Supplementary Information

1. Data Description

a. Word Count

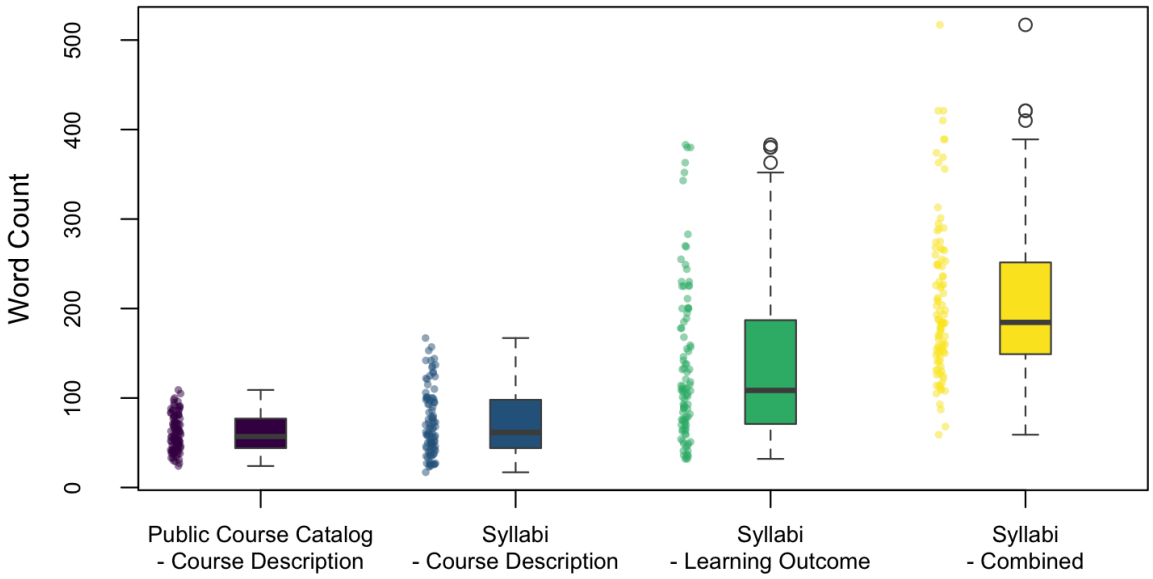

Fig.1. Word Count Boxplot by Course Data Type

The distribution of word counts varies markedly across course materials. Public course catalog descriptions exhibit the lowest median word count and minimal variability, indicating a standardized format. Syllabi course descriptions show slightly higher medians and broader spread, reflecting greater instructor-level customization. Learning outcomes extracted from syllabi display the widest range and highest median values, suggesting more extensive elaboration compared to simple descriptions. The combined section, aggregating description and learning outcomes, predictably exhibits the highest overall word counts and variability, confirming additive effects. Notably, the dispersion is particularly pronounced in the learning outcome and combined categories, with several extreme outliers above 400 words.

b. Field Distribution

In the public course catalog, business management and allied health sciences dominate field representation, each accounting for approximately 14–15 courses. STEM-related fields such as mathematics and science are also prominent, whereas fields associated with the humanities, such as media arts and ethnic studies, exhibit lower course counts. The distribution reflects an institutional emphasis on vocational

and technical domains over liberal arts fields. Notably, nursing and health education are underrepresented despite the prevalence of applied sciences overall.

Syllabi data reveal a more heterogeneous distribution across fields compared to the public catalog. Business remains highly represented, while disciplines such as biology, engineering, and mathematics exhibit substantial increases in course counts. Humanities and arts fields, including French, music, and journalism, are also more evenly represented. The broader dispersion of courses suggests that syllabi capture a wider range of instructional offerings than the official catalog. Applied sciences continue to form a significant component, particularly in nursing and computer science.

