

SP227I INTRODUCTION TO THE SCIENTIFIC LITERATURE

Lecture 4 The art of reading a scientific paper

Dara Richard^a, Linda Sellou^b

^a Center for English Language and Communication (CELC)

^b Special Programme in Science (SPS), Faculty of Science

1. Effective Reading
2. Critical Reading

How to read scientific papers effectively?

When should you read a scientific paper?

You should read a scientific article when:

- You need to stay updated in your field.
- You are tackling a specific research problem or you need inspiration for a new one.
- To understand certain fundamental aspects of a new topic.
- You are conducting literature research for a study or project.
- You are preparing for a presentation or writing a report.

How to pick the right paper?

Section	Description
Title	Topic of study
Abstract	Summary with objectives, methods, key findings, and conclusions.
Introduction	Background, research questions, and objectives.
Methods	Explains data collection and analysis techniques.
Results	Presents key findings using tables and figures.
Discussion	Interprets results, compares with other studies, and discusses implications.
Conclusion	Summarizes main findings and give future directions.
References	Lists sources cited in the paper.

Helps you decide if the paper is relevant to your topic.

How to read a scientific paper?

nature
ecology & evolution

ARTICLES

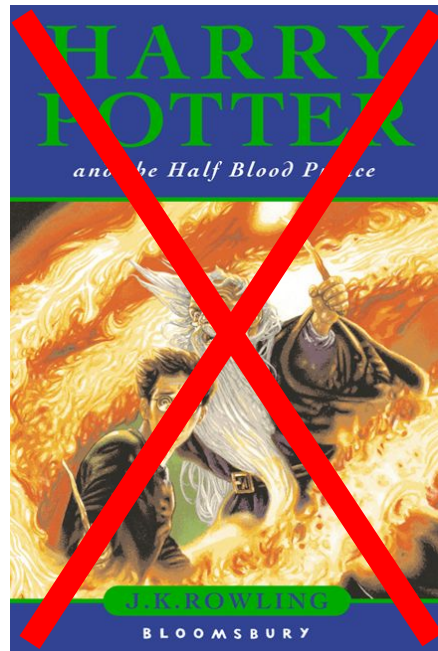
<https://doi.org/10.1038/s41559-018-0475-9>

Origin of spiders and their spinning organs illuminated by mid-Cretaceous amber fossils

Diyang Huang^{1*}, Gustavo Hormiga², Chenyang Cai³, Yitong Su¹, Zongjun Yin¹, Fangyuan Xia⁴ and Gonzalo Giribet^{1,5*}

Understanding the genealogical relationships among the arachnid orders is an onerous task, but fossils have aided in anchoring some branches of the arachnid tree of life. The discovery of Palaeozoic fossils with characters found in both extant spiders and other arachnids provided evidence for a series of extinctions of what was thought to be a grade, Uraraneida, that led to modern spiders. Here, we report two extraordinarily well-preserved Mesozoic members of Uraraneida with a segmented abdomen, multi-articulate spinnerets with well-defined spigots, modified male palps, spider-like chelicerae and a uropygid-like telson. The new fossils, belonging to the species *Chimerarachne yingi*, were analysed phylogenetically in a large data matrix of extant and extinct arachnids under a diverse regime of analytical conditions, most of which resulted in placing Uraraneida as the sister clade of Araneae (spiders). The phylogenetic placement of this arachnid fossil extends the presence of spinnerets and modified palps more basally in the arachnid tree than was previously thought. Ecologically, the new fossil extends the record of Uraraneida 170 million years towards the present, thus showing that uraraneids and spiders co-existed for a large fraction of their evolutionary history.

Huang, Diyang, et al. "Origin of spiders and their spinning organs illuminated by mid-Cretaceous amber fossils." *Nature ecology & evolution* 2.4 (2018): 623-627.



How to read a scientific article?...so that you don't struggle too much

By developing a **strategic reading approach**, you can save time and extract key insights more efficiently.



What strategies to choose?

