STUDY METHODOLOGIES

Research Methodology

The working skill of turning a vague topic into a well-supported answer — framing a real question, evaluating sources, avoiding cherry-picked synthesis, matching method to question, recognizing sampling bias, and citing honestly. Grounded throughout in real, documented cases: the Wakefield MMR fraud, two real Bohannon science-journalism stings, the New Coke launch, the 1936 Literary Digest poll, the Hawthorne effect reanalysis, and the Guttenberg plagiarism scandal — closing with a real, worked capstone research project.

10 Chapters | Philip Osztromok | Generated with Claude

CHAPTER 1: WHY RESEARCH METHODOLOGY MATTERS

Research Methodology

Chapter 1 · Why Research Methodology Matters

This course covers the working skill of turning a vague question into a well-supported answer — framing a real research question, telling primary from secondary sources, judging a source's own credibility, synthesizing findings honestly, and knowing which method actually fits the question being asked. It opens not with a definition, but with a real, extensively documented case of what happens when that skill fails at scale.

A Real, Devastating Case Study

On 28 February 1998, *The Lancet* published a paper by physician Andrew Wakefield and colleagues claiming a link between the MMR (measles, mumps, and rubella) vaccine and autism. The study's entire evidentiary basis was twelve children admitted to London's Royal Free Hospital in 1996-97 — a genuinely tiny sample for a claim that would go on to shape public health policy worldwide.

UNCOVERED BY REAL INVESTIGATIVE JOURNALISM, NOT PEER REVIEW

Journalist Brian Deer's own real investigation found that Wakefield had been paid approximately $750,000 by a lawyer preparing a lawsuit against MMR vaccine manufacturers — a conflict of interest never disclosed to *The Lancet* or the study's own readers. Four or five of the twelve children in the study had been referred through that same litigation-funded legal aid scheme. The British Medical Journal later concluded Wakefield had misrepresented or altered the medical histories of all twelve patients.

The Lancet retracted the paper in February 2010. Britain's General Medical Council found Wakefield "dishonest," "unethical," and "callous," and struck him from the UK medical register in May 2010, ending his ability to practice medicine.

THE REAL, MEASURED CONSEQUENCE

The retraction came twelve years after publication — and by then, the damage was real and lasting. Vaccination rates fell across multiple countries as parents genuinely feared the (nonexistent) risk. Measles, a disease that had been largely brought under control, returned in real, documented

outbreaks: a 2014-2015 outbreak beginning at Disneyland in California, and a 2019 outbreak of 1,274 confirmed cases across 31 US states — the highest single-year total since 1992.

What Actually Went Wrong

Nothing about this story required a reader to personally catch a statistical error. What went wrong was available to anyone who asked the right questions at the time: a sample size of twelve for a population-wide medical claim; an undisclosed financial relationship between the researcher and a party with a direct financial stake in the result; and a claim that spread through media coverage and word of mouth far faster than the actual, much slower process that eventually caught it. This course exists to teach the habits that make those questions automatic — not just for medical claims, but for any research task.

Casual Searching vs. Genuine Research

CASUAL, CONFIRMATION-DRIVEN SEARCHING	GENUINE RESEARCH
Starts with a conclusion, looks for support	Starts with a question, follows the evidence
Stops at the first source that agrees	Actively looks for disagreement and weighs it
Treats "I found an article" as proof	Asks who wrote it, who funded it, and how it was checked
Repeats a claim because it was widely shared	Traces a claim back to its original source before repeating it

What This Course Actually Covers

2
The Research Question
From a vague topic to something answerable

3
Source Types
Primary vs. secondary, across disciplines

4
Credibility
Peer review, predatory journals, real fraud cases

5
Literature Reviews

6
Methods

7
Sampling & Bias

Honest synthesis, not cherry-picking

Qualitative vs. quantitative, matched to the question

Real historical cases of sampling gone wrong

8

Data Collection

Surveys, interviews, experiments, observation

9

Citation & Synthesis

Quoting, paraphrasing, and avoiding plagiarism

10

Capstone

A real research project, start to finish

Hands-On Exercises

EXERCISE 1

Wakefield's own study never proved anything at the moment of publication — the fraud took twelve years to surface. In your own words, explain what specifically about the study's methodology (not its conclusion) should have prompted more scrutiny in 1998, before any investigation had happened.

View solution

EXERCISE 2

Pick a health, nutrition, or productivity claim you've seen shared online recently. Using only the "casual vs. genuine research" table above, write down which column your own first reaction to that claim actually fell into, and why.

View solution

EXERCISE 3

A friend argues that peer review alone would have caught Wakefield's fraud, so individual readers don't need to think critically about published research. Explain, in your own words, why the real twelve-year gap between publication and retraction complicates that argument.

View solution

CHAPTER 1 QUICK REFERENCE

- Wakefield's real 1998 Lancet paper claimed a link between the MMR vaccine and autism, based on just 12 patients
- Brian Deer's real investigative journalism uncovered an undisclosed ~$750,000 conflict of interest and altered patient histories
- The paper was retracted in 2010, twelve years later; Wakefield was struck from the UK medical register the same year
- Real, documented measles outbreaks followed the resulting drop in vaccination rates, including 1,274 US cases in 2019 — the most since 1992
- Genuine research starts with a question and weighs disagreement; casual searching starts with a conclusion and stops at the first source that agrees

CHAPTER 2: FORMULATING A RESEARCHABLE QUESTION

Research Methodology

Chapter 2 · Formulating a Researchable Question

"Is social media bad for teenagers?" isn't a research question — it's a topic wearing a question mark. Nobody can answer it, because it never says which teenagers, which platforms, how much use, or what "bad" would even look like. This chapter covers two real, established tools for turning a topic like that into something an actual research project could answer.

