Consensus

Evidence based recommendations

A one week assignment. Students make one recommendation to a real decision maker, back it with five peer reviewed sources, and determine if evidence applies to the person they are advising.

INSPIRED BY

A humanities professor at Ohio State University and a business professor at Emory who have students find peer reviewed evidence to support recommendations made to real-life organizations.

WHO THIS ASSIGNMENT IS FOR

This assignment is designed for students in professional programs who need to bridge the gap between academic research and real-world application. It teaches students how to:

- **Translate practical problems into research:** Convert a real-world decision into a searchable empirical question, recognizing that the language of decision-makers often differs from academic literature.
- **Set evidence standards:** Determine criteria for evidence (e.g., sample size, study type, country) before searching, rather than simply accepting whatever results appear first.
- **Evaluate transferability:** Judge whether a finding established in one population and setting actually applies to their specific client, and clearly articulate the conditions under which it might not apply.
- **Verify AI-generated summaries:** Critically check AI summaries against the original source papers, specifically looking for nuances like sample size and setting that summaries often omit.
- **Communicate for impact:** Defend a recommendation to a stakeholder by focusing on cost, feasibility, and fit, while honestly acknowledging the limits of the evidence.

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 decision stub

Part 1 The search log

Part 2 The source portfolio

Part 3 The transfer test

Part 4 The brief

Part 5 The stakeholder challenge (optional)

Assignment Overview

Students name a real decision facing a real organization, then spend a week finding out what the research says about it. They collect five peer reviewed sources, read them, and check whether the populations and settings those studies used resemble the organization they are advising. Then they write a one page brief to the person who would actually act on it.

Students must place their recommendation in one of these four categories and defend the placement:

Transfer verdict	What it requires
Transfers directly	Multiple studies establish the effect in populations and settings that resemble the client's, at a scale the client could act at.
Transfers with conditions	A real effect exists, but it depends on conditions the client may or may not meet. Students must name the conditions. This is the correct verdict far more often than students expect.
Does not transfer	Evidence exists and the client's context fails a condition the effect depends on. Students must name the condition and show where the paper establishes it.
Untested in this context	No study has looked at anything like the client's population, sector, or scale. Different from does not transfer, and the difference changes the recommendation.

01 Why this assignment exists

Applied courses ask students to make a recommendation and back it with evidence. What students often produce is a recommendation they already believed, followed by citations collected afterward to decorate it. The citations are real and the papers were published. But the student did not check whether a finding established on 2,896 employees across six countries has anything to do with a 58 person nonprofit.

That check is the assignment. Consensus lets students filter by country, sector, sample size, and study type, which means they can set a standard before seeing results instead of accepting whatever surfaced first. Its extracted study details carry sample and setting, which is exactly the field the transfer question runs on, and exactly the field a summary is most likely to drop.

TIME	One week. Two class sessions of about 40 and 25 minutes, plus roughly 3 hours of student work in between.
COURSE FIT	Management, public policy, social work, education leadership, health administration, family science. Any course with an applied deliverable and a decision maker.
LEVEL	Upper division undergraduate in a professional major, and terminal master's programmes such as MBA, MPA, MSW, MPH, and MEd.
TOOL	Consensus (consensus.app). Check institutional access before you assign it. Free accounts have limited searches.
PREP	About an hour the first time. Run the searches yourself with a client you could actually name. Ten minutes every term after that.

THE RULE THAT GOVERNS THE ASSIGNMENT

No finding travels for free. Every result you cite was established on somebody, somewhere, under conditions the authors chose.

Name that population and those conditions, or do not cite the finding.