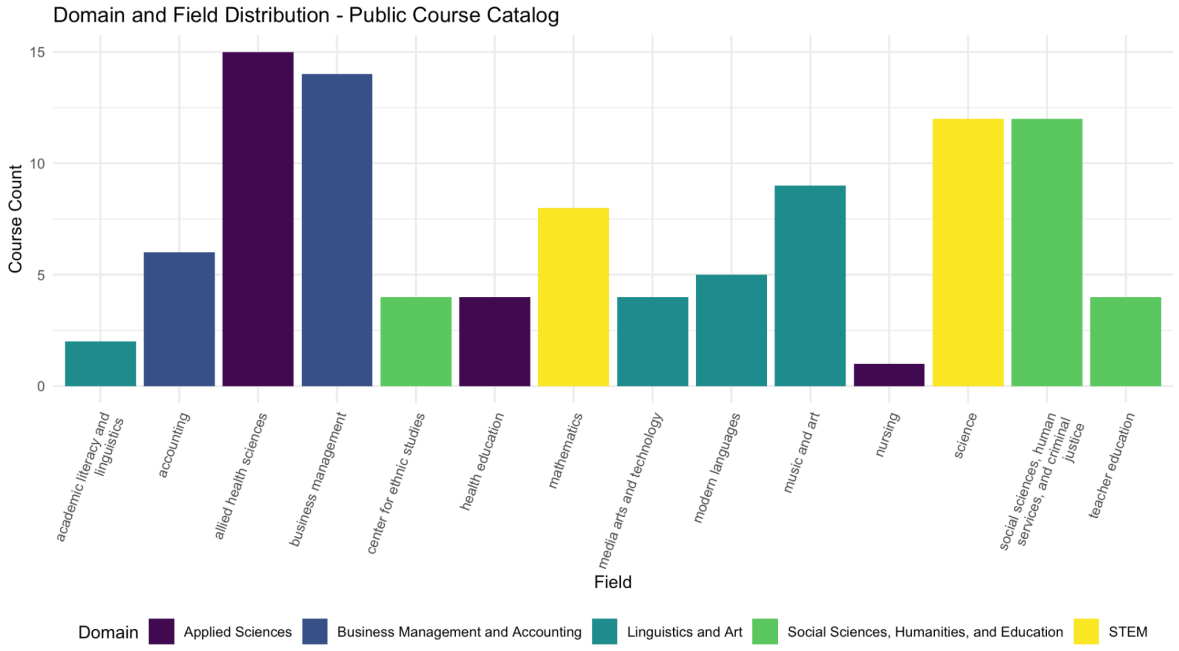

Fig.2. Domain and Field Distribution of Public Course Catalog

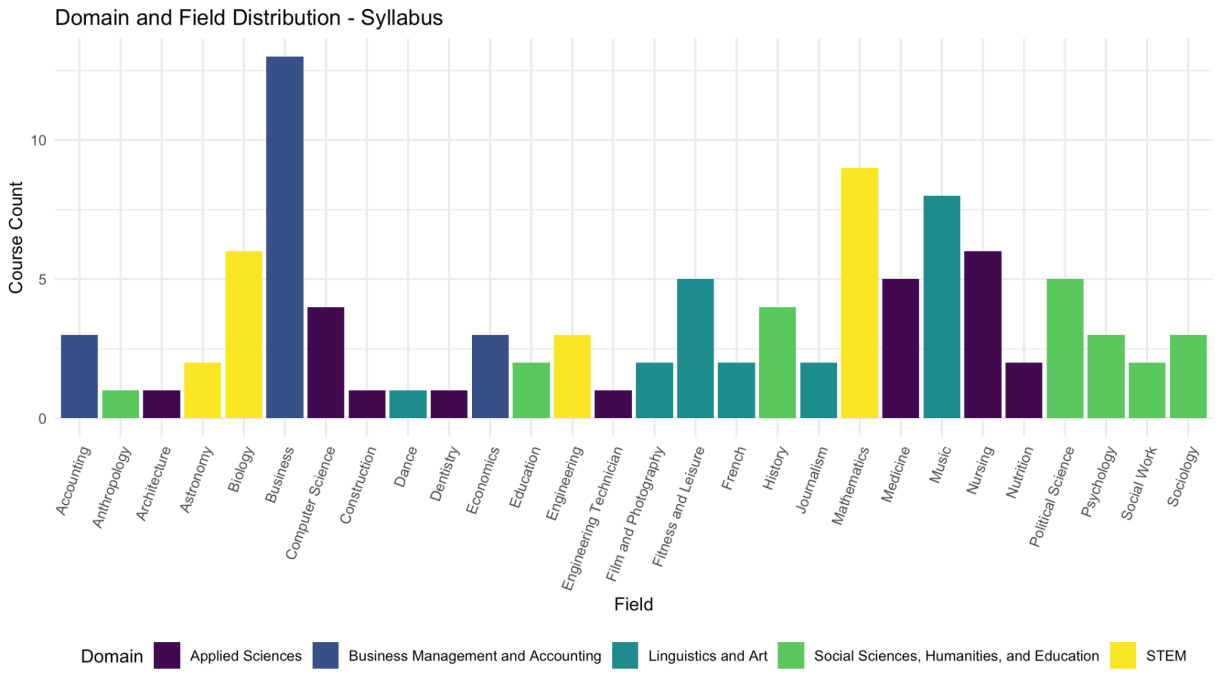

Fig.3. Domain and Field Distribution of Syllabus

c. Data Description & Example

Purpose	Content
Primarily designed for marketing and informational purposes, offering a concise overview of a course's focus, objectives, and appeal. They target a broad audience, including prospective students, the general public, and institutional administrators.	Catalog descriptions are often brief, generalized, and standardized, highlighting the course's subject matter and relevance without delving deeply into specifics.
Example: Engineering mechanics ii: This course is a three-dimensional vector treatment of the kinematics of rigid bodies using various coordinate systems. Topics include relative motion, particle dynamics, Newton's laws, energy, and mechanical vibrations. Elements of design are incorporated in the course.	

Table 1. Purpose, Content and Example of Public Course Catalog - Course Description

Purpose	Content
Tailored to enrolled students and focus on setting clear expectations for the course. They provide more detailed and actionable information about the course content, structure, and learning outcomes to enrolled students and instructors.	Syllabus course descriptions are often more specific, and aligned with the instructor's teaching plan, often including details about topics covered and the course's alignment with broader program goals.