PICO: A Real Framework From Clinical Medicine

REAL, DOCUMENTED ORIGIN

PICO first appears in the research literature in Richardson et al.'s 1995 *ACP Journal Club* article, "The well-built clinical question: a key to evidence-based decisions." It was built for clinical medicine, but its underlying structure — naming exactly who, what, compared to what, and measured how — generalizes well beyond medical research.

PICO breaks a question into four real, separately named parts:

- **P — Population/Problem:** who or what, exactly, is the question about?
- **I — Intervention:** what specific factor, action, or exposure is being examined?
- **C — Comparison:** compared against what alternative or baseline?
- **O — Outcome:** what specific, measurable result is the question actually asking about?

A WORKED, ILLUSTRATIVE EXAMPLE — NOT A CLAIMED REAL FINDING

Applying PICO to the vague social-media question above (this is a worked example of the *method*, not a real cited study):

P: teenagers aged 13-17 · **I:** daily social media use exceeding 3 hours · **C:** daily use under 1 hour · **O:** self-reported anxiety symptoms, measured on a standardized scale

Assembled together: "*Among teenagers aged 13-17, is daily social media use exceeding 3 hours associated with higher self-reported anxiety symptoms, compared to daily use under 1 hour?*"

That's a question a real study could actually be designed around — every vague word in the original topic has been replaced with something specific and measurable.

FINER: A Real Checklist for Question Quality

A question can be perfectly well-structured under PICO and still be a poor choice to actually pursue. Stephen Hulley and colleagues, in the textbook *Designing Clinical Research*, proposed the FINER checklist — five real criteria, now taught well beyond clinical research, for judging whether a well-formed question is actually worth the effort:

LETTER	REAL QUESTION TO ASK
Feasible	Is there realistic access to enough subjects, data, time, and expertise to actually answer it?
Interesting	Would the answer genuinely engage you and the relevant field, not just fill a requirement?
Novel	Does it add something not already well-established?
Ethical	Could it be studied in a way a real ethics review would actually approve?
Relevant	Does the answer matter to real knowledge, decisions, or future research?

A DIAGNOSTIC TOOL, NOT A FORMULA

A question rarely scores perfectly on all five criteria on a first draft — FINER is meant to be applied honestly to reveal where a question needs revision, not treated as a pass/fail gate to satisfy after the fact.

From Topic to Question, Step by Step

STAGE	EXAMPLE
Vague topic	"Is social media bad for teenagers?"
After PICO	"Among teenagers aged 13-17, is daily social media use exceeding 3 hours associated with higher self-reported anxiety symptoms, compared to daily use under 1 hour?"
After FINER review	Checked for feasible access to a real teenage sample, genuine novelty against existing research, and an ethically sound design before moving forward

Hands-On Exercises

EXERCISE 1

Take the vague topic "Does exercise improve mental health?" and apply PICO to turn it into a single, specific, answerable question. Write out each of the four PICO components explicitly before assembling the final question.

View solution

EXERCISE 2

A student proposes the question: "Does drinking exactly 2.3 liters of water per day improve memory in left-handed adults aged 34-36 who live in cities with a population between 200,000 and 210,000?" The question is extremely specific. Using the FINER checklist, explain which criterion (or criteria) this question likely fails, and why extreme specificity alone doesn't make a question a good one.

View solution

EXERCISE 3

PICO was built for clinical medicine, where there's always a clear "intervention" and "comparison." Pick a non-medical topic you're personally interested in and explain, in your own words, how you'd adapt the Intervention and Comparison components to a question that has no obvious "treatment" being tested.

View solution

CHAPTER 2 QUICK REFERENCE

- PICO (Population, Intervention, Comparison, Outcome) first appears in Richardson et al., 1995, ACP Journal Club
- PICO turns a vague topic into a specific, measurable question by naming who, what, compared to what, and measured how
- FINER (Feasible, Interesting, Novel, Ethical, Relevant), from Hulley et al.'s Designing Clinical Research, checks whether a well-formed question is actually worth pursuing

- A question can be perfectly specific under PICO and still fail FINER — the two tools check different things

CHAPTER 3: PRIMARY VS. SECONDARY SOURCES

Research Methodology

Chapter 3 · Primary vs. Secondary Sources

A source's own type doesn't just describe where a fact came from — it determines what claim that source can actually support. This chapter covers the real taxonomy, a genuinely common point of confusion within it, and a real, documented case where getting this distinction wrong let a deliberately flawed study reach millions of readers.

A Real, Cross-Disciplinary Taxonomy

- **Primary source:** original, firsthand material — raw data, an original research article reporting new findings, a historical document, an eyewitness account, an artifact, an interview transcript.
- **Secondary source:** analysis or interpretation built on primary sources — a textbook chapter, a review article synthesizing many studies, a news article reporting on someone else's research.
- **Tertiary source:** a compilation or summary of secondary sources — an encyclopedia entry, a bibliography, many general reference works.

DISCIPLINE	PRIMARY	SECONDARY
History	An original letter, diary, or artifact from the period	A historian's book analyzing that period
Science	A journal article reporting new, original experimental data	A review article or meta-analysis synthesizing many such studies
Journalism	An eyewitness account, an original interview, a press release	A news article reporting on someone else's original reporting

The Same Kind of Article Can Be Either

A genuinely common point of confusion: "peer-reviewed journal article" isn't itself a source type. An original research article reporting new data is a **primary** source. A review article or

meta-analysis, even in the exact same journal, synthesizing dozens of other studies without collecting any new data of its own, is a **secondary** source. Confusing the two matters — a review article can only ever be as reliable as the primary studies it draws on, and citing it as though it were itself original evidence skips a real, necessary evaluation step.