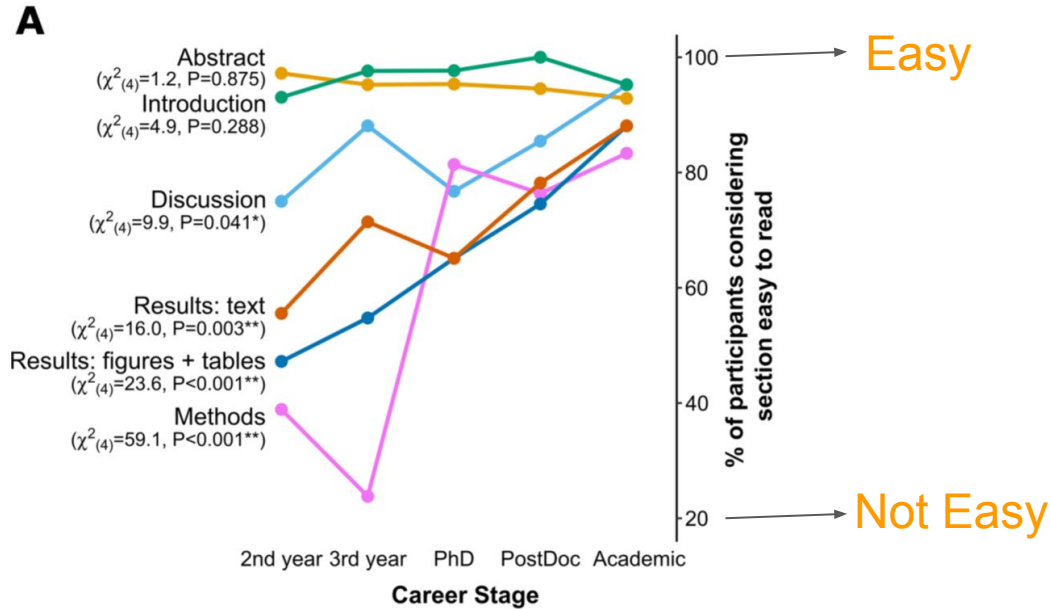


Fig 2. Different sections of scientific papers are considered easy to read and important at different stages of academic careers. A: The proportion of participants considering a section easy to read (presented as 'Somewhat easy', 'easy', 'very easy' combined) as a function of career stage. Results of Chi-square tests are indicated on the left hand side. B: The mean importance rank of sections as a function of career stage. Error bars are omitted from individual points for clarity, with the sole error bar in grey representing the largest 95% confidence interval for any of the data points. Asterisks above data points indicate significant differences in response compared with the previous career stage as determined by Mann-Whitney post-hoc tests.

What strategies to choose?

The method of reading will depend on your **goal** and the **stage** of your research/background.

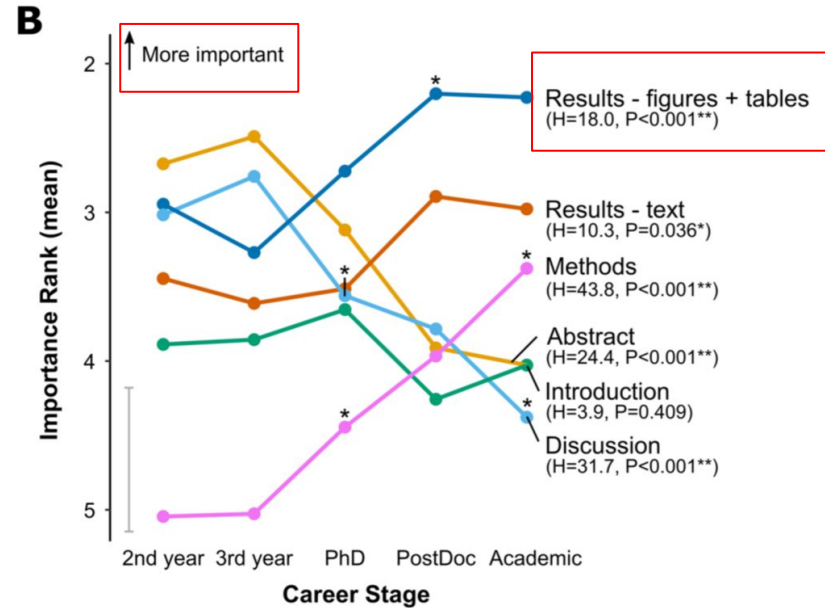


Fig 2. Different sections of scientific papers are considered easy to read and important at different stages of academic careers. A: The proportion of participants considering a section easy to read (presented as 'Somewhat easy', 'easy', 'very easy' combined) as a function of career stage. Results of Chi-square tests are indicated on the left hand side. B: The mean importance rank of sections as a function of career stage. Error bars are omitted from individual points for clarity, with the sole error bar in grey representing the largest 95% confidence interval for any of the data points. Asterisks above data points indicate significant differences in response compared with the previous career stage as determined by Mann-Whitney post-hoc tests.

A proposed strategy

You will typically need to engage with an article **several times**, but in **different ways**.

Start by reading (again) the **title & abstract**: What are the main goals and results of the paper?

Pass 1: **Skim** the article for an overview (10-15 mins)

1. Scan section headings to get familiar with the **structure** of the paper and where information are located.
2. Check the **figures & tables**: Note briefly the information presented (key words).
3. Check the introduction for **references & background information you don't know**.

Pass 2: Detailed reading for **comprehension**

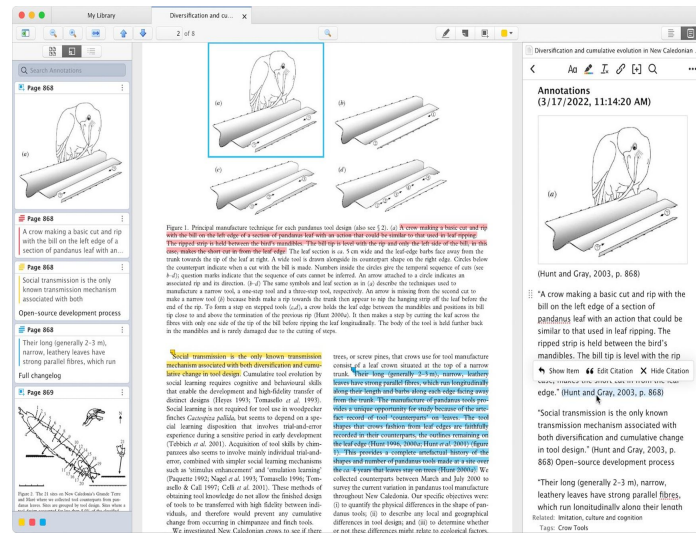
1. Read the introduction to understand the **objectives**.
2. Examine the methods to understand what **experiments** were done.
3. Analyse each **figure & table and try to find out what they are telling you**.
4. Read the discussion for **interpretation of findings**.