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	**Decision Stub**	Six lines written before any searching. The decision maker, the choice, the current default, the population affected, the outcome that matters, and the binding constraint.	Gate, ungraded
1	**Search Log**	Five searches in a table: query as typed, filters used, papers opened versus kept, and what they changed and why. One row must look for evidence from the client's own context.	15 of 100
2	**Source Portfolio**	Five peer reviewed sources. For each one: citation and working link, design and tier, one sentence with n, population, setting, and timeframe, and the finding they will use.	25 of 100

#	Part	What it is	Weight
3	**The Transfer Test**	About 350 words. Three to five conditions the finding depends on, each checked against the paper and marked satisfied, contradicted, or silent.	30 of 100
4	**The Brief**	One page to the named decision maker. The recommendation, the case, what would change it, and the cost.	20 of 100
5	**Stakeholder challenge (optional)**	In class in groups of three. Students present one member's brief and defend it against cost and transfer questions.	10 of 100

Part 3 cannot be written from summaries, because summaries almost never carry the sample and setting detail it runs on. Grade it first, for every student, and everything else reads differently.

03 How the week runs

When	What happens	Due
Before class 1	Students identify a real or realistic decision and write the Decision Stub. You check it at the door.	Stub brought to class
Class 1	About 40 minutes with laptops open. You run three searches live: the client's phrase to a research construct, setting an evidence standard with filters, and one transfer pass against a paper.	Nothing. They leave with a drafted decision question.
Days 2 to 4	Independent work. Search, screen, read, check transfer, write.	Nothing collected
Class 2	Optional stakeholder challenge, 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. Translate a practical decision into a searchable empirical question, and explain why the decision maker's language is not the literature's language.
2. Set an evidence standard before searching, using study type, sample size, and country filters, rather than accepting whatever surfaced first and calling it the evidence.

3. Judge whether a finding established in one population and setting applies to a different one, and state plainly the conditions under which it would not.
4. Check an AI generated summary against the source paper and name what was lost. Summaries keep the headline and drop the sample and setting the headline depends on.
5. Defend a recommendation to people who care about cost and fit rather than methodology, and say what evidence would change their mind.

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 the paper the Transfer Test runs on.

Not permitted: submitting AI generated prose as the case section of the brief, citing a paper they have not opened, or handing in an AI synthesis as their analysis.

Example Submission

Every example follows one student advising one organization. The client is realistic, the papers are real, and the transfer verdict is the one the evidence actually supports.

The client: the executive director of a 58 person environmental nonprofit, deciding whether to trial a four day week, add a retention stipend, or do neither. The outcome she cares about is voluntary turnover over the next twelve months. She cannot increase payroll.

PART 0

The decision stub

Before class 1. Six lines. Ungraded.

Everything downstream is built on a decision. If the decision is vague, the search will be vague and the brief will be a book report. Thirty seconds per student at the door catches the vague ones before they cost a week.

The last two lines matter most. A stub with an unmeasurable outcome produces a search with no target, and a stub with no binding constraint produces a recommendation the decision maker has already rejected.

EXAMPLE SUBMISSION

The decision maker	The executive director. She can authorize a schedule change herself, with board notification.
The decision	Trial a four day week at full pay, add a $2,000 annual retention stipend, or do neither.
The current default	Do nothing. Turnover stays at roughly 22 percent a year, which is where it has been for three years.
The population affected	58 staff. About two thirds are programme staff in the field, one third administrative. Median tenure under three years. United States, single state.
The outcome that matters	Voluntary separations in the next twelve months, as a percentage of headcount. She already tracks this, which is why I picked it over morale.
The binding constraint	No payroll increase. Grant funded salary lines are fixed for the fiscal year, which rules out the stipend unless something else is cut.

PART 1

The search log

15 points. A table, five rows minimum.

Students are graded on how they searched. Two rows are required. One must be a counterweight search, aimed at finding evidence that the recommendation would fail. One must be a context search, aimed deliberately at evidence from a population like the client's. If no such evidence exists, that dead end is one of the most valuable findings in the assignment and it belongs in the brief.