A Real Case Where the Distinction Mattered

A DELIBERATELY FLAWED PRIMARY SOURCE, 2015

Journalist John Bohannon, working with a German documentary crew investigating junk science in diet journalism, deliberately designed and published a real study — "Chocolate with High Cocoa Content as a Weight Loss Accelerator" — in the *International Archives of Medicine*. It measured 15 participants across 18 different variables. Bohannon's own later explanation: "If you measure a large number of things about a small number of people, you are almost guaranteed to get a statistically significant result" — a real, deliberate demonstration of p-hacking.

The published article was, technically, a genuine primary source — it reported original data from a real study. But it was a deliberately unreliable one. That distinction matters: being a primary source describes where a claim came from, not whether the claim is trustworthy.

THE SECONDARY LAYER ADDED ITS OWN DISTORTION

News outlets — including The Daily Star, Cosmopolitan's German edition, and the German and Indian editions of The Huffington Post — picked up the study and ran headlines like "Slim by Chocolate" and "Why You Must Eat Chocolate Daily." None of the secondary coverage that reached millions of readers mentioned the tiny sample size or questioned the finding before publishing. The article was retracted on 10 June 2015, shortly after Bohannon revealed the hoax himself.

THE REAL, PRACTICAL LESSON

Two separate failures happened here, at two separate source layers: a flawed primary source (the study itself), and secondary coverage that repeated its conclusion without evaluating it. Tracing a striking claim back to its actual primary source is necessary — but not sufficient on its own. The primary source itself still needs evaluating, which is exactly what Chapter 4 covers next.

Hands-On Exercises

EXERCISE 1

A friend tells you "I read a peer-reviewed journal article, so it must be a primary source." Explain, in your own words, why this reasoning is incomplete, using the distinction between an original research article and a review article.

View solution

EXERCISE 2

In the Bohannon chocolate case, identify the primary source, the secondary source(s), and explain what specific claim each one could and couldn't actually support, even before the hoax was revealed.

View solution

EXERCISE 3

Choose a topic outside science or history (e.g. a piece of technology, a sport, a piece of music). Name one genuine primary source and one genuine secondary source you could actually consult for that topic, and explain why each one fits its category.

View solution

CHAPTER 3 QUICK REFERENCE

- Primary sources are original/firsthand; secondary sources analyze or interpret primary sources; tertiary sources compile secondary sources
- An original research article is primary; a review article or meta-analysis, even in the same journal, is secondary
- John Bohannon's real 2015 chocolate-weight-loss study (15 participants, 18 measured variables) was a genuine, deliberately p-hacked primary source
- Real media coverage ("Slim by Chocolate") added a further, uncritical secondary layer of distortion before the hoax was revealed and the paper retracted

- Being a primary source describes origin, not reliability — the source still needs to be evaluated, which Chapter 4 covers next

CHAPTER 4: EVALUATING SOURCE CREDIBILITY

Research Methodology

Chapter 4 · Evaluating Source Credibility

Chapter 3 ended on a real gap: knowing a source is primary doesn't mean it's trustworthy. This chapter covers a real, established framework for closing that gap, how peer review actually works (and where it genuinely fails), and a documented case of an entire tier of "peer-reviewed" journals that will publish almost anything for a fee.

The CRAAP Test

REAL, DOCUMENTED ORIGIN

Librarian Sarah Blakeslee coined the CRAAP acronym in 2004 while developing a workshop for first-year students at California State University, Chico. It's since spread well beyond that one campus into information-literacy instruction worldwide.

Currency

How recent is it, and does the topic actually need current information?

Relevance

Does it actually address your specific question, at an appropriate depth?

Authority

Who wrote or published it, and what are their real, checkable qualifications?

Accuracy

Is it supported by evidence, and can it be verified elsewhere?

Purpose

Why does this source exist — to inform, persuade, sell, or entertain?

How Peer Review Actually Works

A submitted paper is first screened by a journal editor, then sent to independent experts in the same field who evaluate its methods, evidence, and conclusions before publication is approved. It's a real, genuinely useful filter — but Chapter 1's own Wakefield case is a direct reminder that

peer review checks whether a paper's stated methods are internally consistent, not whether its underlying data or funding sources are honest. A reviewer generally can't independently re-verify a researcher's own raw records.

Predatory Journals: A Real, Documented Trap

University of Colorado Denver librarian Jeffrey Beall coined the term "predatory publishing" and maintained a real, widely used list of predatory open-access publishers from 2010 until he removed it in January 2017, following legal threats and formal complaints from publishers named on it. These are outlets that charge authors a publication fee while providing minimal or no genuine peer review.

A REAL, MEASURED STING — SAME JOURNALIST AS CHAPTER 3

In October 2013, journalist John Bohannon — the same reporter behind the 2015 chocolate hoax in Chapter 3 — published "Who's Afraid of Peer Review?" in *Science*. He submitted a deliberately flawed fake paper, claiming a lichen extract as a cancer cure, to 304 open-access journals. **157 accepted it**, 98 rejected it, and the rest were still under review or unresponsive. Whenever a journal accepted the paper, Bohannon withdrew it before it entered the real scientific record.

AN HONEST LIMITATION OF THE STING

Bohannon's own study submitted the fake paper only to open-access journals, with no subscription-journal control group — a real, documented critique of the sting's design. It demonstrates that predatory open-access journals accept fatally flawed papers at an alarming real rate; it doesn't, on its own, prove subscription journals would have done any better.

