A NATIONAL LONGITUDINAL DATASET OF SKILLS TAUGHT IN U.S. HIGHER EDUCATION CURRICULA

A PREPRINT

Alireza Javadian Sabet[1]
alj112@pitt.edu.edu

Sarah H. Bana[2,3]
sarah.bana@gmail.com

Renzhe Yu[4,5]
renzheyu@tc.columbia.edu

Morgan R. Frank[1,3,6*]
mrfrank@pitt.edu

[1]Department of Informatics and Networked Systems, University of Pittsburgh, Pittsburgh, PA 15216, USA
[2]Argyros School of Business and Economics, Chapman University, Orange, CA, USA
[3]Digital Economy Lab, Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA 94305, USA
[4]Teachers College, Columbia University, New York, NY 10027, USA
[5]Data Science Institute, Columbia University, New York, NY 10027, USA
[6]Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

April 23, 2024

ABSTRACT

Higher education plays a critical role in driving an innovative economy by equipping students with knowledge and skills demanded by the workforce. While researchers and practitioners have developed data systems to track detailed occupational skills, such as those established by the U.S. Department of Labor (DOL), much less effort has been made to document skill development in higher education at a similar granularity. Here, we fill this gap by presenting a longitudinal dataset of skills inferred from over three million course syllabi taught at nearly three thousand U.S. higher education institutions. To construct this dataset, we apply natural language processing to extract from course descriptions detailed workplace activities (DWAs) used by the DOL to describe occupations. We then aggregate these DWAs to create skill profiles for institutions and academic majors. Our dataset offers a large-scale representation of college-educated workers and their role in the economy. To showcase the utility of this dataset, we use it to 1) compare the similarity of skills taught and skills in the workforce according to the US Bureau of Labor Statistics, 2) estimate gender differences in acquired skills based on enrollment data, 3) depict temporal trends in the skills taught in social science curricula, and 4) connect college majors' skill distinctiveness to salary differences of graduates. Overall, this dataset can enable new research on the source of skills in the context of workforce development and provide actionable insights for shaping the future of higher education to meet evolving labor demands especially in the face of new technologies.

Keywords Skill · Higher Education · Future of Work · Labor Economics · Complexity · O*NET · Workplace Activity

Background & Summary

Skills are essential components of jobs and shape the career outcomes of workers within the labor market. Therefore, quantifying skills and their sources is essential for predicting workers' career trajectories and macro-level workforce dynamics [1, 2, 3]. For example, recent study on remote work [4] and automation [5] explain part of the erosion of

*corresponding author: Morgan R. Frank (mrfrank@pitt.edu)

the U.S. middle class using the relationship between technology and low- or high-skill work. Another study from economists finds increasing demand for social skills for modern, flexible team-based work environments based on required skills in job postings [6]. In response to shifts in skills, employers need to consider skill profiles and skill development in their hiring and training. For instance, employers subjectively perceive the skill content of college majors when determining the requirements to include in online job advertisements [7]. Combined, the focus on skills in the labor market warrants a similar perspective on the sources of skills during workforce development and talent acquisition.

Higher education is arguably the most important source for skill development, which facilitates both economic and social mobility [8]. In the past few decades, empirical studies have consistently demonstrated that college-educated individuals earn higher wages, achieve more extensive professional networks, and collectively experience greater inter-generational upward mobility [9, 10]. Non-college educated workers now engage in markedly less skilled tasks than their counterparts from previous eras [11] and tend to transition into low-wage occupations. On the other hand, the economic returns of higher education vary across fields of study due to differing skill sets imparted by college majors [12, 13]. They also vary along the lines of institutional selectivity [14]. Moreover, students from different demographic and socioeconomic backgrounds are sorted into different educational trajectories due to existing structural inequalities, which may hinder the social mobility that higher education is intended to foster [15]. In recent years, as elevated dropout rates [16] and rising unemployment or underemployment rates of college graduates fuel concerns around the efficacy of higher education [17, 18], it is important to better understand *how* higher education imparts skills and prepares students for the labor market. This process remains under explored due to the challenge of aligning the fine-grained *content* of education with job market demands [19, 20].

Here, we fill this gap by providing a new dataset on the skills taught in American higher education [21]. We first introduce `Syllabus2O*NET`, a novel natural language processing (NLP) framework designed to identify and interpret skills from curricular content, in line with the O*NET taxonomy used by the U.S. Department of Labor (see Figure 1). Applying this algorithmic pipeline to the most extensive dataset of university course syllabi, we then present a longitudinal, national dataset of inferred skill profiles across different institutions, academic majors, and student populations in the United States. To validate this dataset, we perform qualitative and quantitative explorations of the identified skill structures, as well as a benchmarking analysis against existing studies. We further include a handful of descriptive analyses to showcase the potential utility of this dataset, including revealing gender skills gaps, analyzing temporal trends in taught abilities, and quantifying skill-salary correlations.

Overall, our paper makes a handful of intellectual and practical contributions. First, we present a computational framework to describe the content alignment between education and workforce. While we developed the methodology based on two specific document types, it is applicable to other types of documents as well. Second, we provide an essential data source on the skill profiles of institutions, which can facilitate future research in such fields as higher education, labor economics, and future of work especially in an era marked by rapid technological advancements and shifting economic landscapes [22, 23]. Third, we illustrate some macro-level patterns of skills taught in higher education that warrant more in-depth research in the future.

Methods

Materials

Open Syllabus Project Dataset:

The Open Syllabus Project (OSP) Dataset[2] is composed of nearly $8,000,000$ course syllabi worldwide among which $3,162,747$ syllabi across 62 fields of study (FOS) belong to $2,761$ U.S. colleges and universities. Each course syllabus contains features including course description, language, year, field of study, and information about the institution.

O*NET:

The O*NET (Occupational Information Network) [24] is a comprehensive database of worker attributes and job characteristics. Developed under the sponsorship of the U.S. Department of Labor/Employment and Training Administration, O*NET has been used extensively for understanding and analyzing the labor market and workforce [25, 26, 27, 28, 29, 30, 31, 32, 33]. The O*NET database is a vital resource for educators, career counselors, and workforce development professionals. It provides a structured way to analyze job requirements and worker qualifications, aiding in curriculum development, career guidance, and labor market analysis [34].

[2] `https://opensyllabus.org/` (accessed on 28th November 2023)

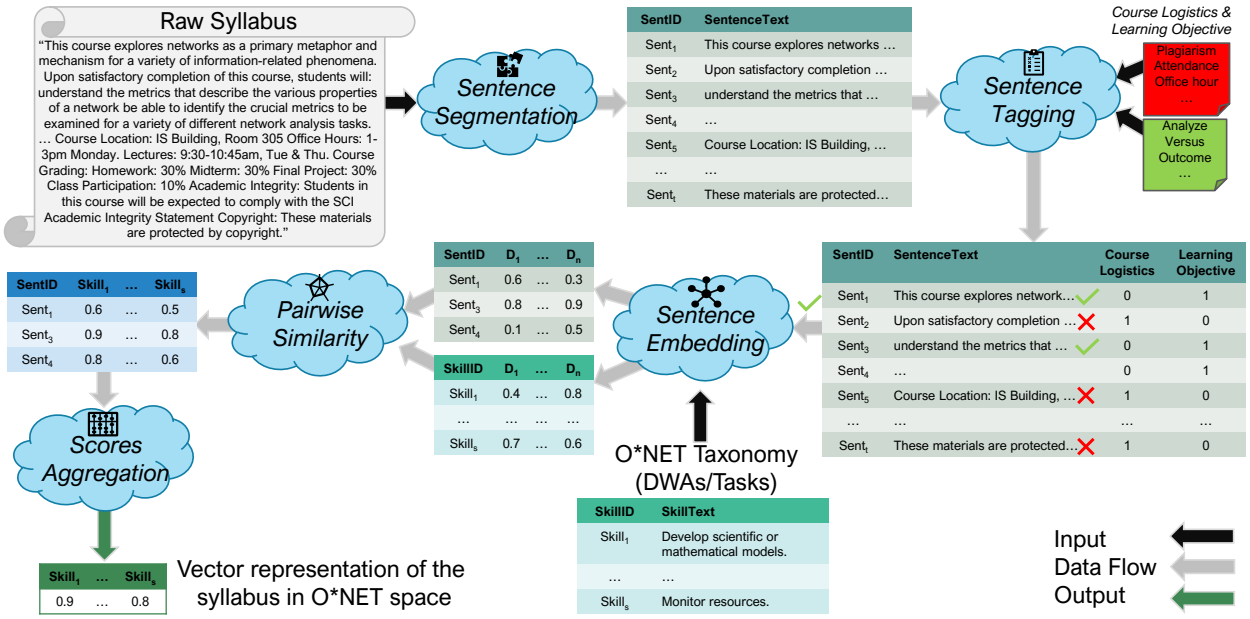


Figure 1: **The `Syllabus2O*NET` skill inference framework.** We start from a course syllabus and employ sentence segmentation. We use a curated bag of words to isolate sentences related to learning outcomes and use sentence embeddings (SBERT) to transform each of these sentences and each O*NET Detailed Work Activity (DWA) or Task (hereafter, "skill") into a high-dimensional space where we compute the pairwise semantic cosine similarity between skills and learning outcome sentences. Finally, we take each skill's maximum similarity score across the learning outcome sentences to produce a skill profile for the syllabus.

O*NET classifies occupations using a standardized hierarchical taxonomy. This taxonomy includes several key components including:

- *Ability:*[3] The performance of individuals is influenced by their enduring abilities, categorized into four key areas: "Cognitive", "Physical", "Psychomotor", and "Sensory".
- *Detailed Work Activity (DWA):*[4] This taxonomy provides more than 2,000 detailed descriptions of work activities specific to various occupations. DWAs are precise and descriptive, outlining the tasks and responsibilities of a particular job. They help in understanding the day-to-day activities and skills required for a particular role.
- *Task Statement (Task):*[5] Finally, a Task is a basic unit of work, the smallest action that constitutes part of a job. Roughly 18,000 Tasks provide the most detailed overview of the job responsibilities.

Integrated Postsecondary Education Data System:

For the analysis presented in the "Gender, Education, and Skills" Section, we employ data from the Integrated Postsecondary Education Data System (IPEDS). IPEDS is a system of surveys conducted annually by the U.S. Department of Education's (DOE) National Center for Education Statistics (NCES). IPEDS collects data from all primary providers of postsecondary education in the United States, which includes universities, colleges, vocational and technical institutions, and other postsecondary institutions. The purpose of IPEDS is to gather information on postsecondary education in the United States, covering a range of areas such as program completions, student financial aid, and institutional finances. The data collected is comprehensive and is used for policy analysis, decision-making, and research in the field of higher education.

In this study, we analyze fields of study or, equivalently, college majors according to the 2-digit Classification of Instructional Programs (CIP) codes used by the DOE. This enables us to connect syllabus data to the number of degrees awarded, disaggregated by demographic group, across fields of study. The dataset encompasses records dating back to 2010.

[3] `https://www.onetonline.org/find/descriptor/browse/1.A` (accessed on 5th February 2024)
[4] `https://www.onetcenter.org/dictionary/20.1/excel/dwa_reference.html` (accessed on 5th February 2024)
[5] `https://www.onetcenter.org/dictionary/20.1/excel/task_statements.html` (accessed on 5th February 2024)

National Employment Records and Skills:

In analyzing the "Dynamic Differences Between Inferred Workplace Activities and Labor Market Workplace Activities" Section, we sourced the national employment share for each occupation from the US Bureau of Labor Statistics (BLS) Occupation Employment and Wage Statistics (OEWS).[6] We focused on occupations requiring at least a Bachelor's, Master's, Associate's, or Doctoral/professional degree. Additionally, we employed the O*NET DWA data per occupation[7] to determine the required DWAs for each occupation.

Skill Inference Framework (`Syllabus2O*NET`):

Figure 1 illustrates a high-level overview of our `Syllabus2O*NET` skill inference framework. Given the text from a course syllabus, we segment sentences using Stanza [35], a tool designed to partition text into individual sentences. We extracted approximately 300 million sentences from the course syllabi in our dataset. On average, each syllabus contains roughly 100 sentences. Each course syllabus starts as unstructured text lacking metadata to distinguish between course logistics (e.g., scheduling and grading rubrics) and learning outcomes (e.g., course content and learning objectives). Thus, we implement a human-in-the-loop approach to remove sentences pertaining to Course Logistics while keeping sentences about Learning Outcomes. To do so, we compiled two distinct lists for labeling sentences. The list of Course Logistic related terms includes 356 phrases that are common but unrelated to the course content (e.g., "Plagiarism," "Attendance," and "Office hours"). The list of Learning Objective related terms includes 51 phrases such as "Analyze," "Versus," and "Outcome." The complete lists can be found on the code's GitHub page. We removed sentences from each syllabus that contained Course Logistic phrases or lacked Learning Objective phrases resulting in the removal of approximately 85% of the sentences. Each syllabus contains nearly 17 Learning Objective related sentences on average (see SI Section "Skills Inference Framework (`Syllabus2O*NET`)" for details on the statistics of Learning Objective sentence counts by FOS.).

