Consensus

Finding a research question

A one week assignment. Students take one published study, build a hypothesis of their own, and find out whether the literature has already answered it.

INSPIRED BY

This assignment is inspired by a psychology professor in Kentucky, who runs a similar assignment with her research methods students to teach them how to create a hypothesis that addresses unanswered questions in their field of choice.

WHO THIS ASSIGNMENT IS FOR

The assignment is designed for students starting a research project or thesis. It teaches students that locating a question in the literature is a distinct skill from simply having an idea. Key learning outcomes include:

- **Deconstructing Empirical Articles:** Learning to separate the actual relationship tested from the claims made in the title or conclusion.
- **Formulating Testable Hypotheses:** Writing hypotheses that clearly define the population, outcome, and predicted direction before beginning research.
- **Navigating Literature:** Establishing whether a research question is truly open, contested, or settled by existing studies.
- **Seeking Contradictory Evidence:** Actively finding and describing evidence that challenges their original hypothesis rather than only assembling sources that agree with them.
- **Iterative Revision:** Revising hypotheses based on evidence discovered during the search process.

IN THIS ASSIGNMENT

Assignment Overview

01 Why this assignment exists

02 The five parts

03 How the week runs

04 Learning outcomes

Example Submission

Part 0 The article teardown

Part 1 The hypothesis

Part 2 The search log

Part 3 The five sources

Part 4 The standing report

Part 5 The proposal defence (optional)

Assignment Overview

Students are given one empirical article. They take it apart, name the relationship it tested, and write a hypothesis of their own that extends or challenges it.

They spend a week finding out where that hypothesis sits in the literature, using Consensus, and collecting five sources that establish its standing. They then assign it into one of the following categories.

Where the hypothesis sits	What it requires, and what it means for the proposal
Settled	Human studies have asked this and converged on an answer. Proposing the main effect again adds nothing. The contribution has to move to a boundary condition, a new population, or a new outcome.
Contested	Asked repeatedly with results that conflict. A direct replication or a test of a moderator is warranted. Students must name the conflict, not just say results are mixed.
Open	Each construct is studied on its own but the relationship between them has not been tested. Students must name the studies that establish each side separately.
Untestable as stated	Cannot be studied as written, because a term is undefined, the outcome is not measurable, or no design could produce evidence either way. Students must name what has to change.

After they assign it into a category, they can revise their hypothesis so that it addresses a contested or open question in the literature.

01 Why this assignment exists

Students in their first research methods course often propose studies that have already been done. Not because they are lazy, but because nobody taught them that locating a question in the literature is a separate skill from having the question. They read one article, get an idea, and write a proposal in a vacuum.

Consensus makes the check fast enough to do before drafting their proposal. Students can search a hypothesis, filter to human studies and meta-analyses, and see in a few minutes whether their hypothesis has already been studied. Doing that check first is the habit this assignment installs.

TIME	One week. Two class sessions of about 40 and 25 minutes, plus roughly 3 hours of student work in between.
COURSE FIT	Research methods, experimental psychology, thesis preparation, honours seminars, any course with a proposal or prospectus deliverable.
LEVEL	Sophomore through senior methods courses. Works as the first graded step of a term-long proposal.
TOOL	Consensus (consensus.app). Check institutional access before you assign it. Free accounts have limited searches.
PREP	About an hour the first time. Pick the seed article and run the searches yourself so you know what students will hit. Ten minutes every term after that.

02 The five parts

Each part below has a description and a worked example of what a real submission looks like. The weights are a suggestion.

#	Part	What it is	Weight
0	**Article Teardown**	Six lines on the seed article, written before any searching. The relationship tested, the sample, the design, the effect, the authors' own stated limitation, and the one thing left unexplained.	Gate, ungraded
1	**The Hypothesis**	One sentence in testable form, plus three lines naming the population, the outcome, and the direction. Written before searching, so it can be graded against what the search then finds.	15 of 100
2	**Search Log**	Five searches in a table: query as typed, filters used, papers opened versus kept, and what they changed and why.	20 of 100
3	**The Five Sources**	Five peer reviewed sources. For each one: citation and working link, design and tier, one sentence with n, population, and result, and what it does to the hypothesis.	25 of 100
4	**The Standing Report**	About 350 words placing the hypothesis in one of the four categories, defending the placement from the five sources, and stating the revised hypothesis.	30 of 100

#	Part	What it is	Weight
5	**Proposal defence (optional)**	In class in groups of three. Students present one member's revised hypothesis and the search that produced it, with the work split three ways.	10 of 100

03 How the week runs

When	What happens	Due
Before class 1	Students read the seed article and write the Article Teardown.	Teardown brought to class
Class 1	About 40 minutes with laptops open. You run three searches live: hypothesis to query, filtering to replications and meta-analyses, and one check of a summary against a paper.	Nothing. They leave with a drafted hypothesis.
Days 2 to 4	Independent work. Search, screen, read, place the hypothesis, revise it.	Nothing collected
Class 2	Optional proposal defence, about 25 minutes in groups of three.	Parts 1 to 4 at the start of class

04 Learning outcomes

By the end of the week, students will be able to:

1. Take an empirical article apart into the relationship it tested, the sample it tested it on, and the claim it can support, and separate that from the claim its title makes.
2. Write a hypothesis in testable form, naming the population, the outcome, and the predicted direction before searching for support.
3. Establish where a hypothesis already sits in the literature, and tell the difference between a question nobody has asked and a question that was asked and answered.
4. Find and describe evidence that points against the study they started from, rather than assembling a set of sources that agree with them.
5. Revise a hypothesis in response to what the search found, and say which source forced the revision.