Legitimate vs. Predatory: Real Warning Signs

LEGITIMATE JOURNAL	PREDATORY JOURNAL
Editorial board is real, named, and reachable	Editorial board is fake, incomplete, or unresponsive
Fees, if any, are disclosed clearly upfront	Fees appear only after acceptance
Indexed in recognized real databases	Claims impressive-sounding but unverifiable indexing

LEGITIMATE JOURNAL	PREDATORY JOURNAL
Genuine, substantive peer review	Acceptance within days, little or no real feedback

Hands-On Exercises

EXERCISE 1

Apply all five CRAAP criteria to a specific article you've recently read online (news, blog, or otherwise). For each criterion, write one sentence explaining whether the source passes or fails it, and why.

View solution

EXERCISE 2

A classmate argues that Bohannon's 2013 sting proves peer review is worthless everywhere. Using the chapter's own "honest limitation" note, explain why this conclusion overreaches what the real study actually showed.

View solution

EXERCISE 3

Using the "Real Warning Signs" table, list three specific, checkable things you could look for on a journal's own website within five minutes to judge whether it's likely predatory.

View solution

CHAPTER 4 QUICK REFERENCE

- The CRAAP test (Currency, Relevance, Authority, Accuracy, Purpose) was coined by Sarah Blakeslee at CSU Chico in 2004
- Peer review checks internal consistency, not underlying data honesty — it did not catch Wakefield's fraud from Chapter 1

- Jeffrey Beall coined "predatory publishing" and maintained a real list of predatory journals from 2010 to 2017
- John Bohannon's real 2013 Science sting found 157 of 304 open-access journals accepted a deliberately flawed fake cancer-cure paper
- The sting's own honest limitation: no subscription-journal control group, so it can't directly compare predatory and legitimate journals' acceptance rates

CHAPTER 5: LITERATURE REVIEWS: SYNTHESIZING WITHOUT CHERRY-PICKING

Research Methodology

Chapter 5 · Literature Reviews: Synthesizing Without Cherry-Picking

A literature review's whole job is to summarize what a body of evidence actually shows — not what a writer already believed before reading it. This chapter covers the real structural difference between a loosely organized review and a rigorously protocoled one, a well-documented reason evidence itself skews toward positive findings, and a real, severe case of what happens when a review leaves out the studies that don't fit.

Narrative vs. Systematic Review

NARRATIVE REVIEW	SYSTEMATIC REVIEW
Sources chosen at the author's own discretion	Sources found via an explicit, pre-defined, reproducible search protocol
No fixed inclusion/exclusion criteria	Inclusion/exclusion criteria set and published in advance
Genuinely useful for broad context and framing	Genuinely stronger for answering one specific, well-defined question
Real risk: selection bias toward the author's own expectations	Real strength: another researcher can, in principle, reproduce the same search

PRISMA: A Real, Structured Protocol

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, first published in 2009 and substantially updated in 2020, gives systematic reviewers a real, standardized checklist and a flow diagram tracking every record from initial search through screening, eligibility review, and final inclusion — making the whole selection process auditable rather than hidden inside a single author's own judgment.

The File Drawer Problem

A REAL, DOCUMENTED BIAS IN THE EVIDENCE ITSELF

Psychologist Robert Rosenthal's 1979 paper "The file drawer problem and tolerance for null results," published in *Psychological Bulletin*, named a real, structural bias: studies finding a significant effect are far more likely to be submitted and accepted for publication than studies finding no effect. Rosenthal's own description: journals fill up with the studies that found something, while the file drawers fill up with the ones that didn't.

This means even a perfectly honest, exhaustive review of published literature can still be systematically skewed — not because anyone cherry-picked, but because the underlying pool of published studies was already tilted before the review ever began.

A Real Case Where Selective Reporting Caused Harm

In 2001, "Study 329" — a clinical trial of the antidepressant paroxetine in adolescents, funded by SmithKline Beecham (now GlaxoSmithKline) — was published reporting the drug was both effective and safe for treating adolescent depression.

THE REAL 2015 REANALYSIS REVERSED BOTH FINDINGS

In 2015, the RIAT (Restoring Invisible and Abandoned Trials) initiative — formed specifically to address selective reporting in clinical trials — published a full reanalysis of the original trial data in *BMJ*. It found paroxetine was **no more effective than placebo**, the opposite of the original paper's own claim — and that the paroxetine group had more than twice as many severe adverse events and four times as many psychiatric adverse events, including suicidal behaviors and self-harm, compared to placebo.

THE REAL, PRACTICAL LESSON

A literature review built only on the original 2001 publication would have synthesized confidently wrong conclusions on both efficacy and safety — not from any error in the review's own writing, but because the underlying source it relied on had itself selectively reported its results. Honest synthesis means actively seeking out reanalyses, raw data, and disconfirming studies, not just the version of a finding that was published first or most widely cited.

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why a systematic review's own explicit, published inclusion/exclusion criteria make it more resistant to cherry-picking than a narrative review — even though both review types could technically cover the exact same set of studies.

View solution

EXERCISE 2

Explain, in your own words, why the file drawer problem can bias a literature review's conclusions even if the reviewer never intentionally cherry-picked anything and read every single published study they could find.

View solution

EXERCISE 3

In the Study 329 case, the 2001 and 2015 papers analyzed the exact same underlying trial data, yet reached opposite conclusions. Explain what this tells you about the real difference between "a study was conducted" and "a study's own published conclusions can be fully trusted."