Finally, we infer the presence of each O*NET DWA or Task (hereafter, "skill") in a syllabus's Learning Outcome sentences based on their semantic similarity. Specifically, we employ Sentence Embedding using the Siamese Bidirectional Encoder Representations from Transformers-Networks (SBERT) [36] utilizing the "all-mpnet-base-v2" model [37] to embed each Learning Outcome sentence and each O*NET DWA or Task into a 768 dimension semantic space. SBERT is well-suited for our project due to its training on a diverse range of more than 1 billion sentences including S2ORC: The Semantic Scholar Open Research Corpus [38], WikiAnswers Corpus [39], PAQ: 65 Million Probably-Asked Questions[40], and GooAQ: Open Question Answering with Diverse Answer Types [41]. The semantic similarity of sentences and skills are calculated from the cosine similarity of their embeddings. For instance, the sentence "understand the metrics that describe the various properties of a network be able to identify the crucial metrics to be examined for a variety of different network analysis tasks." from the example syllabus in Figure 1 is semantically similar to the O*NET DWA "Develop scientific or mathematical models." (i.e., cosine similarity score 0.9). For each skill, we select the maximum similarity score from all the sentences within the syllabus and generate a vector representing the maximum similarity score between a specific course and a corresponding skill. These vectors have 2,070 elements for DWAs and 17,992 elements for Tasks. In addition, we developed another tool to map the O*NET DWAs to O*NET Ability taxonomy (see SI Section "Mapping O*NET DWAs to Abilities (`DWA2Ability`)").

Skill Normalization:

Some DWAs, such as "Maintain student records" and "Document lesson plans", are ubiquitous across fields of study (FOS) and, therefore, do not distinguish the learning outcomes of one FOS from another. To address this issue and control for widespread skills we propose two approaches.

The first approach is applying Revealed Comparative Advantage (RCA) (a.k.a., "location quotient" [42, 28, 43, 44, 45]) according to:

$$rca(m,s) = \frac{onet(m,s) / \sum_{s' \in S} onet(m,s')}{\sum_{m' \in S} onet(m',s) / \sum_{m' \in M, s' \in S} onet(m',s')}, \tag{1}$$

where $m \in M$ denotes a FOS (i.e., a college major) and $s \in S$ denotes a skill (i.e., an O*NET DWA). If $rca(m,s) > 1$, then s is more related to m than would be expected across all DWAs and all FOS; therefore, s is a relatively distinctive skill identifying m more strongly than other FOS.

[6] `https://www.bls.gov/emp/ind-occ-matrix/occupation.xlsx` (accessed on 5th February 2024)

[7] `https://www.onetonline.org/find/all` (accessed on 5th February 2024)

The second approach, which is used in Fig. 4 (top) is to mask the frequently occurring skills based on their prevalence. This was implemented by evaluating the top 100 DWAs for each FOS and masking those that were commonly observed. Note that masked DWAs remain accessible in the published dataset.

Data Records

In this study, we create and release the **DWAs**, **Tasks**, and **Abilities** datasets inferred from the OSP course syllabi. Figure 2 provides the entity relationship diagram of the provided datasets.

Each record in these datasets includes the year, institution name, and UnitID (i.e., the unique identifier assigned by IPEDS to each institution), the geographical location of the institution (i.e., the city and state), and the FOS name along with its CIP code. The provided scores are the average scores of the syllabi belonging to the corresponding year, university, and FOS.

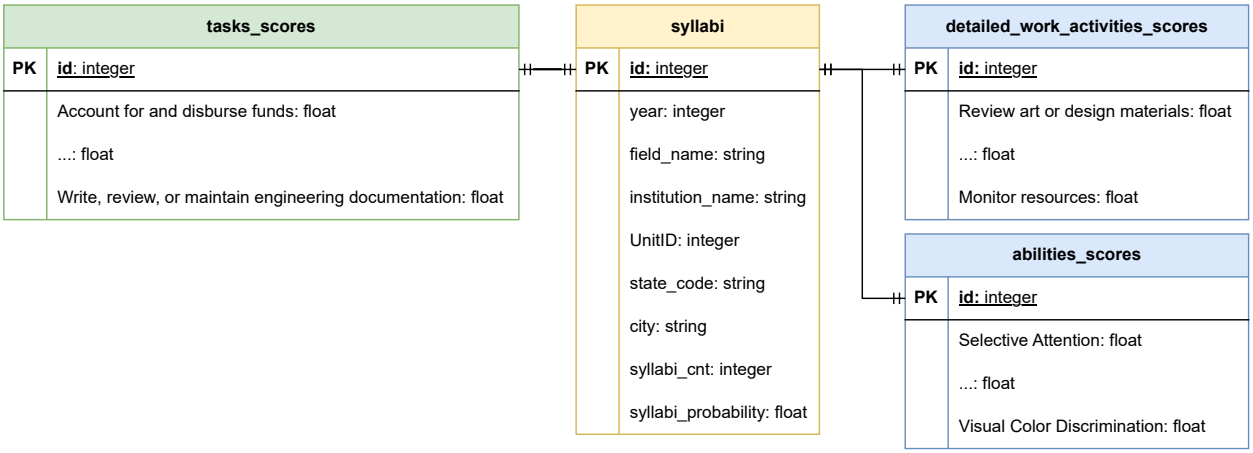


Figure 2: **The entity relationship diagram of the skills extracted from U.S. course syllabi.** In each table, PK represents the table's primary key. Syllabi comprise the main data table encompassing 281,153 records. For each corresponding id from "syllabi" table, "task_scores" and "detailed_work_activites_scores" tables contain the scores for 17,992 tasks and 2,070 DWAs respectively inferred using `Syllabus2O*NET`). "abilities_scores" table contains 52 abilities scores computed using `DWA2Ability` trained models. For brevity, we replaced the remaining tasks, DWAs, and abilities with "...". Lines connecting tables indicate the presence of a relational table.

Figure 3 explores the coverage of the syllabi data. Across each U.S. state, between 32% and 76% of the postsecondary institutions in each state provide at least eight course syllabi (i.e., corresponding to the 25$^{\text{th}}$ percentile. See Fig. 3A). Among the most represented FOS in our data, syllabi counts per year range from 10^3 to $10^{4.5}$ from 2000 through 2016 (see Fig. 3B). Across the entire data set, the majority of universities (nearly 2000) contribute at least 10 syllabi to the

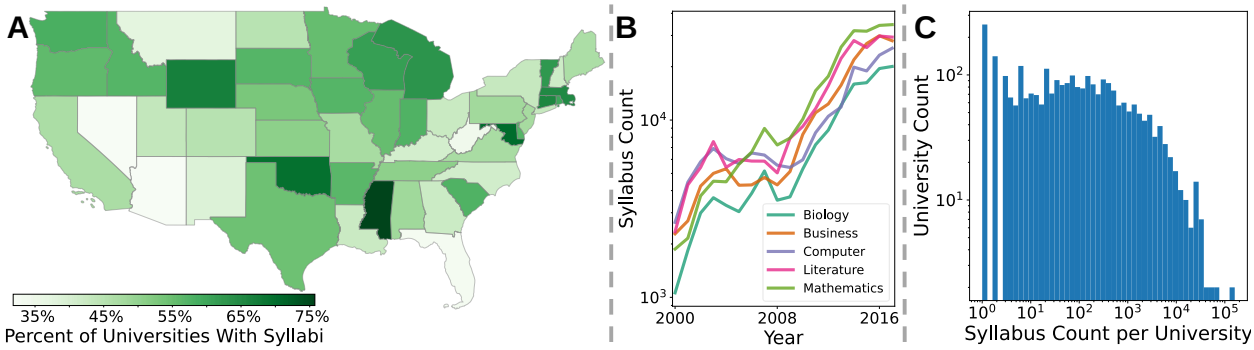


Figure 3: **Descriptive statistics of the Open Syllabus Project (OSP) dataset.** (A) Percentage of the universities with at least eight course syllabi (25$^{\text{th}}$ percentile) available per state. (B) The number of syllabi for the top 5 most represented FOS since 2000. (C) The syllabus count per university across all years and all FOS.

data, but some contribute up to 10^5 syllabi across all FOS (see Fig. 3C). Comprehensive descriptive statistics of these datasets are presented in the "Descriptive Statistics" Section of the SI.

Technical Validation

Qualitative Analysis of the Inferred Workplace Activities

As a face-validity check of `Syllabus20*NET`, we list the ten DWAs that are most strongly associated with three example fields of study (FOS): Agriculture, Biology, and Computer Science (see Fig. 4). Some DWAs (e.g., "Prepare informational or reference materials") are common across many FOS and, therefore, obscure the DWAs that most distinguish one FOS from the others. To normalize for ubiquitous DWAs, we also present the DWAs with the greatest revealed comparative advantage (RCA) in each field (see Section "Skill Normalization"). For instance, "Plant crops, trees, or other plants" emerges as the foremost skill in *Agriculture*, "Research diseases or parasites" is predominant in *Biology*, and "Coordinate software or hardware installation" is leading in *Computer Science* according to RCA scores. SI Section "Top 10 Detailed Work Activities per Field of Study" contains similar results for each FOS.

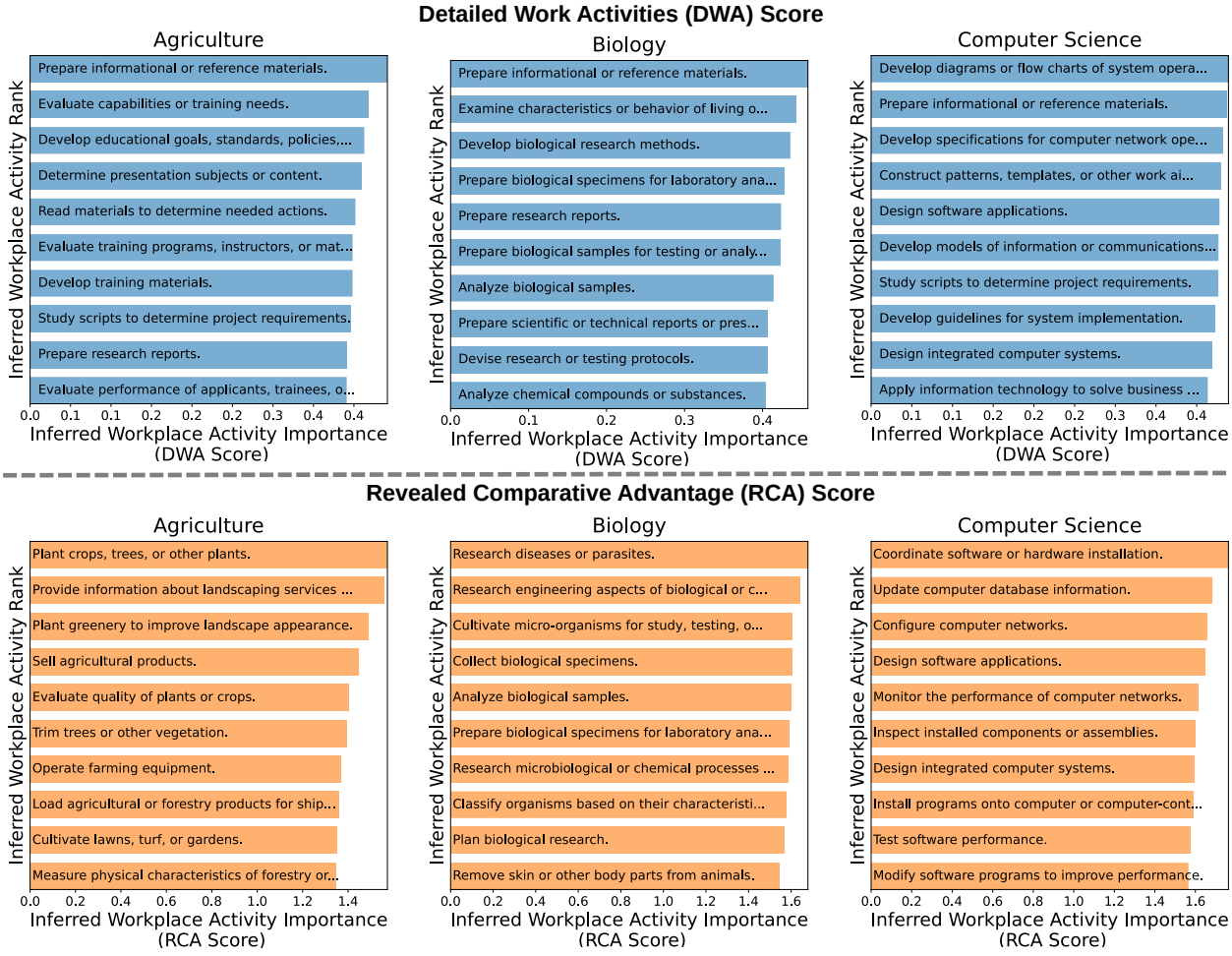


Figure 4: **The DWAs most strongly associated with Agriculture, Biology, and Computer Science.** (*top*) Top 10 inferred workplace activities with the highest DWA scores. (*bottom*) Top DWAs according to their RCA scores.

