Scientific Writing: A journey from the idea to the published paper

A guide for students, researchers and practitioners

Prof. Dr. Andreas Theissler

June 24, 2024

Outline

In this document it will be shown

- how to find a research problem and a research question,
- how to structure and write a paper,
- how to assess your paper with the proposed STRaWBERRY checklist,
- and how the publication and review process works.

A **fictitious(!) research problem** will be used in some examples and exercises: *the cake-cutting problem*.
With a touch of humour, we define the cake-cutting problem as follows:

Definition

The **cake-cutting problem** is the problem of cutting a cake C into a set of pieces $\mathcal{P} = \{P_1, ...P_N\}$ such that the trade-off between the cutter's effort $\mathcal{E}_{cutter}$ and the cake-eaters satisfaction level $\mathcal{S}_{eater}$ is optimized. The number of pieces is denoted as N and one piece is referred to as P_i. Furthermore, $\bigcup_{i=1}^{N} P_i = C$ has to hold.

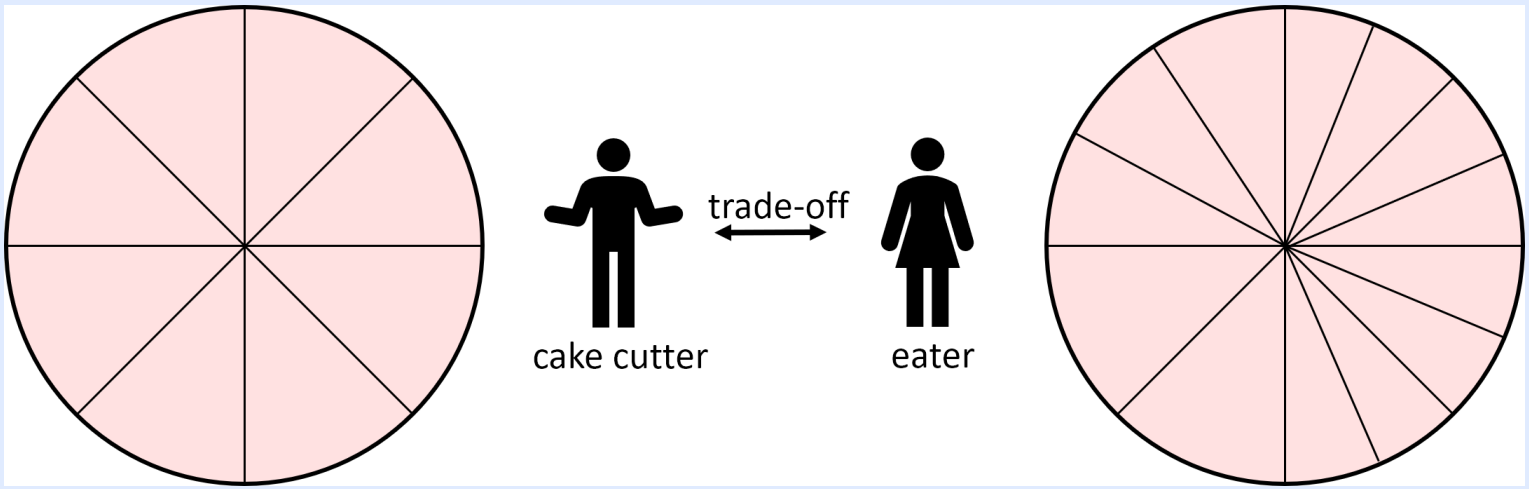


Figure: The fictitious(!) cake-cutting problem: How to cut a cake C into N pieces?

Finding a research problem

Finding a research problem

Finding a research problem and formulating a research question or hypothesis is the starting point of writing a scientific paper.

It is highly important and at the same time highly difficult.