View solution

CHAPTER 5 QUICK REFERENCE

- Narrative reviews use author discretion; systematic reviews follow explicit, published inclusion/exclusion criteria
- PRISMA (2009, updated 2020) gives systematic reviewers a real, standardized checklist and flow diagram for transparency
- Rosenthal's real 1979 "file drawer problem" describes how published literature skews toward positive results, even without deliberate cherry-picking
- GSK's Study 329 (2001) claimed paroxetine was effective and safe for adolescents; a real 2015 RIAT reanalysis of the same data found neither was true

- Honest synthesis requires actively seeking reanalyses and disconfirming evidence, not just the first or most-cited published version of a finding

CHAPTER 6: QUALITATIVE VS. QUANTITATIVE METHODS

Research Methodology

Chapter 6 · Qualitative vs. Quantitative Methods

Choosing a research method isn't about which one is more rigorous — it's about which one actually answers the question being asked. This chapter covers the real distinction between quantitative and qualitative approaches, and a genuinely striking real case where a company had both kinds of evidence available and trusted the wrong one.

Two Different Kinds of Evidence

QUANTITATIVE	QUALITATIVE
Numeric data, statistical analysis	Words, observations, images — non-numeric data
Larger samples, structured instruments (surveys, experiments)	Smaller samples, less structured (interviews, focus groups, case studies)
Fits "how much," "how many," "is there a correlation"	Fits "why," "how do people experience this"
Genuinely strong for measuring the size of an effect	Genuinely strong for understanding the meaning behind it

Mixed Methods: A Real, Established Third Category

Researchers John Creswell and Vicki Plano Clark's real, widely used textbook *Designing and Conducting Mixed Methods Research* lays out a third real approach: deliberately combining both types within one study — using qualitative work to explain a quantitative pattern, or quantitative data to test a pattern a qualitative study first surfaced. Neither method is treated as inherently superior; each is treated as answering a genuinely different part of the same question.

A Real Case Where the Wrong Evidence Was Trusted

REAL DATA: 200,000 TESTS

Before launching "New Coke" in 1985, Coca-Cola conducted over 200,000 blind taste tests, surveys, and focus groups. The core quantitative result: roughly 53% of tested consumers preferred the new formula's taste over the original, in a blind comparison. On that basis, the company confidently discontinued the original formula entirely.

THE QUALITATIVE SIGNAL WAS THERE — AND WAS OVERRIDDEN

Coca-Cola's own focus groups — qualitative research, not quantitative — had genuinely signaled that switching the formula entirely would provoke widespread dissatisfaction, separate from how the new taste scored in a blind sip test. The quantitative preference numbers were trusted over that qualitative warning. The blind taste test measured a narrow question — "which tastes better in an isolated sip" — it never asked how people would feel about losing the original Coca-Cola altogether, a question only the qualitative research had actually been positioned to answer.

The backlash was real and immediate. Seventy-nine days after launch, on 11 July 1985, Coca-Cola announced the return of the original formula as "Coca-Cola Classic."

THE REAL, PRACTICAL LESSON

This wasn't a case of skipping qualitative research — Coca-Cola had it. The failure was treating quantitative data as automatically more authoritative when the two methods produced genuinely conflicting signals, rather than recognizing that the qualitative research was answering the more relevant question for the actual decision being made: not "which formula tastes better," but "how attached are people to the existing brand."

Hands-On Exercises

EXERCISE 1

A company wants to know both "what percentage of our users are unhappy with a new feature" and "why the unhappy users feel that way." Explain which method (quantitative, qualitative, or both) fits each of these two sub-questions, and why.

View solution

EXERCISE 2

In the New Coke case, explain specifically what real-world question the blind taste test's 53%-vs-47% result could and couldn't answer, and why that gap mattered for the actual business decision being made.

View solution

EXERCISE 3

A classmate argues quantitative data is always more objective and trustworthy than qualitative data, since it involves numbers. Using the New Coke case, explain why "more numeric" doesn't automatically mean "answers the actual question better."

View solution

CHAPTER 6 QUICK REFERENCE

- Quantitative methods measure size/frequency with numeric data; qualitative methods explore meaning and experience with non-numeric data
- Mixed methods (Creswell & Plano Clark) deliberately combine both within one study, treating them as complementary rather than competing
- Coca-Cola's real 1985 New Coke launch relied on a 200,000-test quantitative taste preference (~53%/47%) that genuinely favored the new formula
- Real focus-group (qualitative) data had already signaled a loyalty backlash risk — a different, more relevant question than raw taste preference
- The company reverted to the original formula as "Coca-Cola Classic" just 79 days after launch

CHAPTER 7: SAMPLING & BIAS

Research Methodology

Chapter 7 · Sampling & Bias

A larger sample isn't automatically a better one. This chapter covers real sampling methods, the most famous real polling failure in the history of statistics — and an honest reanalysis decades later showing even that famous story was more complicated than it's usually told.

Probability vs. Non-Probability Sampling

PROBABILITY SAMPLING	NON-PROBABILITY SAMPLING
Every member of the population has a known chance of selection	Selection isn't random — some members have no real chance of being included
Simple random, stratified, cluster, systematic	Convenience, quota, snowball
Genuinely supports statistical generalization to the full population	Faster and cheaper, but real generalization is genuinely limited

A Real, Canonical Case of Selection Bias

1936: THE LITERARY DIGEST'S REAL PREDICTION

The magazine *Literary Digest* mailed roughly 10 million ballots — drawn from telephone directories, car registration lists, and its own subscriber rolls — and received about 2.3 million responses. It confidently predicted Alf Landon would beat Franklin Roosevelt, 57% to 43%. Roosevelt actually won, 62% to 38%.

Phones and cars were luxury items during the Great Depression, so the sample frame itself skewed toward wealthier respondents — a group genuinely less likely to support Roosevelt's New Deal policies than Landon's platform. George Gallup, using a real, much smaller but properly constructed sample, correctly predicted a Roosevelt win — a result now treated as the founding moment of modern scientific polling.