Relating Fields of Study from Skill Similarity

How similar are fields of study based on their skills? Following existing work [27], we cluster [46] FOS based on the similarity of their inferred DWAs (see Fig. 5). The resulting dendrogram offers another face-validity check as

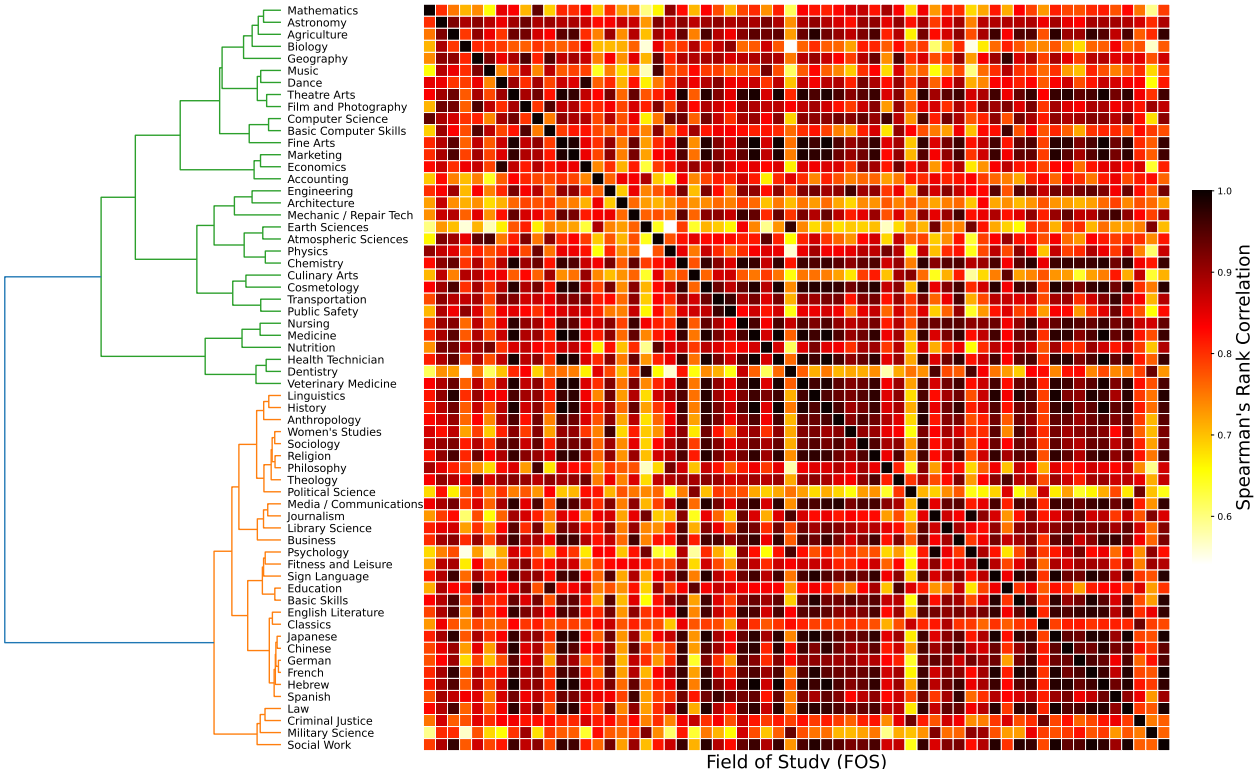


Figure 5: **The similarity of FOS according to their DWA profiles**. The heatmap shows the Spearman's rank correlation between the fields of study and the dendrogram represents the results of hierarchically clustering similar FOS.

similar FOS (e.g., STEM majors) tend to require similar DWAs. For instance, Marketing and Economics are closely related, as are Linguistics and History. Notably, just before the final clustering step, which amalgamates all majors (indicated in blue), two predominant clusters are discernible: one representing technical majors including those in STEM (in green) and the other humanities-based majors (in orange). For example, although Film and Photography is not a STEM-designated program, it requires skills that are common in STEM fields, such as "Draw detailed or technical illustrations" and "Design video game features or details" (see SI Section "Top 10 Detailed Work Activities per Field of Study").

Dynamic Differences Between Inferred Workplace Activities and Labor Market Workplace Activities

How responsive are the skills taught in higher education to the skills required in the US labor market? A "skills mismatch" may occur if higher education fails to adapt [47, 48, 49] (e.g., by teaching more theoretical skills than practical skills [47]). For the first time, our data set offers a direct comparison between the skills taught across higher education and the skills in the labor market using the skill taxonomy deployed by the US Department of Labor to describe occupations.

For comparison, one existing study [50] uses similar syllabus data to compare the skills taught in Computer Science (CS) with the skill demands of CS-related jobs in the US labor market. The study chooses an external measure of skills, from Burning Glass Technologies, which represents skills required in a job posting. On the other hand, our measure of skills uses the DWA scores required by each occupation. Specifically, we weight by the national employment share associated with each occupation according to the US BLS OEWS and cover a longer period with our data. Across all FOS, the skills inferred from course syllabi are increasingly similar to the skills in the US labor market over time (see Fig. 6A). For instance, skills (i.e., DWAs) from the 2000-2003 course syllabi are the least similar to labor force skills from the same period but are most similar to those from the latest period, 2012-2016. These results suggest that taught skills are forward-looking.

However, going beyond existing research, our dataset enables us to make direct comparisons between specific FOS and labor market dynamics for individual occupations. For example, motivated by earlier analysis of CS syllabi and CS-related job postings [50], we compare CS syllabi to Computer and Mathematical occupations (i.e., Standard

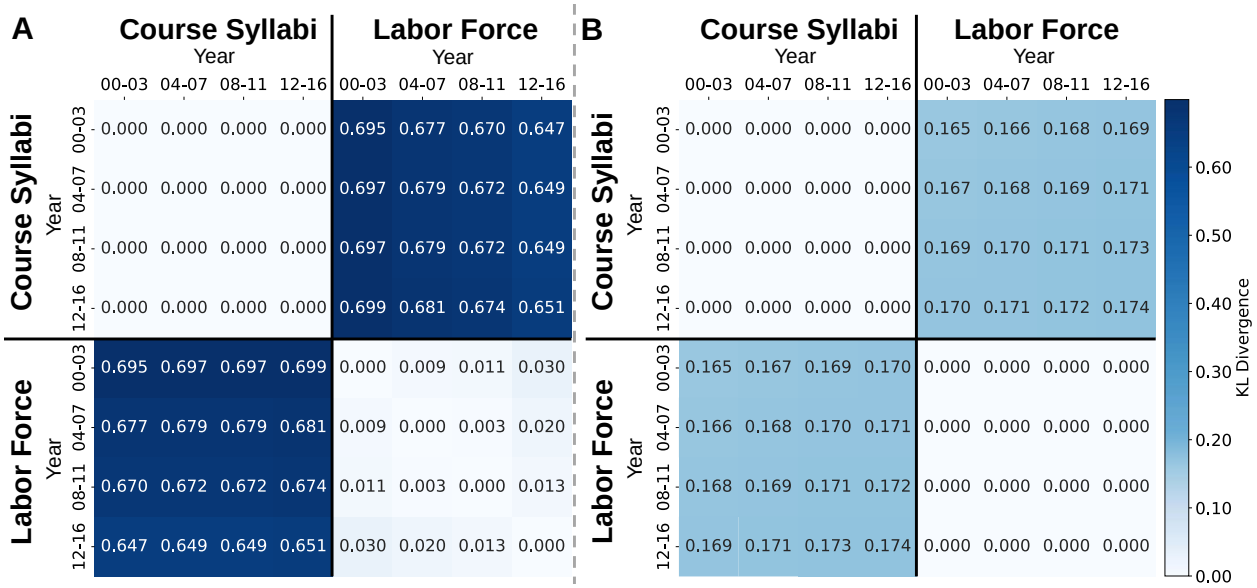


Figure 6: **Dynamic differences between skill (DWA) distributions in course syllabi and the labor force.** The matrix reports the KL divergence between course syllabi and the labor force for (A) syllabi from all FOS and employment-weighted O*NET DWA profiles for occupations trequiring a university degree, and (B) Computer Science and Mathematics course syllabi and employment-weighted skill O*NET DWA profiles for Computer and Mathematical Occupations (SOC 15-0000).

Occupation Classification code: 15-0000. See Fig. 6B). Despite the trend across all FOS, over time, the labor force skill distribution becomes increasingly dissimilar to older course syllabi which confirms the rapidly changing nature of such domains. Comparing the KL Divergence scores of the syllabi among different periods (top left box of Fig. 6B), we observe that syllabi are staying stagnant, and as a result, they are moving away from the frontier of knowledge required in the labor force.

Usage Notes

Our dataset offers a versatile tool for addressing a variety of research questions pertinent to education and workforce development across multiple domains. In the subsequent sections, we illustrate potential analytical approaches utilizing this dataset, such as exploring differences in skill sets across gender profiles in U.S. higher education, the trend of abilities in teaching activities, and utilizing skill scores for salary estimation. Lastly, we discuss our data's limitations.

Gender, Education, and Skills

Existing research finds that males and females tend to possess different workplace skills on aggregate [51, 52] which may correspond to gender stereotypes shaping careers [53, 54]. But are these differences the result of education or labor market outcomes? In general, these questions can only be studied through enrollment and graduation statistics from the US Department of Education without taking into account the granular differences in taught skills across different majors and institutions. However, our dataset enables the study of this heterogeneity and, thus, creates an opportunity to explain career outcomes from the differences in skills taught during higher education—even differences within a given FOS based on varied enrollment across educational institutions.

For example, recent research [55] from the Organisation for Economic Cooperation and Development (OECD) examined the difference in skills by gender. This difference is most pronounced in "Healthcare & Welfare", which has higher female employment, and "Engineering, Manufacturing & Construction", which has higher male employment. Our data set allows us to explain the skill differences between these majors based on course syllabi. We first weight the DWA scores for each combination of FOS, university, and year, using the number of graduates from the IPEDS dataset. Subsequently, we calculate the RCA score to pinpoint distinctive skills for each gender group and observe results in agreement with the OECD report (see Table 1). Healthcare-related skills, such as "Prescribe medications," are more

Detailed Work Activity	Women Rank	Detailed Work Activity	Men Rank
Administer blood or other fluids intravenously.	1	Measure stock or liquid levels in sustainable fuel ...	1
Apply bandages, dressings, or splints.	2	Reshape small metal components for precision assembly.	2
Prepare official health documents or records.	3	Design industrial processing systems.	3
Administer non-intravenous medications.	4	Calculate shipping costs.	4
Give medications or immunizations.	5	Design control systems for mechanical or other ...	5
Administer anesthetics or sedatives to control pain.	6	Compact materials to create level bases.	6
Prescribe treatments or therapies.	7	Move large objects using heavy equipment.	7
Prescribe medications.	8	Estimate costs or terms of sales.	8
Order medical diagnostic or clinical tests.	9	Design environmental control systems.	9
Refer patients to other healthcare practitioners or ...	10	Weigh parcels to determine shipping costs.	10

Table 1: **The inferred workplace activity importance scores inferred from syllabi reveal key differences among women and men.** Top 10 inferred workplace activities with the highest RCA scores inferred using `Syllabus2O*NET` framework from course syllabi weighted by the number of degrees awarded according to IPEDS.

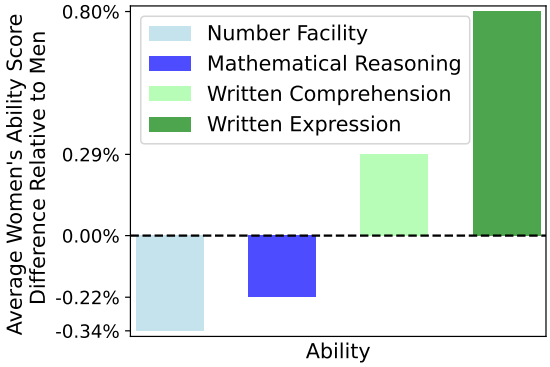


Figure 7: **The numeracy and literacy abilities inferred from syllabi reveal key differences among male and female students.** For numeracy (number facility and mathematical reasoning) and literacy (written comprehension and written expression) abilities, we compare female students' exposure to related DWAs compared to males by combining syllabi data with enrollment statistics from IPEDS.

indicative of females and engineering-related skills, such as "Design industrial processing systems," are more indicative of males.