Pass 3: Deep dive for **critical analysis** (several hours)

1. Critically evaluate the study's **assumptions and hypotheses**.
2. Compare findings with **other literature**.
3. **Assess methodology** robustness and limitations.

Practical tips for efficient reading

- Do not read linearly: use a strategic approach.
- Take active notes: Highlight terms/jargons you need to check, references you need to read, questions you have, important key words or results. Use your own words.
- Discuss with your peers/friends/mentors.
- Revisit papers : Some articles require multiple readings for full comprehension.
- Use citation management tools: e.g. **Zotero**



Guiding questions for critical analysis

Try to answer while reading/annotating the paper.

Abstract

- What did the researchers do?
- Why is this significant?
- What were the techniques used?
- What were the results?
- Was the research question answered?

Materials & Methods

- How many data points/samples did the researchers collect/work on?
- Were there controls, and were these controls suitable?
- What techniques were used? What other tools/tests did they use? Do they make sense?
- Were the methods detailed enough for others to do the same?
- Draw a simple flowchart for what they did.

Introduction

- What is currently known in the field?
- Why did the researchers do what they did?
- Were the researchers trying to extend or challenge existing knowledge, or describe a new technique?
- What is the research question/hypothesis?

Results & Discussion

- What is each of the figures saying?
- Is the data reliable? Are there huge error bars?
- How do the controls compare to the test samples?
- Does the data attempt to answer the research question?
- Does the data *support* their conclusions?
- Did their methods have any limitations?

Critical Reading

What will you learn...

- The difference between critique and criticism
- Steps for critical reading

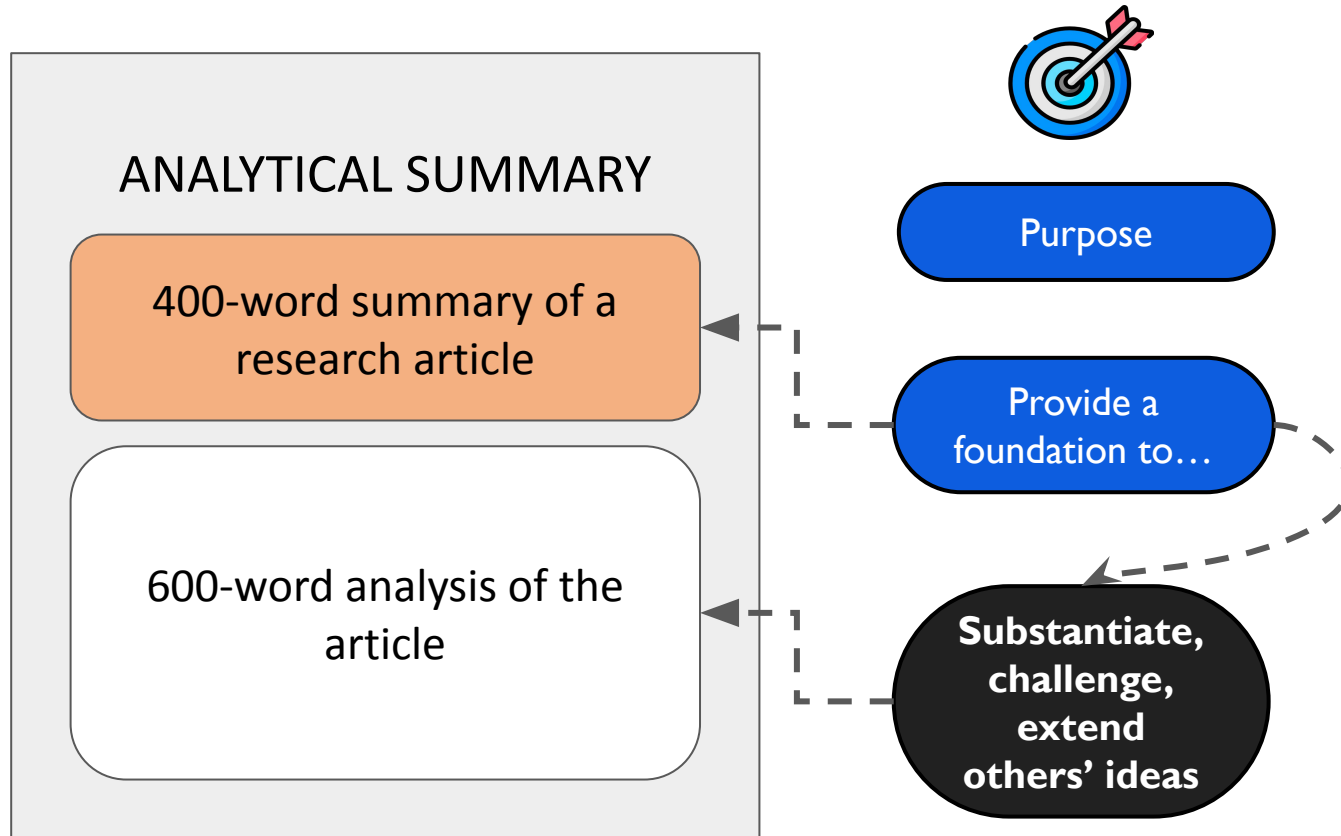
Why should you watch...

