CODEBOOK · EDITION 1.1

Mid-Atlantic European Union Simulation, 2015–2020

Nicholas Clark, Susquehanna University
John Scherpereel, James Madison University

September 11, 2026

Edition 1.1 (2015 control-group duplicates removed) — figures computed from meus_long.csv

About this codebook

This codebook documents a harmonized dataset assembled from the pre- and post-simulation student surveys fielded around the Mid-Atlantic European Union Simulation between 2015 and 2020. The file holds 924 survey responses from 13 institutions, of which 544 are matched into 272 pre/post pairs. It is built directly from the original survey exports rather than from any analysis file, and every figure in this document is computed from the released data at build time.

The document is organized so that a reader who has never seen the study can use the data without asking the authors anything, and so that a reader who intends to publish from it knows precisely where it is thin. Section 11 is the honest account of the limits, and it should be read before the data are used, not after.

1 What this dataset is

The Mid-Atlantic European Union Simulation is a three-day simulation of European Union decision-making held each November in Washington, D.C. Undergraduates from participating institutions are assigned to the Commission, the Parliament, the Council of the European Union and the European Council, each playing a named European political figure — an "alter ego" — with each school representing a member state.

Students completed a survey at the start of the semester and a second at the end. The design is therefore a pre/post panel with a non-equivalent control group: students who did not attend the simulation, some of whom were enrolled in a European Union course and some of whom were not, completed the same instrument on the same schedule.

The surveys were built around four constructs, which this codebook uses throughout as the organizing vocabulary: Knowledge, Engagement, Skills and Empathy.

2 Coverage

Year	Pre	Post	Responses	Pairs	Paired, as a share of the smaller wave
2015	100	75	175	38	51%
2016	137	84	221	72	86%
2017	125	112	237	65	58%
2018	75	52	127	43	83%
2019	61	68	129	43	70%
2020	24	11	35	11	100%
All	522	402	924	272	68%

TABLE 2.1. *Responses and matched pairs by year.*

Unmatched respondents are retained in the file. A dataset containing only successful matches would hide its own selection problem: students who complete both waves are not a random subset of those who complete one, and the difference between the two groups is itself analysable. Every row carries a matched flag and, where matched, the method by which the pairing was made.

3 How the file is organized

The unit of observation is the student-wave: one row for each survey a student completed. A student who completed both waves appears twice, sharing a student_id. This shape keeps the unmatched responses in the same file as the matched ones and lets a user derive a wide, one-row-per-student file for change scores without losing anything on the way.

The identifying columns are meus_id (unique to the row), student_id (shared by the two rows of a pair, missing when unmatched), year, and wave. The variables that describe the pairing — matched, match_method, match_score, pair_complete — are documented in section 5.

4 The instrument and its three eras

The questionnaire was not constant. It falls into three eras, and the boundaries matter more than any other single fact about this dataset.

Era	Administration	Knowledge items	Matching key	What differs
2015	Paper, transcribed	Six	Favorite book or movie, plus favorite color	No skills items; no ideological or alter-ego placements
2016	Qualtrics, choice text	Six	Maternal grandmother's first name, plus high-school street	Skills present; placements asked in the post wave only
2017–2020	Qualtrics, numeric	Ten	Maternal grandmother's first name, plus high-school street	Full instrument: skills and placements in both waves

TABLE 4.1. *The three instrument eras.*

The knowledge battery is the consequential change. A six-item index and a ten-item index are not the same variable, and pooling them produces an artefact that looks like a finding. Section 8 gives the three knowledge measures the dataset carries and says which is safe for which comparison.

5 Matching

Students generated their own identifier so that the two waves could be linked without collecting a name. From 2016 the key was the first name of the maternal grandmother together with the name of the street the student lived on in the final year of high school. In 2015 it was the favorite book or movie together with the favorite color.

Students are inconsistent about which grandmother they name, and about whether they give a first name or a surname. Requiring both halves of the key to agree therefore loses roughly a quarter of the pairs that a human reader would make without hesitation. Matching runs instead in tiers of decreasing confidence, and every pair records the tier on which it was made, so a cautious user can restrict the file to exact matches.

