Syllabus Codebook and Pre-Analysis Plan
MMCPSR: Mapping Methods in Contemporary Political Science Research
Tranae Hardy, Diana Kapiszewski, and Daniel Solomon
Last updated 12 May 2023

This document outlines a supplementary pre-analysis plan for the project "Mapping Methods in Contemporary Political Science Research." The project surveys the conduct of political science research published in ten top journals from 1998 – 2018, based on original data collected from a sample of 1,926 research articles between July 2020 and May 2021. This pre-analysis plan documents our analysis of original data of 986 graduate political science methods syllabi from departments in the United States, Canada, and Europe, which we collected and summarized between March 2022 and March 2023.

Section 1: Syllabus collection and variables in the dataset

Sources: We collected syllabi from two main sources. We collected the original syllabi from two main sources: (1) previous studies and data-collection efforts about graduate methods training (e.g., Emmons and Moravcsik 2019; Hardt et al. 2019); and (2) mass outreach to Directors of Graduate Studies in political science programs in the United States, Canada, and Europe. The syllabi in category (1) were principally associated with qualitative or multi-method courses; the syllabi in category (2) were principally quantitative.

For the latter effort, we initially contacted the Director of Graduate Studies at 103 political science PhD programs to request syllabi from quantitative methods courses taught between 1998 and 2018. A number of the respondents referred us to methods faculty in their department. In our follow up correspondence with the faculty, we requested referrals for additional contacts inside of the department. In total, we contacted 110 institutions and received quantitative syllabi from 62 institutions, with a response rate of 56 percent. We conducted this outreach from January – July 2022.

After we collected these syllabi, an undergraduate research assistant sorted the syllabi into categories focused on qualitative and quantitative methods. They also compiled an initial list of syllabi in our database, including information about (1) the institution that hosted the course; (2) the last name of the instructor; (3) the semester and year associated with the course; and (4) meta-data about the location of the syllabi in our folder directory on Box. We excluded non-political science and non-methods syllabi from our database.

With the syllabi in hand, we collected information about the following variables. We drew this information from explicit references to methods of data collection or analysis in the syllabi.

General

- *method*: This category indicates the particular method of data collection or analysis associated with the course. We primarily use the course title and description to inform this category, even if some of the specific course material diverges from the method in question. This is a categorical variable that includes the following mutually-exclusive

values: Quant General, Quant Collection, Quant Analysis, Qual General, Qual Collection, Qual Analysis, Mixed Collection, Mixed Analysis, and Unclear. See the project's broader Methodological Appendix for further discussion of these categories.

- *exercises_problem_sets*: This is a binary variable that indicates whether the course uses exercises or problem sets as instructional material for the course, where 1 indicates that the course uses exercises and a 0 indicates that it does not. The syllabus should refer explicitly to exercises or problem sets that the instructor assigns to model core course skills. We operationalize exercises or problem sets as bounded activities that require students to replicate a series of technical instructions. We exclude conventional "research design" assignments from this category, such as case justifications, concept mapping, research questions, literature reviews, typologies, prospecti, and grant-proposal writing.

- *replication_exercise*: This is a binary variable that indicates whether the syllabus requires students to undertake a replication of a quantitative or qualitative research project, where 1 indicates that the syllabus requires a replication exercise and a 0 indicates that it does not.

Methods of analysis

- *simple_probability*: This is a binary variable that indicates that the syllabus provides instruction about using basic probability methods to identify statistically significant differences or relationships, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *regression*: This is a binary variable that indicates that the syllabus provides instruction about multivariate regression methods, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *stats_id_strategy*: This is a binary variable that indicates that the syllabus provides instruction about multivariate regression methods involving and identification strategy, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *machine_learning*: This is a binary variable that indicates that the syllabus provides instruction about machine learning approaches, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *process_tracing*: This is a binary variable that indicates that the syllabus provides instruction about process tracing methods, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader

Methodological Appendix for a more detailed description of our operational definition.