- Critique is essential for Assignments 1, 2 & 3
- Critical reading is essential in academia

Think about this...

What questions would I ask?

RECAP: Structure of Analytical Summary assignments



The difference between critique and criticism

Criticism



- Fault finding
- Focused on creator

Critique



- Analytical
- Focused on work

Balanced

Evidence-based

Constructive

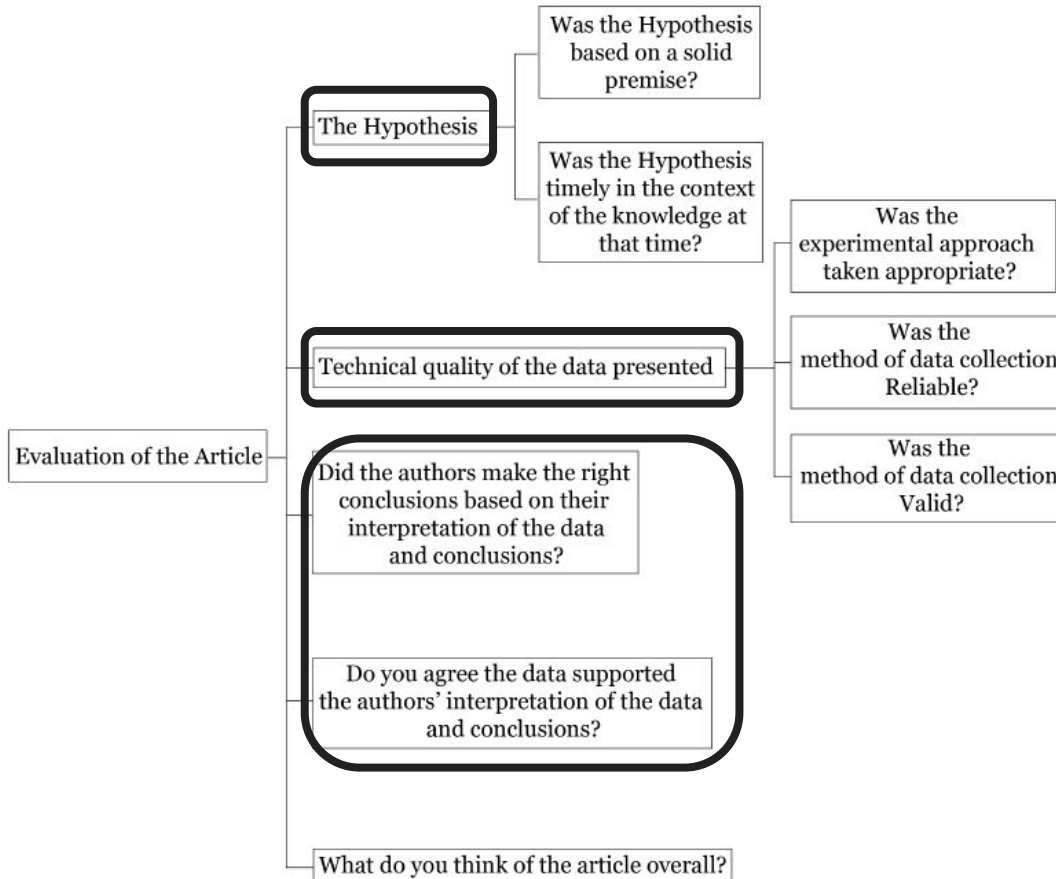
1st step: Understanding the article's main points

Table 1.1 Several Points Students could Bear in Mind as They Read Research Articles for the First Time

- (I) What was the current state of knowledge in the field that the research work was described in the article? This information will be found in the **Introduction** section of the article.
- (II) What was the idea tested in the article? This information will be found **at the end of the Introduction section** of the article.
- (III) What were the key techniques used in the experiments described? This information will be found in the **Materials and Methods** section of the article.
- (IV) What were the data obtained in the study? This information will be found in the **Results** section of the article.
- (V) What were the conclusions drawn by the authors from their data? This information will be found in the **Discussion** section of the article.