Some options to find a research problem:

1. Becoming an expert in some field: "Knowing the Frontier of Knowledge" [Reis and Reis, 2013]
2. Discussing with experts
3. Identifying research gaps in literature:
   - often limitations and future work are discussed towards the end of a paper
   - critically evaluating papers trying to identify open issues
   - some papers (e.g. surveys) explicitly report open research problems
4. Starting from a practical problem

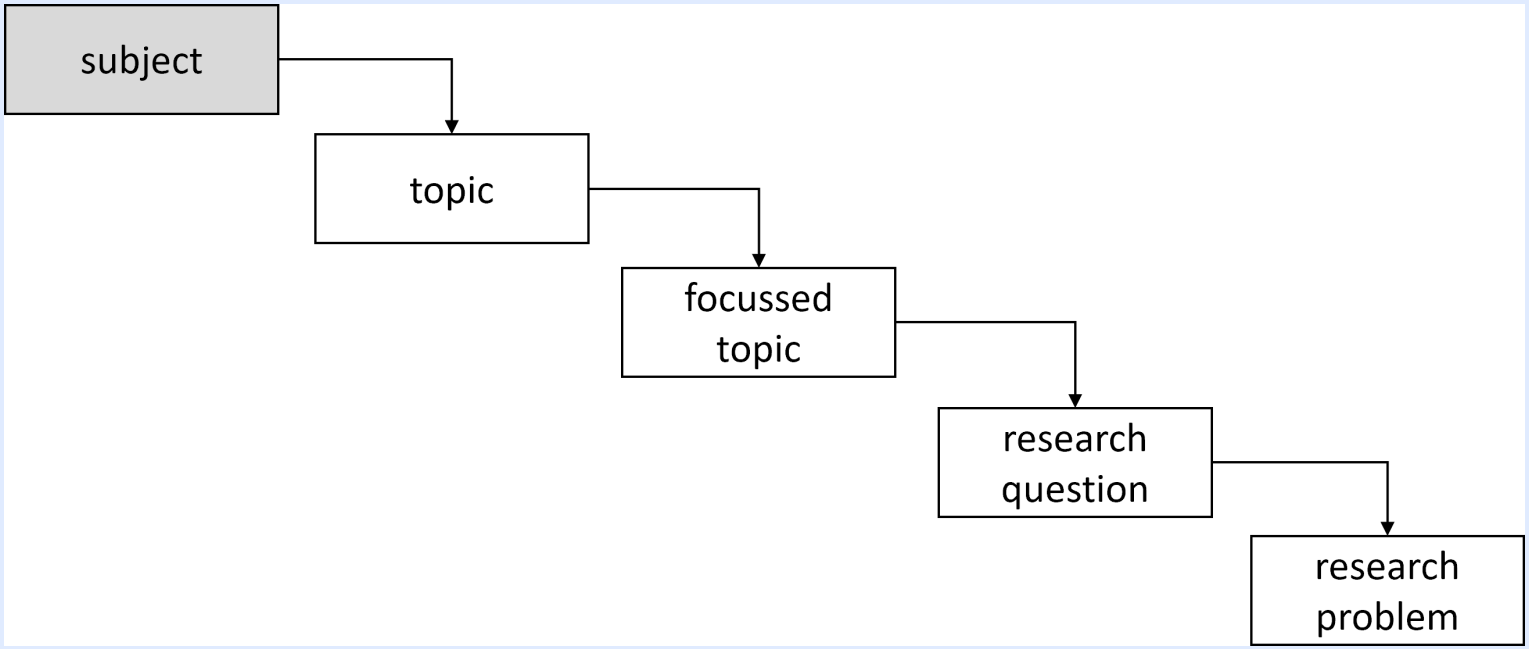


Figure: Finding a research problem by narrowing down from a subject (e.g. "machine learning"), to a topic (e.g. "interpretable machine learning"), etc.(taken from [Booth et al., 2016])

