# Turning Course Descriptions into Skill Evidence: A Case Study of Two Undergraduate Programs

Xianghui Meng[1], Xian Chen[2], Jionghao Lin[1*]

[1] Faculty of Education, The University of Hong Kong

[2] The University of Manchester

[*] Corresponding author: jionghao@hku.hk

***Abstract:*** *Course descriptions often state broad descriptions of what students will learn without clearly naming specific skills (e.g., data analysis, policy analysis) or tools (e.g., Python, SQL). We present a course-to-skill extractor that converts course description into Tools, Hard Skills, and Soft Skills while extracting only labels explicitly supported by the course-description text. Using Qwen2.5 (dense), we analyze 60 course descriptions from two undergraduate programs: one in data science in a social science context and one in information systems and management. The extraction workflow preserves traceability by storing each extracted skill label with its source course, allowing program-level summaries to be checked against the original course descriptions. Across both programs, most labels appear only once, while explicit tool names are rare.*

**Keywords:** curriculum analysis, curriculum mapping, large language models, higher education, skill extraction

## 1. Introduction

When choosing courses, students rely on official course descriptions to assess program skills. However, these texts often use less detailed language. For example, a job posting lists Python or SQL directly, but a course description may say "analytical methods." This mismatch can lead students to over-infer or miss relevant skills entirely, because a broad phrase such as "analytical methods" does not specify whether the course involves tools such as Excel, Python, or SQL, or methods such as statistical analysis or data modeling. Prior work has explored curriculum analysis through natural language processing (Rashtian et al., 2020) and large-scale skill mapping across higher education curricula (Sabet et al., 2024), yet fewer studies focus on extracting student-facing skill evidence from short course descriptions in ways that can be checked against the original descriptions. We therefore propose a structured extraction workflow that identifies explicitly stated skill labels, categorized into *Tools*, *Hard Skills*, and *Soft Skills*, with every label linked to its source course for full auditability.

## 2. Method

The proposed workflow consisted of five steps. *First*, we collected officially published course descriptions for the two undergraduate programs and treated each course description as one unit of analysis. *Second*, we processed each course description independently with Qwen2.5 (dense). *Third*, the model extracted only skill labels explicitly stated in the course-description text and organized them into three categories: Tools, Hard Skills, and Soft Skills. *Fourth*, we stored each extracted skill label together with its source course to preserve traceability. *Fifth*, we aggregated the course-level outputs into program-level summaries and compared the two programs mainly through recurring skill labels and relative label frequencies.

The dataset contains 60 officially published course descriptions from a university in Hong Kong in the 2025–2026 academic year: 41 from Program 1 (data science in a social science context) and 19 from Program 2 (information systems and management). Each record includes a program label, course code, course title, and description text. The descriptions are similarly short across both programs (mean length about 73 words), reducing the likelihood that observed differences are driven primarily by description length. Each description was processed independently with `Qwen2.5` (dense). We

used a structured prompt with a fixed JSON schema to guide Qwen2.5 in generating skill extractions (tools, hard skills, and soft skills) from course descriptions. The prompt enforced a fixed JSON schema with three output lists: *Tools* (named technologies such as software or programming languages), *Hard Skills* (domain-specific technical methods or competencies), and *Soft Skills* (transferable capabilities such as communication or teamwork). The model was instructed to extract only explicitly stated labels and return empty lists when no supporting evidence was available. A low temperature setting (0.2) was used to improve stability, and outputs were replaced with empty lists if JSON parsing failed. Course-level outputs were then converted into a long-format course-to-skill table for counting, aggregation, and auditing. For program comparison, we treated extracted skills as curriculum-as-described evidence, interpreting cross-program differences mainly through recurring labels and size-normalized mention frequencies.

