SoftwareX 36 (2026) 103235



Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx



CurricuMap: an open-source Python pipeline for converting registrar transcript exports into auditable student-by-competency-domain matrices

Tarık Küçükdeniz [a,*], Tuncer Can [b]

[a] Department of Industrial Engineering, Faculty of Engineering, İstanbul University-Cerrahpaşa, İstanbul, Türkiye
[b] Department of English Language Teaching, Hasan Ali Yücel Faculty of Education, İstanbul University-Cerrahpaşa, İstanbul, Türkiye

ARTICLE INFO

Keywords:
Learning analytics
Curriculum mapping
Competency assessment
Data provenance

ABSTRACT

Competency-level learning analytics needs a student-by-competency-domain matrix, but institutions hold registrars' transcript exports, and the conversion is usually an undocumented spreadsheet step. CurricuMap is an open-source Python library and command-line tool that performs it. The course-to-domain mapping is a declarative, schema-validated YAML specification of ordered keyword and regular-expression rules and explicit overrides; every assignment records its provenance, unmatched courses are flagged, and an audit reports coverage and reliability. For the 143 course titles of a Turkish teacher-education programme, a taxonomy written by language-model agents from the labelled course list of the national curriculum, unedited and frozen before testing, assigned every title as its labelling convention prescribes, as did the best baselines given the same list.

Required metadata

Current code version

Table 1
Code metadata.

Nr.	Code metadata description	Please fill in this column
C1	Current code version	v0.2.0
C2	Permanent GitHub link to code/repository used for this code version	https://github.com/tkucukdeniz2/curricumap/tree/v0.2.0; archived on Zenodo, https://doi.org/10.5281/zenodo.22140689 (all versions)
C3	Legal Code Licence	MIT
C4	Code versioning system used	git
C5	Software code languages, tools, and services used	Python (≥3.10); continuous integration on GitHub Actions on Linux (Python 3.10, 3.12, 3.13) and macOS (Python 3.13), including a test of the installed wheel; 206 tests, 99% statement coverage
C6	Compilation requirements, operating environments & dependencies	Python ≥3.10, tested on Linux and macOS. Dependencies declared in `pyproject.toml`: pandas ≥2.0, numpy ≥1.24, pyyaml ≥6.0, jsonschema ≥4.0, rapidfuzz ≥3.0, openpyxl ≥3.1, jinja2 ≥3.1. `requirements.txt` pins the versions used for the illustrative example and benchmarks (Python ≥3.12), `validation/requirements.txt` those of the validation
C7	If available Link to developer documentation/manual	https://github.com/tkucukdeniz2/curricumap#readme; https://github.com/tkucukdeniz2/curricumap/blob/v0.2.0/docs/architecture.md
C8	Support email for questions	tkdeniz@iuc.edu.tr

* Corresponding author.
Email address: tkdeniz@iuc.edu.tr (T. Küçükdeniz).

https://doi.org/10.1016/j.softx.2026.103235
Received 28 August 2026; Received in revised form 26 October 2026; Accepted 11 November 2026
2352-7110/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Motivation and significance

Learning-analytics studies increasingly ask questions at the level of competency domains rather than individual courses: whether curricular knowledge areas are separable, or how coursework in one domain predicts performance in another. Answering them requires a student-by-competency-domain matrix (hereafter, student-by-domain matrix), in which each cell summarises a student's attainment in one domain. Registrars do not store data in this form. They export transcripts as flat grade records, one row per student per course, with inconsistent column names, mixed-language course titles, retaken courses, administrative sentinel grades, and courses outside any competency domain. Converting such an export into an analysis-ready matrix is a routine prerequisite for secondary analysis of existing records, and it is usually done by hand in a spreadsheet.

These preprocessing decisions matter. Which course belongs to which domain, how retakes are resolved, which grades are real, and which students have enough coverage all change the matrix and therefore the results, just as analytic choices change other learning-analytics findings [1]. Yet such choices are rarely reported in enough detail to be rerun, and reproducibility reviews of learning-analytics and educational-data-mining papers find that few analyses can be reproduced [2,3].

CurricuMap makes this step explicit. It does not propose a new classification algorithm; its classifier is conventional by design, because which course belongs to which domain is a substantive decision that domain experts must make and a methods section must report. CurricuMap's contribution is to turn that decision into a declarative, versioned, schema-validated specification that colleagues can read, cite, test, and reuse, and to execute it deterministically, with per-course provenance, documented registrar cleaning, and an audit of the resulting matrix. The specification may be written by hand or drafted with a large language model (LLM); either way the decision is recorded in a reviewable file rather than in spreadsheet formulas or model weights. Section 3.2 tests the bundled taxonomy on a real registrar catalogue against natural language processing (NLP) and LLM classifiers.

Table 2 compares CurricuMap with the closest tools. Commercial assessment-management platforms support curriculum mapping, the alignment of courses with programme outcomes [11], and collect outcome-tagged rubric scores and student work from the learning-management system (LMS) [4–6]; some also integrate with student-information systems or issue comprehensive learner records of assessed skills [4,5]. Their public documentation describes curriculum maps built inside the platform, not rules applied to a registrar's transcript export, and documents neither per-course mapping provenance nor a reliability audit of a student-by-domain matrix. Moodle's open-source competency frameworks rate students on competencies linked to courses inside the LMS [7]. Among research software, ouladFormat prepares a public dataset [8], and Curricular Analytics measures the prerequisite structure of curricula [9]. NLP and LLM studies map course descriptions or learning outcomes to skills [12,13] or classify transcript titles into a national course taxonomy [14], but they provide no reusable pipeline from a registrar export to an audited student-by-domain matrix. A search of PyPI, CRAN, GitHub, and the literature on 1 October 2026 found no open-source, general-purpose package for this conversion.