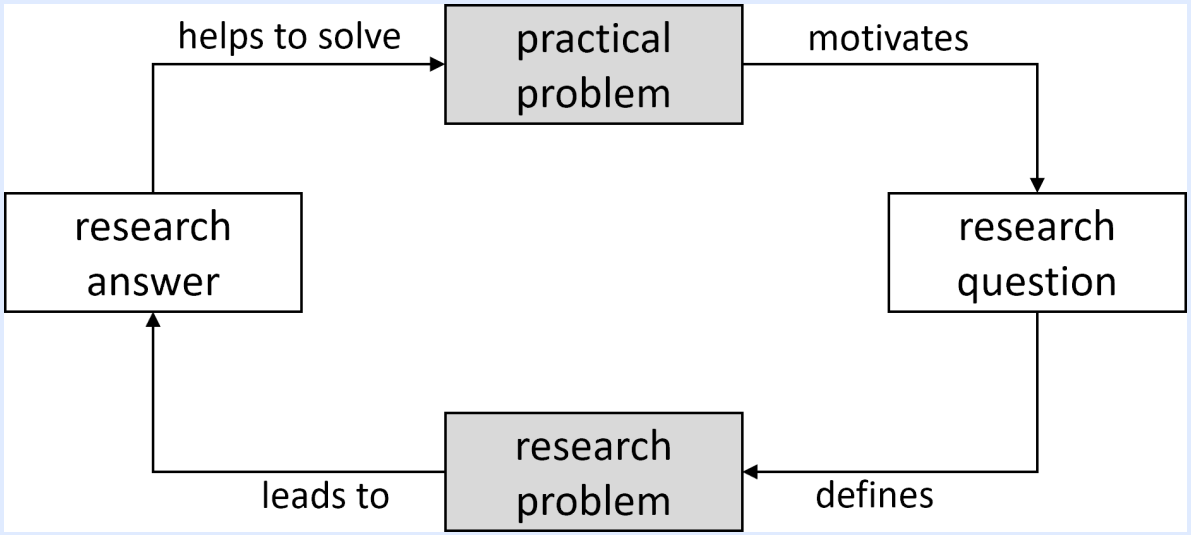


Figure: Finding a research problem, starting with a practical problem. (taken from [Booth et al., 2016])

Further options to identify a research problem:

- Identifying and connecting topics
- Taking a different perspective compared to existing work
- Using a different research method than done in existing work
- Identifying research gaps from the body of published literature

A pragmatic way to identify potential research problems:

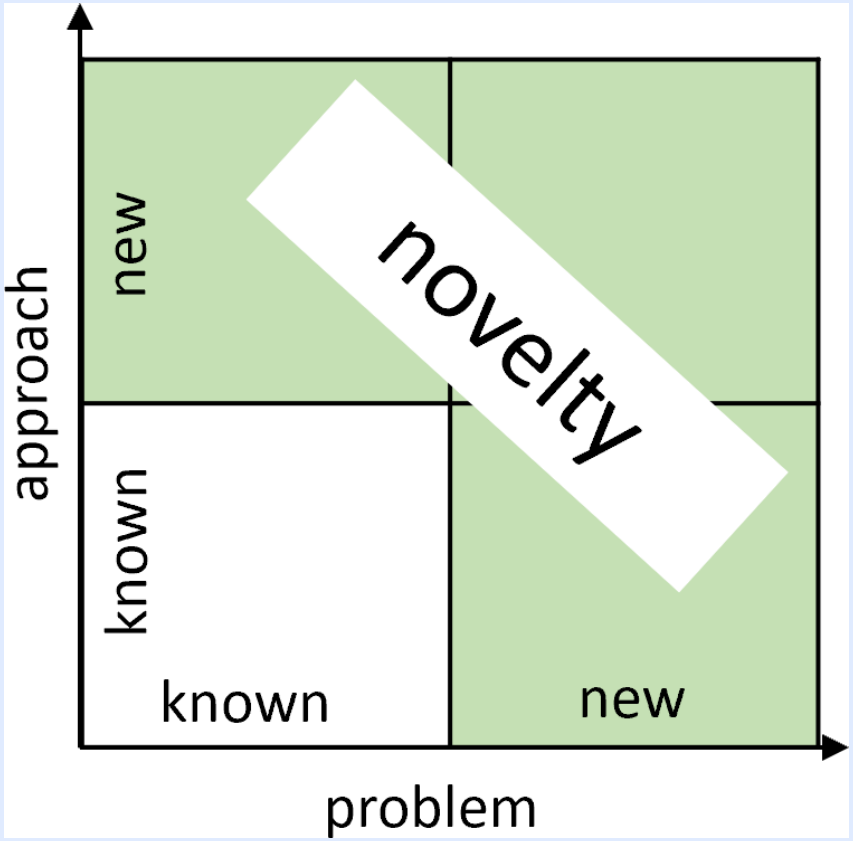


Figure: A new approach applied to an known problem, an known approach applied to a new problem, or a new approach applied to a new problem *may* present an interesting research problem.