- *qca*: This is a binary variable that indicates that the syllabus provides instruction about Qualitative Comparative Analysis or fuzzy-set methods, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *congruence*: This is a binary variable that indicates that the syllabus provides instruction about congruence analysis methods, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *case_comparison*: This is a binary variable that indicates that the syllabus provides instruction about approaches to multiple-case comparison, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *case_comparison*: This is a binary variable that indicates that the syllabus provides instruction about approaches to multiple-case comparison, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

- *case_comparison*: This is a binary variable that indicates that the syllabus provides instruction about approaches to multiple-case comparison, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. See the project's broader Methodological Appendix for a more detailed description of our operational definition.

Methods of collection

- *transform_quant_data*: This is a binary variable that indicates that the syllabus provides instruction about approaches to changing or manipulating one or more pre-existing datasets collected by another party for use in one's own analysis, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. This category includes approaches such as scaling, factor analysis, and item response theory, as well as references to manipulation, imputation, or missing data.

- *build_quant_text*: This is a binary variable that indicates that the syllabus provides instruction about approaches to creating a dataset using a pre-existing textual, archival, or secondary source, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. This category includes discourse and text analysis if that analysis results in numerical data. The category also includes topic

modeling.

- *build_quant_interview*: This is a binary variable that indicates that the syllabus provides instruction about approaches to creating a dataset using a pre-existing textual, archival, or secondary source, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not.

- *fieldwork*: This is a binary variable that indicates that the syllabus provides instruction about approaches to collecting qualitative data at a research site located away from the researcher's home institution, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not.

- *qual_data_gen*: This is a binary variable that indicates that the syllabus provides instruction about approaches to creating non-numerical data such as archival research, interviews, focus groups, ethnography, participant observation, and oral histories, where 1 indicates that the syllabus provides instruction about this method and a 0 indicates that it does not. This category includes discourse or text analysis if that analysis results in non-numerical data.

Ethics

- *ethics_research_design*: This is a binary variable that indicates that the syllabus provides instruction about ethical questions in research design, where 1 indicates that the syllabus provides instruction about this set of questions and a 0 indicates that it does not. This category includes ethics in the design of puzzles or research questions and problems of deciding the topics on which researchers should work.

- *ethics_collection*: This is a binary variable that indicates that the syllabus provides instruction about ethical questions in data collection, where 1 indicates that the syllabus provides instruction about this set of questions and a 0 indicates that it does not.

- *ethics_analysis*: This is a binary variable that indicates that the syllabus provides instruction about ethical questions in data analysis, where 1 indicates that the syllabus provides instruction about this set of questions and a 0 indicates that it does not. This category includes ethics in quantification, or the reduction of complex political phenomena into numerical measures, as well as problems of drawing subjective conclusions in interpretive research.

- *positionality*: This is a binary variable that indicates that the syllabus provides instruction about ethical questions associated with the social position or identity of researchers, where 1 indicates that the syllabus provides instruction about this set of questions and a 0 indicates that it does not. This category includes ethics associated with common researcher identity categories such as race, gender, nationality, and citizenship.

- *transparency*: This is a binary variable that indicates that the syllabus provides instruction about ethical questions associated the transparency or replicability of research, where 1

indicates that the syllabus provides instruction about this set of questions and a 0 indicates that it does not.

- *irb*: This is a binary variable that indicates that the syllabus provides instruction about Institutional Review Boards (IRBs), where 1 indicates that the syllabus provides instruction about IRBs and a 0 indicates that it does not.

- *irb*: This is a binary variable that indicates that the syllabus requires that students complete their Collaborative Institutional Training Initiative (CITI) certification in human subjects research, where 1 indicates that the syllabus requires that students complete their CITI or equivalent certification, and a 0 indicates that it does not.

An undergraduate research assistant also collected information about the author(s) and publication date of the first 10 books listed as "required" or "recommended" at the top of each syllabus. In the syllabus database, we list these as *textbook_1*, *textbook_2*, and so on, until *textbook_10*.

Section 2: Analysis – Descriptive

We present the following descriptive information about the syllabi in the sample:

- *Number of institutions surveyed:* The count of all institutions who were contacted to be in the sample.

- *Number of institutions in the sample:* The count of all institutions in the final sample.

- *Proportion of institutions in sample:* The proportion of institutions in the sample out of the total number institutions survey.

- *Number of syllabi by institution*: The count of syllabi from each institution.

- *Number of syllabi by year*: The count of syllabi by year taught from 1998 to 2018.

