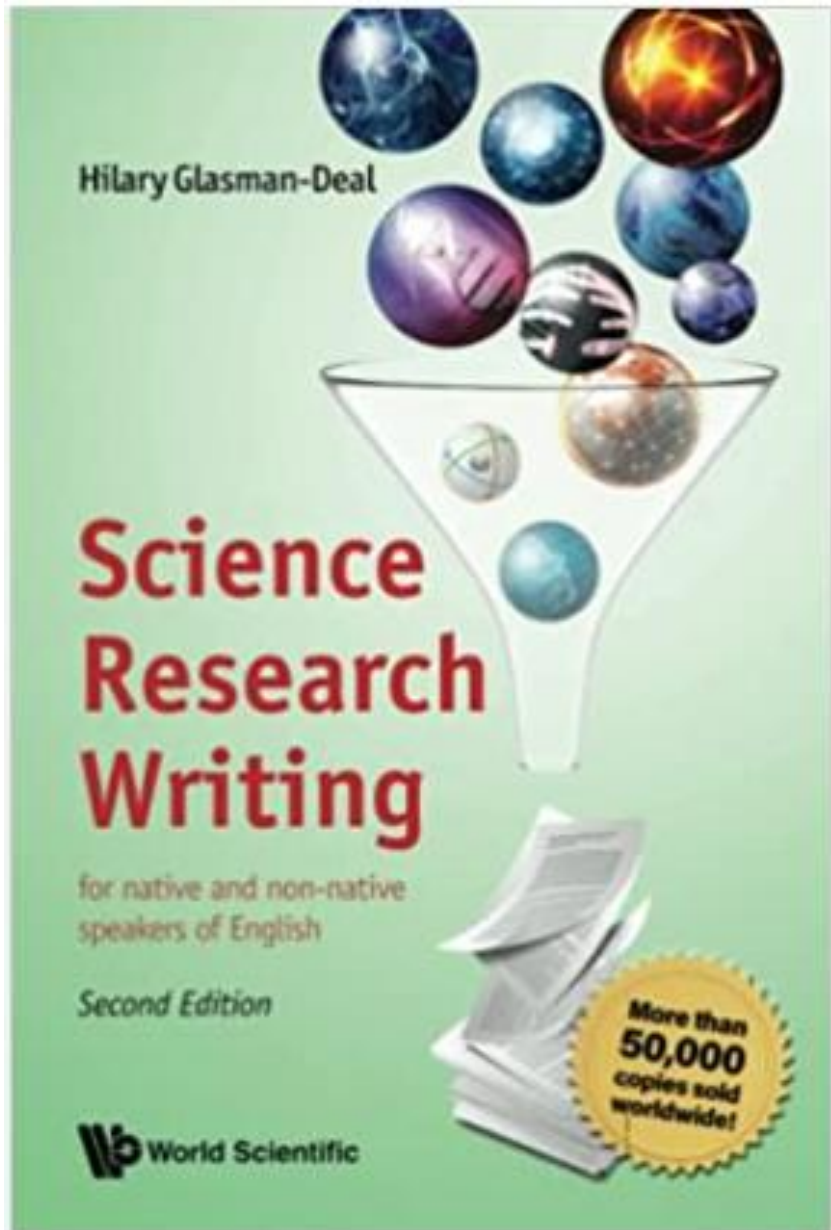
- Tuesdays 12-3pm: Top Floor TLC
- Wednesdays 12-3pm: Bill Bryson Library
- Thursdays 12-3pm: Top Floor TLC

Online

- Wednesdays 12-3pm: via Teams

**Drop-in for
quick advice**





(Glasman-Deal,
2010)

Contains lots of useful language sections to avoid repeating favourite expressions and phrases!

A Useful Resource (e-copies in the library)



References

- Booth, Wayne C. et.al. (1995) *The Craft of Research*, University of Chicago Press, Chicago.
- O'Leary, Zara (2018) *Little Quick Fix: Research Question*, Sage Publications, London.