Transcripts are personal data under Türkiye's Personal Data Protection Law No. 6698 (KVKK) [15] and the EU General Data Protection Regulation (GDPR) [16], and learning analytics carries ethical obligations beyond legal compliance, including transparency about how student data are processed [17]. CurricuMap responds with design choices: it runs locally and transmits nothing; it writes every course assignment to a provenance table; and it flags rather than guesses a course that matches no rule. For testing and peer review without real records, it includes a seeded generator of synthetic transcripts. The generator is parameterised only by a taxonomy, a random seed, and size settings, and it never reads real records, so its output does not relate to any identified or identifiable person: it is anonymous information, which falls outside the GDPR (Recital 26) and outside the KVKK's definition of personal data (Art. 3). This differs from synthetic data derived from confidential microdata, whose disclosure risk and analytic utility must be evaluated [18,19]. The generator aims at functional realism, injecting the irregularities the pipeline must handle (missing enrolments, retakes, sentinel zeros), not at reproducing the statistics of any real cohort, in the spirit of reference generators for learning-analytics infrastructure [20].

The design targets teacher education first. The 2018 national English Language Teaching (ELT) programme of the Turkish Council of Higher Education (YÖK) [21] was introduced as a common curriculum from the 2018–2019 intake; since a 2020 decision, universities may adapt courses, curricula, and credits within its three course categories [22]. CurricuMap ships five taxonomies: an ELT taxonomy written from the national programme and tested on real data (Section 3.2), and four compact illustrations, not tested on real exports, for mathematics education, engineering accreditation, medicine (CanMEDS), and the descriptors of the Bologna Process and the European Qualifications Framework. This paper describes CurricuMap v0.2.0.

Table 2
CurricuMap and the closest tools, from the cited documentation (accessed 1 October 2026) and, for ouladFormat and Curricular Analytics, their source code. Mapping provenance: which rule or decision assigned each course; matrix audit: coverage and reliability statistics of the student-by-domain matrix; "not documented": the public documentation does not describe the feature.

Tool	Form	Student data used	Course-to-domain mapping	Mapping provenance	Matrix audit	Synthetic data
Assessment platforms: HelioCampus [4], eLumen [5], Watermark [6]	proprietary platforms	outcome-tagged rubric scores and student work from the LMS; some integrate with student-information systems	curriculum maps built in the platform	not documented	outcome dashboards; no reliability audit documented	not documented
Moodle competencies [7]	open-source LMS feature	teacher ratings or activity completion inside the LMS	teachers link competencies to courses and activities	not documented	not documented	not documented
ouladFormat [8]	open-source R package	one public dataset (the Open University Learning Analytics Dataset)	none	none	none	none
Curricular Analytics [9,10]	open-source Julia package	none (curricula and degree plans)	none (prerequisite graphs)	none	none	student-flow simulation
CurricuMap	**open-source Python package and command-line tool**	**CSV or XLSX registrar exports, one row per grade record**	**declarative YAML rules and overrides**	**per course**	**coverage, descriptives, Cronbach's alpha**	**seeded generator**

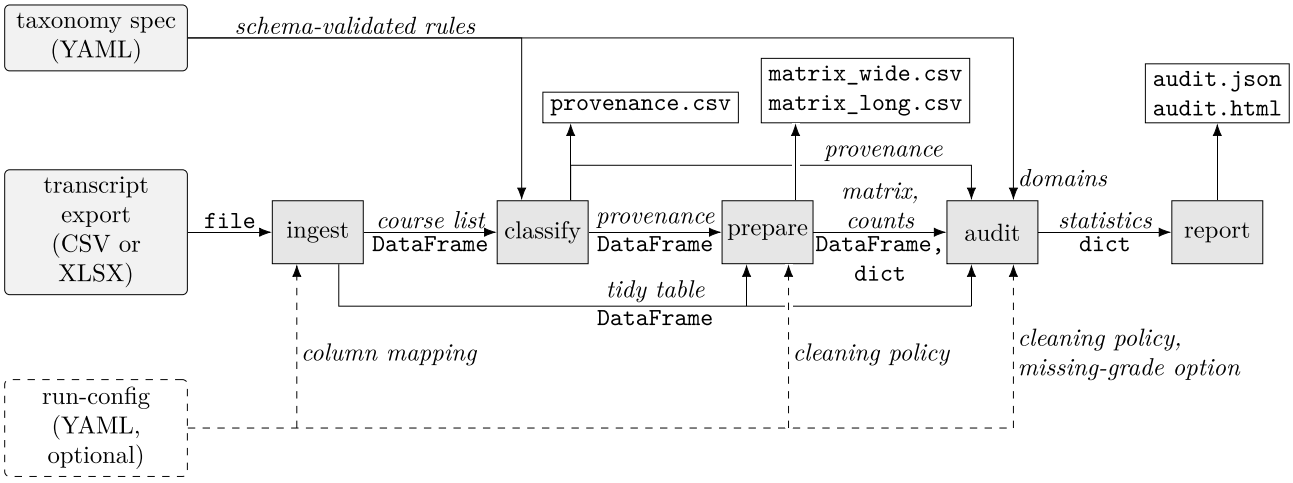


Fig. 1. Data flow. A transcript export and a taxonomy specification are required; a run-configuration file (dashed) is optional and supplies the column mapping, the cleaning policy, and the audit's missing-grade option. Arrow labels give what each arrow carries (italic) and its in-memory format (monospace). Classify reads only the course columns of the tidy table; prepare and audit read all of it, and audit re-applies the cleaning so that Cronbach's alpha uses the course grades.