Working with literature

Working with literature is an essential part of paper writing. Other work is referenced, summarized, analyzed and discussed.

- In some cases, the research question is answered by critically analysing published papers (in reviews).
- If prototypes are developed and experiments are conducted the state-of-the-art needs to be reported
- The own work needs to be contrasted to published work.
- The own work is connected to the research field – "embedding the own work in related literature is one of the essential parts of research writing" [Derntl, 2014]

Plagiarism

The writer of a scientific paper needs to make clear which ideas are new, which are taken from other sources and which are adapted from other sources. Failing to do so is called **plagiarism**.

There are various forms of plagiarism. Some cases:

1. Copying text without citation
2. Copying text with citation but failing to mark it as quote
3. Slightly adapting text with citation: Minor editing of grammar is not of scientific value.

EXAMPLE 3.1: Text from a published paper:

While computer vision is undeniably an important research field of machine learning, we argue that there might be a bias in explainable AI (XAI) research toward image data.

EXAMPLE 3.2: Example of properly citing other work:

(Theissler et al., 2022) point out that in XAI research, there might be a bias towards image data, while acknowledging that computer vision is an important research field. This is confirmed by (Maier et al., 2023), stating that... As opposed to that, (Johnson et al., 2023) point out ...

Writing a scientific paper

Scientific papers typically contain four main parts, commonly referred to as the **IMRaD** format:

IMRaD: **I**ntroduction, **M**ethod, **R**esults, **a**nd **D**iscussion.

(does not necessarily correspond to paper sections in a 1:1 manner)

Common components of a research paper – not necessarily using the exact section headers:

Common components of a research paper – not necessarily using the exact section headers:

- Title
- Abstract
- Introduction
- Related work
- (Background – optional, not necessarily a separate section)
- Method/Methods or Methodology
- Results
- Discussion
- Conclusion
- Bibliography

EXAMPLE 4.1: Bad titles for the fictitious cake-cutting research problem

(to be discussed...)

1. Cutting cakes
2. The cake-cutting problem
3. Optimization using Hierarchy Regularization and Negotiation Simulation for Cutting Algorithms of Cakes with Height Estimation and Content Synthesis
4. Optimizing cake-cutting under the RDN-condition using AD and LT
5. A literature review of research papers addressing the cake-cutting problem
6. A review of the cake-cutting problem: Cake-cutting algorithms and cake-cutting user studies

[INFO]

The importance of a good title:

- Based on the title, a potential reader decides whether to have a look at the paper, then checking the Abstract for a final decision whether to read the paper.
- The title is one important component for the paper to be included in search results.
- The title is a first criterion that determines if other researchers reference the paper in own papers, e.g. in literature reviews.

[HINT]

Pragmatic test, introduced in [Alley, 2018]:
Run a web search for papers in the area of your paper. Then imagine your paper title to be among them and ask yourself how your candidate title compares to the other titles in the field.
Would your title attract the desired readers?

[HINT]

To easily memorize important properties of a good title, without claiming to be an exhaustive list, the **SPICE** properties are proposed. A title should be:

- **S**pecific
- **P**recise
- **I**nteresting
- **C**ompact
- **E**asy to understand

EXAMPLE 4.2: Bad example of an Abstract for the fictitious cake-cutting research problem

(to be discussed...)

Manufacturing, selling and consumption of cakes is a major industry branch and has recently been addressed by various researchers. The Association of Cake Science (ACS) estimates the amount of consumed cake per year to be 10 million tons and the generated annual turn-over to be 50 billion US dollars [1]. As stated in [2], the major stakeholders are cake-eaters, manufacturers, cake-sellers as well as logistic companies.
A problem setting frequently encountered by cake-eaters and cake-sellers is the cake-cutting problem which was first defined by Miller et al. as the problem setting of cutting a cake C into pieces P_i such that the cutter's effort is optimized [3].