AN HONEST COMPLICATION — THE POPULAR STORY ISN'T THE WHOLE STORY

A real 1988 reanalysis by political scientist Peverill Squire found that sample skew alone doesn't fully explain the error — even accounting for who was mailed a ballot, the poll would still have predicted Roosevelt correctly if everyone who received one had actually responded. The real, larger factor was non-response bias: Landon supporters returned their ballots at a meaningfully higher rate than Roosevelt supporters. The famous "the sample frame was rigged" story is real, but the size of its actual effect is smaller than commonly told — non-response was the bigger real driver.

Confirmation Bias

Psychologist Peter Wason's 1960 study introduced the "2-4-6 task": participants were told a rule governed which number triples were valid, shown one valid example, and asked to discover the rule by proposing their own triples. Most participants formed an overly narrow hypothesis and then only tested triples that would confirm it, rarely trying an example designed to prove themselves wrong. Wason coined the term **confirmation bias** for this real, well-documented tendency to favor evidence that supports an existing belief over evidence that challenges it.

THE REAL, PRACTICAL LESSON

Sampling bias and confirmation bias are two different failure points that reinforce each other. A biased sample can produce a confidently wrong result on its own — but a researcher who's already confident in a conclusion is also less likely to go looking for the kind of disconfirming case (a Gallup-style smaller, representative sample; a genuinely skeptical reanalysis like Squire's) that would catch the error. Both require the same real discipline: actively seeking out the evidence that could prove you wrong, not just the evidence that agrees with you.

Hands-On Exercises

EXERCISE 1

A researcher wants to survey "public opinion on a new city park proposal" by standing outside the park itself and asking passersby. Identify what kind of sampling this is, and explain, in your own words, why the resulting sample likely can't represent the opinions of city residents who never visit the park.

View solution

EXERCISE 2

Explain, in your own words, why Squire's 1988 reanalysis is a genuinely important addition to the Literary Digest story, rather than just a minor footnote — specifically, what does it change about which lesson a researcher should actually take from the case?

View solution

EXERCISE 3

Describe a real or hypothetical situation from your own life or work where you noticed yourself only looking for information that confirmed something you already believed. Explain what a genuine, confirmation-bias-resistant version of that same search would have looked like.

View solution

CHAPTER 7 QUICK REFERENCE

- Probability sampling gives every population member a known chance of selection; non-probability sampling doesn't, limiting real generalization
- The 1936 Literary Digest poll (10M mailed, 2.3M returned) wrongly predicted Landon 57%-43%; Roosevelt actually won 62%-38%
- George Gallup's smaller, properly constructed sample correctly predicted the real result — the founding moment of modern scientific polling
- Squire's real 1988 reanalysis found non-response bias, not just sample-frame skew, was the larger real driver of the error
- Peter Wason's 1960 study coined "confirmation bias" — the tendency to seek confirming rather than disconfirming evidence

CHAPTER 8: DATA COLLECTION METHODS

Research Methodology

Chapter 8 · Data Collection Methods

Every data collection method distorts what it measures in some real, characteristic way. This chapter covers four common methods and two real, well-documented cases — one showing how people misreport their own behavior, and one showing how a famous "textbook fact" about observation itself turned out to be far shakier than its own reputation.

Surveys

Scalable, structured — but relies entirely on honest, accurate self-report

Interviews

Genuine depth and follow-up — but small samples and real interviewer-effect risk

Experiments

Strongest real causal claims — but real observation itself can change behavior

Observational Studies

Real, naturalistic behavior — but no controlled comparison, and real observer bias risk

Surveys: A Real Social-Desirability Case

A REAL, DIRECT HEADCOUNT VS. SELF-REPORT

Sociologists C. Kirk Hadaway, Penny Long Marler, and Mark Chaves, in a 1993 *American Sociological Review* study, physically located and counted attendance at every church in a rural Ohio county, comparing it directly against self-reported church attendance from public opinion surveys. The real headcount came out to roughly **half** the attendance rate people reported on surveys.

The authors attributed most of the gap to social desirability bias — people overreporting behavior they believe is socially valued. Later researchers debated exactly how much of the gap comes from overreporting versus incomplete headcounts and other measurement issues — a real, honest reminder that even a well-designed comparison study can leave its own precise mechanism genuinely disputed, even when the headline gap itself is well established.

Experiments and the Real Hawthorne Effect Story

Between 1924 and 1932, researchers at Western Electric's Hawthorne Works — later reinterpreted by a Harvard team led by Elton Mayo — varied factory lighting and other conditions, and reported that worker productivity rose regardless of the change, supposedly because workers were changing their behavior simply from knowing they were being observed. That story became the textbook "Hawthorne effect."

A REAL, DOCUMENTED REANALYSIS

Economists Steven Levitt and John List (2011), building on an earlier reanalysis by Jones (1992), went back to the actual original data and found the dramatic "output rose no matter what changed" pattern largely collapses once real day-of-week patterns and pay-period incentives are properly controlled for. The Hawthorne effect isn't fully debunked — but the version repeated in most textbooks is a genuinely weaker piece of evidence than its own decades-long reputation suggests.

Four Methods, Compared

METHOD	REAL STRENGTH	REAL WEAKNESS
Survey	Scales to large samples cheaply	Relies on accurate self-report — real social desirability bias risk
Interview	Real depth, follow-up questions possible	Small samples, real interviewer-effect risk
Experiment	Strongest support for real causal claims	Observation itself can change the behavior being measured
Observational study	Captures real, naturalistic behavior	No controlled comparison; real observer bias risk

THE REAL, PRACTICAL LESSON

Every method distorts its own data in a specific, predictable direction — surveys toward social desirability, experiments toward observation effects. Choosing a method well means anticipating its own characteristic distortion in advance, not discovering it after the data's already collected.