2. Software description

2.1. Software architecture

CurricuMap is library-first: every processing step is an importable function, and a thin command-line interface (CLI) wraps the common workflows. The package has six modules: `taxonomy` (specification loading, schema validation, and text matching), `classify`, `prepare` (ingest, cleaning, and aggregation), `audit` (statistics and reports), `synth`, and `cli`. Fig. 1 shows the data flow. The ingest stage reads a CSV or XLSX export into a tidy table [23] with one row per grade record (student, course, and attempt) and the required columns `student_id`, `course_name`, and `grade`. An optional run-configuration file maps any registrar's column names onto this schema, sets the cleaning policy, and chooses how the audit treats missing grades, so no source file is edited by hand.

2.2. Software functionalities

Classification with provenance. For every course, CurricuMap records the method (`override`, `rule`, `fuzzy`, or `unmatched`), the pattern and the text it matched, a score, and whether rules for more than one domain matched, so each matrix cell can be traced to the rules behind it: a derivation record of the kind that data-provenance standards formalise [24]. An optional fuzzy fallback (via `rapidfuzz`) assigns near-misses above a threshold. A course that matches nothing is flagged and listed in the audit rather than being given a default domain, and a course code that the export uses for more than one title is reported, because its grades are pooled under the domain of its first title.

Registrar cleaning. The `prepare` stage drops declared sentinel grades and records without a numeric grade, resolves retaken courses under a configurable policy (keep the maximum, the mean, or the latest attempt by date), and removes students who cover fewer domains than a minimum, reporting how many records and students each step removed. It then aggregates grades into per-student, per-domain means or medians and writes the matrix in wide and long forms.

Audit. The `audit` stage reports coverage (unmapped and ambiguous courses, courses per domain), descriptive statistics per domain, and Cronbach's alpha [25], computed with the domain's courses as items on the cleaned grades of the students in the delivered matrix. A course a student did not take is a missing item. By default it is replaced by the course mean, which lowers alpha as the share of missing grades grows; listwise or pairwise treatment can be chosen, and the report gives each domain's share of missing grades. Alpha assumes essentially tau-equivalent items and is a lower bound to reliability rather than evidence that a domain is unidimensional [26–28]; the report states these assumptions and shows alpha without a colour grade when it rests on fewer than 30 students, where even unrelated grades can reach.70. Results are written as `audit.json` and as a self-contained `audit.html` (Fig. 2).

Reproducibility. Classification, cleaning, and aggregation are deterministic, and the generator is seeded. Following common practice for scientific software [29], 206 tests with 99% statement coverage are run in continuous integration, including regression tests that run Listing 2 verbatim and pin the results of Section 3.2.

2.3. Sample code snippets analysis

The reuse contract is the taxonomy specification (Listing 1). It names the domains and an ordered list of rules. Each rule lists substring keywords (`any`) and/or regular expressions (`regex`), which are matched against the case-folded course title, and the first matching rule wins, so specific rules can precede general ones. Overrides assign exact titles and take precedence over rules. CurricuMap validates the file against a JSON Schema that rejects unknown keys, checks that every rule and override refers to a defined domain and that every regular expression compiles, and `curricumap validate-taxonomy` reports any errors. Adopting the tool for a new curriculum means writing or adapting a specification, not code.

3. Illustrative examples

3.1. A synthetic walkthrough

The bundled ELT taxonomy has five domains: Language Skills, Pedagogy, Linguistics, Literature, and General Education. Because real transcripts are personal data, the walkthrough uses the synthetic generator, so it can be reproduced exactly on any machine. After `pip install .` in a clone of the repository, two commands run the full pipeline from any directory (Listing 2): the first synthesises a transcript export for 60 students, and the second classifies, cleans, and audits it. A bundled taxonomy is named rather than given as a path.

With seed 7 the generator emits 878 grade records across 15 courses. Its sentinel zeros always accompany a real grade, so the default retake policy (keep the highest grade) removes them with the retakes; declaring the sentinel value in the run-configuration file matters when a zero is a student's only grade in a course, as in the bundled sample transcript. The run produces a 60-by-5 student-by-domain matrix. All 15 courses are

```yaml
taxonomy:
  id: elt_yok_2018
  label: "ELT Teacher Education Curriculum (YÖK 2018)"
  # ... locale: Turkish (language: tr)
  domains:
    - { id: language_skills, label: "Language Skills" }
    # ... pedagogy, linguistics, literature, general_ed
  # ... match: fuzzy matching off
  rules:
    # ... rule 1: general_ed (Atatürk, Türk Dili, ...)
    - domain: literature
      any:
        - edebiyat
      regex:
        - '\bedeb(?![iı]l)'
        # ... one more
    # ... linguistics, language skills (2), pedagogy (öğretim)
    - domain: pedagogy
      any:
        - eğitimde
        # ... four more
      regex:
        - '(?<!beden )(?<!m[üu]z[iı]k )\be[gğ][iı]t[iı]m(e|de|[iı]n)?\b'
        # ... six more
    # ... a third pedagogy rule; general_ed electives
  overrides: []
  # ... defaults: unmatched courses are flagged
```