Tier	Rows	What it means
both_exact	306	Both halves of the key identical
both_fuzzy	16	Both halves close (0.85 or better on a sequence ratio)
street_only	108	Street identical, grandmother name different; schools not contradictory
name_only	30	Grandmother name identical, street different; schools not contradictory
hand_2015	76	Adopted from the 2015 hand pairing; not reproducible by rule
workbook_manual	8	Adopted from the 2021 analysis workbook; required human judgment
unmatched	380	No partner found

TABLE 5.1. *Matching tiers, in descending confidence.*

The one-sided tiers are the ones to watch. They pair on a single half of the key and are included because they recover pairs the published analyses used, but a user who wants only unambiguous links should keep the first two rows of Table 5.1 and discard the rest.

Two further flags matter. duplicate_submission marks the 39 rows where a student completed the same survey twice; the fullest response is used for matching and the others are retained rather than deleted. pair_complete is 1 for every pair in this edition: the two pairs that were incomplete in edition 1.0 were both transcription defects in the 2015 workbook — the first pairing row had been skipped by the build, and one post-survey identifier was typed LVCO014 for LVCP014 — and both are repaired at source in the build script, so the flag is kept as a standing check rather than because any pair fails it. control_single_wave marks the 32 responses of 2015 from students who were not in the simulation and answered the survey once — the Susquehanna group in October, the James Madison group in December — which the transcription had copied into both the pre and the post sheet; each is carried once, in the wave it was actually taken, and they have no pair by design.

6 Variable dictionary

Variables are grouped by the construct they serve. The availability column gives the years in which the variable has any non-missing value; a variable absent from a year was not asked that year, not merely unanswered.

6.1 Administration

Variable	What it asks	Coding	Available
school	Institution, coded (see section 10)		all years
year	Year of the simulation		all years
wave	pre or post		all years
student_id	Links the two rows of a matched pair		all years
matched	Whether the row is part of a pair	1 Row is part of a matched pair, 0 otherwise	all years
match_method	Tier the pair was made on (Table 5.1)		all years
pair_complete	Whether both rows of the pair are present	1 Both rows of the pair are present, 0 otherwise	all years
duplicate_submission	Whether the student submitted this survey more than once	1 Student submitted this survey more than once, 0 otherwise	all years
control_single_wave	Whether this is a 2015 non-participant response taken once (no pair by design)	1 Non-participant surveyed once in 2015 (no pair by design), 0 otherwise	all years
sim_mode	in person, or virtual in 2020		all years
fielding_season	fall, or spring in 2020		all years

6.2 Background

Variable	What it asks	Coding	Available
age	Age in years, banded at 25 and above		all years
gender	Harmonized from free text: female, male, other/unspecified		all years
graduation_year	Intended year of graduation		2017–2020

Variable	What it asks	Coding	Available
officer	Are you currently serving as an executive officer (e.g., president, vice-president, treasurer, secretary) for a club or other student organization at your college or university?	1 Is an officer of a club or organization, 0 otherwise	2016–2020
visited_europe	Have you ever visited Europe? (If yes, note how many times you have visited and how long you stayed each time)	1 Has visited Europe, 0 otherwise	all years
eu_class_now	Currently taking a course on the EU	1 Currently taking an EU course, 0 otherwise	all years
eu_class_before	Took one in an earlier term	1 Took an EU course in an earlier term, 0 otherwise	all years
eu_class_never	Has never taken one	1 Has never taken an EU course, 0 otherwise	all years
sim_before	Before this year (2018), had you ever participated in an EU simulation?	1 Had taken part in a simulation before this year, 0 otherwise (original codes kept in sim_before_raw)	all years
sim_now	Took part in the simulation this year	1 Took part in the simulation this year, 0 otherwise	2015, 2017–2019