To make a more direct comparison with the OECD report, we quantify these differences at the O*NET Ability level using our `DWA2Ability` trained models (see Section SI "Mapping O*NET DWAs to Abilities (`DWA2Ability`)"). The OECD [55] Survey of Adult Skills (PIAAC) reports that female students engage 15% less in advanced mathematics and numeracy in daily life compared to male students. In general agreement with this result, we find that females learn fewer numeracy skills according to the courses they enroll in (see Fig. 7). However, contrary to the OECD report, female students report a smaller gender gap in literacy use, around 5% less, while we find that female students enroll in classes teaching literacy in greater rates than male students. This inconsistency might suggest a mismatch between the skills taught at universities and the daily activities involving literacy skills (see SI Section "Gender, Education, and Skills" for the complete list).

As automation and AI continue to reshape job markets [56], the contrast in skills between genders might influence how each group adapts to new technologies both during education and once they enter the workforce. As a result, educational institutions and policymakers may need to consider these differences to ensure that both genders have equal opportunities to acquire a diverse range of skills, preparing them for various roles in the future job market [57]. In addition, the observed differences might also reflect traditional gender roles and stereotypes in education and career choices [58]. Initiatives aimed at diversifying skills across genders could be beneficial in breaking down these stereotypes.

Temporal Trends in Teaching Activities

Our dataset enables the study of skill differences within and across majors and universities over time. Taking active learning in social sciences as an example, a recent critique of active learning and the employability agenda in higher education within the social sciences [59] identified an inadvertent neglect of key skills including reading, listening, and note-taking. Utilizing our `DWA2Ability` trained models (see SI Section "Mapping O*NET DWAs to Abilities (`DWA2Ability`)"), we mapped the DWAs from the social sciences syllabi (CIP code=45) to the corresponding O*NET abilities. The abilities of Written Comprehension, Written Expression, and Oral Comprehension are analogous to reading, writing (analogous to note-taking), and listening, respectively. Active learning, as defined in [60], "refer[s] to an assortment of techniques where students do more than simply listen to a lecture." Thus, it encompasses a range of techniques that extend beyond passive listening in a lecture setting.

We observe a positive trend over time for the average annual scores of Social Sciences syllabi for the abilities of writing, reading, and listening, which suggests an increased emphasis on these skills in social sciences education (see Figure 8A and C). Conversely, the ability of Time Sharing, essential for active learning, displays a negative trend (see Figure 8B and C). Time Sharing, according to O*NET,[8] is "th[e] ability to shift back and forth between two or more activities or sources of information (such as speech, sounds, touch, or other sources)." This divergence in trends suggests a potential gap in fostering Time Sharing ability within the social sciences, a skill that might be the missing piece for the effective integration of writing, reading, and listening as recommended in [59]. Our findings highlight the trends in key abilities within the social sciences, provide a critical foundation in our dataset to pave the way for further investigation into how educational strategies can be developed to effectively balance traditional academic skills with the competencies essential for active learning environments.

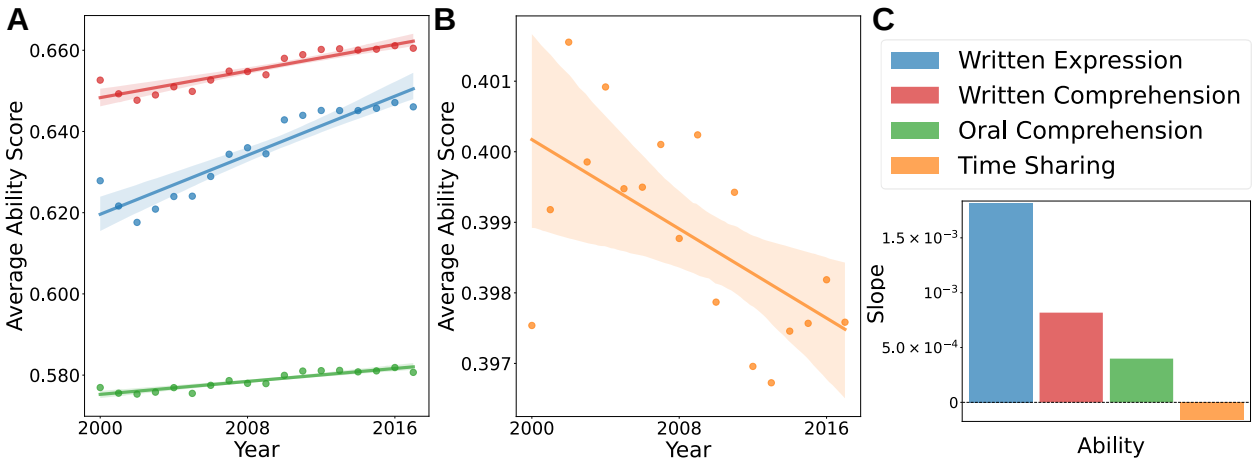


Figure 8: **Abilities over time within Social Sciences.** The average annual scores of Social Sciences syllabi for (A) Written Expression (writing), Written Comprehension (reading), and Oral Comprehension (listening), (B) Time Sharing for Social Science between 2000 and 2017. (C) The slope of the abilities' trend over time (all significant at $p = 0.01$).

Taught Skill Distinctiveness and Labor Market Outcomes

Research has consistently shown a strong relationship between college majors, skills, and wages [61, 62, 63]. For example, some majors may offer more diverse skill sets with more general skills profiles that lead to adaptable careers after a student graduates and enters the workforce [7]. We test this hypothesis by comparing the distinctiveness of FOS's taught skills with earnings of 25- to 29-year-old bachelor's degree holders reported by the National Center for Education Statistics (NCES).[9] We measure skill distinctiveness by noting the 90TH percentile skill RCA score within each FOS; under this measure, FOS with higher diversity scores teach more distinctive skills compared to other FOS. Graduates of FOS with more distinctive skills (i.e., greater RCA scores) tend to earn higher salaries with differences in RCA scores accounting for 29.4% of the variation in graduate earnings (see Fig. 9). As a robustness check, we observe similar results for modest changes to the RCA cut-off (see SI Section "Taught Skill Distinctiveness and Labor Market Outcomes").

[8] `https://www.onetonline.org/find/descriptor/result/1.A.1.g.2` (accessed on 8th March 2024)

[9] `https://nces.ed.gov/programs/digest/d19/tables/dt19_505.10.asp` (accessed on 12th February 2024)

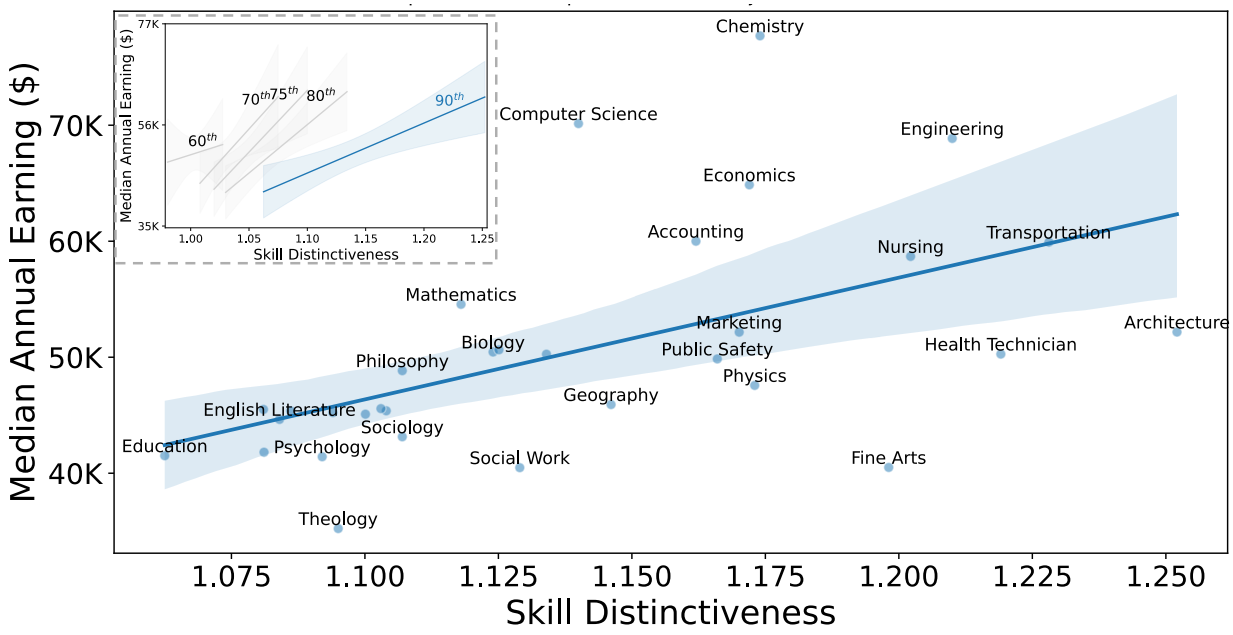


Figure 9: **Skill distinctiveness versus salary.** The x-axis shows the skill distinctiveness as RCA value at the 90th percentile within each field of study and the y-axis shows the corresponding median annual earnings (salary) obtained from NCES. The line represents a regression line with $R^2 = 0.294$ (p-*value* = 0.001). The subplot provides the same trends at the 60th, 70th, 75th, 80th, and 90th percentile RCA cut-off.

The salary gap among majors is multifaceted, involving factors like labor market demands [64], institutional focus on specific fields [65], major distribution's effect on gender wage disparities [62], and the impact of gender segregation in the labor market [66]. Addressing this gap requires a blend of strategies such as adjusting recruitment incentives, reevaluating institutional priorities, and fostering gender equality in education and occupations.

Limitations This study produces a novel large-scale data set reflecting the skills taught to US college and university students across majors. While useful for understanding one of the major pathways for workforce development in the US, there are some limitations to the current data set. First, the syllabus dataset is, to our knowledge, the largest collection of university syllabi available, but it is not comprehensive. The syllabus data has a sampling bias towards specific majors and universities; we describe the coverage of this data in Section "Data Records". Second, we propose a new approach for inferring taught skills (i.e., O*NET DWAs and Tasks) from syllabus text, but it is difficult to confirm the effectiveness of our approach without wide-scale comprehensive exams to test the skills that students actually obtain during a course. Such an effort would be extremely cumbersome because each student would ideally be assessed on over 2,000 DWAs; it's not clear how to empirically validate each of these DWAs and implementing such an examination across universities and majors throughout the US would be an immense undertaking. Effectively, it is crucial to acknowledge that teaching does not necessarily equate to learning. Third, our approach relies on the O*NET database which is designed to describe workers in the US workforce, and not explicitly designed to describe learning outcomes. While O*NET serves as a standardized taxonomy for communicating results to policymakers, its coverage across all occupations, and by extension, academic majors, is not uniformly comprehensive.

Code availability

The source code for `Syllabus2O*NET` is available at: `https://github.com/AlirezaJavadian/Syllabus-to-ONET` and the dataset is available at `https://doi.org/10.6084/m9.figshare.25632429`

References

[1] Spenner, K. I. Skill: Meanings, methods, and measures. *Work and Occupations* **17**, 399–421 (1990). https://doi.org/10.1177/0730888490017004002.

[2] Attewell, P. What is skill? *Work and Occupations* **17**, 422–448 (1990). https://doi.org/10.1177/0730888490017004003.

[3] Warhurst, C., Mayhew, K., Finegold, D. & Buchanan, J. *The Oxford handbook of skills and training* (Oxford University Press, 2017).

[4] Garrote Sanchez, D. *et al.* Who on earth can work from home? *The World Bank Research Observer* **36**, 67–100 (2021).

[5] Acemoglu, D. & Restrepo, P. Tasks, automation, and the rise in us wage inequality. *Econometrica* **90**, 1973–2016 (2022).

[6] Deming, D. J. The value of soft skills in the labor market. *NBER Reporter* **4**, 7–11 (2017).

[7] Hemelt, S. W., Hershbein, B., Martin, S. M. & Stange, K. M. College majors and skills: Evidence from the universe of online job ads. Tech. Rep., National Bureau of Economic Research (2021).

[8] Haveman, R. & Smeeding, T. The role of higher education in social mobility. *The Future of children* 125–150 (2006).

[9] Chetty, R., Friedman, J. N., Saez, E., Turner, N. & Yagan, D. Mobility report cards: The role of colleges in intergenerational mobility. Tech. Rep., national bureau of economic research (2017).

[10] Kerckhoff, A. C. Education and social stratification processes in comparative perspective. *Sociology of education* 3–18 (2001).

[11] Acemoglu, D. & Autor, D. Chapter 12 - skills, tasks and technologies: Implications for employment and earnings. vol. 4 of *Handbook of Labor Economics*, 1043–1171, https://doi.org/10.1016/S0169-7218(11)02410-5 (Elsevier, 2011).

[12] Altonji, J., Arcidiacono, P. & Maurel, A. Chapter 7 - the analysis of field choice in college and graduate school: Determinants and wage effects. vol. 5 of *Handbook of the Economics of Education*, 305–396, https://doi.org/10.1016/B978-0-444-63459-7.00007-5 (Elsevier, 2016).

