DL 901 — Research Methods I: Design and Measurement

The first course of the doctoral sequence. It answers the question every dissertation begins with: "What is worth studying here, and how would you study it?"

Level: doctoral · **Prerequisites:** admission to the program (the master's foundation is assumed) · **Companion:** DL 902 (statistics and causal inference) · **Credits:** 3 · **Format:** one 3-hour seminar per week, one term.

This course is the **design half** of the methods core. It teaches a doctoral student to move from a practitioner hunch ("I think HUBZone status matters more than we admit") to a defensible research design ("I will compare win rates of HUBZone-certified and non-certified firms on set-aside competitions, holding size, NAICS, and value constant, using award records from the public data environment"). DL 902 supplies the statistics that make the comparison defensible; this course supplies the design and the measurement.

Course description

Research in capture management, government business development, and business structuring is research about **decisions under rules**. Organizations decide what to pursue, how to score it, whom to team with, and what to bid — inside a highly published rule structure. The researcher's job is to study those decisions and their outcomes without fooling themselves about what the evidence shows.

This course covers the foundations that make that possible:

- the **philosophy of applied research** — what counts as knowledge in a practice discipline;
- the **research cycle** — from question to hypothesis to design to measurement to inference;
- the **catalogue of designs** — observational, quasi-experimental, experimental, and case designs — and which fit which question;
- **measurement** — turning the field's slippery constructs into measurable variables with evidence of validity and reliability;
- **sampling and the evidence environment** — the public data the field runs on, and the field's central sampling hazard (only pursued bids resolve);
- **ethics and provenance** — the rules for studying organizational decisions and public records.

Everything is taught **stack-agnostic** (doctrine/08): the course is about research design, and no lesson requires operating any product.

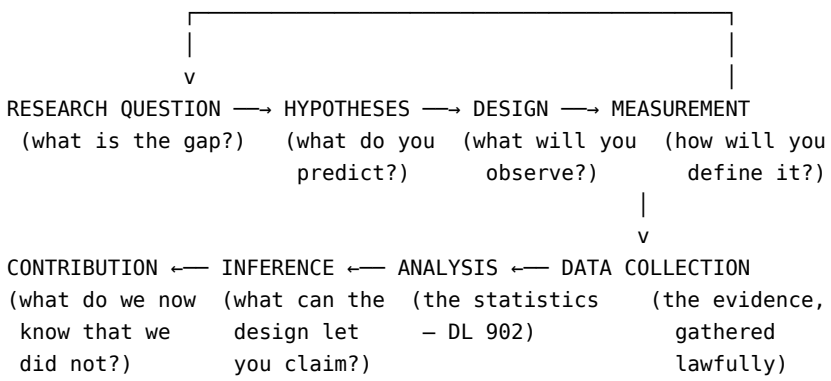
Learning outcomes

By the end of this course, a student can:

1. **Pose** a researchable question about the pursuit discipline that is specific enough to design for and important enough to be worth a dissertation.
2. **Select** an appropriate design for a given question, and defend why the design can (and cannot) support the intended inference.
3. **Operationalize** a discipline construct — probability of win, win rate, gate discipline, value, capture effort, portfolio posture — into a measurable variable, and argue its validity and reliability.
4. **Identify** the sampling and selection hazards that threaten a study of pursuit outcomes, and design mitigations for them.
5. **Write** the design half of a research proposal: question, hypotheses, design, measurement plan, sampling plan, and ethics plan.

The course's spine: the research cycle

Every design decision in this course traces back to one cycle. A researcher who can walk this cycle for their question can design almost any study:



The cycle is not one-way: measurement problems force design revisions, and data realities force question revisions. The skill of research is learning to move around the cycle without breaking it. Each unit below is one station on the cycle.

Unit 1 — What counts as knowledge in this field

The question. The discipline of capture is practiced by brilliant, experienced people who disagree about almost everything that matters. Why should their disagreement be settled by research rather than by seniority?

The ideas.