6.3 Engagement

Variable	What it asks	Coding	Available
interest_politics	To what extent would you say you are interested in politics?	1 Very; 2 Somewhat; 3 A little; 4 Not at all	all years
interest_eu	To what extent would you say you are interested in the European Union?	1 Very; 2 Somewhat; 3 A little; 4 Not at all	all years
discuss_local	Discusses Local political matters	1 Frequently; 2 Occasionally; 3 Never	all years
discuss_national	Discusses National political matters	1 Frequently; 2 Occasionally; 3 Never	all years
discuss_intl	Discusses International political matters	1 Frequently; 2 Occasionally; 3 Never	all years
media_tv	Watch television on a TV set	1 Every day/almost every day; 2 2-3 times a week; 3 About once a week; 4 2-3 times a month; 5 Less often; 6 Never; 7 No access to this medium	all years

Variable	What it asks	Coding	Available
media_tv_internet	Watch television via the Internet	1 Every day/almost every day; 2 2-3 times a week; 3 About once a week; 4 2-3 times a month; 5 Less often; 6 Never; 7 No access to this medium	all years
media_radio	Listen to the radio	1 Every day/almost every day; 2 2-3 times a week; 3 About once a week; 4 2-3 times a month; 5 Less often; 6 Never; 7 No access to this medium	all years
media_press	Read the written press	1 Every day/almost every day; 2 2-3 times a week; 3 About once a week; 4 2-3 times a month; 5 Less often; 6 Never; 7 No access to this medium	all years
media_internet	Use the Internet	1 Every day/almost every day; 2 2-3 times a week; 5 Less often	all years
media_social	Use online social networks	1 Every day/almost every day; 2 2-3 times a week; 3 About once a week; 4 2-3 times a month; 5 Less often; 6 Never	all years

6.4 Knowledge

Variable	What it asks	Coding	Available
know1	Switzerland is a member of the European Union	1 correct, 0 incorrect. Correct answer: FALSE	all years
know2	The European Union has 28 member states	1 correct, 0 incorrect. Correct answer: TRUE	all years
know3	Every country in the European Union elects the same number of representatives to the European Parliament	1 correct, 0 incorrect. Correct answer: FALSE	all years
know4	Every six months, a different member state becomes president of the Council of the European Union	1 correct, 0 incorrect. Correct answer: TRUE	all years
know5	The European Union has a single seat on the United Nations Security Council	1 correct, 0 incorrect. Correct answer: FALSE	all years
know6	The European Union has primary authority to legislate in the area of education	1 correct, 0 incorrect. Correct answer: FALSE	all years

Variable	What it asks	Coding	Available
know7	The European Parliament initiates all EU legislative proposals	1 correct, 0 incorrect. Correct answer: FALSE	2017–2020
know8	The president of the European Commission is directly elected	1 correct, 0 incorrect. Correct answer: FALSE	2017–2020
know9	To become a law, EU legislative proposals must be supported by every member state	1 correct, 0 incorrect. Correct answer: FALSE	2017–2020
know10	It is extremely rare for one party group to control more than 50% of the seats in the European Parliament	1 correct, 0 incorrect. Correct answer: TRUE	2017–2020

6.5 Skills

Variable	What it asks	Coding	Available
speaking	Public speaking	1 Significantly below average; 2 Slightly below average; 3 Average; 4 Slightly above average; 5 Significantly above average	2016–2020
leadership	Leadership	1 Significantly below average; 2 Slightly below average; 3 Average; 4 Slightly above average; 5 Significantly above average	2016–2020
negotiation	Negotiation	1 Significantly below average; 2 Slightly below average; 3 Average; 4 Slightly above average; 5 Significantly above average	2016–2020

6.6 Empathy

Variable	What it asks	Coding	Available
leftright_self	In political matters, people talk of "the left" and "the right". What is your position? Please use a scale from 0 to 10, where '0' means "left" and '10' means "right". Which number best describes your	0 to 10 on the left/right scale, corrected from the raw storage (section 7)	2016–2020

Variable	What it asks	Coding	Available
eupos_self	In political matters, people often talk of "europhiles" (people who support European integration) and "euroskeptics" (people who oppose European integration). What is your position? Please use a scale	0 to 10 on the europhile/eurosceptic scale, corrected from the raw storage (section 7)	2016–2020
leftright_alter	In political matters, people talk of "the left" and "the right". What is the position of your most recent simulation alter ego (e.g., the politician you represented in the EU simulation indicated abov	0 to 10 on the left/right scale, corrected from the raw storage (section 7)	2016–2020
eupos_alter	In political matters, people often distinguish "europhiles" (people who support European integration) and "euroskeptics" (people who oppose European integration). What is the position of your most rec	0 to 10 on the europhile/eurosceptic scale, corrected from the raw storage (section 7)	2016–2020

6.7 Open-ended responses, and the scores derived from them

The response text is a variable in its own right and is released unchanged. The scores derived from three of these items are documented in sections 8.4 and 8.5; the columns are listed here so that every variable in the file has an entry in this dictionary.