Listing 1. Excerpt of the bundled ELT taxonomy (`elt_yok_2018.yaml`); lines starting with `#...` replace omitted lines, and inline comments are removed. Rules are evaluated in order; the pedagogy expression accepts *eğitim* bare or with the endings *-e*, *-de*, and *-in*, but not the possessive form in *Beden Eğitimi* (physical education). The full file has 9 rules with 53 keywords and 63 regular expressions (Table 5).

```
curricumap synth --taxonomy elt_yok_2018 --out data \
    --n-students 60 --seed 7
curricumap run --input data/transcript.csv --taxonomy elt_yok_2018 \
    --out out
```

Listing 2. Running the illustrative example.

classified by rule, none are ambiguous and none are unmapped, and every decision, with its matched pattern, is recorded in `provenance.csv`. With $k = 3$ courses per domain and 6.7–15.0% of course grades missing, Cronbach's alpha is.769 for Language Skills,.767 for Pedagogy,.761 for Linguistics,.819 for Literature, and.781 for General Education (Fig. 2).

Table 3 reports run time and peak memory of `curricumap run` on synthetic transcripts from 60 to 100,000 students, with catalogues of 15 or 150 courses. Below about 10^5 grade records the time is dominated by interpreter start-up; the largest transcript, with 1.5 million records, is processed in 3.19 s using 412 MB of memory, so a programme's full history fits on a laptop.

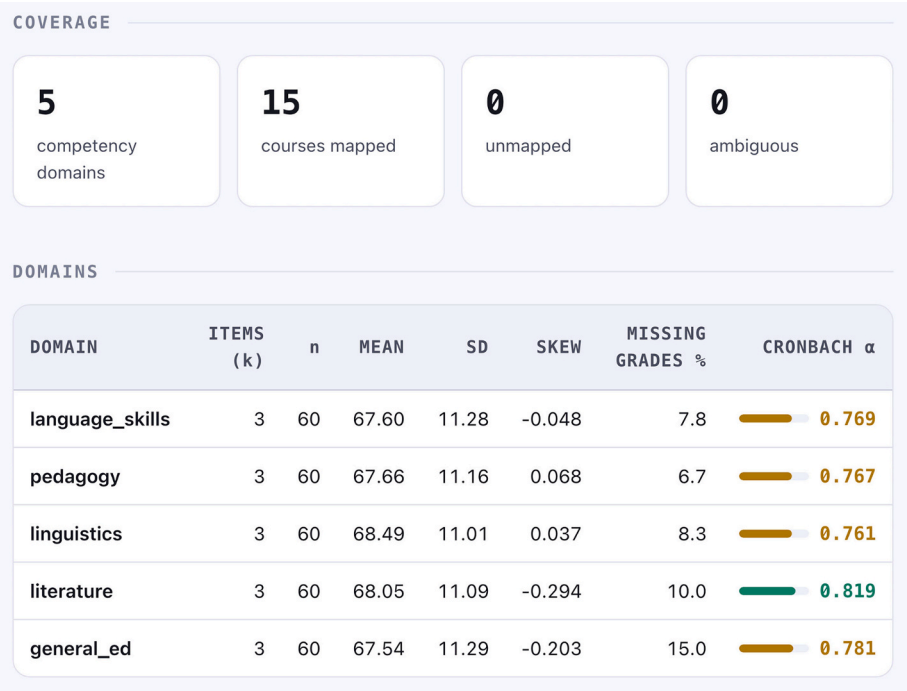
Table 3
Run time (median of three runs) and peak resident memory for `curricumap run` on synthetic transcripts (generator seed 1). Each run is a fresh Python process, so the times include interpreter start-up and imports (about 0.5 s). Hardware: Apple M2 Max with 32 GB of memory, macOS 26, Python 3.13, Pandas 3.0.

Students	Courses	Grade records	Run time (s)	Peak memory (MB)
60	15	872	0.51	81
1000	15	14,716	0.53	87
10,000	15	146,892	0.77	138
100,000	15	1,469,156	3.19	412
60	150	8778	0.53	86
1000	150	145,912	0.71	124
10,000	150	1,459,146	2.42	385

3.2. Validation on a real registrar catalogue

We tested how a taxonomy written from curriculum documents assigns the courses from a real registrar export, what its rules add to a lookup of the labelled official titles, and how NLP and LLM classifiers given the same information compare; data, prompts, model outputs, and scripts are in the repository's `validation` directory.