- The **two layers** of the doctrine (doctrine/08): durable concepts vs. swappable tools. A doctoral researcher studies the concept layer and treats the tool layer as an obstacle course, not the subject.
- **Positivist, interpretive, and critical** research postures. Positivist research asks "what is the effect of X on Y?" — the bulk of this program's methods. Interpretive research asks "how do the people in this world make sense of what they do?" — indispensable for understanding gate decisions that never get written down. Critical research asks "who benefits, and who is excluded?" — the equity lens the MPA tier already touches. A strong field needs all three; a strong dissertation usually commits to one and respects the others.
- The **young-field opportunity**. A field with little published research does not need incremental contributions; it needs the *first* rigorous studies of foundational questions. The bar for "contribution" is real but reachable: be the person who measures something the field has only argued about.

Seminar activity. Each student brings a practitioner claim they have heard ("the incumbent always wins," "HUBZone is overrated," "color-team reviews catch real errors") and we classify each claim as *testable, arguable-but-untestable, or not-yet-defined*. The exercise makes the difference between a research question and a dinner argument concrete.

Reading. doctrine/08; doctrine/05 (the score as arithmetic); the glossary entries for *pWin, win rate, gate, ORBITAL*.

Deliverable. A one-page "question sketch" — the student's candidate dissertation question, stated as a gap in knowledge, with the doctrine claim it would test.

Unit 2 — The research cycle and hypothesis formation

The question. Once you have a question, how do you make it *researchable* — specific enough that a design can answer it?

The ideas.

- **Descriptive vs. relational vs. causal questions.** *Descriptive:* "What share of small-business set-asides go to HUBZone firms?" *Relational:* "Do HUBZone-certified firms win set-asides at a higher rate than non-certified firms, all else equal?" *Causal:* "Does HUBZone certification *cause* a higher win rate, or do certified firms differ in other ways?" The design burden rises sharply across the three. Most dissertations in a young field are descriptive or relational at the core, with causal claims carefully bounded.
- **Hypotheses as falsifiable predictions.** A hypothesis is not a hope; it is a prediction that could be wrong, stated so the data can make it wrong. "HUBZone certification is associated with a higher win rate on HUBZone set-asides, holding firm size and NAICS constant" is falsifiable. "HUBZone matters" is not.

- **The distinction between the research question and the decision question.** Practitioners want to know "should we pursue this?" Researchers answer "what predicts winning?" The researcher's answer informs the practitioner's decision but does not replace it — the same division the doctrine draws between arithmetic and judgment (doctrine/05).

Seminar activity. Take the question sketches from Unit 1 and rewrite them as *falsifiable hypotheses*, each with a stated "what the data would have to show to make me abandon this."

Reading. doctrine/03 (the pipeline — the natural map of where research questions live).

Deliverable. A two-page hypothesis memo: the research question, the falsifiable hypothesis, and a sentence stating the *opposite* finding the student would be willing to publish.

Unit 3 — The design catalogue

The question. What designs exist, and which ones fit which questions in this field?

The ideas. A full graduate treatment of design families, each illustrated with a pursuit-discipline example:

- **Cross-sectional observational studies.** A snapshot across many firms or pursuits at one time. *Example:* comparing win rates across certification categories using a year of award records. Cheap, broad, weak on causality (no time order, unmeasured confounders).
- **Longitudinal and panel designs.** The same units followed over time. *Example:* a firm's portfolio of pursuits and outcomes followed over five years. Stronger for ordering and change; the panel is the natural shape of "does the organization learn?"
- **Quasi-experimental designs.** A treatment occurs, but assignment is not random; the researcher exploits the structure. *Examples:* a **difference-in-differences** study of win rates before vs. after a firm obtains HUBZone certification, compared against firms that never obtained it; an **interrupted time series** of an agency's award patterns before vs. after a change in its review criteria. These are the workhorses of the field's causal questions, because no one will randomize a firm's certification status for you.
- **Randomized and field experiments.** The gold standard when ethics and feasibility allow. *Example:* an experiment randomizing which of two proposal-drafting *processes* (not products) teams use, then measuring review scores — feasible in a teaching or training context, rarely feasible on live federal competitions. The seminar treats field experiments honestly: rare in this field, but powerful where they fit (e.g., in the design of training, templates, and review protocols).