Variable	What it holds	Coding	Available
open_eu_purpose	What is the purpose of the European Union? Coded as themes, not scored	Free text, unedited	all years
open_maastricht	What was the major change introduced by the Treaty of Maastricht?	Free text, unedited	all years
open_maastricht_2	The 2016 pre survey asked the Maastricht question a second time	Free text, unedited	2016
open_olp	Describe the ordinary legislative procedure	Free text, unedited	2017–2020
sim_experience_open	The simulation experience and its benefits. Not scored: it asks for reflection, not knowledge	Free text, unedited	2016–2020
maastricht_score	Score for open_maastricht	1 no substantive answer or wrong; 2 one correct change, named; 3 characterized, or two or more correct	all years

Variable	What it holds	Coding	Available
maastricht_score_2	Score for the 2016 second asking	Same 1–3 scale; kept apart from maastricht_score so that variable means the same thing in every year	2016
maastricht_asked_twice	The question was asked twice on this survey	1 yes, 0 no	2016
olp_score	Score for open_olp	1 no substantive answer or wrong; 2 one component; 3 two components; 4 three or more including joint agreement	2017–2020
purpose_economic	Theme in open_eu_purpose: trade, market, currency, prosperity	1 present, 0 absent	all years
purpose_peace	Theme: peace, stability, preventing war	1 present, 0 absent	all years
purpose_political	Theme: common governance or policy-making	1 present, 0 absent	all years
purpose_mobility	Theme: free movement, open borders	1 present, 0 absent	all years
purpose_themes	How many of the four themes are present	0 to 4	all years
maastricht_nonanswer	Response declined or made no attempt	1 yes, 0 no	all years
olp_nonanswer	Response declined or made no attempt	1 yes, 0 no	2017–2020
purpose_nonanswer	Response declined or made no attempt	1 yes, 0 no	all years
maastricht_coder_disagreed	The two coders differed before adjudication	1 yes, 0 no	all years
olp_coder_disagreed	The two coders differed before adjudication	1 yes, 0 no	2017–2019
purpose_coder_disagreed	The two coders differed before adjudication	1 yes, 0 no	all years
maastricht_needs_review	The rubric supported neither coder	1 yes, 0 no	2017
olp_needs_review	The rubric supported neither coder	1 yes, 0 no	never set
purpose_needs_review	The rubric supported neither coder	1 yes, 0 no	2017
score_by_rule	Punctuation-only response, scored by rule rather than read by a coder	1 yes, 0 no	2015–2018, 2020

TABLE 6.7. *The open-ended responses and their derived scores.*

7 Original codings are kept

No variable was overwritten. Where a response required correction or recoding, the original is retained under the same name with a _raw suffix and the corrected version carries the plain name. A reader who disagrees with a decision made here can therefore rebuild from the original values without returning to the source exports.

This matters most for the placement scales. Qualtrics stored the 0–10 scales as choice identifiers 1–11 from 2017 onward, while the 2016 export stored the displayed 0–10 value directly. The corrected columns are on the true 0–10 scale in every year; the column placement_raw_convention records which coding each row came from.

8 Derived measures

Each of the four constructs has an index. Every index is built from the harmonized variables in section 6 and can be rebuilt from them; none is a black box.