A typical abstract contains

1. A brief introduction into the paper's topic.
2. The problem statement or research question.
3. Current solutions for 2., together with their limitations, or a statement about an identified research gap.
4. The proposed approach, solution, etc., stating its novelty and methodology used.
5. Brief summary of the evaluation and results.

[HINT]

To check your abstract, **The-5-S** quality criteria are proposed:

- **S**hort
- **S**elf-contained
- **S**pecific
- **S**tatetement of novelty
- **S**ummary of most important results

EXAMPLE 4.3: Bad example of a paper introduction section for the fictitious cake-cutting research problem

(to be discussed...)
It is widely known that the segmentation of a cake C into $|P|$ pieces is highly relevant. Segmentation into $N \geq N_{\text{min}}$ pieces of static sizes and with shape S is a common approach. However, the quality of current research in the field is not sufficient.
Our approach to solve the cake-cutting problem is novel and makes significant scientific contributions which will revolutionize the research in the field.
The cake-cutting problem is the problem of cutting a cake C into a set of pieces $\mathcal{P} = \{P_1, ...P_N\}$ such that the trade-off between the cutter's effort $\mathcal{E}_{cutter}$ and the cake-eaters satisfaction level $\mathcal{S}_{eater}$ is optimized. The number of pieces is denoted as N and one piece is referred to as P_i.
For cake-cutting we have to minimize the cake cutter's effort. It has been shown that this allows to cut more cakes in less time. So the question is, if an algorithm can minimize the cutter's effort.

A typical introduction section may look as follows:

1. An introduction into the paper's topic.
2. The motivation, for example shown by a research gap or shortcoming of current research in a field.
3. Research questions, hypotheses, and/or research problem
4. The research method, i.e. showing how the research question will be answered.
5. the paper's novelty
6. the scientific contributions
7. possibly: a brief description of the paper's structure

[HINT]

The importance of a good introduction section

Readers should have a clear understanding of what will be presented in the paper, why it is worth reading, the novelty, and the scientific contributions.

[Booth et al., 2016] have an interesting view: one can work with readers saying "I do not agree" but not with readers saying "I do not care".

[HINT]

The motivation, research gap or shortcoming show **"Why"** the research was conducted, the research method shows **"How"**, and the novelty and scientific contributions show **"What's new"**. Hence, the **WHWN-questions** should be answered by an introduction section:

- **W**hy ?
- **H**ow ?
- **W**hat's **n**ew ?

EXAMPLE 4.4: Bad example 1 of a Related Work section for the fictitious cake-cutting research problem

(to be discussed...)

Artificial Intelligence (AI) was introduced in the 1950s where the idea of a Perceptron was published by Rosenberg [1]. The idea is that it mimics the human brain in order to make decisions [2]. The human brain consists of neurons and synapses, these are reflected by the architectures of artificial neural networks.
Machine learning is a subfield of AI. Machine Learning comprises tasks like classification [3], clustering [4], forecasting [5], ...
Recently the importance of AI was shown by text generation in the form of ChatGPT [6] and image generation [7]. Furthermore, it achieved great success in playing games against human experts [8].
In this paper we address the cake-cutting problem with machine learning.

EXAMPLE 4.5: Bad example 2 of a Related Work section for the fictitious cake-cutting research problem

(to be discussed...)

There are almost no scientific publications on the cake-cutting problem. The only papers that could be found are [1] and [2].

[INFO]

Discussing related work should convince the reader that

- the **research problem is relevant** – this is shown by the fact that various other papers have been published
- the own paper **makes a scientific contribution** – this is shown by contrasting the own paper to previous work (showing the research gap)

There are several common ways to include related work in a paper:

1. In an own section following the introduction
2. In the introduction within the text or as a subsection
3. Towards the end of the paper *(rarely used in papers)*.
4. Distributed throughout the paper, i.e. discussed where needed for example in the method section *(sometimes used in papers)*.

While all of the aforementioned options are valid, this document's author recommends to use either option 1 or 2.

[HINT]

To quickly validate the Related Work, the **RICK** questions are proposed:

- **R**ecent papers cited ?
- **I**mprovement of own work w.r.t. related work shown ?
- **C**ontrasted to own work ?
- **K**nowledge gap deduced ?