[13] Altonji, J. G., Blom, E. & Meghir, C. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics* **4**, 185–223 (2012). https://doi.org/10.1146/annurev-economics-080511-110908.

[14] Triventi, M. The role of higher education stratification in the reproduction of social inequality in the labor market. *Research in social stratification and mobility* **32**, 45–63 (2013).

[15] Lovenheim, M. & Smith, J. Chapter 4 - returns to different postsecondary investments: Institution type, academic programs, and credentials. vol. 6 of *Handbook of the Economics of Education*, 187–318, https://doi.org/10.1016/bs.hesedu.2022.11.006 (Elsevier, 2023).

[16] Shapiro, D. *et al.* Completing college: A national view of student attainment rates by race and ethnicity–fall 2010 cohort (signature 12 supplement). *National Student Clearinghouse* (2017).

[17] Bonvillian, W. B. & Sarma, S. E. *Workforce education: a new roadmap* (MIT Press, 2021).

[18] Taylor, P. *et al. Is college worth it?* (Pew Social and Demographic trends, 2011).

[19] Brauer, S. Towards competence-oriented higher education: a systematic literature review of the different perspectives on successful exit profiles. *Education+ Training* **63**, 1376–1390 (2021).

[20] Li, X., Linde, S. & Shimao, H. Major complexity index and college skill production. *Available at SSRN 3791651* (2021).

[21] Javadian Sabet, A., Bana, S. H., Yu, R. & Frank, M. A national longitudinal dataset of skills taught in u.s. higher education curricula, 10.6084/m9.figshare.25632429.v1 (2024).

[22] Akour, M. & Alenezi, M. Higher education future in the era of digital transformation. *Education Sciences* **12**, 784 (2022).

[23] Brasca, C., Krishnan, C., Marya, V., Owen, K. & Sirois, J. How technology is shaping learning in higher education. *McKinsey & Company. June* **15** (2022).

[24] O*NET OnLine, (2024). Available at https://www.onetonline.org/.

[25] Lim, J., Aklin, M. & Frank, M. R. Location is a major barrier for transferring us fossil fuel employment to green jobs. *Nature Communications* **14**, 5711 (2023).

[26] Frank, M. R., Sun, L., Cebrian, M., Youn, H. & Rahwan, I. Small cities face greater impact from automation. *Journal of the Royal Society Interface* **15**, 20170946 (2018).

[27] Chau, H., Bana, S. H., Bouvier, B. & Frank, M. R. Connecting higher education to workplace activities and earnings. *Plos one* **18**, e0282323 (2023).

[28] Alabdulkareem, A. *et al.* Unpacking the polarization of workplace skills. *Science advances* **4**, eaao6030 (2018).

[29] Frank, M. R. *et al.* Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences* **116**, 6531–6539 (2019).

[30] Moro, E. *et al.* Universal resilience patterns in labor markets. *Nature communications* **12**, 1972 (2021).

[31] Frank, M. R. *et al.* Network constraints on worker mobility. *Nature Cities* **1**, 94–104 (2024).

[32] Agnihotri, A. & Misra, R. K. Managerial competencies: A comparative study of us-india employer's needs. *Global Business and Organizational Excellence* **43**, 92–106 (2024).

[33] Cabell, A. Supporting the career development of black adults during the covid-19 pandemic. *Journal of Employment Counseling* **60**, 62–71 (2023).

[34] Council, N. R. *et al.* A database for a changing economy: Review of the occupational information network (o*net) (2010).

[35] Qi, P., Zhang, Y., Zhang, Y., Bolton, J. & Manning, C. D. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082* (2020).

[36] Reimers, N. & Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[37] Reimers, N. & Inui, K. Sentence-transformers/all-mpnet-base-v2 · hugging face (2019).

[38] Lo, K., Wang, L. L., Neumann, M., Kinney, R. & Weld, D. S2ORC: The semantic scholar open research corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4969–4983 (Association for Computational Linguistics, Online, 2020).

[39] Fader, A., Zettlemoyer, L. & Etzioni, O. Open Question Answering Over Curated and Extracted Knowledge Bases. In *KDD* (2014).

[40] Lewis, P. *et al.* Paq: 65 million probably-asked questions and what you can do with them. *Transactions of the Association for Computational Linguistics* **9**, 1098–1115 (2021).

[41] Khashabi, D. *et al.* Gooaq: Open question answering with diverse answer types. *arXiv preprint* (2021).

[42] Hoen, A. R. & Oosterhaven, J. On the measurement of comparative advantage. *The Annals of Regional Science* **40**, 677–691 (2006).

[43] Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proceedings of the national academy of sciences* **106**, 10570–10575 (2009).

[44] Hausmann, R. & Hidalgo, C. A. The network structure of economic output. *Journal of economic growth* **16**, 309–342 (2011).

[45] Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M. & Hidalgo, C. A. Linking economic complexity, institutions, and income inequality. *World development* **93**, 75–93 (2017).

[46] Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967).

[47] Aloysius, O. I., Ismail, I. A., Suandi, T. & Arshad, M. Enhancing university's and industry's employability-collaboration among nigeria graduates in the labor market. *International Journal of Academic Research in Business and Social Sciences* **8**, 32–48 (2018).

[48] Pujol-Jover, M., Riera-Prunera, C. & Abio, G. Competences acquisition of university students: Do they match job market's needs? *Intangible Capital* **11**, 612–626 (2015).

[49] Jackson, D. & Tomlinson, M. The relative importance of work experience, extra-curricular and university-based activities on student employability. *Higher Education Research & Development* **41**, 1119–1135 (2022).

[50] Börner, K. *et al.* Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences* **115**, 12630–12637 (2018).

[51] Christl, M. & Köppl-Turyna, M. Gender wage gap and the role of skills and tasks: evidence from the austrian piaac data set. *Applied Economics* **52**, 113–134 (2020).

[52] Azmat, G. & Ferrer, R. Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy* **125**, 1306–1355 (2017).

[53] Hausmann, R., Tyson, L. D., Zahidi, S. *et al.* The global gender gap report 2012 (World Economic Forum Geneva, 2012).

[54] Fana, M., Villani, D. & Bisello, M. Gender gaps in power and control within jobs. *Socio-Economic Review* **21**, 1343–1367 (2022). https://academic.oup.com/ser/article-pdf/21/3/1343/50969983/mwac062.pdf.

[55] Encinas-Martin, M. & Cherian, M. Gender, education and skills: The persistence of gender gaps in education and skills. *OECD Skills Studies* 1–54 (2023).

[56] Eloundou, T., Manning, S., Mishkin, P. & Rock, D. Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130* (2023).

[57] Li, L. Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond. *Information Systems Frontiers* 1–16 (2022).

[58] Balducci, M. Linking gender differences with gender equality: A systematic-narrative literature review of basic skills and personality. *Frontiers in Psychology* **14**, 1105234 (2023).

[59] David, M. & Maurer, H. Reclaiming agency: skills, academics and students in the social sciences. *European Political Science* 1–17 (2022).

[60] Ishiyama, J. Frequently used active learning techniques and their impact: A critical review of existing journal literature in the united states. *European Political Science* **12**, 116–126 (2013).

[61] Grogger, J. & Eide, E. Changes in college skills and the rise in the college wage premium. *Journal of Human Resources* 280–310 (1995).

[62] Eide, E. College major choice and changes in the gender wage gap. *Contemporary Economic Policy* **12**, 55–64 (1994).

[63] Long, M. C., Goldhaber, D. & Huntington-Klein, N. Do completed college majors respond to changes in wages? *Economics of Education Review* **49**, 1–14 (2015).

[64] Metcalf, D. University salaries: faculty differentials. *Economica* 362–372 (1970).

[65] Johnson, J. A. & Taylor, B. J. Academic capitalism and the faculty salary gap. *Innovative Higher Education* **44**, 21–35 (2019).

[66] Doseděl, T. Poorly paid jobs or study fields? gender pay gap of tertiary-educated employees in contemporary european labour markets. *Population Review* **61**, 55–71 (2022).

Acknowledgments

This work received funding from Russell Sage Foundation and is supported by the University of Pittsburgh Center for Research Computing.

Author contributions statement

A.J.S. processed the data, performed all calculations, and produced all figures and statistics. All authors designed the research, analyzed the results, and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Information:
A national longitudinal dataset of skills taught in U.S. higher education curricula

Alireza Javadian Sabet[1], Sarah H. Bana[2,3], Renzhe Yu[4,5], Morgan R. Frank[1,3,6]

[1]Department of Informatics and Networked Systems, University of Pittsburgh, Pittsburgh, PA 15216, USA
[2]Argyros School of Business and Economics, Chapman University, Orange, CA, USA
[3]Digital Economy Lab, Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA 94305, USA
[4]Teachers College, Columbia University, New York, NY 10027, USA
[5]Data Science Institute, Columbia University, New York, NY 10027, USA
[6]Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Contents

1 Descriptive Statistics

The Open Syllabus Project (OSP) Dataset[1] is composed of nearly $8,000,000$ course syllabi worldwide among which $3,162,747$ syllabi across 62 fields of study (FOS) belong to $2,761$ U.S. colleges and universities. Tables 1 and 2 list the frequency of syllabi per FOS and year respectively. Table 3 details the geographical coverage of the OSP dataset. Number of educational institutions within each state is obtained from the Carnegie Classification of Institutions of Higher Education (CCIHE).[2] For example, Texas with $865,973$ syllabi has the largest number of syllabi (27.85%) in the dataset. According to CCIHE, there are 226 universities and educational institutions located in Texas, among which 54.42% have at least 8 syllabi (25^{th} percentile) in the OSP dataset. Figure 1 details the number of syllabi per FOS and year between 2000 and 2017 (see Figure 2 for FOS based on 2-digit Classification of Instructional Programs (CIP)). Figure 3 shows the joint distribution of the syllabus count and university count across all years and all FOS. Table 4 lists the syllabus count per top 100 universities across all years and all FOS.

Field of Study	**# Syllabi**	**Field of Study**	**# Syllabi**	**Field of Study**	**# Syllabi**
Mathematics	258,160	Accounting	51,984	Religion	14,440
English Literature	232,065	Sociology	46,836	French	14,305
Business	201,100	Physics	44,802	Journalism	12,712
Computer Science	184,649	Film and Photography	42,690	Nutrition	11,883
Biology	140,187	Criminal Justice	39,805	Dentistry	10,367
Education	140,182	Spanish	39,650	Culinary Arts	9,430
Fitness and Leisure	131,262	Health Technician	38,268	Sign Language	8,665
Psychology	122,387	Social Work	36,745	German	8,385
History	107,676	Philosophy	35,583	Classics	7,813
Media / Communications	85,561	Agriculture	35,305	Cosmetology	7,291
Music	82,329	Marketing	31,430	Astronomy	7,286
Fine Arts	75,722	Law	31,421	Transportation	7,121
Basic Skills	73,362	Theatre Arts	29,087	Japanese	5,456
Engineering	70,084	Theology	24,584	Women's Studies	5,237
Political Science	69,111	Public Safety	23,931	Chinese	5,054
Basic Computer Skills	68,028	Earth Sciences	21,870	Linguistics	4,859
Nursing	63,603	Anthropology	21,509	Military Science	3,202
Mechanic / Repair Tech	62,423	Library Science	20,234	Atmospheric Sciences	2,231
Chemistry	61,280	Dance	19,694	Veterinary Medicine	2,105
Economics	56,157	Architecture	19,379	Hebrew	1,674
Medicine	55,161	Geography	17,935		

Table 1: Frequency of syllabi per field of study (FOS).