Example: Theory of Automatic Controls: This course is an introduction to control systems. We first learn to write basic mathematical models, manipulate the transfer function block diagrams, and study their stability, and then predict the performance of a closed loop control systems through the systematic study of the servomechanism, including steady state and transient analysis, single loop systems analysis, root-locus and frequency response techniques, and design methods to meet performance specifications. Towards the last part of the course we will include methods for closed loop compensation. MATLAB computer simulations will be introduced to the students.	

Table 2. Purpose, Content and Example of Syllabus - Course Description

Purpose	Content
Learning outcomes articulate the specific skills, knowledge, and abilities students are expected to acquire by the end of the course, guiding teaching strategies, assessment design, and shaping student expectations.	Contain specific, measurable statements focused on what students will learn or be able to do, emphasizing student achievements over course content, typically categorized into knowledge-based, skills-based, and ability-based outcomes, and are written as clear, action-oriented statements.
Example:	

Deterministic Operations Research: Express a wide variety of problems as linear programs. Understand the simplex algorithm. Use sensitivity analysis to determine the range of applicability of solutions. Understand the relationship between a linear program and its dual. Model problems as integer programs. Understand the branch and bound technique for integer programs. Model problems as network flows. Understand flow augmentation algorithms and the max-flow min-cut algorithm. Understand game theory and decision theory models.