- **Case study and comparative case designs.** Deep, contextual study of one or a few organizations or pursuits. *Examples:* a comparative case study of three firms' gate governance; a single-case longitudinal study of one pursuit from sense to award. The best designs in this family use *pattern matching* against rival explanations, not storytelling.
- **Mixed designs.** A quantitative core with a qualitative layer — e.g., a statistical analysis of win predictors *plus* interviews with the capture teams behind the outliers. The mixed design is unusually well suited to this field, because the quantitative record rarely explains *why* a gate decision happened.

Seminar activity. For each of four real research questions (provided by the instructor from the research agenda), teams pick a design family, then argue what the design can *and cannot* claim. The discipline of naming the limit is the point.

Reading. doctrine/03, doctrine/04 (gates as the field's natural "treatments" and "outcomes").

Deliverable. A one-page design-selection memo for the student's own question: the chosen design family, two alternative families, and why the chosen one fits better.

Unit 4 — Measurement: turning constructs into variables

The question. The field's key ideas — probability of win, gate discipline, win rate, value, capture effort — are real but slippery. How do you measure them so the measurement is defensible?

The ideas.

- **Constructs vs. variables.** A construct is the idea (probability of win). A variable is the operational definition (the composite score from a seven-factor rubric, normalized to [0,1]). The gap between them is measurement error, and measurement error is the quiet killer of dissertations.
- **Validity, the four faces of it:** *construct validity* (does the variable actually capture the idea?), *internal validity* (does the design support the causal claim?), *external validity* (does the finding travel beyond this sample?), and *statistical conclusion validity* (is the analysis sensitive enough to detect the effect?). A doctoral student should be able to name which of the four is the weakest point of their own design.
- **Reliability.** Would the same measurement procedure produce the same value twice? Inter-rater reliability (two coders scoring the same proposals), test-retest reliability (the same firm's score at two times), and the reproducibility of a measurement protocol. The field's scoring rubrics are reliability problems as much as validity problems: a pWin rubric that two capture managers apply to the same opportunity and get wildly different numbers is a measurement instrument in need of study.

- **A worked example: measuring "probability of win."** The discipline already models this as a **seven-factor composite** — past performance, technical capability, price competitiveness, incumbent/relationship advantage, team strength, customer knowledge, and strategic alignment — each scored on a scale and weighted into a 0-to-1 estimate (doctrine/05). For a researcher, that composite is a *measurement instrument*: what is its construct validity (does it predict awards?), its reliability (do two scorers agree?), and its calibration (does 0.7 mean 70%)? Those questions are dissertation-sized, and the empirical-design seminar (DL 903) builds on them directly.
- **Measuring the outcome side.** The field's ground-truth outcomes — won, lost, withdrawn — are clean in concept and messy in practice. "Withdrawn" is not "lost" but is often treated as a non-win. Award value is public and precise in the record. Win rate over a period is computable but sensitive to *which* pursuits enter the denominator — a measurement decision with big consequences.

Seminar activity. In teams, take "gate discipline" and produce two different operational definitions (e.g., "share of scored pursuits that were declined below threshold" vs. "existence of written gate criteria before pursuit start"). Discuss what each definition captures and misses. The exercise demonstrates that measurement choices *are* substantive choices.

Reading. doctrine/05 in full (the composite and the threshold as measurement objects).

Deliverable. A measurement memo for the student's own study: the construct, the operational variable, the source of the data, and a one-paragraph validity-and-reliability argument with the weakest face of validity named.

Unit 5 — Sampling, the evidence environment, and the survivorship hazard

The question. Where does the evidence come from, and what can quietly make it lie?

The ideas.

- **The public evidence environment.** The discipline runs on an extraordinary public record: solicitations published in advance; award records with dollar values and winners; agency program data; and the review criteria agencies publish for their competitions. This environment is the researcher's birthright — it is free, it is lawfully public, and it is the ground truth the whole field argues about. A doctoral student should know the shape of this environment (published awards, solicitation archives, agency data portals, congressional spending records) as well as a historian knows the archive.