8.1 Knowledge

Year	total_correct	total_correct_common6	knowledge_pct
2015	4.779	4.779	0.803
2016	4.343	4.343	0.733
2017	6.828	4.363	0.686
2018	6.667	4.415	0.668
2019	7.336	4.571	0.737
2020	6.032	3.613	0.621

TABLE 8.1. *Knowledge means by year, on the three available measures.*

The first column is the reason this dataset carries more than one knowledge measure. total_correct is the era index: a count out of six through 2016 and out of ten from 2017. It jumps by more than two points at the era boundary, and the jump is an artefact of the battery lengthening, not a change in what students know.

total_correct_common6 counts only the six items asked in every year — and they are the same six, word for word, not merely six of the ten. On that measure the series is flat, which is the point. total_correct_common5 additionally drops item 2, and total_correct_common6 should not be compared into 2020 without it; section 11 explains why.

8.2 Engagement

engagement_interest is the mean of the two interest items and engagement_discussion the mean of the three discussion items, both on reversed scales so that a higher value means more. The reversal is deliberate: in the raw data "very interested" is coded 1, and an index that runs backwards invites misreading. engagement_index combines the two, each rescaled to run from 0 to 1.

8.3 Skills

skills_index is the mean of the self-ratings for public speaking, leadership and negotiation, each on a five-point scale from significantly below average to significantly above average. skills_index01 rescales it to 0–1, and skills_items_answered records how many of the three a respondent answered. The items were not asked in 2015.

8.4 The open-ended answers, scored

Three of the open-ended questions have knowledge content and have been scored. The scores are derived variables like any other in this section: the response text is kept unchanged beside them, so a user who distrusts the scoring can discard it and code the text again.

Two of the three have defensible right answers and carry an ordinal score. The third does not. Asked what the European Union is for, a student who answers "to prevent another war" is not less correct than one who answers "to make trade easier", so scoring that item for correctness would be scoring sophistication under a misleading name. It is coded instead as four themes, each present or absent.

Variable	Scale	What it scores	Responses
maastricht_score	1–3	The major change introduced by the Treaty of Maastricht. 1 no substantive answer or wrong; 2 one correct change, named; 3 a correct change characterized accurately, or two or more correct changes.	668
olp_score	1–4	The ordinary legislative procedure. Scored on four components — the Commission proposes; Parliament and Council are both involved; both must agree and may amend; disagreement goes to further readings or conciliation. 4 requires three components including joint agreement.	376
purpose_economic, _peace, _political, _mobility	0/1	The purpose of the European Union, coded as four independent themes rather than scored. purpose_themes counts how many are present.	697

TABLE 8.2. *The scored open-ended items.*

8.5 How the scoring was done, and how far to trust it

The full rubric, with anchor examples and the judgment calls settled in advance of any scoring, is released with the data as rubric.md. Three features of the procedure matter more than the rubric's content.

It was blind. Coders saw the response text and nothing else — no year, no wave, no school, no pairing. This is not a formality. These items are the outcome measures in a pre/post design, and a coder who can see that a response came after the simulation will grade it more generously; the contamination is undetectable afterward. Identical responses were collapsed to one unit before scoring, so the same answer could not be graded twice in different company.

It was done twice, independently. Two coders scored every response from the same rubric, neither seeing the other's work. Agreement was computed before any disagreement was resolved.

Variable	Kappa	Exact	Within 1	Adjudicated
maastricht_score	0.95	95.4%	100%	26
olp_score	0.98	93.3%	100%	24
purpose_economic	0.99	99.3%	—	5
purpose_peace	0.97	98.9%	—	8

Variable	Kappa	Exact	Within 1	Adjudicated
purpose_political	0.91	95.7%	—	30
purpose_mobility	0.98	99.7%	—	2

TABLE 8.3. *Inter-coder agreement before adjudication. Kappa is quadratic-weighted Cohen's kappa on the ordinal scales and unweighted on the theme binaries. No disagreement on either ordinal scale exceeded one level.*

Disagreements were then resolved against the rubric by an adjudicator who produced neither pass, and who was restricted to choosing between the two submitted values rather than introducing a third — so a case the rubric did not settle is marked rather than quietly decided. Every response the coders differed on is flagged in the data (maastricht_coder_disagreed and its counterparts), and the three cases the adjudicator could not resolve within the rubric carry needs_review. A user who wants only the uncontested scores can drop the flagged rows and lose about five per cent of them.