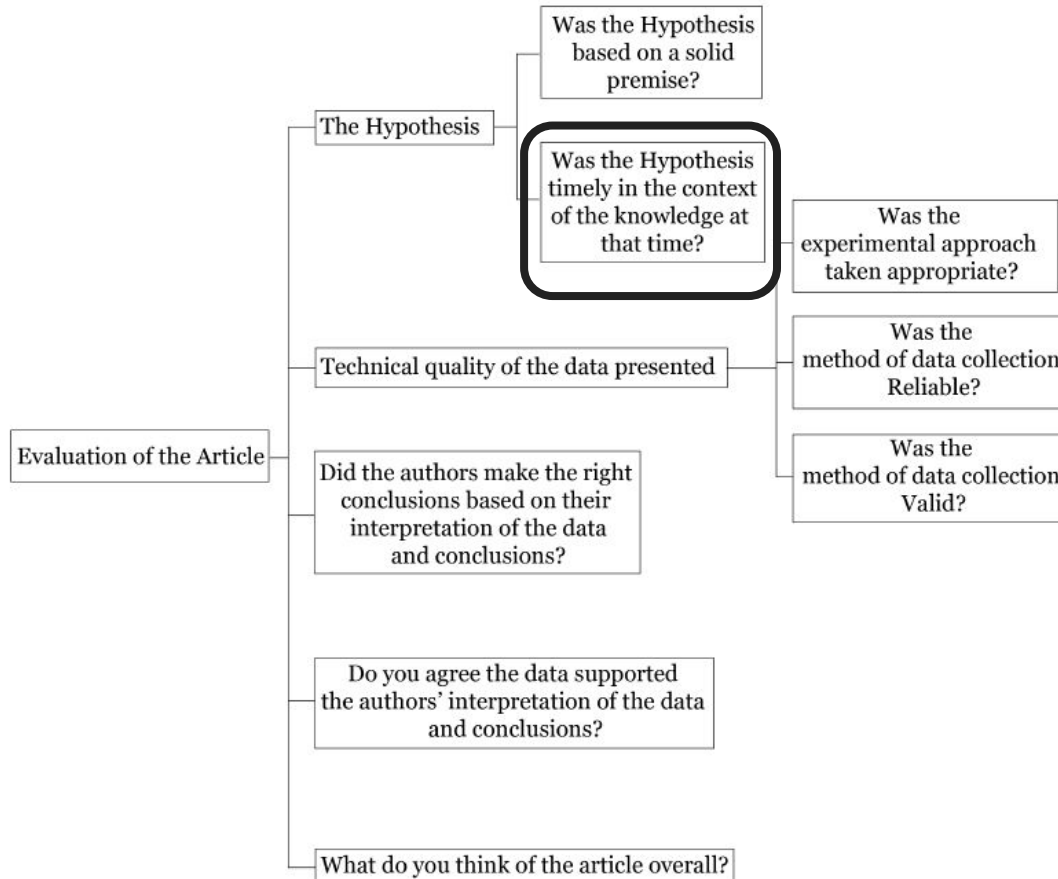


2nd step: Evaluating the article



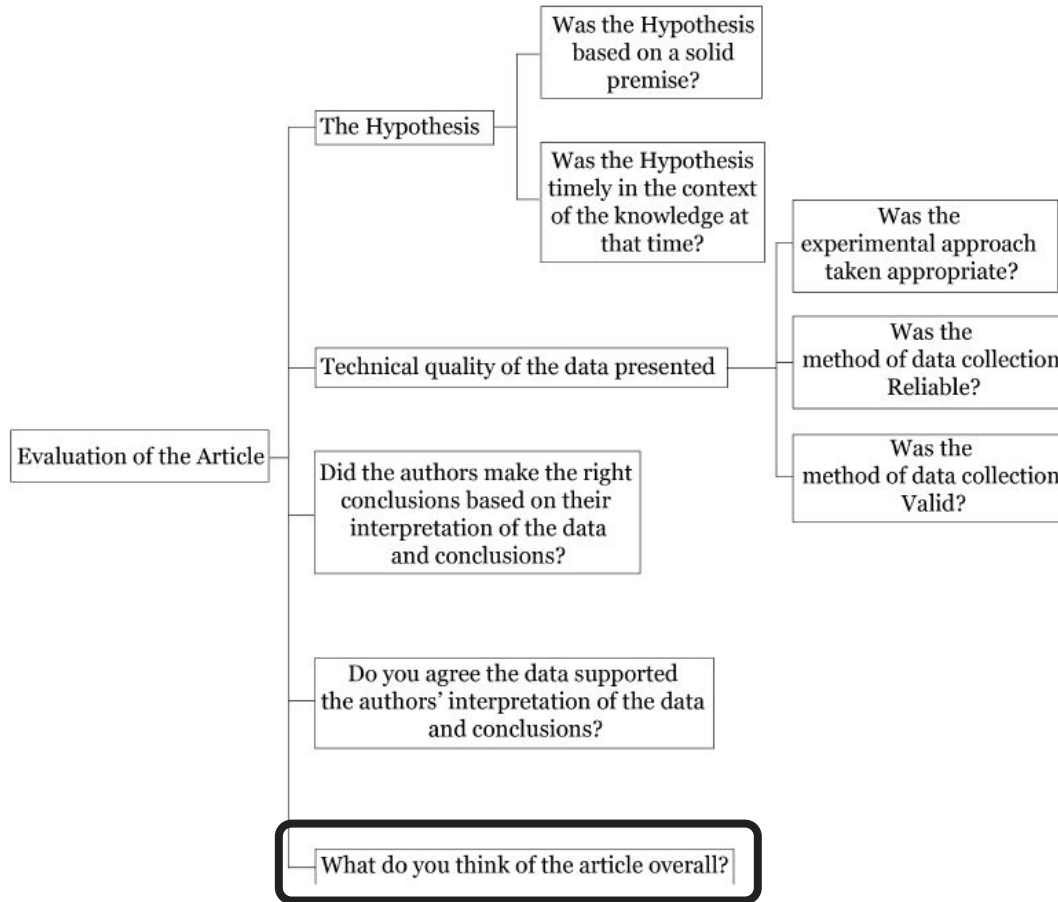
Yeong, F.M. (2014). *How to read and critique a scientific research article*. New Jersey: World Scientific. P. 4