- **The sampling frame problem — the field's central hazard.** Here is the hard fact that shapes every study of pursuit outcomes: **only pursued bids resolve.** You can observe what happened to the pursuits an organization *chose* to pursue; you cannot observe what would have happened to the opportunities it declined. The universe of possible pursuits is not the same as the sample of pursued bids. This is the **survivorship / selection problem**, and it is not a footnote — it is the design constraint that separates serious capture research from naive capture research.
- **Mitigations.** An **observation floor** (do not infer from a handful of outcomes); **stratification** (analyze within funder family or agency, where behavior is more homogeneous); **treating withdrawal as a non-win** by a stated policy, applied consistently; **modeling the selection decision** (why was this pursued?) where possible; and **bounded claims** — never generalizing from pursued bids to all opportunities.
- **Sample size and power at the design stage.** A study must be designed with enough expected terminal outcomes to detect the effect it cares about. Power analysis appears in DL 902; at the design stage, the student must at least ask "how many resolved pursuits will I plausibly observe, and is that enough to answer the question?"
- **The ethics of sampling.** Organizational outcome data is sensitive even when lawful. Public award records are fair game; internal decision logs, budgets, and deliberations require consent and confidentiality protocols (see Unit 6).

Seminar activity. Given a described study ("I will use the award record to measure whether color-team review improves win rates"), teams list every way the sample could be biased, then redesign the study to reduce the three worst biases.

Reading. literacy/where-the-money-flows.md (the public record as the evidence base); doctrine/03.

Deliverable. A sampling-plan memo: the target population, the sampling frame, the expected sample size, the three most serious selection hazards, and the mitigation for each.

Unit 6 — Ethics, provenance, and the integrity standard

The question. How do you study real organizations' decisions without harming them — and without corrupting the evidence?

The ideas.

- **The two data regimes.** *Public data* (award records, solicitations, published criteria) needs no consent, but demands **provenance discipline**: every record sourced, every extraction documented, every transformation logged, so another researcher can verify the data is what you say it is. *Organizational data* (decision logs, budgets, deliberations, win-probability

estimates at decision time) is sensitive even when not secret: it needs informed consent, confidentiality, anonymization where feasible, and — for a working professional studying their own firm — a clear boundary between the researcher hat and the practitioner hat.

- **The non-fabrication standard.** The field's own research corpora model the rule: documented facts and clearly-labeled illustrative material are kept rigorously apart, and no one attributes invented specifics to real organizations. A dissertation is held to the same standard. Fabricated awards, invented dollar figures, or misattributed data fail the milestone outright.
- **Institutional review and the IRB.** Studies involving human subjects — interviews, surveys, or analysis of identifiable organizational decision-makers — require institutional review. The course treats the IRB not as paperwork but as the field's own gate discipline applied to research: criteria written in advance, an owner, and a review.
- **Conflict of interest.** A researcher who works in the industry must disclose the interest and design so the finding cannot be an advertisement. The integrity of the field's young empirical base depends on this.

Seminar activity. Teams are given three study scenarios (one pure public-data, one interview-based, one a professional studying their own firm) and produce the ethics + provenance plan for each: consent, confidentiality, IRB route, provenance log.

Reading. doctrine/04 (gates and governance as the model for research governance).

Deliverable. An ethics-and-provenance section for the student's study design: data regime, consent/confidentiality plan, IRB route, provenance log, and the disclosure statement.

Assessment

Layer	Weight	What it is
Weekly memos	30%	The six unit deliverables, each a concrete piece of the eventual design (question → hypotheses → design → measurement → sampling → ethics). Graded on the 4-point rubric: correctness, completeness, doctrine use, clarity.
Seminar participation and peer critique	20%	Prepared, honest critique of classmates' designs — the research version of the field's review discipline.
The design proposal	50%	A complete study design for the student's own research question, assembled from the six memos: question, hypotheses, design, measurement, sampling, ethics, and the intended contribution. Presented and defended in the final seminar session.

The design proposal is the course's proof of mastery and the natural first chapter of a dissertation proposal.

The course in two sentences

A dissertation is not a longer term paper; it is a research design that can answer a question the field needs answered. This course gives you the design and the measurement — DL 902 gives you the statistics that make the answer defensible.

Reflection questions

1. Which of the four faces of validity is most likely to be your design's weak point — and what would you do about it now?
2. What is the survivorship hazard in your own study, and what is your mitigation?

3. If a skeptic read only your design proposal, would they know exactly what you are claiming — and what you are *not* claiming?