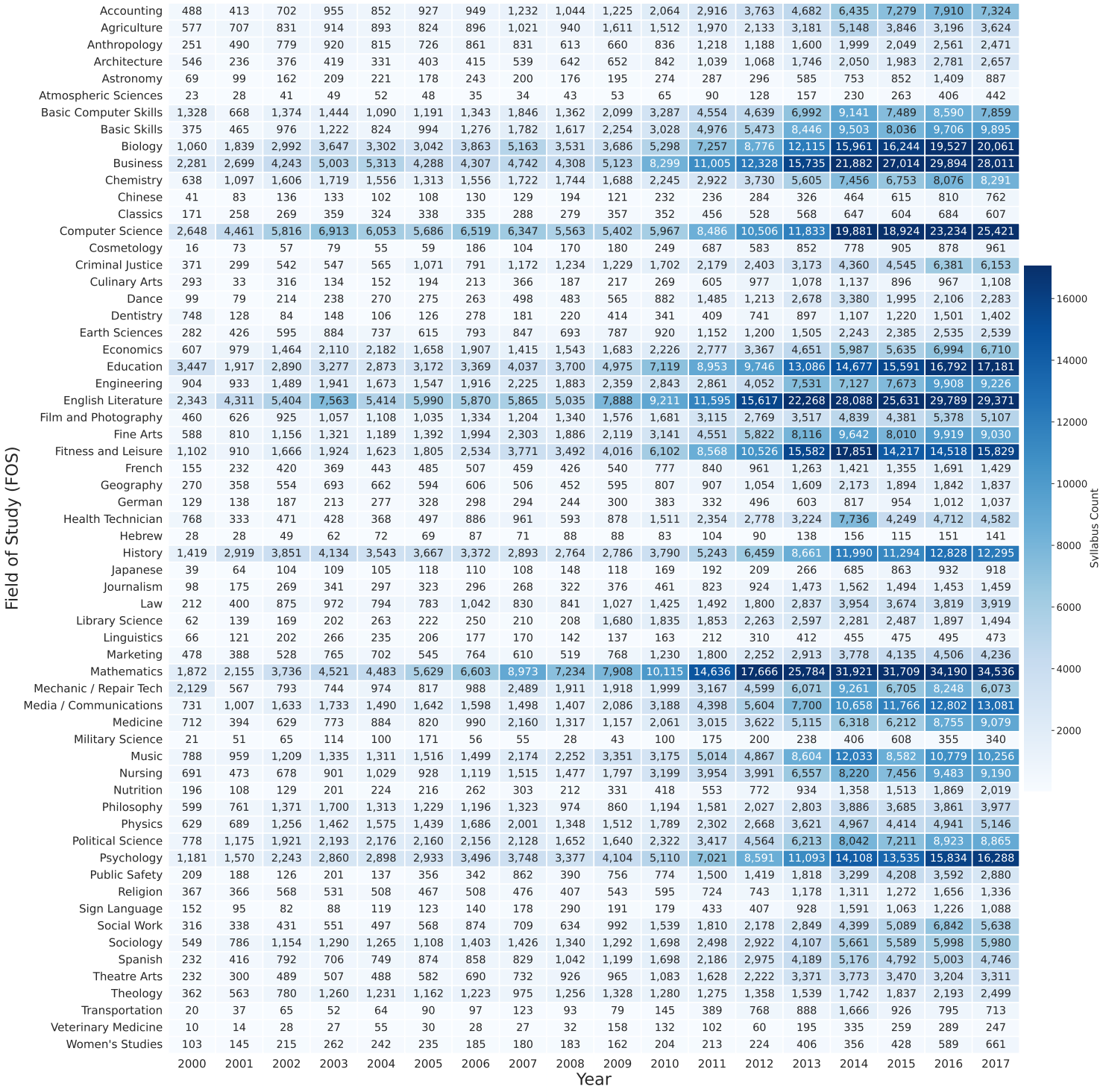
Year	**# Syllabi**	**Year**	**# Syllabi**	**Year**	**# Syllabi**	**Year**	**# Syllabi**	**Year**	**# Syllabi**
2017	406,981	2006	80,568	1995	4,075	1984	632	1973	3,692
2016	413,215	2005	71,937	1994	2,122	1983	473	1972	1,371
2015	360,353	2004	71,248	1993	1,931	1982	382	1971	83
2014	380,261	2003	77,695	1992	3,488	1981	2,099	1970	237
2013	290,702	2002	65,177	1991	1,488	1980	341	1969	121
2012	207,899	2001	43,519	1990	2,768	1979	214	1968	124
2011	174,492	2000	38,359	1989	1,183	1978	231	1967	97
2010	127,618	1999	18,509	1988	2,454	1977	148	1966	103
2009	96,799	1998	15,206	1987	830	1976	217		
2008	80,451	1997	11,642	1986	794	1975	217		
2007	92,128	1996	5,116	1985	574	1974	383		

Table 2: Frequency of syllabi per year.

[1] https://opensyllabus.org/ (accessed on 28^{th} November 2023)
[2] https://carnegieclassifications.acenet.edu/ (accessed on 18^{th}Februaryy 2023)

Name	Abbreviation	# Syllabi	% Syllabi	# Inst.	% Covered Inst.
Alabama	AL	66,548	2.14%	60	48.33%
Alaska	AK	3,271	0.11%	8	50.00%
Arizona	AZ	16,476	0.53%	66	31.82%
Arkansas	AR	9,018	0.29%	53	56.60%
California	CA	553,589	17.80%	425	46.82%
Colorado	CO	18,718	0.60%	62	45.16%
Connecticut	CT	8,695	0.28%	38	65.79%
Delaware	DE	1,737	0.06%	7	57.14%
District of Columbia	DC	13,356	0.43%	22	36.36%
Florida	FL	78,441	2.52%	161	32.92%
Georgia	GA	72,955	2.35%	107	42.06%
Hawaii	HI	19,016	0.61%	17	52.94%
Idaho	ID	40,312	1.30%	14	57.14%
Illinois	IL	65,455	2.10%	152	55.26%
Indiana	IN	22,389	0.72%	66	57.58%
Iowa	IA	20,861	0.67%	56	57.14%
Kansas	KS	14,598	0.47%	63	50.79%
Kentucky	KY	49,872	1.60%	59	40.68%
Louisiana	LA	24,547	0.79%	52	42.31%
Maine	ME	8,252	0.27%	30	46.67%
Maryland	MD	157,830	5.08%	54	70.37%
Massachusetts	MA	36,102	1.16%	106	65.09%
Michigan	MI	120,455	3.87%	87	63.22%
Minnesota	MN	68,253	2.19%	85	55.29%
Mississippi	MS	7,174	0.23%	33	75.76%
Missouri	MO	47,594	1.53%	93	47.31%
Montana	MT	6,929	0.22%	24	37.50%
Nebraska	NE	4,609	0.15%	34	52.94%
Nevada	NV	16,541	0.53%	22	31.82%
New Hampshire	NH	3,066	0.10%	24	41.67%
New Jersey	NJ	38,225	1.23%	82	47.56%
New Mexico	NM	28,830	0.93%	36	38.89%
New York	NY	77,699	2.50%	295	43.05%
North Carolina	NC	76,644	2.46%	134	41.04%
North Dakota	ND	4,120	0.13%	20	45.00%
Ohio	OH	80,494	2.59%	160	42.50%
Oklahoma	OK	35,158	1.13%	46	69.57%
Oregon	OR	24,273	0.78%	50	56.00%
Pennsylvania	PA	72,515	2.33%	193	48.19%
Puerto Rico	PR	40	0.00%	86	2.33%
Rhode Island	RI	5,322	0.17%	15	60.00%
South Carolina	SC	37,259	1.20%	66	57.58%
South Dakota	SD	2,880	0.09%	21	57.14%
Tennessee	TN	23,144	0.74%	83	50.60%
Texas	TX	865,973	27.85%	226	54.42%
Utah	UT	18,612	0.60%	23	43.48%
Vermont	VT	5,722	0.18%	16	62.50%
Virginia	VA	41,186	1.32%	108	47.22%
Washington	WA	45,493	1.46%	72	58.33%
West Virginia	WV	5,836	0.19%	41	34.15%
Wisconsin	WI	23,880	0.77%	67	61.19%
Wyoming	WY	19,570	0.63%	9	66.67%

Table 3: **Geographical coverage of the OSP dataset.** For each U.S. state "# Syllabi" and "% Syllabi" detail the number and percentage of course syllabi from the universities and institutions located in that state (the sum of "% Syllabi" equals to 100%)). "# Inst." specifies the number of institutions located in the given state based on the CCIHE. "% Covered Inst." specifies the percentage of the number of universities with at least 8 course syllabi (25^{th} percentile) in the OSP dataset.



Field of Study (FOS)	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Accounting	488	413	702	955	852	927	949	1,232	1,044	1,225	2,064	2,916	3,763	4,682	6,435	7,279	7,910	7,324
Agriculture	577	707	831	914	893	824	896	1,021	940	1,611	1,512	1,970	2,133	3,181	5,148	3,846	3,196	3,624
Anthropology	251	490	779	920	815	726	861	831	613	660	836	1,218	1,188	1,600	1,999	2,049	2,561	2,471
Architecture	546	236	376	419	331	403	415	539	642	652	842	1,039	1,068	1,746	2,050	1,983	2,781	2,657
Astronomy	69	99	162	209	221	178	243	200	176	195	274	287	296	585	753	852	1,409	887
Atmospheric Sciences	23	28	41	49	52	48	35	34	43	53	65	90	128	157	230	263	406	442
Basic Computer Skills	1,328	668	1,374	1,444	1,090	1,191	1,343	1,846	1,362	2,099	3,287	4,554	4,639	6,992	9,141	7,489	8,590	7,859
Basic Skills	375	465	976	1,222	824	994	1,276	1,782	1,617	2,254	3,028	4,976	5,473	8,446	9,503	8,036	9,706	9,895
Biology	1,060	1,839	2,992	3,647	3,302	3,042	3,863	5,163	3,531	3,686	5,298	7,257	8,776	12,115	15,961	16,244	19,527	20,061
Business	2,281	2,699	4,243	5,003	5,313	4,288	4,307	4,742	4,308	5,123	8,299	11,005	12,328	15,735	21,882	27,014	29,894	28,011
Chemistry	638	1,097	1,606	1,719	1,556	1,313	1,556	1,722	1,744	1,688	2,245	2,922	3,730	5,605	7,456	6,753	8,076	8,291
Chinese	41	83	136	133	102	108	130	129	194	121	232	236	284	326	464	615	810	762
Classics	171	258	269	359	324	338	335	288	279	357	352	456	528	568	647	604	684	607
Computer Science	2,648	4,461	5,816	6,913	6,053	5,686	6,519	6,347	5,563	5,402	5,967	8,486	10,506	11,833	19,881	18,924	23,234	25,421
Cosmetology	16	73	57	79	55	59	186	104	170	180	249	687	583	852	778	905	878	961
Criminal Justice	371	299	542	547	565	1,071	791	1,172	1,234	1,229	1,702	2,179	2,403	3,173	4,360	4,545	6,381	6,153
Culinary Arts	293	33	316	134	152	194	213	366	187	217	269	605	977	1,078	1,137	896	967	1,108
Dance	99	79	214	238	270	275	263	498	483	565	882	1,485	1,213	2,678	3,380	1,995	2,106	2,283
Dentistry	748	128	84	148	106	126	278	181	220	414	341	409	741	897	1,107	1,220	1,501	1,402
Earth Sciences	282	426	595	884	737	615	793	847	693	787	920	1,152	1,200	1,505	2,243	2,385	2,535	2,539
Economics	607	979	1,464	2,110	2,182	1,658	1,907	1,415	1,543	1,683	2,226	2,777	3,367	4,651	5,987	5,635	6,994	6,710
Education	3,447	1,917	2,890	3,277	2,873	3,172	3,369	4,037	3,700	4,975	7,119	8,953	9,746	13,086	14,677	15,591	16,792	17,181
Engineering	904	933	1,489	1,941	1,673	1,547	1,916	2,225	1,883	2,359	2,843	2,861	4,052	7,531	7,127	7,673	9,908	9,226
English Literature	2,343	4,311	5,404	7,563	5,414	5,990	5,870	5,865	5,035	7,888	9,211	11,595	15,617	22,268	28,088	25,631	29,789	29,371
Film and Photography	460	626	925	1,057	1,108	1,035	1,334	1,204	1,340	1,576	1,681	3,115	2,769	3,517	4,839	4,381	5,378	5,107
Fine Arts	588	810	1,156	1,321	1,189	1,392	1,994	2,303	1,886	2,119	3,141	4,551	5,822	8,116	9,642	8,010	9,919	9,030
Fitness and Leisure	1,102	910	1,666	1,924	1,623	1,805	2,534	3,771	3,492	4,016	6,102	8,568	10,526	15,582	17,851	14,217	14,518	15,829
French	155	232	420	369	443	485	507	459	426	540	777	840	961	1,263	1,421	1,355	1,691	1,429
Geography	270	358	554	693	662	594	606	506	452	595	807	907	1,054	1,609	2,173	1,894	1,842	1,837
German	129	138	187	213	277	328	298	294	244	300	383	332	496	603	817	954	1,012	1,037
Health Technician	768	333	471	428	368	497	886	961	593	878	1,511	2,354	2,778	3,224	7,736	4,249	4,712	4,582
Hebrew	28	28	49	62	72	69	87	71	88	88	83	104	90	138	156	115	151	141
History	1,419	2,919	3,851	4,134	3,543	3,667	3,372	2,893	2,764	2,786	3,790	5,243	6,459	8,661	11,990	11,294	12,828	12,295
Japanese	39	64	104	109	105	118	110	108	148	118	169	192	209	266	685	863	932	918
Journalism	98	175	269	341	297	323	296	268	322	376	461	823	924	1,473	1,562	1,494	1,453	1,459
Law	212	400	875	972	794	783	1,042	830	841	1,027	1,425	1,492	1,800	2,837	3,954	3,674	3,819	3,919
Library Science	62	139	169	202	263	222	250	210	208	1,680	1,835	1,853	2,263	2,597	2,281	2,487	1,897	1,494
Linguistics	66	121	202	266	235	206	177	170	142	137	163	212	310	412	455	475	495	473
Marketing	478	388	528	765	702	545	764	610	519	768	1,230	1,800	2,252	2,913	3,778	4,135	4,506	4,236
Mathematics	1,872	2,155	3,736	4,521	4,483	5,629	6,603	8,973	7,234	7,908	10,115	14,636	17,666	25,784	31,921	31,709	34,190	34,536
Mechanic / Repair Tech	2,129	567	793	744	974	817	988	2,489	1,911	1,918	1,999	3,167	4,599	6,071	9,261	6,705	8,248	6,073
Media / Communications	731	1,007	1,633	1,733	1,490	1,642	1,598	1,498	1,407	2,086	3,188	4,398	5,604	7,700	10,658	11,766	12,802	13,081
Medicine	712	394	629	773	884	820	990	2,160	1,317	1,157	2,061	3,015	3,622	5,115	6,318	6,212	8,755	9,079
Military Science	21	51	65	114	100	171	56	55	28	43	100	175	200	238	406	608	355	340
Music	788	959	1,209	1,335	1,311	1,516	1,499	2,174	2,252	3,351	3,175	5,014	4,867	8,604	12,033	8,582	10,779	10,256
Nursing	691	473	678	901	1,029	928	1,119	1,515	1,477	1,797	3,199	3,954	3,991	6,557	8,220	7,456	9,483	9,190
Nutrition	196	108	129	201	224	216	262	303	212	331	418	553	772	934	1,358	1,513	1,869	2,019
Philosophy	599	761	1,371	1,700	1,313	1,229	1,196	1,323	974	860	1,194	1,581	2,027	2,803	3,886	3,685	3,861	3,977
Physics	629	689	1,256	1,462	1,575	1,439	1,686	2,001	1,348	1,512	1,789	2,302	2,668	3,621	4,967	4,414	4,941	5,146
Political Science	778	1,175	1,921	2,193	2,176	2,160	2,156	2,128	1,652	1,640	2,322	3,417	4,564	6,213	8,042	7,211	8,923	8,865
Psychology	1,181	1,570	2,243	2,860	2,898	2,933	3,496	3,748	3,377	4,104	5,110	7,021	8,591	11,093	14,108	13,535	15,834	16,288
Public Safety	209	188	126	201	137	356	342	862	390	756	774	1,500	1,419	1,818	3,299	4,208	3,592	2,880
Religion	367	366	568	531	508	467	508	476	407	543	595	724	743	1,178	1,311	1,272	1,656	1,336
Sign Language	152	95	82	88	119	123	140	178	290	191	179	433	407	928	1,591	1,063	1,226	1,088
Social Work	316	338	431	551	497	568	874	709	634	992	1,539	1,810	2,178	2,849	4,399	5,089	6,842	5,638
Sociology	549	786	1,154	1,290	1,265	1,108	1,403	1,426	1,340	1,292	1,698	2,498	2,922	4,107	5,661	5,589	5,998	5,980
Spanish	232	416	792	706	749	874	858	829	1,042	1,199	1,698	2,186	2,975	4,189	5,176	4,792	5,003	4,746
Theatre Arts	232	300	489	507	488	582	690	732	926	965	1,083	1,628	2,222	3,371	3,773	3,470	3,204	3,311
Theology	362	563	780	1,260	1,231	1,162	1,223	975	1,256	1,328	1,280	1,275	1,358	1,539	1,742	1,837	2,193	2,499
Transportation	20	37	65	52	64	90	97	123	93	79	145	389	768	888	1,666	926	795	713
Veterinary Medicine	10	14	28	27	55	30	28	27	32	158	132	102	60	195	335	259	289	247
Women's Studies	103	145	215	262	242	235	185	180	183	162	204	213	224	406	356	428	589	661