2nd step: Evaluating the article



Yeong, F.M. (2014). *How to read and critique a scientific research article*. New Jersey: World Scientific. P. 4

2nd step: Evaluating the article



Yeong, F.M. (2014). *How to read and critique a scientific research article*. New Jersey: World Scientific. P. 4

In-video quiz

Read the summary. Choose True or False for each statement on the next slide.

1. Jiang et al. validated an existing blood-based panel to diagnose Alzheimer's Disease (AD).
2. Jiang et al. used proximity extension assay (PEA) to identify plasma biomarkers associated with AD.
3. The panel was highly accurate because the area under curve (AUC) was greater than 1.

In-video quiz

Read the summary. Choose True or False for each statement on the next slide.

1. Jiang et al. **devised and** validated an existing blood-based panel to diagnose Alzheimer's Disease (AD). **FALSE**
2. Jiang et al. used proximity extension assay (PEA) to identify plasma biomarkers associated with AD. **TRUE**
3. The panel was highly accurate because the area under curve (AUC) was ~~greater than~~ **close to 1**. **FALSE**
 - a. Training group (AUC=0.9891)
 - b. Verification group (AUC=0.9690)

Example from an SP227I student: Evaluating claims of novelty

“Recent advances in ultrasensitive and high-throughput protein measurement technologies have enabled large-scale proteomic profiling of the blood, which have been widely adopted to study cardiovascular diseases and aging, consequently identifying novel biomarkers and providing biological annotations for disease stages. Accordingly, in this study, we used proximity extension assay (PEA) technology to systematically evaluate the protein profiles of AD plasma.” (Jiang et al. (2021) p. 89)

Was the experimental approach taken appropriate?



????

In-video quiz

“Recent advances in ultrasensitive and high-throughput protein measurement technologies have enabled large-scale proteomic profiling of the blood, which have been widely adopted to study cardiovascular diseases and aging, consequently identifying novel biomarkers and providing biological annotations for disease stages. Accordingly, in this study, we used proximity extension assay (PEA) technology to systematically evaluate the protein profiles of AD plasma.” (Jiang et al. (2021) p. 89)

Was the experimental approach taken appropriate?



Choose the best option.

- A. What are the most advanced technologies to identify plasma biomarkers?
- B. How does PEA work?
- C. Why is PEA better than other available options?
- D. Other

In-video quiz

“Recent advances in ultrasensitive and high-throughput protein measurement technologies have enabled large-scale proteomic profiling of the blood, which have been widely adopted to study cardiovascular diseases and aging, consequently identifying novel biomarkers and providing biological annotations for disease stages. Accordingly, in this study, we used proximity extension assay (PEA) technology to systematically evaluate the protein profiles of AD plasma.” (Jiang et al. (2021) p. 89)

Was the experimental approach taken appropriate?



Choose the best option.

- A. What are the most advanced technologies to identify plasma biomarkers?
- B. How does PEA work?
- C. **Why is PEA better than other available options?**
- D. Other

Example from an SP227I student: Substantiate

Firstly, the study¹ aptly used the new PEA technology in the development of their AD-biomarker panel.

... because...



PEA enabled the comprehensive screening of 1160 proteins, whereas previous blood-based biomarker studies typically screened less than 300 proteins.²⁻⁶ Moreover, PEA can detect very low protein concentrations, down to 1pM.^{7,8} This addressed the challenge of detecting small quantities of blood-based biomarkers.⁹ For instance, the plasma amyloid levels were found to be 5-10 times lower than in the [cerebrospinal fluid] CSF,¹⁰ due to the blood-brain barrier obstructing diffusion. Therefore, earlier studies that used enzyme-linked immunosorbent assay (ELISA) could not detect minute changes in the plasma biomarker levels.^{3,11,12} Comparatively, PEA is high-throughput and more sensitive.

Example from an SP227I student: Evidence-based evaluation

- (2) Plasma Proteomics for the Identification of Alzheimer's Disease.
- (3) Classification and Prediction of Clinical Alzheimer's Diagnosis Based on Plasma Signaling Proteins.
- (5) Multiplex Proteomics Identifies Novel CSF and Plasma Biomarkers of Early Alzheimer's Disease.
- (7) Homogeneous Antibody-Based Proximity Extension Assays Provide Sensitive and Specific Detection of Low-Abundant Proteins in Human Blood.
- (8) Our PEA technology - Olink <https://www.olink.com/our-platform/our-pea-technology/>
- (11) Identification of a Blood-Based Biomarker Panel for Classification of Alzheimer's Disease.