The third feature is the limitation, and it is real: both coders were language models. Two models working from one rubric share failure modes that two human coders would not, so the agreement above is evidence that the rubric is unambiguous, not evidence that the scores are right. A human double-coding a random subsample is the check that would settle it, and it has not been done. Anyone publishing from these three variables should do it, and should disclose the machine coding.

Two mechanical conventions are worth knowing. Responses containing no letter or digit at all — a lone full stop or question mark, twenty-four of them — were never shown to a coder; they are scored at the bottom of the scale with the non-answer flag set and score_by_rule set to 1, so they stay separable from both blanks and coded responses. And the 2016 pre survey asked the Maastricht question twice: the first asking carries maastricht_score, so the variable means the same thing in every year, and the second is kept apart as maastricht_score_2 with maastricht_asked_twice marking the rows.

9 Empathy: two measures, and they disagree

Empathy is the construct where this dataset departs from what has been published, and the departure is deliberate enough to need its own section.

The published analysis measured empathy as movement in the student's own position: it compared self-placement on the left/right and europhile/eurosceptic scales before and after the simulation. That measure is retained here as self_shift_lr and self_shift_eu so that the published result can be reproduced exactly.

But the instrument was built to support a different measure. Students place their alter ego on the same two scales as themselves, which makes it possible to ask whether a student moved toward the position of the figure they played. That is what the construct means in the framework, and the alter-ego variables were collected for it — yet they have never been used. The dataset therefore also carries

$$\text{empathy\_lr} = |\,\text{pre self} - \text{alter}\,| - |\,\text{post self} - \text{alter}\,|$$

and the same for the European dimension, with empathy_index averaging the two. A positive value means the student ended the simulation closer to the position of the person they were playing. The component distances are kept as divergence_lr_pre and divergence_lr_post so that a reader can see what the difference is made of.

The two measures are almost entirely unrelated: across the 109 pairs where both exist, they correlate at r = -0.033. Moving one's own position and moving toward one's alter ego are not the same thing, and a study that measures the first should not describe itself as having measured the second.

A first look at the convergence measure, offered as description and not as a result: across 109 pairs the mean is +0.225 (standard deviation 1.220); 49 students moved closer to their alter ego, 31 moved away and 29 did not move. The direction is the one the framework predicts and the magnitude is small. This construction has never been validated, and the codebook presents it as new work rather than as replication.

Two constraints limit it. The alter-ego placements are asked only after the simulation, so the measure needs a matched pair. And 2016 asked its self-placements in the post wave only, so that year has an alter-ego position but no starting point to move from. The convergence measure therefore exists for 2017 through 2020; 2015 has neither variable.

10 De-identification

The study was approved on the representation that responses are not recorded in a way that identifies a subject, directly or through linked identifiers, and participants were told they would be identified in the research records by a code so that their responses could not be linked to them. The released file is built to that standard.

Field	Treatment
Self-generated matching key	Removed entirely. Not published even as a hash: a hash of a first name and a street name is open to a dictionary attack.
Institution	Coded (SCH01, SCH02, …). The participating institutions are listed in the publications; the mapping from code to name is not released, so a reader cannot place a respondent at a named small college.
Age	Banded: every respondent aged 25 or older is recorded as 25.
Graduation date	Reduced to a year.
Qualtrics metadata	IP address, latitude, longitude and the recipient name and email fields are dropped from every export.
Gender	Harmonized from free text to female, male, other/unspecified. The raw text is not released.

TABLE 10.1. *De-identification applied to the released file.*

Two things remain for the data owner to decide before any public posting. The open-ended responses have not been screened, and free text can carry self-identifying detail. And small cells persist: the least-represented institutions contribute only a handful of respondents each, so institution by year by a demographic or two is thin enough to warrant a suppression rule.

11 An honest account of the limits

What follows is not a list of caveats appended for form. Each item changes what can be concluded, and several of them were discovered only by rebuilding the data from source.

11.1 The knowledge battery changed length

Six items through 2016, ten from 2017. Use total_correct_common6 or knowledge_pct for any comparison that crosses that boundary, and never the raw count.

11.2 Brexit invalidated a knowledge item

Item 2 asserts that the European Union has 28 member states. That was true when the item was written and false from 31 January 2020. The 2020 post-simulation survey was fielded in late April and early May, after the United Kingdom had left. The variable know2_key_valid marks the item as ungradable in 2020, and total_correct_common5 excludes it.