EXAMPLE SUBMISSION

#	Query, as typed	Filters	Opened / kept	What I changed and why
1	`how to improve employee retention`	none	`5 / 0`	My client's phrase. Results spanned three unrelated literatures and none of it was about a schedule change. Picked one intervention and one countable outcome instead.
2	`four day work week and employee turnover`	none	`8 / 3`	Now I am getting studies. Noticed immediately that most measure wellbeing, not turnover, which turned out to be the whole problem.
3	`Does a four day work week reduce voluntary turnover?`	Human studies	`7 / 2`	Yes or no phrasing gave me a Meter. Weaker than the media coverage led me to expect.
4	`four day work week systematic review`	Systematic review, Meta-analysis	`4 / 2`	Top of the hierarchy. Found a review of 31 academic articles spanning 50 years, which is where my turnover evidence actually came from.
5	`four day workweek small nonprofit organization`	none	`3 / 1`	My context search. Found exactly one study in a small nonprofit. Its sample is five people. Logging that because the absence is itself a finding.
6	`four day work week problems intensified monitoring`	Human studies	`4 / 2`	My counterweight search. Looked for evidence the change backfires. Found a critical case study where performance monitoring intensified. Kept it.

Counts are from one student's session. What is graded is the reasoning in the last column. Row 5 returning almost nothing is a result, not a failure.

PART 2

The source portfolio

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

Four constraints.

1. At least one source must complicate, qualify, or contradict the others. A portfolio where everything agrees was assembled to agree.
2. At least one must come from a population or setting materially unlike the client's, with the difference named in one sentence.
3. At least one must be a primary empirical study whose methods section they read.
4. No two sources may share a first author, because one lab is not a literature and one consultancy is not a field.

The sample and setting field is the one the Transfer Test runs on. If a student cannot fill it in, they have not opened the paper.

EXAMPLE SUBMISSION, TWO OF FIVE SOURCES SHOWN

Citation and link	Fan, W. et al. (2025). Work time reduction via a 4-day workweek finds improvements in workers' well-being. Nature Human Behaviour. [Consensus link pasted here]
Design and tier	Multi-organization trial with a control group. The strongest design available on this question, and human research.
Sample and setting	2,896 employees across 141 organizations in Australia, Canada, Ireland, New Zealand, the UK and the USA. Six month trial, no reduction in pay, compared against 12 control companies. Crucially, organizations went through pre-trial work reorganization to improve efficiency before the trial began.
The finding I will use	Improvements in burnout, job satisfaction, mental health and physical health, not seen in the controls. Larger individual level hour reductions were associated with larger wellbeing gains. Note what is not here: turnover is not an outcome in this study.

Citation and link	Barnett, B. et al. (2025). "I think that the world needs this": results of a reduced 4-day workweek pilot in a small non-profit organization. Employee Relations: The International Journal. [Consensus link pasted here]

Design and tier	Qualitative evaluation, focus groups and interviews. Low tier for causal claims. I am including it because it is the only study I found in anything like my client's organization.
Sample and setting	Five full time staff at a small nonprofit in Canada. One year trial, 35 hours over five days reduced to 32 hours over four, no salary change.
The finding I will use	Staff and the leader reported improvements in work life balance, efficiency, and workplace attitudes, with no reported change in the quantity or quality of work. They recommend a pilot period and a review of policies and boundaries first. This is my closest match on context and my weakest source on evidence, and holding both of those facts at once is the whole assignment.

PART 3

The transfer test

30 points. About 350 words. The heaviest part.

Students pick the single source their recommendation leans on hardest, not the most interesting one. They list three to five conditions that must hold for the finding to apply to their client, then check each one against the paper and mark it:

Verdict	What it means
Satisfied	The study's sample and setting meet this condition. Cite the sentence in the methods that told you so.
Contradicted	The study's sample and setting fail this condition. Say by how much, in numbers.
Silent	The paper does not report enough to tell. This is the most common answer and the most useful one, because it names exactly what you do not know.

A condition like "the organizations must be similar" is a placeholder, not a condition. Students must also name one thing the AI summary layer got right that saved real time, and one thing it flattened that mattered for transfer.