Identification, Classification, Detection = identif* OR classif* OR detect*

Example from an SP227I student: Evaluating data interpretation

“discovery cohort comprised 180
... validation cohort comprised
97”
(Jiang et al. p 89)

“[Proteins] PRDX1, VAMP5, and
GAMT ... were consistently
dysregulated in ... discovery and
validation cohorts ... the above
inconsistencies are likely due to the
small sample size of our validation
cohort. Therefore, replication
studies examining the 19-protein
biomarker panel in a larger
population are required”
(Jiang et al. p 100)

Was the
method of data collection
Reliable?

Was the
method of data collection
Valid?



???

In-video quiz

“discovery cohort comprised 180
... validation cohort comprised 97”

(Jiang et al. p 89)

“[Proteins] PRDX1, VAMP5, and GAMT ... were consistently dysregulated in ... discovery and validation cohorts ... the above inconsistencies are likely due to the small sample size of our validation cohort. Therefore, replication studies examining the 19-protein biomarker panel in a larger population are required”
(Jiang et al. p 100)

Was the
method of data collection
Reliable?

Was the
method of data collection
Valid?



Choose the best option(s).

- A. What else could a small sample size have caused?
- B. What would be the implications of consistently dysregulated proteins?
- C. How large of a population would be needed in a replication study?
- D. Other

In-video quiz

“discovery cohort comprised 180
... validation cohort comprised
97”
(Jiang et al. p 89)

“[Proteins] PRDX1, VAMP5, and
GAMT ... were consistently
dysregulated in ... discovery and
validation cohorts ... the above
inconsistencies are likely due to the
small sample size of our validation
cohort. Therefore, replication
studies examining the 19-protein
biomarker panel in a larger
population are required”
(Jiang et al. p 100)

Was the
method of data collection
Reliable?

Was the
method of data collection
Valid?



Choose the best option(s).

- A. What else could a small sample size have caused?
- B. What would be the implications of consistently dysregulated proteins?
- C. How large of a population would be needed in a replication study?
- D. Other

Example from an SP227I student: Challenge

The small sample size could also impact the panel's impressive diagnostic accuracy.

Nevertheless, biomarker studies like this one are susceptible to overfitting bias.¹⁴ Overfitting occurs when the model of a panel not only fits the variation between two groups but also the noise within individual samples. This results in a high published accuracy that is unable to be reproduced.^{17,18} Overfitting effects were prominent when the size of a panel is more than one-tenth of the sample size.¹⁹ In this case, there are 19 biomarkers, with 180 and 97 samples in the two cohorts respectively – risking both cohorts to overfitting bias (panel/sample=0.11 and 0.20 respectively). Although it is commendable that the authors validated the panel, the sample size of 97 could have been too small to assess the panel accurately. A small sample size reduces statistical power, decreasing the likelihood of a statistically-significant result to represent a true outcome. Hence, the 19-protein panel should be further evaluated to possibly narrow it down, or replicated in larger cohorts.



Example from an SP227I student: Challenge

The small sample size could also impact the panel's impressive diagnostic accuracy.

Nevertheless, biomarker studies like this one are susceptible to overfitting bias.¹⁴ Overfitting occurs when the model of a panel not only fits the variation between two groups but also the noise within individual samples. This results in a high published accuracy that is unable to be reproduced.^{17,18} Overfitting effects were prominent when the size of a panel is more than one-tenth of the sample size.¹⁹ In this case, there are 19 biomarkers, with 180 and 97 samples in the two cohorts respectively – risking both cohorts to overfitting bias (panel/sample=0.11 and 0.20 respectively). Although it is commendable that the authors validated the panel, the sample size of 97 could have been too small to assess the panel accurately. A small sample size reduces statistical power, decreasing the likelihood of a statistically-significant result to represent a true outcome. Hence, the 19-protein panel should be further evaluated to possibly narrow it down, or replicated in larger cohorts.



Example from an SP227I student: Challenge

The small sample size could also impact the panel's impressive diagnostic accuracy.

Nevertheless, biomarker studies like this one are susceptible to overfitting bias.¹⁴ Overfitting occurs when the model of a panel not only fits the variation between two groups but also the noise within individual samples. This results in a high published accuracy that is unable to be reproduced.^{17,18} Overfitting effects were prominent when the size of a panel is more than one-tenth of the sample size.¹⁹ In this case, there are 19 biomarkers, with 180 and 97 samples in the two cohorts respectively – risking both cohorts to overfitting bias (panel/sample=0.11 and 0.20 respectively). Although it is commendable that the authors validated the panel, the sample size of 97 could have been too small to assess the panel accurately. A small sample size reduces statistical power, decreasing the likelihood of a statistically-significant result to represent a true outcome. Hence, the 19-protein panel should be further evaluated to possibly narrow it down, or replicated in larger cohorts.



Example from an SP227I student: Challenge

The small sample size could also impact the panel's impressive diagnostic accuracy.