The background material can be introduced in different parts of the paper, not necessarily in a separate section:

- In the introduction section (recommended if it is very short)
- In a separate background section (has the benefit that (a) readers familiar with the topic can skip it and (b) it can be referenced from text later in the paper)
- In the method section, i.e. where the specific background knowledge is needed to understand the ideas. (Drawback: own ideas are mixed with previously known knowledge)

Shape of the method section depends on type of paper, research method, proposed novelty and contributions.
One high-level categorization of research methods, see e.g. [McNeill, 2006], is **quantitative and qualitative research methods**.

One possible categorization of paper types [Derntl, 2014]:

- **Empirical papers:** address some research question or hypothesis for example with experiments, expert interviews or user studies.
- **Case study papers:** apply some methods or theories to a (real-world) problem.
- **Methodology papers:** propose some new approach, etc. (quite frequently found in data science papers)
- **Theory papers:** report new theory in a purely theoretical way

In addition: **literature reviews/surveys** [Snyder, 2019, Wee and Banister, 2016, Renner et al., 2022]

[HINT]

According to the **BURNS** properties, the described research method should be:

- **B**acked by acknowledged research methods
- **U**nderstandable
- **R**eproducible
- **N**ovel
- **S**ystematic

- Results should be reported in a way that allows to evaluate their validity [Cuschieri et al., 2019, Davidson and Delbridge, 2012, Antić, 2009].
- The results must be linked to the research method and must **address the research question or hypothesis**.

[HINT]

To assess the quality of a Results section, the **ELVIRA** properties are proposed. The results should (be):

- **E**xplained
- **L**abelled
- **V**isually supported
- **I**nspire discussion
- **R**eproducible
- **A**ddress the research question or hypothesis

- The discussion section is the place to **analyse, evaluate, discuss and possibly interpret the results**
- From the plain results, **findings should be deduced**.
- Some papers discuss results in the results section. Having the discussion in a separate section can avoid mixing plain reporting of results with interpretation and discussions.

[HINT]

The Discussion section can be checked with **REFLOW**:

- **R**esults discussed
- **E**valuation conducted
- **F**indings stated
- **L**imitations stated[a]
- **O**pen issues identified[a]
- **W**ork critically evaluated

---

[a] Sometimes limitations and open issues (future work, identified research gaps, future research directions, etc.) are presented in the Discussion section, sometimes they are discussed in the Discussion section and summarized in the Conclusion.

EXAMPLE 4.6: Bad example of a Conclusion for the fictitious cake-cutting problem

(to be discussed...)

In this paper the cake-cutting problem was addressed. This problem was first addressed by Miller et al. and describes the segmentation of a cake C into $|P|$ pieces. A machine learning approach was proposed. Machine learning is the research field of computer programs learning from data in order to make autonomous decisions. The research question was stated and addressed systematically. An approach was proposed and evaluated successfully.

- The Conclusion section is typically the last part that is read. Hence, it significantly contributes to the impression a paper.
- It should be kept short, however, there is no fixed limit on the length. It is often a paper's shortest section, however, longer than the Abstract.

- A brief repetition of the paper's **research question, hypothesis or goal**.
- A **brief(!) summary** of what was done in the paper.
- The **findings**, i.e. overriding results, insights, conclusions.
- The **implications**, i.e. how does it influence the research field, what are possible applications, etc.
- The **limitations**[1],
- Potential **future work**[1]

[1] Sometimes limitations and future work are discussed in the Discussion prior to the conclusion

[HINT]

To evaluate a Conclusion, the **RIB**-check is proposed. In the Conclusion there should be:

- **R**esearch summary
- **I**nsights summary
- **B**eginning of new research

[HINT]

While the Conclusion is the end of your paper, it is best to view it as the beginning of new research (for yourself or other researchers). You should:

- broaden the paper's scope suggesting the applicability to related problem settings or to other application domains
- suggest future research directions

This will help your paper to make an impact.