HOW STUDENTS MAY AND MAY NOT USE AI

Consensus is required here. Its use is graded. The skill being assessed is using an AI research tool well, which is different from using it at all.

Permitted: searching, filtering, screening abstracts, and using summary and study extraction features to decide which papers are worth their time.

Required: they open and read every paper they cite, and they read the methods section of every source that changes the hypothesis.

Not permitted: submitting an AI generated hypothesis or an AI generated literature summary as their own analysis, or citing a paper they have not opened.

Example Submission

Every example follows one student through one seed article. The article, the papers, and the outcome of the search are all real.

The seed article: Mueller, P. A. and Oppenheimer, D. M. (2014). The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking. Psychological Science, 25(6), 1159 to 1168.

PART 0

The article teardown

Before class 1. Six lines. Ungraded.

The point is to separate what the article tested from what its title claims. Students who skip this write hypotheses about the title. Require the effect size or the test statistic in line four, because a student who cannot find it has read the abstract only.

Line six is the one that generates the hypothesis. Students who cannot name anything the article left unexplained have not read the discussion section, where authors almost always say it themselves.

EXAMPLE SUBMISSION

The relationship tested	Note taking medium (longhand or laptop) as a cause of performance on factual and conceptual test questions about a lecture.
The sample	College undergraduates across three studies. Lab based. Study 1 used a TED talk lecture and a quiz taken after a short delay.
The design	Experimental. Participants were randomly assigned to a note taking medium, so a causal claim is available here, unlike most of what I read this term.
The effect	Laptop note takers took more notes overall and more verbatim strings, and scored worse on conceptual questions. The authors attribute this to shallower processing rather than to distraction, since laptops were used only for notes.
The authors' own stated limitation	They note the lab setting and the short delay before testing, and that they did not test what happens when students study their notes before being tested.

The one thing left unexplained	If verbatim transcription is the mechanism, then anything that reduces verbatim transcription should erase the effect. Nothing in the study tests that. Also nobody in these studies was distracted, which is the actual condition in a real classroom.

PART 1

The hypothesis

15 points. One sentence plus three lines. Written before any searching.

This part is graded on form, not on whether it survives. A hypothesis must name a population, a measurable outcome, and a predicted direction, and it must be possible to be wrong. Students lose points here for hypotheses that cannot fail.

Collect this before the search so you can grade Part 4 against it. That sequence is the whole design of the assignment. If students write the hypothesis after searching, they will write one they already know the answer to, and Part 4 becomes unscoreable.

EXAMPLE SUBMISSION

Hypothesis	Undergraduates who take lecture notes by hand will score higher on conceptual questions than undergraduates who type their notes, when both groups are tested after a one week delay.
Population	Undergraduate students in a lecture course. I am extending from lab participants to an actual class.
Outcome	Percentage correct on conceptual questions, scored the way the seed article scored them.
Direction	Longhand higher than laptop. I predict the gap widens with the longer delay, because if longhand encodes more deeply the advantage should grow, not shrink.

PART 2

The search log

20 points. A table, five rows minimum.

Tell students at assignment time that dead ends earn credit, then actually award it. “Nothing came up” is not a reason. “The first page was all commentary on the original paper rather than new data, so I filtered to RCTs” is.

EXAMPLE SUBMISSION

#	Query, as typed	Filters	Opened / kept	What I changed and why
1	`pen is mightier than the keyboard`	none	4 / 1	Searched the title. Got the original plus commentary about it. No new data, so I searched the relationship instead of the paper.
2	`longhand versus laptop note taking and test performance`	none	7 / 2	Now I am getting studies. Mixed in with opinion pieces about laptops in classrooms.
3	`Does handwriting notes improve test performance compared to typing?`	Human studies	9 / 3	Yes or no phrasing gave me a Meter, and it was far less one sided than I expected given how confident the original paper is.
4	`replication of Mueller and Oppenheimer note taking`	Human studies, RCT	6 / 3	My required search against the seed article. This is the row that broke my hypothesis. Two direct replications, neither of which found what the original found.
5	Same relationship as row 2	Meta-analysis, Systematic review	4 / 2	Went to the top of the hierarchy. Found a meta-analysis of 77 effect sizes with a pooled effect of essentially zero. At this point my hypothesis was finished.
6	`note taking medium digital distraction lecture learning`	Human studies	5 / 2	Followed the moderator the meta-analysis pointed at. This is where my revised hypothesis came from.