Hands-On Exercises

EXERCISE 1

A company wants to know how often employees actually take real lunch breaks (versus working through them). Explain why a self-report survey on this specific topic carries a real social-desirability risk, using the church attendance case as a model, and propose one alternative or supplementary method that would reduce that risk.

View solution

EXERCISE 2

Explain, in your own words, what specifically the Levitt and List reanalysis changed about the real, appropriate lesson to draw from the Hawthorne studies — not "the Hawthorne effect never happened," but something more precise.

View solution

EXERCISE 3

You need to know both "what percentage of customers are satisfied" and "what specifically frustrates the dissatisfied ones." Choose one method from this chapter for each sub-question, and explain why a single method couldn't answer both well.

View solution

CHAPTER 8 QUICK REFERENCE

- Surveys, interviews, experiments, and observational studies each distort data in their own characteristic way
- Hadaway, Marler & Chaves' real 1993 study found self-reported church attendance roughly double the real, physically counted rate
- The exact cause of that gap (overreporting vs. measurement issues) remains genuinely debated, even though the headline gap itself is well established

- The classic 1924-1932 Hawthorne studies' "observation alone boosts output" story is real but, per Levitt & List (2011) and Jones (1992), weaker evidence than its textbook reputation suggests
- Choosing a data collection method well means anticipating its own predictable distortion in advance

CHAPTER 9: CITATION, SYNTHESIS & AVOIDING PLAGIARISM

Research Methodology

Chapter 9 · Citation, Synthesis & Avoiding Plagiarism

Citation isn't a formatting requirement bolted onto research — it's what actually lets a reader trace a claim back to Chapter 3's own primary source. This chapter covers real citation conventions, a genuinely common misconception about paraphrasing, a real named gray zone between honest paraphrase and plagiarism, and a real, high-stakes case of what happens when citation is skipped at scale.

Citation Styles: A Real, Brief Overview

STYLE	COMMONLY USED IN
APA	Sciences and social sciences
MLA	Humanities, literature
Chicago	History, and much of trade and academic publishing

The specific formatting rules differ, but every style solves the same underlying problem: letting a reader locate the exact source a claim came from, without having to take the writer's word for it.

Quoting vs. Paraphrasing — A Real, Common Misconception

Quoting reproduces a source's exact wording, in quotation marks, when the wording itself matters. Paraphrasing restates an idea in your own words to integrate it into your own analysis. A genuinely common misconception: paraphrasing is often assumed to remove the need for a citation, since the exact words changed. It doesn't — the idea still came from somewhere else, and a paraphrase without a citation still claims someone else's idea as though it were your own.

Patchwriting: A Real, Named Gray Zone

A REAL, COINED TERM — REBECCA MOORE HOWARD, 1993

Writing researcher Rebecca Moore Howard coined "patchwriting" to describe a specific, real pattern: "copying from a source text and then deleting some words, altering grammatical structures, or plugging in one-for-one synonym substitutes." It keeps the source's own original structure and much of its wording while technically avoiding a direct quote — and it still counts as plagiarism, even when it's unintentional, and even when (as Howard herself argued) it often reflects a genuine, developmental struggle to engage with unfamiliar material rather than deliberate dishonesty.

A Real, High-Stakes Case

In February 2011, anonymous online reviewers began comparing German Defense Minister Karl-Theodor zu Guttenberg's 2007 doctoral thesis against its cited (and uncited) sources, coordinating their findings on a crowd-sourced wiki, VroniPlag. The investigation found more than 400 separate instances of plagiarism — ultimately, more than half of the 475-page thesis contained large sections copied without proper attribution.

THE REAL, DOCUMENTED CONSEQUENCES

On 23 February 2011, the University of Bayreuth rescinded Guttenberg's doctorate. He resigned as Germany's Defense Minister on 1 March, and from the Bundestag on 3 March — a real, rapid collapse of a prominent political career, triggered by a crowd-sourced citation audit rather than any single formal academic review process.

THE REAL, PRACTICAL LESSON

The Guttenberg case wasn't caught by an institution's own internal review — it was caught by outside readers doing exactly what Chapter 3 described: tracing specific claims back to their real, original sources and checking whether the sources actually said what the thesis claimed they said. Honest citation is what makes that kind of check possible in the first place.

Hands-On Exercises

EXERCISE 1

A classmate says, "I didn't plagiarize — I changed the wording, so it's my own writing now." Using the chapter's own definition of patchwriting, explain why this reasoning is incomplete.

View solution

EXERCISE 2

Explain, in your own words, why the Guttenberg case being uncovered by an anonymous, crowd-sourced wiki — rather than his university's own formal review process — matters for how seriously the case should be taken as a research-integrity lesson.

View solution

EXERCISE 3

Take any single sentence from an earlier chapter of this course and write both a legitimate paraphrase (with citation) and a patchwritten version (without citation, following Howard's own definition) of the same idea. Explain what makes the two different.

View solution

CHAPTER 9 QUICK REFERENCE

- APA, MLA, and Chicago are real, discipline-specific citation styles that all solve the same problem: letting a reader trace a claim to its source
- Paraphrasing still requires a citation — changing the wording doesn't remove the need to credit the original idea
- Rebecca Moore Howard's real 1993-coined "patchwriting" describes altering a source's wording while keeping its structure, without proper citation — still plagiarism, even when unintentional
- Karl-Theodor zu Guttenberg's real 2011 scandal found more than half of his 475-page thesis plagiarized, uncovered by the crowd-sourced VroniPlag wiki

- His doctorate was rescinded and he resigned as Defense Minister within days of the findings becoming public

CHAPTER 10: CAPSTONE — CONDUCTING A REAL RESEARCH PROJECT

Research Methodology

Chapter 10 · Capstone: Conducting a Real Research Project

This capstone runs one real, chosen question through every tool this course has covered, in order — **"Does caffeine improve endurance exercise performance?"** — grounded in real, verified evidence rather than an invented example.