11.3 The placement scales were stored differently in different years

From 2017 the 0–10 scales were stored as Qualtrics choice identifiers 1–11, so an uncorrected reading is a full point high. The published means on those scales carry that shift. Because the published tests compare the same variable before and after, the shift cancels and no significance test is affected — but any reported level is one point too high. The corrected columns in this file are on the true scale.

11.4 Matching is incomplete and not random

544 of 924 responses are matched. Students who complete both waves differ from those who complete one, and the unmatched are retained precisely so that this can be examined rather than assumed away.

11.5 2020 is not a normal year

The simulation ran virtually because of the pandemic, and the survey moved from a fall administration to spring. The instrument did not change, but the intervention did. The variables sim_mode and fielding_season carry this so that a pooled analysis can see it. The year is also small: eleven pairs.

11.6 The 2020 post-simulation export no longer exists in raw form

The only surviving record of that wave is a hand-assembled workbook whose post rows were pasted under the pre-survey header, displacing every value from one question onward by four columns. The displacement was measured rather than assumed, and the realignment reproduces the analysis workbook exactly for nine of the ten knowledge items and for the interest, discussion and media batteries. One variable — interest in politics — reproduces at no offset and is therefore set missing for those eleven rows rather than guessed.

11.7 2015 and 2016 are partial on two of the four constructs

2015 asked no skills items and no placements at all, so it contributes to Knowledge and Engagement and to nothing else. 2016 asked the placements in the post wave only. Neither gap is sparseness that more data would fill; the questions were not asked.

11.8 The 2015 pairing cannot be reproduced

That year used a different key, recorded only in the pre-survey sheet, and the pairing was made by hand at the time. It is adopted as given and flagged hand_2015. It cannot be checked against the key, because the post-survey sheet does not carry one.

11.9 Some 2015 variables resist harmonization

Whether the student had visited Europe, taken a European Union course, or previously taken part in a simulation were recorded that year as multi-select strings — values such as "1 and 2" and "1,2". They are carried as raw text with no derived version. Forcing them into a clean coding would invent precision that the source does not contain.

11.10 The 2015 control group was transcribed twice

The thirty-two 2015 students who were not in the simulation answered the survey once — sixteen at Susquehanna in October, sixteen at James Madison in December — but the transcription entered each of those responses in both the pre and the post sheet, with the same identifier and the same answers. An earlier edition of this file carried every one of them twice (956 responses rather than 924). Each is now carried once, in the wave it was taken, flagged control_single_wave, with sim_now = 0; pairs are unaffected because these students were never paired. Found on 11 September 2026 by checking the James Madison exports against the transcription. The same check settled the two pairs that edition 1.0 reported as incomplete: pair 1 lost its pre-survey row because the build skipped the first row of the hand-pairing sheet, and pair 15 lost its post-survey row to a typed identifier (LVCO014 for LVCP014). Both rows are now in their pairs; matched rows 542 to 544.

11.11 Institutions are self-reported

Students typed the name of their institution, producing spellings that had to be harmonized. One response ("Economics") is not an institution at all and is left unassigned.

11.12 The open-ended scores were produced by machine coders

Sections 8.4 and 8.5 give the procedure and the agreement statistics. The limitation is that both coders were language models working from one rubric, which is not the same as two human coders, and no human has double-coded a subsample. The scores are offered as a usable first pass with every disagreement flagged, not as a validated instrument. The response text is kept unchanged beside every score.

12 Sources

Year	Source
2015	Transcription of the paper surveys, pre and post sheets, with the hand pairing from the comparison sheet
2016	Qualtrics exports, pre and post, in choice-text form
2017	Qualtrics exports, pre and post, numeric
2018	Qualtrics exports, pre and post, numeric
2019	Qualtrics exports, pre and post, numeric, with the choice-text twins used to recover the value labels
2020	Qualtrics export for the pre wave; the post wave from the hand-assembled pairs file described in 11.6
Archival	The 09.01.21 All Years workbook, used for two purposes only: to recover the knowledge answer key, and to adopt the hand pairings recorded there that no rule in section 5 reproduces (match_method workbook_manual)

TABLE 12.1. *Source of each year.*

The value labels used throughout were not transcribed from a questionnaire. They were recovered by joining the 2019 and 2020 numeric exports to their choice-text twins, which gives the code-to-label mapping as evidence rather than as recollection, and that mapping was then used to decode the 2016 choice-text export onto the same coding. The knowledge answer key was recovered the same way, by joining raw responses to the already-scored analysis workbook across roughly 290 responses per item, with complete agreement.