Figure 1: Frequency of syllabi per field of study and year between 2000 and 2017.

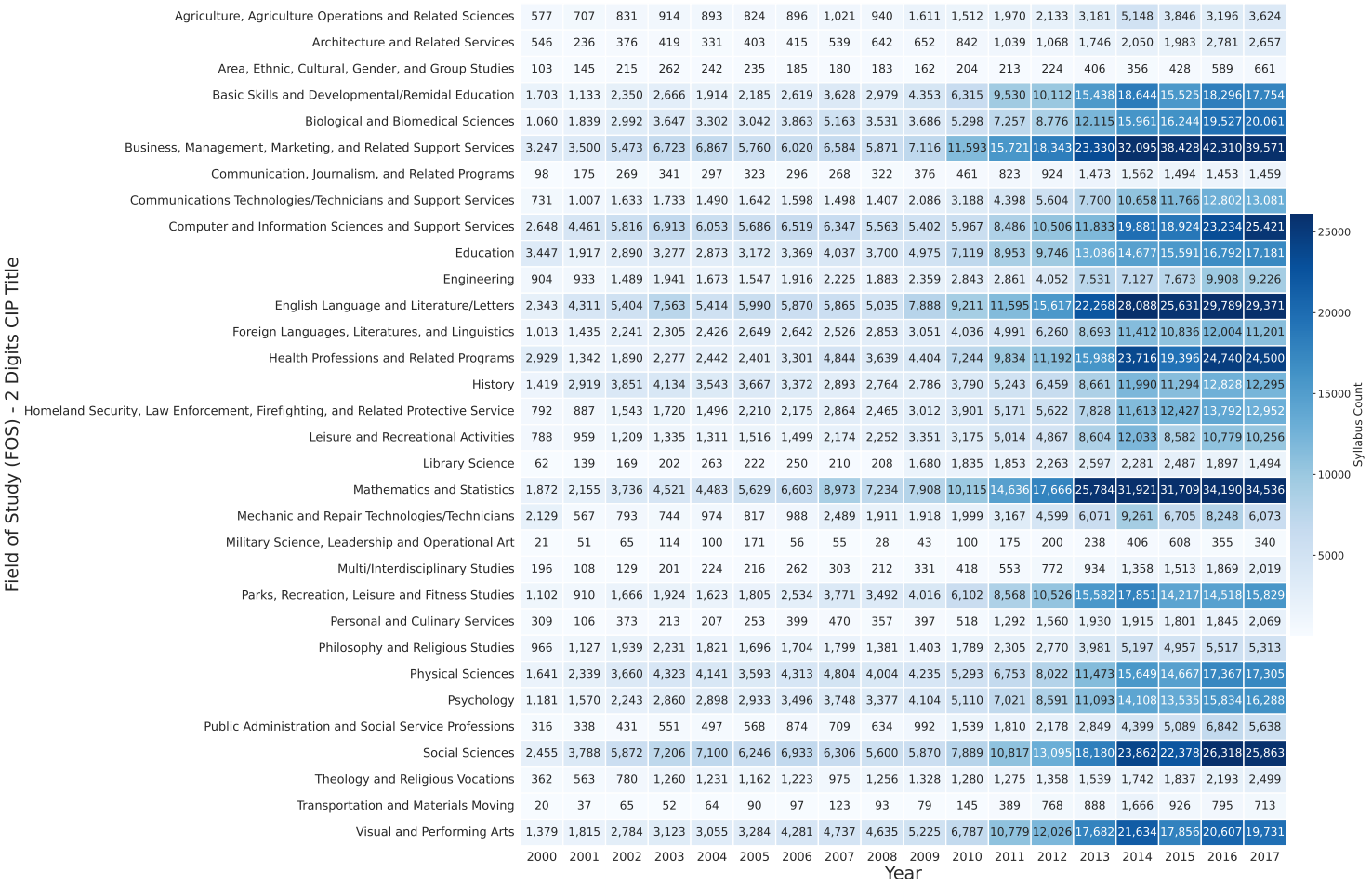


Figure 2: Frequency of syllabi per field of study based on 2-digit CIP and year between 2000 and 2017.

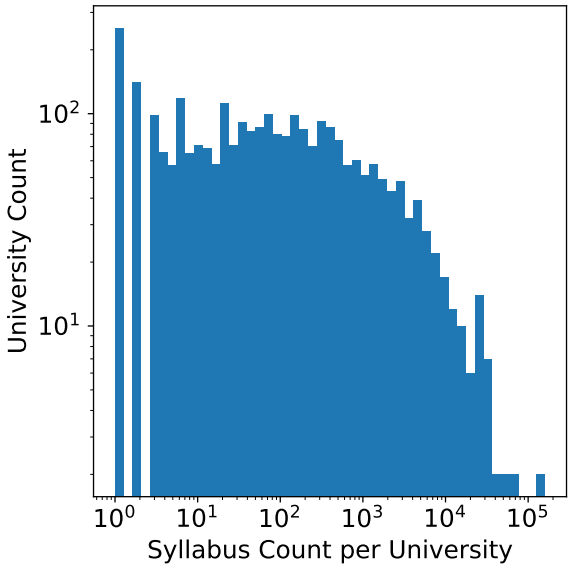


Figure 3: The syllabus count per university across all years and all FOS.

Institution Name	Count	Institution Name	Count
Alamo Colleges	160,041	South Texas College	12,009
University of Maryland University College	137,257	Dallas County Community College District	11,641
Amarillo College	75,198	Monterey Peninsula College	11,258
Lansing Community College	63,945	University of Minnesota System	11,220
University of Alabama, Tuscaloosa	54,278	South Plains College	11,184
Texas State Technical College	48,291	Wilkes University	11,036
Clark State Community College	46,094	Bellevue College	10,935
Houston Community College System	45,401	Reedley College	10,785
Santa Rosa Junior College	35,621	Modesto Junior College	10,780
Texas A&M University	33,579	University of Texas Rio Grande Valley	10,653
Rowan-Cabarrus Community College	33,403	University of Southern California	10,480
North Idaho College	31,292	Kentucky Community and Technical College System	10,264
University of Texas at Dallas	30,922	Santa Barbara City College	10,232
University of Georgia	30,349	Fullerton College	9,993
Texas State University–San Marcos	30,051	Pennsylvania State University	9,303
University of Texas at Arlington	28,651	Lewis and Clark Community College	9,103
San Diego Community College District	27,768	Southwestern Community College	9,004
Western Kentucky University	27,684	Chaminade University of Honolulu	8,866
Park University	27,096	Great Basin College	8,837
Sam Houston State University	26,523	University of Akron	8,787
Stephen F. Austin State University	26,170	University of California, San Diego	8,746
University of Michigan–Ann Arbor	25,883	University of Washington	8,707
Midwestern State University	25,460	Nova Southeastern University	8,555
Fresno City College	24,786	University of Colorado Boulder	8,553
George Mason University	24,046	Rutgers University	8,494
Oral Roberts University	23,950	University of South Florida	8,193
San Jose State University	23,505	San Mateo County Community College District	8,124
Minnesota State Colleges and Universities System	23,112	Palomar College	8,112
Texas Tech University	22,812	Westmont College	8,068
McLennan Community College	22,561	Mt. San Jacinto College	8,015
University of California, Irvine	22,505	Stony Brook University	7,982
Galveston College	22,077	New York University	7,959
Tyler Junior College	20,986	Victoria College	7,919
Clemson University	20,453	Iowa State University	7,885
University of Texas at Austin	20,387	University of Maryland, College Park	7,842
Collin College	17,608	Butte College	7,728
New Mexico Junior College	16,635	Merced College	7,388
Hartnell College	16,522	Ventura County Community College District	7,368
University of Texas at El Paso	15,842	Alvin Community College	7,347
Loyola University New Orleans	15,830	Ohlone College	7,240
University of North Texas	15,680	Imperial Valley College	7,174
University of Florida	14,919	Chaffey College	7,076
Casper College	14,665	Santa Monica College	6,870
University of Texas at San Antonio	14,595	California State University, San Marcos	6,865
University of Houston–Clear Lake	14,386	Palm Beach State College	6,742
Angelo State University	13,663	Carnegie Mellon University	6,702
Excelsior College	13,517	El Paso Community College	6,624
Texas A&M University–Commerce	13,264	Massachusetts Institute of Technology	6,329
Princeton University	13,182	Florida International University	6,287
Santa Ana College	12,969	Napa Valley College	6,224

Table 4: **Most frequent universities.** The syllabus count per top 100 universities across all years and all FOS.

2 Skills Inference Framework (Syllabus2O*NET)

Given the text from a course syllabus, we segment sentences using Stanza [4], a tool designed to partition text into individual sentences. We extracted 322,473,524 sentences from the course syllabi in our dataset. On average, each syllabus contains 101.96 sentences (median 83). Each course syllabus starts as unstructured text lacking metadata to distinguish between course logistics (e.g., scheduling and grading rubrics) and learning outcomes (e.g., course content and learning objectives). Thus, we implement a human-in-the-loop approach to remove sentences pertaining to Course Logistics while keeping sentences about Learning Outcomes. To do so, we compiled two distinct lists for labeling sentences. The list of Course Logistic related terms includes 356 phrases that are common but unrelated to the course content (e.g., "Plagiarism," "Attendance," and "Office hours"). The list of Learning Objective related terms includes 51 phrases such as "Analyze," "Versus," and "Outcome." The complete lists can be found on the code's GitHub page.[3] We removed sentences from each syllabus that contained Course Logistic phrases or lacked Learning Objective phrases resulting in the removal of 85.82% of the sentences. Each syllabus contains nearly 17.61 Learning Objective related sentences on average (median = 12) (see Table 5 for details on the statistics of Learning Objective sentence counts by FOS.).