## 3. Results

The workflow produced complete outputs for all 60 courses and extracted 348 skill items in total: 239 from Program 1 and 109 from Program 2. As shown in Table 1, mean yield was highly similar across programs (5.83 vs. 5.74 items per course), suggesting that the observed differences are not simply a result of one program having longer or denser descriptions. The table also shows a strong singleton pattern in both programs and limited explicit tool naming, with only 19.5% of Program 1 courses and 15.8% of Program 2 courses naming tools. Together, these results suggest that short course descriptions provide many one-off labels but relatively little direct information about specific tools.

**Table 1**

*Summary of results by program. Details of the metrics can be found in project repository.*[1]

| **Metrics** | **Program 1** | **Program 2** |
|---|---|---|
| Courses analyzed | 41 | 19 |
| Total extracted skill items | 239 | 109 |
| Mean extracted skill item per course | 5.83 | 5.74 |
| Distinct extracted skill labels | 188 | 100 |
| Singleton share(%) | 87.2 | 96.0 |
| Courses with at least one tool (%) | 19.5 | 15.8 |

The extracted skill labels showed a strong singleton pattern. Program 1 contained 188 distinct skill labels and Program 2 contained 100, but 87.2% and 96.0% of these labels, respectively, appeared only once within each program. This suggests that comparing programs using all extracted labels may overstate how broad or different the programs appear. We therefore compared the programs mainly through skill labels that appeared in more than one course. When labels were matched without distinguishing uppercase and lowercase letters, the two programs shared 56 labels. Within this shared set, Program 1 more often documented “Critical Thinking,” “Policy Analysis,” and “Data Analysis,” whereas Program 2 more often documented “Information Management.” A complete list of distinct skill labels for both programs is available in the project repository.[1] The Appendix in the repository provides the details for each reported metric, including exact definitions and equations. Explicit tool naming was limited in both programs, with tools named in 19.5% of Program 1 courses and 15.8% of Program 2 courses. This means that the course descriptions often did not clearly state which tools were used; it does not mean that tools were unimportant in teaching. When we compared the programs using relative

frequencies instead of raw counts, we reached the same conclusion: the clearest differences between the programs came from skill labels that appeared repeatedly across courses, rather than from labels that appeared only once.

## 4. Discussion and Conclusion

Our study shows that short official course descriptions can be converted into structured skill evidence without overstating what the text supports. The main strength of this workflow is not deep interpretation, but careful evidence handling. It extracts only information explicitly stated in the text, keeps the source course for each label, and compares programs based on labels supported repeatedly across courses. For students, the extracted skill summaries make heterogeneous course descriptions easier to inspect and compare. For program teams, these summaries can reveal where course descriptions do not clearly specify expected tools or rely on broad wording with limited practical guidance. These results describe only what is explicitly written in the course descriptions. They should be used to support follow-up review of course documents, not to draw conclusions about what was actually taught, what students mastered, or how well the curriculum matches labor-market needs. The study also shows why it is important to extract only skill labels that are explicitly supported by the course-description text.: allowing inferred skills into the dataset can produce summaries that are not strongly supported by the text. Future work could align near-synonymous labels through a controlled ontology, add human review for program differences, and analyze job advertisements with course descriptions to compare curriculum and labor-market skill profiles using the same categories.

**Acknowledgements**

This work was supported by the Teaching Development Grant (Project No. 1119) from the University of Hong Kong.

## References

Hilliger, I., Aguirre, C., Miranda, C., Celis, S., & Pérez-Sanagustín, M. (2022). Lessons learned from designing a curriculum analytics tool for improving student learning and program quality. *Journal of Computing in Higher Education*, 34(3), 633–657.

Rashtian, H., Hashemi, A., & Macfadyen, L. (2020). Harnessing natural language processing to support curriculum analysis. *In ICERI2020 Proceedings* (pp. 1779–1784). IATED.

Sabet, A. J., Bana, S. H., Yu, R., & Frank, M. R. (2024). Course-Skill Atlas: A national longitudinal dataset of skills taught in US higher education curricula. *Scientific Data*, 11(1), 1086.