PART 3

The five sources

25 points. Five peer reviewed sources, four fields each.

Three constraints.

1. At least one source must point against the seed article. This is the constraint students resist and learn the most from.

2. At least one must be a primary empirical study whose methods section they read.
3. No two sources may share a first author, because one lab is not a literature.

EXAMPLE SUBMISSION, TWO OF FIVE SOURCES SHOWN

Citation and link	Urry, H. L. et al. (2021). Don't Ditch the Laptop Just Yet: A Direct Replication of Mueller and Oppenheimer's (2014) Study 1 Plus Mini Meta-Analyses Across Similar Studies. Psychological Science. [Consensus link pasted here]
Design and tier	Direct pre-registered replication of Study 1, plus exploratory meta-analyses across eight similar studies. Higher tier than the original for my purposes, because it was designed to test the original rather than to discover something.
Who, and what happened	Laptop n = 74, longhand n = 68. As in the original, laptop participants recorded more verbatim words and more words overall. Unlike the original, the two groups did not differ on the quiz. The authors conclude that the results do not support the idea that longhand note taking improves immediate learning through better encoding.
What it does to the hypothesis	It contradicts the direction I predicted. The mechanism the original proposed, verbatim transcription, replicated. The learning outcome did not. That split is the most interesting thing I found all week, because it means the story was half right.

Citation and link	Voyer, D. et al. (2021). The Effect of Notetaking Method on Academic Performance: A Systematic Review and Meta-Analysis. Contemporary Educational Psychology. [Consensus link pasted here]
Design and tier	Multilevel meta-analysis. Top of the hierarchy.
Who, and what happened	77 effect sizes from 39 samples across 36 articles. Mean estimated g = -0.008, 95% CI -0.18 to 0.16, not significantly different from zero. Moderator analysis found that the apparent longhand advantage in some earlier studies can be explained at least partly by distraction from other applications, which is only possible on a digital device.
What it does to the hypothesis	It kills the main effect and hands me a replacement. If the medium does nothing but distraction does something, then the real question is about distraction, not handwriting. My revised hypothesis comes directly from this sentence.

PART 4

The standing report

30 points. About 350 words. The heaviest part.

Students place their hypothesis in one of the four categories from page one, defend the placement using their five sources by name, and then state the revised hypothesis the search produced. Three things must appear: the category and why, the single source that most constrains the original hypothesis, and the revised hypothesis.

Ask for the name of the source that forced the revision, and the ones who did no real search cannot supply it.

> **EXAMPLE SUBMISSION, ABOUT 350 WORDS**
>
> **Standing: Contested, and drifting toward settled in the null direction.**
>
> My hypothesis was that longhand note takers would outscore typists on conceptual questions. The literature does not support this, and I did not know that when I wrote it.
>
> The seed article found the effect across three studies. But two direct replications did not. Urry and colleagues (2021) replicated Study 1 with 142 participants and found the verbatim difference but no quiz difference. Morehead, Dunlosky and Rawson (2019) ran a direct replication plus extensions with eWriter and no-notes groups, and performance did not consistently differ between any groups, including the group that took no notes at all. Their meta-analysis of direct replications found small effects favouring longhand that were not statistically significant, and they concluded that deciding which method is superior is premature.
>
> The aggregate is stronger than either replication. Voyer and colleagues (2021) pooled 77 effect sizes from 39 samples and found a mean g of -0.008, with a confidence interval running from -0.18 to 0.16. That interval contains zero comfortably. On the main effect I proposed, this is close to settled, and settled against me.
>
> The source that most constrains my hypothesis is the Voyer meta-analysis, and not only because of the null. Its moderator analysis found that the longhand advantage reported in earlier work is explained at least partly by distraction from other applications on digital devices. Flanigan and colleagues (2020) point the same way, reporting that digital distraction predicted lecture learning more meaningfully than note taking method did.
>
> Revised hypothesis. Among undergraduates in a lecture course, note taking medium will not affect conceptual test performance, but access to non-note applications during the lecture will, such that students on unrestricted laptops score lower than students on laptops restricted to a notes application and lower than longhand note takers.
>
> This is a moderator test rather than a main effect test, it is not settled, and unlike my first hypothesis it could come out either way.

PART 5

The proposal defence (optional)

10 points. Groups of three, about 25 minutes, in class.

Each student arrives with a completed sprint. The group spends five minutes comparing the three revised hypotheses and picks one to present. They then present that one together, screen shared, with the work split three ways.

Section	Time	The job
The original hypothesis and the teardown	5 min	State the seed article, the relationship it tested, and the hypothesis the group member started with. Say what was reasonable about it.
The search	5 min	Pull up the actual search log on screen. Show the query that found the disconfirming evidence and the filters that got there.
Standing and revision	5 min	Name the category, the source that forced the revision, and the revised hypothesis.

Whoever did not originally run this sprint needs to understand their section well enough to present it, not read it off a groupmate's document. That is the part that cannot be faked, and it is the reason the group presents one hypothesis rather than three.