13 How the file was built, and how to rebuild it

The dataset is the output of a single script. Nothing in it was assembled by hand in a spreadsheet, and no figure in this codebook was typed: every count, percentage and range printed above was computed from the released file at the moment this document was generated. A reader who reruns the build and regenerates the codebook should get the same numbers, and if they do not, the discrepancy is itself informative.

13.1 The order of operations

The build proceeds in five stages, in this order.

Stage	What happens
1 Read	Each of the year-and-wave source files is read in its own right, with the two Qualtrics metadata rows above the data removed where present, and its own question identifiers mapped onto the common variable names. The 2016 map is kept separate from the 2017-and-later map, and the 2016 pre and post maps are kept separate from each other, because the same question identifier does not mean the same thing across those files.
2 Harmonize	Responses are put onto one coding. The 2016 choice-text export is decoded onto the numeric coding recovered from the later files; the placement scales are corrected for the change of convention described in section 7; knowledge items are scored against the recovered answer key; and the derived measures of section 8 are computed.
3 De-identify	Names, e-mail addresses and any other direct identifier are removed from the analysis file and written instead to the restricted crosswalk, keyed on meus_id. Open-ended responses are carried through unaltered and are the one place where identifying detail could survive; see section 10.
4 Match	Pre and post responses are paired by the tiers described in section 5, in order, with each tier applied only to responses the previous tier left unmatched.
5 Write	The long file and the two restricted crosswalks are written out, and a set of internal checks is run: that no student_id carries more than one pre and one post response, that every row flagged as matched carries an identifier, that no derived scale falls outside its declared range, that meus_id is unique, and that the released file contains none of the identifier columns. A failed check raises rather than letting the build report success.

TABLE 13.1. *The build, in order.*

13.2 What is released and what is not

File	Status	What it is
meus_long.csv	Released	The analysis file. 924 rows, 145 columns, one row per survey response.
meus_long.dta	Released	The same file as Stata, with a variable label on every variable and value labels on every coded one. The labels are the ones documented in section 6, not a separate set.
meus_RESTRICTED_crosswalk.csv	Restricted	Links meus_id to the identifying information removed in stage 3. Held by the principal investigator and not distributed.
meus_RESTRICTED_school_crosswalk.csv	Restricted	Links the school codes used in the released file to institution names.
rubric.md	Released with the data	The coding manual for the scored open-ended items, with the anchor examples and the judgment calls settled before any response was read (sections 8.4 and 8.5).
build.py	Released with the data	The script that produces the released file and the crosswalks from the sources listed in section 12; a second script writes the labeled Stata version from it.
MEUS Codebook.docx	Released with the data	This document, generated from the released file.

TABLE 13.2. *The files.*

The restricted crosswalks are the only place identifiers survive. They exist because the pairing in section 5 cannot be checked or corrected without them, and because a student who withdraws consent has to be findable. They are not part of any release, and a request to reproduce the matching should be met by rerunning the build rather than by sharing them.

13.3 Rebuilding

Place the source files of Table 12.1 in a directory named raw, together with the archival workbook named in that table, and run build.py against it. Beyond pandas and numpy the script needs only the standard library and the readers pandas uses for Excel files. It writes the files of Table 13.2, prints the matching table reproduced as Table 5.1, and finishes with the checks of stage 5. Regenerating this codebook is a second script, run against the file the first one produced.

One consequence of building this way is worth stating plainly. Correcting an error means correcting the script and rerunning it, not editing the released file. A dataset that has been edited by hand after the fact cannot be reproduced from its sources, and the moment that happens the documentation in this codebook stops being a description and becomes a claim.