> EXAMPLE SUBMISSION, ABOUT 350 WORDS
>
> **Source the recommendation leans on: Fan et al. (2025).**
>
> Condition 1: the outcome studied must be the outcome my client cares about. Contradicted. This trial measured burnout, job satisfaction, and health. It did not measure turnover. My client asked about

> voluntary separations. My strongest source does not speak to her question at all, and every popular article I read about this trial implied that it did.
>
> Condition 2: the effect must appear in organizations at my client's scale. Silent. The paper reports 141 organizations across six countries but I could not establish from it how many were under 100 people, or whether effects differed by size. It does report that individual level hour reductions mattered more than company level ones, which is mildly encouraging for a small employer, but that is not the same finding.
>
> Condition 3: the organization must be able to do the pre-trial work reorganization the trial required. Contradicted, or close to it. Participating organizations restructured work to improve efficiency and collaboration before the six month trial started. My client has no operations staff and no consultant budget. This is the condition I would have missed entirely if I had read only the summary.
>
> Condition 4: the effect must persist past six months. Silent here, and answered elsewhere. The trial ran six months. Campbell's (2023) review of 31 academic articles notes that benefits may fade over time, which matters because my client is asking about a twelve month outcome.
>
> What the AI layer got right. The extracted details gave me n, the six countries, the trial length, and the presence of controls in about fifteen seconds, and all four checked out against the paper.
>
> What it flattened. The pre-trial reorganization never appeared in anything I read except the paper itself. It is one clause in the abstract and it is arguably the active ingredient. A summary that drops it turns a structured organizational intervention into "give people Fridays off."
>
> What I will do differently. Before I cite a study, I will write down its outcome variable and check it against the outcome my client named. I assumed a wellbeing study was a turnover study for four days.

PART 4

The brief

20 points. One page, roughly 450 words, written to the decision maker and not to you.

Section	Size	What it has to do
Recommendation	`2 sentences`	Name the option and the transfer verdict. It must be something the client could start on Monday. If the recommendation is the current default, say so and defend it as a choice.
The case	`~200 words`	Organized by argument, not by source. Every claim carries a source and every source carries its tier.
What would change this	`~130 words`	The finding that would reverse the recommendation, and where to look for it. Include the condition the Transfer Test marked silent.
Cost and feasibility	`~80 words`	Address the binding constraint from Part 0. A recommendation that ignores it has already been rejected.

> **EXAMPLE SUBMISSION, EXCERPT**
>
> **Recommendation. Run a three month four day week pilot with two teams rather than the whole organization, and do not add the stipend. The evidence transfers with conditions, and a pilot is how you test the conditions that the research is silent on.**
>
> The case. The strongest evidence for a four day week is a controlled trial across 141 organizations, and it found real improvements in burnout and job satisfaction that control companies did not show. But it measured wellbeing, not turnover. The turnover evidence is older and weaker: a systematic review of 31 studies over 50 years reports reduced turnover among mostly favourable findings, while also noting that benefits can fade and that performance monitoring often intensifies. One qualitative study in a small nonprofit like ours reported success, and it had five participants.
>
> What would change this. Two things. First, the trial's organizations restructured their work before shortening the week, and we have no capacity to do that, so if the pilot teams simply compress the same workload into four days I would expect it to fail. Watch for after hours email volume in month two. Second, if voluntary separations in the pilot teams do not move by month three while wellbeing does, that tells us we bought morale rather than retention, which is worth knowing and is not what you asked for.
>
> Cost and feasibility. No payroll increase, which is why this beats the stipend on your constraint. The real cost is coverage on Fridays for field programmes, and that is the thing to solve before starting, not after.

PART 5

The stakeholder challenge (optional)

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

Each student arrives with a completed brief. The group spends five minutes comparing all three, then picks one and presents it while the other two attack it from fixed angles.

Section	Time	The job
The brief	5 min	Present the decision, the recommendation, the transfer verdict, and the single condition the Transfer Test marked silent.
The cost question	5 min	Attack cost and practical significance, not methodology. What does that effect mean in this client's units, how many people or dollars does it move, and what does it cost against the constraint? "Statistically significant" is not an answer to any of those.
The transfer question	5 min	Attack fit. Who was this studied on, and why do we think we are like them? Then run the core claim as a yes or no question in Consensus, live, and say what the Meter does and does not establish.

Whoever did not write this brief has to argue it well enough to be pressed on it. Presenting someone else's transfer verdict is the part that cannot be faked.