Data and reference. The test set is the 143 distinct course titles in the 2017–2026 transcript export of the ELT programme at İstanbul University-Cerrahpaşa, spelled as exported, in upper case with numerals and abbreviations. After CurricuMap's normalisation, 128 titles are identical to an entry of the 2018 national programme [21] or its 2006 predecessor [30], 10 are other renderings of an entry, and 5 are in neither programme (three local or elective courses and two versions of



DOMAIN	ITEMS (k)	n	MEAN	SD	SKEW	MISSING GRADES %	CRONBACH α
language_skills	3	60	67.60	11.28	-0.048	7.8	0.769
pedagogy	3	60	67.66	11.16	0.068	6.7	0.767
linguistics	3	60	68.49	11.01	0.037	8.3	0.761
literature	3	60	68.05	11.09	-0.294	10.0	0.819
general_ed	3	60	67.54	11.29	-0.203	15.0	0.781

Fig. 2. Coverage tiles and domain table of the `audit.html` from the illustrative run (ELT taxonomy, seed 7, 60 synthetic students; labels enlarged for print). The bar and value of α are green for $\alpha \geq .80$, amber for $.70 \leq \alpha < .80$, and red for $\alpha < .70$; a negative α shows an empty bar with a note, and an α resting on fewer than 30 students is shown in grey.

an official course for international students). Reference domains follow a fixed convention. Teaching-profession courses are Pedagogy, and general-culture courses are General Education; these two rules follow from the official category, which the programmes or the registrar's course codes record, and settle 79 titles. Field courses go to the domain of their subject matter, and English-language-teaching methodology courses to Pedagogy; these 62 decisions are ours. A 2006 second-foreign-language course is General Education like its 2018 successor, and a university-wide elective is General Education, a rule we added when labelling the test set. One official title, *Türk Eğitim Tarihi* (History of Turkish Education), is a teaching-profession course in 2018 and a general-culture course in 2006; a title-level specification can follow only one programme, and the test set labels it by 2018. The convention, the reference labels, and the domain definitions were drafted with the help of an AI assistant (Claude, Anthropic).

Systems. The bundled taxonomy was drafted by an LLM coding assistant (Claude Opus 5.5 in Claude Code) from the convention and the 158 labelled entries (150 distinct titles) of the two national programmes, critiqued by three further instances (for over-fitting, false positives, and fidelity to the convention), and revised by a fifth, in 85 min; by design, no author edited or reviewed the rules before they were frozen. The released tool-call logs, which replay to the frozen file, show that no agent opened the catalogue, which had been moved out of the repository. Their instructions, however, were written by an assistant who had seen it: they quote one registrar variant verbatim and name two hypothetical courses resembling catalogue titles, and the reviser probed a second variant verbatim, so neither the variant nor the unseen titles is a fully blind test. The drafter, which saw none of these, had already assigned all exact and variant titles correctly, and three of the five unseen ones. Claude Code also placed a memory index and a file listing from the first author's private workspace in each agent's context; neither contains a course title, and no agent opened a listed file. The specification was frozen and applied once. The alternatives receive the labelled official titles (L), definitions of the five domains (D), or both. An exact-title lookup runs the labelled titles in CurricuMap as overrides; string similarity (`rapidfuzz`) and multilingual-E5 embeddings [31,32] assign the label of the most similar title; a character 2–5-gram logistic regression [33] is trained on them; E5 zero-shot picks the most similar definition. Two LLMs classify every title zero-shot (D) or few-shot (L + D): Gemma 4 12B [34], an open-weight model run locally in 4-bit quantisation, greedily and in five sampled runs, and Claude Opus 5.5 [35], a proprietary model run five times through its command-line client on 1 October 2026. The definitions were also written with knowledge of the catalogue, which can only favour the systems that receive them.

Results. The frozen taxonomy assigned all 143 titles by rule, each to its reference domain (Table 4). For 10 compound titles, such as "Dil ve Edebiyat Öğretimi" (teaching language through literature), a subject rule and a teaching rule both matched; rule order resolved each correctly, and the ambiguity flag marked them for review. The exact lookup assigned the 128 exact titles and flagged the other 15 without error, so the rules contributed the 10 registrar variants (a lookup that also ignored punctuation, spaces, and diacritics would recover four of them) and the five unseen titles, whose general stems (*öğretim*, *Atatürk*, *kültür*) placed them correctly but without a flag. The other methods given the labelled titles reached 98.6–100.0%, none significantly different from CurricuMap ($p \geq 0.500$). Without them, E5 zero-shot reached 68.5% and the open-weight LLM 91.6% (sampled runs 88.1–94.4%, with identical labels for only 90.9% of the titles), both below CurricuMap ($p < 0.001$ greedy; $p \leq 0.008$ for each sampled run). The proprietary LLM reached 98.6% zero-shot ($p = 0.500$ against CurricuMap), with identical labels in all five runs; its two errors were general-culture courses of the 2006 programme, such as "Drama", whose category only the programme documents can settle. The taxonomy released with v0.1.0, written for synthetic data, assigned only 49.0% of the titles and 40.6% correctly.

Specification size. The validated ELT taxonomy has 9 rules with 53 keywords, 63 regular expressions, and no overrides in 159 lines of YAML (Table 5). Keywords alone correctly assign 113 of the 158 official entries. Regular expressions assign all but one of the rest, mostly because a plain stem would also take in look-alike courses (*eğitim*, education, which also occurs in *Beden Eğitimi*, physical education); the remaining entry, the

Table 4
Classification of the 143 registrar titles. Input: labelled official titles (L), domain definitions (D), both, or neither (–). Cov.: titles assigned a domain (CurricuMap flags the rest). Acc.: titles assigned the reference domain (flags count as errors), with a Wilson 95% interval; with the 114 title stems as units (numbered courses share one), the bound for 100% is 96.7%. Correct titles per stratum (n): exact official titles, other registrar renderings, titles in neither programme. p: exact McNemar test against the first row; a difference of five titles or fewer cannot reach $p < 0.05$. String similarity breaks ties by the plain ratio, a rule set after tie errors in its first run were seen. Gemma: greedy decoding; Claude: five runs with identical labels.