- *Average number of associated books, by course type:* The average number of required or recommended books[1] per course disaggregated by course type, qualitative or quantitative, for each year from 1998 to 2018.

- *Average number of associated qualitative books, by focus, by year*: The average number of required or recommended books per qualitative course disaggregated by focus (i.e., data generation or analysis) for each year from 1998 to 2018

- *Existence of textbooks for teaching qualitative analytic methods*: The books used most consistently across time / departments for teaching qualitative analytic methods are not standard walk-through-the-steps textbooks.

[1] A maximum of ten required or recommended books were recorded for each syllabi

https://docs.google.com/spreadsheets/d/1eX9nebcJh-Xj9NAVxxutrM6To3DzD5e5qIkqdyOfWHU/edit?pli=1#gid=1794849230

- *Number of syllabi including quantitative causal methods by year, by educational setting:* The count of syllabi including quantitative causal identification strategies for each year from 1998 to 2018 by educational setting, graduate programs or ICPSR.

- *Number of qualitative syllabi, by focus:* The count of syllabi for qualitative courses disaggregated by focus, data generation only, data analysis only, or a combination of data generation and analysis.

- *Number of syllabi featuring replication exercises, by course type*: The count of syllabi that include replication exercises by course type, quantitative or qualitative.

- *Proportion of institutions with syllabi including causal identification strategies, by year:* The proportion of institutions with at least one syllabi that include causally identified methods, statistics with an identification strategy or qualitative process tracing methods, for each year from 1998 to 2018.

Section 3: Analysis – Correlations / Bivariate regression

We also estimate a series of bivariate correlations and probit regressions to describe the relationships between multiple variables in the sample. For each set of dependent and independent variables discussed in this section we outline (1) hypotheses about their anticipated relationship; (2) how we operationalize the variables; (3) the statistical strategy that we use to test these hypotheses; and (4) the sample on which we conduct these analyses.

Sample: The analysis discussed below is performed on the full set of coded syllabi. In describing the analyses below, we use "All Syllabi" to indicate this sample. We subset subsequent samples using this All Syllabi sample.

33. Time > Causally identification strategy.

- H33: Causal ID strategies did not begin to appear on syllabi / be taught in graduate-level quant meth classes until after 2010.
- Dependent variable: Instruction on causal identification strategies. We test this hypothesis using two separate measures, as follows; for each, 1 indicates that the use of statistics with an identification strategy appears on the syllabus, and 0 indicates that it does not:
- Independent variable: Year of syllabus
- Estimation strategy: OLS, probit.
- Expected coefficient value: Positive
- Sample: The subset of all syllabi that use quantitative methods.

34. Time > Explicit qualitative methods

- H34: Syllabi dated after the qualitative renaissance are more likely to include explicit qualitative methods

- Dependent variable: Inclusion of "explicit" qualitative methods, exclusively or in tandem with other methods, where 1 indicates that the syllabi lists (1) process tracing; (2) structured case comparison; (3) QCA; (4) congruence analysis; and / or (5) counterfactual analysis, and 0 indicates that it does not.
- Independent variable: Year of syllabus
- Estimation strategy: OLS, probit
- Expected coefficient value: positive
- Sample: The subset of all syllabi that use qualitative methods.

35. Syllabi for qualitative courses> Data generation

- H35: Qualitative methods courses are more likely than quantitative methods courses to focus on strategies for data generation
- Dependent variable: Syllabi for a qualitative course where 1 indicates that the syllabus is for a course on qualitative methods and 0 indicates the syllabus was not.
- Independent variable: Syllabus includes data generation methods, where 1 indicates that the syllabus includes data generation methods and 0 indicates that it does not.
- Estimation strategy: OLS, probit
- Expected coefficient value: positive
- Sample: All Syllabi

36. Syllabi for quantitative courses > Data analysis

- H36: Quantitative methods courses are more likely than qualitative methods courses to focus on strategies data analysis
- Dependent variable: Syllabi for a quantitative course where 1 indicates that the syllabus is for a course on quantitative methods and 0 indicates the syllabus was not.
- Independent variable: Syllabus includes data analysis methods, where 1 indicates that the syllabus includes data analysis methods and 0 indicates that it does not.
- Estimation strategy: OLS, probit
- Expected coefficient value: positive
- Sample: All Syllabi