The Bibliography

- contains the **complete list of work** referenced throughout the paper.
- should be proof-read like the other sections
- should contain work from acknowledged sources to ensure they are of appropriate quality

[HINT]

According to the proposed **YEAR**-criteria, the Bibliography should contain:

- **Y**our full list of cited work
- **E**mbedding of own work in related work
- **A**cknowledged sources
- **R**elevant papers

- In the previous sections, criteria for a quality check of the paper's components were proposed.
- These are now unified into the **STRaWBERRY-checklist**.
- The **STRaWBERRY-checklist** does not represent the full list of content in a section, but rather focuses on essential properties allowing for a high-level quality check.
- Recommended to be used during writing and prior to submission.

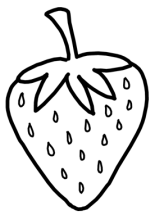
Title: **SPICE** **S**	Abstract: **The-5-S** **T**	Related Work: **RICK** **R**	**a**	Introduction: **WHWN** **W**
☐ **S**hort ☐ **P**recise ☐ **I**nteresting ☐ **C**ompact ☐ **E**asy to understand	☐ **S**hort ☐ **S**elf-contained ☐ **S**pecific ☐ **S**tatement of novelty ☐ **S**ummary of most important results	☐ **R**ecent papers cited ☐ **I**mprovement of own work w.r.t. related work shown ☐ **C**ontrasted to own work ☐ **K**nowledge gap deduced		☐ **W**hy? ☐ **H**ow? ☐ **W**hat's **n**ew?
Methods: **BURNS** **B**	Results: **ELVIRA** **E**	Discussion: **REFLOW** **R**	Conclusion: **RIB** **R**	Bibliography: **YEAR** **Y**
☐ **B**acked by acknowledged research methods ☐ **U**nderstandable ☐ **R**eproducible ☐ **N**ovel ☐ **S**ystematic	☐ **E**xplained ☐ **L**abelled ☐ **V**isually supported ☐ **I**nspire discussion ☐ **R**eproducible ☐ **A**ddress the research question or hypothesis	☐ **R**esults discussed ☐ **E**valuation conducted ☐ **F**indings stated ☐ **L**imitations stated ☐ **O**pen issues identified ☐ **W**ork critically evaluated	☐ **R**esearch summary ☐ **I**nsights summary ☐ **B**eginning of new research	☐ **Y**our full list of cited papers ☐ **E**mbedding own work in related work ☐ **A**cknowledged sources ☐ **R**elevant papers

Figure: The **STRaWBERRY**-checklist: an easy-to-remember, component-wise checklist to critically evaluate a paper draft.

Thank you!

Good luck with your paper.

References I

Alley, M. (2018).
The craft of scientific writing (4th ed.).
Springer.

Antić, Z. (2009).
Some implications for teaching scientific medical writing.
Acta Facultatis Medicae Naissensis, 26(1):55–60.

Booth, W. C., Colomb, G. G., Williams, J. M., Bizup, J., and Fitzgerald, W. T. (2016).
The Craft of Research, (4th ed.).
University of Chicago Press.

References II


Cuschieri, S., Grech, V., and Savona-Ventura, C. (2019).
WASP (write a scientific paper): Structuring a scientific paper.
Early Human Development, 128:114–117.


Davidson, A. and Delbridge, E. (2012).
How to write a research paper.
Paediatrics and Child Health, 22(2):61–65.


Derntl, M. (2014).
Basics of research paper writing and publishing.
International Journal of Technology Enhanced Learning, 6(2):105–123.

References III


McNeill, P. (2006).
Research methods.
Routledge.


Reis, S. R. N. and Reis, A. I. (2013).
How to write your first scientific paper.
In *2013 3rd Interdisciplinary Engineering Design Education Conference*, pages 181–186. IEEE.


Renner, A., Müller, J., and Theissler, A. (2022).
State-of-the-art on writing a literature review: An overview of types and components.
In *2022 IEEE Global Engineering Education Conference (EDUCON)*, pages 1895–1902. IEEE.

References IV


Snyder, H. (2019).
Literature review as a research methodology: An overview and guidelines.
Journal of business research, 104:333–339.


Wee, B. V. and Banister, D. (2016).
How to write a literature review paper?
Transport reviews, 36(2):278–288.