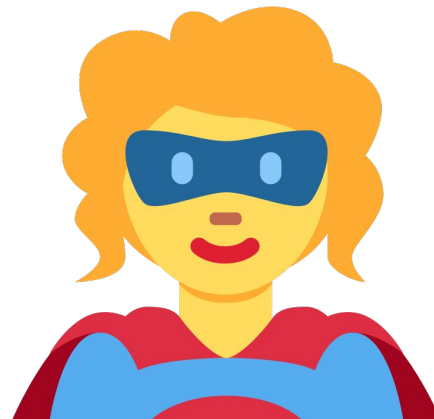
Nevertheless, biomarker studies like this one are susceptible to overfitting bias.¹⁴ Overfitting occurs when the model of a panel not only fits the variation between two groups but also the noise within individual samples. This results in a high published accuracy that is unable to be reproduced.^{17,18} Overfitting effects were prominent when the size of a panel is more than one-tenth of the sample size.¹⁹ In this case, there are 19 biomarkers, with 180 and 97 samples in the two cohorts respectively – risking both cohorts to overfitting bias (panel/sample=0.11 and 0.20 respectively). Although it is commendable that the authors validated the panel, the sample size of 97 could have been too small to assess the panel accurately. A small sample size reduces statistical power, decreasing the likelihood of a statistically-significant result to represent a true outcome. Hence, the 19-protein panel should be further evaluated to possibly narrow it down, or replicated in larger cohorts.



Example from an SP227I student: References that help

★ Review articles! ★

- (14) Decade of Blood Biomarkers for Alzheimer's Disease Research: An Evolving Field, Improving Study Designs, and the Challenge of Replication. *J. Alzheimer's Dis.* **2018**
- (17) Points of Significance: Model Selection and Overfitting. *Nat. Methods* **2016**.
- (18) Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience. *Nat. Rev. Neurosci.* **2013**.
- (19) Autoantibodies for Cancer Detection: Still Cause for Excitement? *Cancer Biomark.* **2010**



Example from an SP227I student: Evaluating application

“Our findings not only serve as a foundation for the development of a high-performance, blood-based test for AD screening and monitoring in clinical settings but also provide abundant protein targets for future therapeutic development.”
(Jiang et al. p 100)



How could this research be applied?

Example from an SP227I student: Extend

The researchers' test probably has a long journey before it could be deployed in clinical practice.

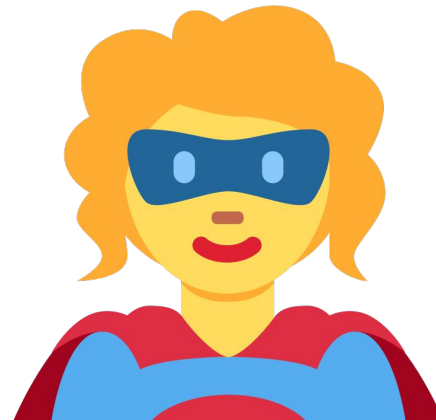


Finally, despite research on AD biomarkers totalling around 8000 publications over the last 5 years,²³ translation into clinical practice has been slow.²⁴ Clinical adoption of this 19-protein panel as a diagnostic tool will require [1] synergy across research groups to replicate the panel in large, independent cohorts; [2] “lock-down” all the steps in the process from collection to analysis^{24,25}; followed by [3] clinical trials and [4] commercial development of the panel. In conclusion, this is a good exploratory study that should be further validated and standardised for potential clinical applications.

Example from an SP227I student: More superhero reviews

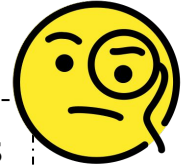
★ Review articles! ★

- (23) Advantages and Pitfalls in Fluid Biomarkers for Diagnosis of Alzheimer's Disease. *J. Pers. Med.* **2020**
- (24) Strategic Roadmap for an Early Diagnosis of Alzheimer's Disease Based on Biomarkers. *Lancet. Neurol.* **2017**
- (25) 2020 Update on the Clinical Validity of Cerebrospinal Fluid Amyloid, Tau, and Phospho-Tau as Biomarkers for Alzheimer's Disease in the Context of a Structured 5-Phase Development Framework. *Eur. J. Nucl. Med. Mol. Imaging* **2021**



Example from an SP2271 student: Overall evaluation

– What do you think of the article overall?



Finally, despite research on AD biomarkers totalling around 8000 publications over the last 5 years,²³ translation into clinical practice has been slow.²⁴ Clinical adoption of this 19-protein panel as a diagnostic tool will require [1] synergy across research groups to replicate the panel in large, independent cohorts; [2] “lock-down” all the steps in the process from collection to analysis^{24,25}; followed by [3] clinical trials and [4] commercial development of the panel. In conclusion, this is a **good exploratory study that should be further validated and standardised for potential clinical applications.**

Summary of Critical Reading Guide

1. Get a good understanding of the paper's main ideas
 - see Table 1.1 in the textbook.
2. Evaluate the paper with critical questions
 - see Figure 1.1 in the textbook.
3. Search the library
 - look for review articles
4. Return to the paper



Thank you for watching!

Dara Richard

dara.richard@nus.edu.sg