System	Input	Cov. (%)	Acc. (%)	95% CI (%)	Correct titles			p
					Exact (128)	Variant (10)	Unseen (5)	
CurricuMap rules (v0.2.0)	L	100.0	100.0	97.4–100.0	128	10	5	–
CurricuMap exact lookup	L	89.5	89.5	83.4–93.5	128	0	0	<0.001
CurricuMap, v0.1.0 taxonomy	–	49.0	40.6	32.9–48.8	50	4	4	<0.001
String similarity	L	100.0	98.6	95.0–99.6	128	10	3	0.500
Character n-grams	L	100.0	100.0	97.4–100.0	128	10	5	1.000
E5, nearest title	L	100.0	100.0	97.4–100.0	128	10	5	1.000
E5, zero-shot	D	100.0	68.5	60.5–75.6	92	3	3	<0.001
Gemma 4 12B, zero-shot	D	100.0	91.6	85.9–95.1	118	8	5	<0.001
Gemma 4 12B, few-shot	L+D	100.0	100.0	97.4–100.0	128	10	5	1.000
Claude Opus 5.5, zero-shot	D	100.0	98.6	95.0–99.6	126	10	5	0.500
Claude Opus 5.5, few-shot	L+D	100.0	100.0	97.4–100.0	128	10	5	1.000

Table 5
Size of the bundled taxonomy specifications; YAML lines exclude blank and comment lines. `elt_yok_2018_v0.1.0` is the ELT taxonomy released with v0.1.0.

Taxonomy	Domains	Rules	Keywords	Regexes	Overrides	YAML lines
`bologna`	3	3	14	0	0	19
`canmeds`	7	7	27	0	0	31
`elt_yok_2018`	5	9	53	63	0	159
`engineering_mudek`	5	5	40	0	0	25
`math_ed_yok`	4	4	23	0	0	22
`elt_yok_2018_v0.1.0`	5	5	34	0	0	25

2006 *Türk Eğitim Tarihi*, needs an override. Most other regular expressions accept registrar renderings, such as ASCII spellings (`OGRETIM`) and abbreviations (`ÖĞR.`).

What the comparison shows. Given the labelled official list, every method that generalises beyond exact titles was about equally accurate on this catalogue (98.6–100.0%); only the exact lookup, which flags every title it has not seen, fell short (89.5%). Accuracy therefore does not separate CurricuMap from those methods; they differ in what they leave behind. CurricuMap executes and records whichever mapping is chosen: rules generalise to registrar renderings at a cost in readability, overrides record single decisions, and an LLM or a nearest-neighbour lookup can propose either. A hosted LLM can change or be withdrawn, identical requests can return different answers [36,37], and its stated reasons are not guaranteed to be faithful or reproducible. A reviewed specification, versioned with the study, states every decision, runs identically without sending data anywhere, and is cited with the software.

Limitations. The test set comes from one institution that follows the national curriculum closely: 128 of its 143 titles are official titles in upper case, which any method given the labelled list recognises, and only five are outside both programmes. The reference for the 62 field courses reflects our convention; no independent experts labelled the titles, so expert agreement was not measured. Rules match stems, so an off-curriculum course containing one is assigned rather than flagged: *Rus Dili Dilbilgisi* (Russian grammar) would go to Language Skills through *dilbilgisi*, and the provenance table, not the absence of flags, is what a reviewer must check. The rules were written by LLM agents and should be reviewed by a curriculum expert before reuse. The evaluation scores course titles, not the matrices built from them; matrix construction is tested on synthetic data.

4. Impact

CurricuMap removes a bottleneck in competency-level learning analytics: turning records that institutions already hold into a matrix that can be analysed and reported. For existing research questions, the main gain is that the preprocessing becomes part of the evidence: a study can cite its taxonomy file and software version, and reviewers can inspect every course assignment in the provenance table. Because a specification is a small file, a study can also rebuild its matrix under alternative course mappings and report how much its findings depend on them, in the spirit of multiverse analyses of learning-analytics pipelines [38]. Programmes that follow the same national curriculum can share a taxonomy and add only their local courses: the bundled taxonomy assigned to all 143 course titles of one institution without any local addition (Section 3.2).

Some questions may become practical, such as comparing competency profiles across institutions that share a national programme, or following the same domains across a curriculum reform. The bundled taxonomy covers both the 2006 and the 2018 Turkish ELT programmes except for the one course whose category changed between them, which needs an override for 2006 cohorts.

CurricuMap grew out of the preprocessing for two ongoing studies of English-language-teacher education by the authors, to be reported separately. Its repository was made public in July 2026, and v0.1.0 was released in August 2026. The repository, its Zenodo archive, and the open validation data allow every result here to be rerun or, for the hosted LLM, recomputed from the released outputs. We invite curriculum and learning-analytics researchers to contribute taxonomies for their own programmes.