STEP 1 — CHAPTER 1: AVOIDING THE CASUAL-SEARCHING TRAP

Starting point: rather than searching for "does caffeine help exercise" and stopping at the first confident-sounding article, this project starts from the question itself and follows real evidence, actively looking for disagreement rather than only confirmation — the exact distinction Chapter 1 drew between casual searching and genuine research.

STEP 2 — CHAPTER 2: FRAMING WITH PICO

P: adults engaging in endurance exercise · **I:** pre-exercise caffeine ingestion · **C:** a placebo or no caffeine · **O:** measured endurance performance (time to exhaustion, performance time). Assembled: "Among adults performing endurance exercise, does pre-exercise caffeine ingestion improve measured performance compared to a placebo?" — genuinely feasible, novel enough to be worth checking personally, ethical, and relevant (FINER).

STEP 3 — CHAPTER 3: IDENTIFYING SOURCE TYPES

The real evidence located is a meta-analysis of 21 randomized controlled trials on caffeine and endurance running performance. Per Chapter 3, the meta-analysis itself is a **secondary** source — it synthesizes 21 **primary** sources (the individual RCTs) rather than collecting new data of its own.

STEP 4 — CHAPTER 4: EVALUATING CREDIBILITY

Applying CRAAP: Currency (a recent meta-analysis, appropriate for an active research area); Relevance (directly addresses the PICO question above); Authority (published, peer-reviewed sports science research); Accuracy (reports a specific, checkable effect size and p-value rather than a vague claim); Purpose (informing, not selling a supplement) — the source passes on all five.

STEP 5 — CHAPTER 5: A SYSTEMATIC REVIEW, NOT A NARRATIVE ONE

Because this source is a systematic review of 21 RCTs with explicit inclusion criteria — not one author's own curated selection — Chapter 5's own narrative-vs-systematic distinction gives real reason to trust it more than a single cherry-picked study or a narrative blog summary would deserve.

THE REAL, VERIFIED FINDING

The meta-analysis found caffeine improved time to exhaustion with a medium effect size ($g = 0.392$, $p < 0.001$) across the pooled RCTs, at doses between 3 and 9 mg/kg — a real, statistically significant result.

STEP 6 — CHAPTER 6: QUANTITATIVE FITS THIS QUESTION

This question is inherently quantitative — it asks "how much," measured in effect size and percentage improvement. A qualitative study (e.g. interviewing athletes about perceived exertion) could meaningfully complement this finding, but couldn't answer the core PICO question on its own — matching Chapter 6's own principle of matching method to question.

STEP 7 — CHAPTER 7: A REAL, HONEST SAMPLING COMPLICATION

The same real evidence also reports substantial individual variability: a mean performance improvement of 3.2% ± 4.3%, ranging from -0.3% to as high as 17.3% across different studies and participants — some people benefit far more than others, and a small number see close to no benefit at all. Reporting only the average effect size, without this real variability, would risk the same kind of

overgeneralization Chapter 7 warned against with the Literary Digest case — a real, aggregate result doesn't mean the effect applies identically to every individual.

STEP 8 — CHAPTER 8: WHY RCTS ARE THE RIGHT METHOD HERE

The 21 underlying studies are randomized controlled trials — real experiments, per Chapter 8's own comparison — genuinely well suited to this question specifically because it's a causal one ("does X cause a change in Y"), the exact real strength experiments have over surveys or observational studies.

STEP 9 — CHAPTER 9: A PROPERLY CITED SYNTHESIS

Final synthesized answer: Based on a systematic review and meta-analysis of 21 randomized controlled trials, pre-exercise caffeine ingestion (3-9 mg/kg) is associated with a real, statistically significant improvement in endurance time to exhaustion (effect size $g = 0.392$, $p < 0.001$; mean improvement 3.2% ± 4.3%). The effect is genuine but shows substantial individual variability (range: -0.3% to 17.3%), meaning the real, honest answer is "caffeine improves average endurance performance in controlled trials, with a meaningfully different effect size for different individuals" — not a uniform guarantee for any one person.

Chapter-Attribution Table

STEP	CHAPTER	TOOL APPLIED
1	Ch. 1	Methodology-first mindset over casual searching
2	Ch. 2	PICO question framing + FINER check
3	Ch. 3	Primary vs. secondary source identification
4	Ch. 4	CRAAP credibility evaluation
5	Ch. 5	Systematic vs. narrative review distinction
6	Ch. 6	Matching quantitative method to a quantitative question

STEP	CHAPTER	TOOL APPLIED
7	Ch. 7	Reporting real variability, avoiding overgeneralization
8	Ch. 8	Recognizing RCTs as the appropriate method for a causal question
9	Ch. 9	A properly cited, honestly synthesized final answer

COURSE COMPLETE

Research Methodology closes at 10/10 chapters — the first course in the Study Methodologies Subject. Critical Thinking, Scientific Methodology, and Mathematical Proofs remain as reserved sibling ideas for this same Subject, not yet fleshed out.

CHAPTER 10 QUICK REFERENCE

- The capstone applied all nine prior chapters' own tools, in sequence, to one real question: does caffeine improve endurance performance?
- A real, verified 21-RCT meta-analysis found a statistically significant effect ($g = 0.392$, $p < 0.001$)
- Real, honest reporting includes the substantial individual variability (3.2% ± 4.3%, range -0.3% to 17.3%), not just the average
- This closes the Research Methodology course and opens the door to further Study Methodologies Subject courses