Table 3. Purpose, Content and Example of Syllabus - Learning Outcome

Purpose	Content
Integrates the course description and learning outcomes to provide a comprehensive view of both the course's thematic scope and its intended learning objectives. It captures the course's purpose, content, and expected student achievements in a unified narrative.	Reflects both what the course offers and what students are expected to gain, enabling richer analysis of course intent and alignment with educational goals. It serves as a holistic representation of course design for both instructional and research purposes.
Example: Statistics: Topics include probability, random variables, hypothesis testing, confidence intervals, correlation, linear regression, small sample\nmethods, and non\xadparametric statistics. Define population and sample. Describe a set of quantitative data by computing measures of central tendency (mean, median, and mode) and measures of dispersion (range,\nvariance, standard deviation). Calculate probabilities using the addition law of probability, law of complement, multiplication law, and conditional probability. Solve probability problems using counting techniques. Construct a probability distribution for a discrete random variable and compute the expected value (mean), variance, and standard deviation of\nthis distribution. Construct a binomial probability distribution and compute the mean, variance and standard deviation. Determine probabilities using the z\xadtables (Z is a standard normal variable). Use the Central Limit Theorem to find the mean and variance of the sampling distribution of the sample mean. Use a hypothesis test to determine whether to accept or reject the null hypothesis: a statement, assertion, or claim about the nature of a\npopulation. Calculate the coefficient of linear regression (correlation	

coefficient) for bivariate data. Construct a least squares regression line to bivariate data. Use the equation of the regressions line to predict a particular value of y for a specific value of x. Utilize random samples of data to draw inferences and conclusions from an analysis of the data and utilize these inferences and conclusions in\ndecision\xadmaking.

Table 4. Purpose, Content and Example of Syllabus - Combined

2. **Prompt**
    a. **Alignment task**
        i. **Zero-shot**

“You are a skilled assistant for aligning course descriptions with real-world job tasks.
Your task:

> *1. For the given course description, identify the top 10 real-world job tasks that best align with the course's goal of developing students' skills and preparing them for their careers after completing the class.*
> *2. Only return the task IDs for the top 10 tasks.*

Output format: [Task ID1, Task ID2, ..., Task ID10]
Please ensure you output exactly 10 IDs for each course.
Course Description: {course_description}
DWA Titles: {dwa_titles_text}”

        ii. **RAG**

“Course Description: {course["description"]}
Retrieved Skills: {retrieved_skills}
Based on the course description and the retrieved skills, identify the 10 most relevant skills.
*Output the skill IDs **only** in the format: ["skill_1", "skill_2", ..., "skill_10"].”*

    b. **LLM evaluation**

        i. **Few shot**

“You are a skilled assistant for evaluating the alignment between course descriptions and real-world job tasks.
Task: Evaluate the alignment between the following course description and real-world task using a 5-level rubric:

5 - Direct Match: The real-world task is a core learning objective of the course and is explicitly covered in the course description.
4 - Strongly Related: The real-world task aligns with the course content, and students should be able to perform this task upon completing the course.
3 - Indirectly Related but transferable: The task is not explicitly covered in the course, but students may develop transferable skills that can be applied to it.
2 - Unrelated but Within the Same Domain: The task falls within the same discipline or broader field but is not relevant to the course content.
1 - Completely Unrelated: The task is outside the scope of the course and belongs to a different domain.
Here are some examples:

Example 1:
description: this course is a training program to provide the students with the necessary basic skills and knowledge to deal with a broad spectrum of illness and injuries in the pre hospital care phase of emergency medicine upon successful completion of the course students will take the new york state emergency medical technical certification examination once certified and upon completion of certain fundamental core courses the student will be eligible to take the advanced paramedic level courses of the program the course will be offered in the fall and spring semesters only
Real-world Task:
Treat medical emergencies, Rating:5
Develop emergency response plans or procedures. Rating:4
Provide medical or cosmetic advice for clients. Rating:3
Teach health management classes. Rating:2
Sell products or services. Rating:1

Example 2:
description: this is a course in fundamental engineering drawing and industrial drafting room practice lettering orthographic projection auxiliary views sessions and conventions pictorials threads and fasteners tolerances detail drawing dimension ing and electrical drawing introduction to computer aided graphics are covered
Real-world Task:
Create graphical representations of mechanical equipment. Rating:5
Document technical design details. Rating:4
Develop equipment or component configurations. Rating:3
Develop tools to diagnose or assess needs. Rating:2
Plan community programs or activities for the general public. Rating:1

Now, please evaluate this case:

Course Description: {course_description}
Real-world Task: {job_task}
Please provide only the numerical rating (1-5) without any explanation."