5. Conclusions

CurricuMap converts a registrar transcript export and a declarative YAML taxonomy into an analysis-ready student-by-domain matrix, with per-course provenance, documented registrar cleaning, and an audit of coverage and reliability. Its contribution is not a new classifier but an explicit, versioned, and testable record of the course-to-domain mapping, executed the same way every time on local data. In a real registrar catalogue, an LLM-drafted taxonomy frozen before testing assigned all 143 course titles as its conventions prescribe, as did the best alternatives given the same labelled course list, and each decision can be traced to a rule that a reader can check. Future work will validate taxonomies for further programmes against independent expert labels and support LLM-assisted drafting in which an expert approves each rule.

CRediT authorship contribution statement

Tarık Küçükdeniz: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Tuncer Can:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The source code, the five taxonomies, and the validation data (course titles, registrar course codes, official categories, reference labels, prompts, raw model outputs, and all predictions) are openly available under the MIT licence (Table 1). Student-level transcript records are confidential and are not shared; the synthetic generator reproduces every example without them.

References

[1] Kovanović V, Gašević D, Dawson S, Joksimović S, Baker RS, Hatala M. Does time-on-task estimation matter? Implications for the validity of learning analytics findings. J Learn Anal 2015;2(3):81–110. https://doi.org/10.18608/jla.2015.23.6

[2] Haim A, Shaw S, Heffernan N. How to open science: A principle and reproducibility review of the Learning. In: Analytics and Knowledge Conference, in: LAK23: 13th International Learning Analytics and Knowledge Conference. New York, NY, USA: ACM; 2023. p. 156–64. https://doi.org/10.1145/3576050.3576071

[3] Haim A, Gyurcsan R, Baxter C, Shaw ST, Heffernan NT. How to open science: Debugging reproducibility within the Educational Data Mining Conference. In: Feng M, Käser T, Talukdar P, editors. Proceedings of the 16th International Conference on Educational Data Mining. Bengaluru, India: International Educational Data Mining Society; 2023. p. 114–24. https://doi.org/10.5281/zenodo.8115651

[4] HelioCampus. Assessment management; curriculum mapping; 2026 https://www.heliocampus.com/, https://www.heliocampus.com/products/assessment-management, https://www.heliocampus.com/use-case/curriculum-mapping [accessed 1 October 2026].

[5] eLumen. Insights for Canvas outcomes; 2026 https://www.elumenconnect.com/insights [accessed 1 October 2026].

[6] Watermark Insights. Assessment projects; planning & self-study; curriculum strategy; 2026 https://www.watermarkinsights.com/solutions/outcomes-assessment-software/, https://www.watermarkinsights.com/solutions/planning-and-self-study/ https://www.watermarkinsights.com/solutions/curriculum-strategy/ [accessed 1 October 2026].

[7] Moodle C. MoodleDocs (Moodle 5.2 documentation); 2026 https://docs.moodle.org/en/Competencies [accessed 1 October 2026].

[8] Howard E. ouladFormat R package: Preparing the Open University Learning Analytics Dataset for analysis, version 2. arXiv:2501.08366. 2025. https://doi.org/10.48550/arXiv.2501.08366 [16 January 2025].

[9] Heileman GL, Abdallah CT, Slim A, Hickman M. Curricular analytics: A framework for quantifying the impact of curricular reforms and pedagogical innovations. arXiv:1811.09676. 2018. https://doi.org/10.48550/arXiv.1811.09676

[10] Heileman GL, Free H. CurricularAnalytics.jl: A toolbox for studying and analyzing academic program curricula; 2026 https://github.com/CurricularAnalytics/CurricularAnalytics.jl [Accessed 1 October 2026]. version 2.1.0, AGPL-3.0.

[11] Smith HR, Chittams J. Defining best practices and validation for curriculum mapping. Cogent Educ 2024;11(1):2342662. https://doi.org/10.1080/2331186X.2024.2342662

[12] Zamecnik A, Barthakur A, Wang H, Dawson S. Mapping employable skills in higher education curriculum using LLMs. In: Mello RF, Rummel N, Jivet I, Pishtari G, Valiente JAR, editors. Technology Enhanced Learning for Inclusive and Equitable Quality Education: 19th European Conference on Technology Enhanced Learning, EC-TEL 2024, Proceedings, Part II, Vol. 15160 of Lecture Notes in Computer Science. Cham: Springer; 2024. p. 18–32. https://doi.org/10.1007/978-3-031-72312-4_2

[13] Xu Z, Guan X, Shi C, Chen Q, Yu R. Evaluating 21st-century competencies in postsecondary curricula with large language models: Performance benchmarking and reasoning-based prompting strategies. J Learn Anal 2026;13(1):7–30. https://doi.org/10.18608/jla.2026.9127

[14] Paulson A, Stange K, Flaster A. Classifying courses at scale: A text as data approach to characterizing student course-taking trends with administrative transcripts, EdWorkingPaper 24-1042. Annenberg Institute at Brown University; 2024. https://doi.org/10.26300/7fpa-s433

[15] Republic of Türkiye. Personal Data Protection Law No. 6698 [Kişisel Verilerin Korunması Kanunu], Resmî Gazete (Official Gazette of the Republic of Türkiye), No. 29677, 2016 https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=6698&MevzuatTur=1&MevzuatTertip=5 [7 April 2016].

[16] European Parliament and Council of the European Union. Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Off J Eur Union 2016;L 119, pp. 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj [4 May 2016].

[17] Slade S, Prinsloo P. Learning analytics: Ethical issues and dilemmas. Am Behav Sci 2013;57(10):1510–29. https://doi.org/10.1177/0002764213479366

[18] Drechsler J. Synthetic datasets for statistical disclosure control: theory and implementation, Vol. 201 of Lecture Notes in Statistics. New York, NY: Springer; 2011. https://doi.org/10.1007/978-1-4614-0326-5

[19] Liu Q, Khalil M, Jovanovic J, Shakya R. Scaling while privacy preserving: A comprehensive synthetic tabular data generation and evaluation in learning analytics. In: Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24). ACM; 2024. p. 620–31. https://doi.org/10.1145/3636555.3636921

[20] Berg AM, Mol ST, Kismihók G, Sclater N. The role of a reference synthetic data generator within the field of learning analytics. J Learn Anal 2016;3(1):107–28. https://doi.org/10.18608/jla.2016.31.7

[21] Council of Higher Education (YÖK). İngilizce öğretmenliği lisans programı [English language teaching undergraduate programme], in: Yeni Öğretmen Yetiştirme Lisans Programları [New Teacher Training Undergraduate Programmes], Council of Higher Education, Türkiye; 2018 https://egitim.yok.gov.tr/tr/document/2440 [Accessed 1 October 2026]. in Turkish. Implemented from the 2018–2019 academic year.

[22] Council of Higher Education (YÖK). Pedagojik formasyon eğitimi: Öğretmen yetiştirme geçmişi [Pedagogical formation: History of teacher training]; 2020 https://egitim.yok.gov.tr/tr/page/496 [Accessed 1 October 2026]. in Turkish. Records the YÖK General Assembly decision of 10 August 2020 authorising higher education institutions to determine the courses, curricula and credits of teacher-education programmes.

[23] Wickham H. Tidy data. J Stat Softw 2014;59(10):1–23. https://doi.org/10.18637/jss.v059.i10

[24] Moreau L, Missier P. PROV-DM: The PROV data model, W3C Recommendation; 2013 https://www.w3.org/TR/prov-dm/ [30 April 2013].

[25] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297–334. https://doi.org/10.1007/BF02310555

[26] Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993;78(1):98–104. https://doi.org/10.1037/0021-9010.78.1.98

[27] Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 2009;74(1):107–20. https://doi.org/10.1007/s11336-008-9101-0

[28] McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2018;23(3):412–33. https://doi.org/10.1037/met0000144

[29] Wilson G, Aruliah DA, Brown CT, Hong NPC, Davis M, Guy RT, Haddock SHD, Huff KD, Mitchell IM, Plumbley MD, Waugh B, White EP, Wilson P. Best practices for scientific computing. PLOS Biol 2014;12(1):e1001745. https://doi.org/10.1371/journal.pbio.1001745

[30] Council of Higher Education (YÖK). İngilizce öğretmenliği lisans programı [English language teaching undergraduate programme]. In: Eğitim Fakültesi Öğretmen Yetiştirme Lisans Programları (2006–2007), Council of Higher Education, Türkiye, in Turkish. Adopted on 21 July 2006 and applied from the 2006–2007 academic year; final version approved on 2 February 2007. (2006. https://egitim.yok.gov.tr/tr/document/522 [Accessed 1 October 2026].

[31] Wang L, Yang N, Huang X, Yang L, Majumder R, Wei F. MultilingualE5text embeddings: A technical report. arXiv:2402.05672. 2024. https://doi.org/10.48550/arXiv.2402.05672

[32] Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics. Hong Kong, China; 2019. p. 3982–92. https://doi.org/10.18653/v1/D19-1410

[33] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher

M, Perrot M, Duchesnay É. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12(85):2825–30.

[34] Gemma Team. Gemma 4 technical report. arXiv:2607.02770. 2026. https://doi.org/10.48550/arXiv.2607.02770

[35] Anthropic. System card: Claude Opus 5.5; 2026 https://www.anthropic.com/claude-opus-5-5-system-card [Accessed 1 October 2026]. dated 22 September 2026.

[36] Atılıl B, Aykent S, Chittams A, Fu L, Passonneau RJ, Radcliffe E, Rajagopal GR, Sloan A, Tudrej T, Ture F, Wu Z, Xu L, Baldwin B. Non-determinism of "deterministic" LLM system settings in hosted environments. In: Proceedings of the 5th Workshop on Evaluation and Comparison of NLP Systems, Association for Computational Linguistics. Mumbai, India; 2025. p. 135–48. https://doi.org/10.18653/v1/2025.eval4nlp-1.12

[37] Reiss MV. Testing the reliability of ChatGPT for text annotation and classification: A cautionary remark. arXiv:2304.11085. 2023. https://doi.org/10.48550/arXiv.2304.11085

[38] Tang Y, Harvey E, Yao C, Yu R, Kizilcec RF, Brooks C. Understanding predictive models of student success with a multiverse analysis. In: Mills C, Alexandron G, Taibi D, Bosco GL, Paquette L, editors. Proceedings of the 18th International Conference on Educational Data Mining. Palermo, Italy: International Educational Data Mining Society; 2025. p. 518–25. https://doi.org/10.5281/zenodo.15870276