Field of Study	**# Sent.** (median)	**# L. Sent.** (median)	**% L. Sent.** (median)	**Field of Study**	**# Sent.** (median)	**# L. Sent.** (median)	**% L. Sent.** (median)
Accounting	101.0	15.0	14.61%	Hebrew	65.0	7.0	11.32%
Agriculture	60.0	8.0	15.11%	History	88.0	9.0	10.62%
Anthropology	86.0	13.0	14.29%	Japanese	84.0	10.0	13.12%
Architecture	73.0	14.0	20.45%	Journalism	105.0	13.0	12.22%
Astronomy	76.0	10.0	12.5%	Law	60.0	7.0	12.12%
Atmospheric Sciences	66.0	11.0	14.29%	Library Science	107.0	13.0	12.04%
Basic Computer Skills	88.0	11.0	13.21%	Linguistics	72.0	9.0	11.43%
Basic Skills	76.0	11.0	14.29%	Marketing	106.0	16.0	15.79%
Biology	85.0	12.0	13.25%	Mathematics	75.0	10.0	12.73%
Business	97.0	15.0	15.22%	Mechanic / Repair Tech	62.0	10.0	17.95%
Chemistry	83.0	11.0	13.16%	Media / Communications	107.0	14.0	13.73%
Chinese	60.0	8.0	12.1%	Medicine	95.0	14.0	15.79%
Classics	47.0	4.0	9.3%	Military Science	71.0	11.0	14.35%
Computer Science	67.0	10.0	14.29%	Music	62.0	8.0	13.04%
Cosmetology	85.0	12.0	15.79%	Nursing	95.0	16.0	16.95%
Criminal Justice	79.0	12.0	15.38%	Nutrition	85.0	12.0	14.46%
Culinary Arts	61.0	14.0	22.22%	Philosophy	76.0	9.0	12.5%
Dance	59.0	11.0	17.54%	Physics	61.0	8.0	14.19%
Dentistry	69.0	16.0	25.93%	Political Science	96.0	12.0	12.23%
Earth Sciences	59.0	9.0	14.29%	Psychology	105.0	16.0	15.09%
Economics	74.0	11.0	13.92%	Public Safety	60.0	10.0	16.03%
Education	121.0	23.0	19.74%	Religion	84.0	11.0	12.61%
Engineering	51.0	9.0	16.28%	Sign Language	78.0	12.0	15.69%
English Literature	97.0	11.0	10.99%	Social Work	139.0	26.0	19.28%
Film and Photography	68.0	11.0	14.88%	Sociology	100.0	14.0	13.92%
Fine Arts	79.0	13.0	15.75%	Spanish	95.0	10.0	10.6%
Fitness and Leisure	54.0	10.0	17.19%	Theatre Arts	63.0	11.0	17.68%
French	69.0	7.0	10.45%	Theology	113.0	16.0	14.29%
Geography	70.0	10.0	14.29%	Transportation	94.0	12.0	13.46%
German	71.0	8.0	11.39%	Veterinary Medicine	70.0	10.0	11.39%
Health Technician	75.0	11.0	16.22%	Women's Studies	81.0	13.0	15.19%

Table 5: **Sentence statistics per FOS.** The table presents the median of the number of sentences (# Sent.), number of identified Learning Objective related sentences (# L. Sent.), and the percentage of the identified Learning Objective related sentences (% L. Sent.) per FOS.

[3] https://github.com/AlirezaJavadian/Syllabus-to-ONET

3 Mapping O*NET DWAs to Abilities (`DWA2Ability`)

In the "Gender, Education, and Skills" and "Temporal Trends in Teaching Activities" sections, we use our data to explore results from the literature. But existing studies are often forced to use less specific descriptions of skills, capabilities, or knowledge compared to the specificity of O*NET DWAs and Tasks (see O*NET's website for examples of DWAs[4] and Tasks)[5]. We facilitate a fairer comparison between our data and existing research using the O*NET Abilities taxonomy (see O*NET's website *Ability*[6] for examples). However, O*NET does not provide a standardized cross-walk linking DWAs, tasks, and abilities. To bridge this gap, we introduce `DWA2Ability`.

We start with the O*NET database[7] profiles of DWAs for each occupation. Next, we extract importance scores of each O*NET ability within each occupation.[8] We formulate a map between the two sets of occupation profiles as a regression using DWAs as independent variables and ability scores as dependent variables. We train a Random Forest Regressor [3] for each ability and fine-tune hyperparameters via Grid Search and 5-fold cross-validation. This approach yielded 52 models (i.e., one per O*NET ability), each achieving mean squared error of at most 0.025 (see Table 6 for details on model performance). Using the trained models, we map syllabis' DWA scores to abilities.

Ability	MSE	Ability	MSE
Arm-Hand Steadiness	0.025	Number Facility	0.012
Auditory Attention	0.014	Oral Comprehension	0.010
Category Flexibility	0.016	Oral Expression	0.011
Control Precision	0.019	Originality	0.014
Deductive Reasoning	0.011	Perceptual Speed	0.012
Depth Perception	0.014	Peripheral Vision	0.013
Dynamic Flexibility	0.007	Problem Sensitivity	0.010
Dynamic Strength	0.016	Rate Control	0.019
Explosive Strength	0.016	Reaction Time	0.020
Extent Flexibility	0.019	Response Orientation	0.012
Far Vision	0.011	Selective Attention	0.008
Finger Dexterity	0.015	Sound Localization	0.025
Flexibility of Closure	0.014	Spatial Orientation	0.017
Fluency of Ideas	0.014	Speech Clarity	0.008
Glare Sensitivity	0.017	Speech Recognition	0.014
Gross Body Coordination	0.013	Speed of Closure	0.010
Gross Body Equilibrium	0.018	Speed of Limb Movement	0.020
Hearing Sensitivity	0.012	Stamina	0.019
Inductive Reasoning	0.010	Static Strength	0.024
Information Ordering	0.017	Time Sharing	0.013
Manual Dexterity	0.023	Trunk Strength	0.017
Mathematical Reasoning	0.011	Visual Color Discrimination	0.019
Memorization	0.013	Visualization	0.017
Multilimb Coordination	0.022	Wrist-Finger Speed	0.024
Near Vision	0.018	Written Comprehension	0.012
Night Vision	0.015	Written Expression	0.014

Table 6: **Abilities model training performance**. The mean squared error (MSE) obtained from the 5-fold cross-validation (CV) for finding the best model of ability.

[4] https://www.onetcenter.org/dictionary/20.1/excel/dwa_reference.html (accessed on 5th February 2024)
[5] https://www.onetcenter.org/dictionary/20.1/excel/task_statements.html (accessed on 5th February 2024)
[6] https://www.onetonline.org/find/descriptor/browse/1.A (accessed on 5th February 2024)
[7] https://www.onetonline.org/find/all (accessed on 5th February 2024)
[8] https://www.onetonline.org/find/descriptor/browse/1.A (accessed on 5th February 2024)

4 Top 10 Detailed Work Activities (DWA) per Field of Study (FOS)

Table 7 lists the top 10 inferred workplace activities with the highest DWA scores per FOS. We mask the frequently occurring skills based on their prevalence. This was implemented by evaluating the top 100 DWAs for each FOS and masking those that were commonly observed. Note that masked DWAs remain accessible in the published dataset (see Section "Skill Normalization").

Detailed Work Activity (DWA)	Rank
Accounting	
Analyze budgetary or accounting data.	1
Analyze financial records to improve budgeting or planning.	2
Analyze financial information.	3
Develop financial analysis methods.	4
Prepare financial documents, reports, or budgets.	5
Implement financial decisions.	6
Analyze financial records to improve efficiency.	7
Manage financial activities of the organization.	8
Analyze financial records or reports to determine state of operations.	9
Reconcile records of sales or other financial transactions.	10
Agriculture	
Document technical specifications or requirements.	1
Send information, materials or documentation.	2
Establish standards for products, processes, or procedures.	3
Estimate supplies, ingredients, or staff requirements for food preparation activities.	4
Determine types of equipment, tools, or materials needed for jobs.	5
Prepare materials or solutions for animal or plant use.	6
Analyze data to assess operational or project effectiveness.	7
Develop plans to manage natural or renewable resources.	8
Evaluate environmental or sustainability projects.	9
Evaluate potential of products, technologies, or resources.	10
Anthropology	
Conduct anthropological or archaeological research.	1
Research issues related to the environment or sustainable business practices.	2
Send information, materials or documentation.	3
Analyze environmental data.	4
Interview people to obtain information about actions or status of individuals.	5
Conduct surveys in organizations.	6
Assist other educational professionals with projects or research.	7
Document events or evidence, using photographic or audiovisual equipment.	8
Teach social science courses at the college level.	9
Analyze geological or geographical data.	10
Architecture	
Design structures or facilities.	1
Plan facility layouts or designs.	2
Plan layout of construction, installation, or repairs.	3
Develop artistic or design concepts for decoration, exhibition, or commercial purposes.	4
Determine construction project layouts.	5
Design civil structures or systems.	6
Collaborate with others to develop or refine designs.	7
Review art or design materials.	8
Create construction or installation diagrams.	9
Discuss designs or plans with clients.	10

Astronomy

Measure physical or chemical properties of materials or objects.	1
Inspect sets or exhibits.	2
Research new technologies.	3
Analyze geological or geographical data.	4
Measure radiation levels.	5
Analyze geological samples.	6
Calibrate scientific or technical equipment.	7
Collaborate on research activities with scientists or technical specialists.	8
Stay informed about current developments in field of specialization.	9
Send information, materials or documentation.	10

Atmospheric Sciences

Conduct climatological research.	1
Analyze environmental data.	2
Research hydrologic features or processes.	3
Record information about environmental conditions.	4
Compile environmental or climatological data.	5
Measure environmental characteristics.	6
Analyze geological or geographical data.	7
Monitor environmental conditions to detect hazards.	8
Inspect condition of natural environments.	9
Collect environmental data or samples.	10

Basic Computer Skills

Develop procedures for data entry or processing.	1
Train others in computer interface or software use.	2
Teach others to use computer equipment or hardware.	3
Design software applications.	4
Develop specifications for computer network operation.	5
Design integrated computer systems.	6
Develop models of information or communications systems.	7
Create databases to store electronic data.	8
Implement security measures for computer or information systems.	9
Develop procedures for data management.	10

Basic Skills

Assist other educational professionals with projects or research.	1
Manage documentation to ensure organization or accuracy.	2
Attend events to develop professional knowledge.	3
Maintain current knowledge related to work activities.	4
Maintain professional knowledge or certifications.	5
Evaluate performance of educational staff.	6
Evaluate program effectiveness.	7
Conduct employee training programs.	8
Encourage patients or clients to develop life skills.	9
Counsel clients regarding educational or vocational issues.	10

Biology

Prepare biological specimens for laboratory analysis.	1
Prepare biological samples for testing or analysis.	2
Analyze biological samples.	3
Research genetic characteristics or expression.	4
Plan biological research.	5
Analyze laboratory findings.	6
Research microbiological or chemical processes or structures.	7

Analyze laboratory specimens to detect abnormalities or other problems.	8
Supervise laboratory work.	9
Collect biological specimens.	10

Business

Establish business management methods.	1
Develop organizational policies or programs.	2
Establish organizational guidelines or policies.	3
Analyze data to inform personnel decisions.	4
Communicate organizational information to customers or other stakeholders.	5
Represent the organization in external relations.	6
Communicate with clients about products, procedures, and policies.	7
Confer with managers to make operational decisions.	8
Establish operational policies.	9
Establish interpersonal business relationships to facilitate work activities.	10

Chemistry

Prepare compounds or solutions for products or testing.	1
Measure physical or chemical properties of materials or objects.	2
Prepare biological samples for testing or analysis.	3
Prepare biological specimens for laboratory analysis.	4
Research engineering aspects of biological or chemical processes.	5
Test materials, solutions, or samples.	6
Collect samples of materials or products for testing.	7
Prepare chemicals for work application.	8
Research microbiological or chemical processes or structures.	9
Supervise laboratory work.	10

Chinese

Attend events to develop professional knowledge.	1
Present business-related information to audiences.	2
Teach humanities courses at the college level.	3
Maintain current knowledge related to work activities.	4
Teach online courses.	5
Assist other educational professionals with projects or research.	6
Accompany individuals or groups to activities.	7
Record information from meetings or other formal proceedings.	8
Supervise trainees.	9
Communicate with others to coordinate material handling or movement.	10

Classics

Teach humanities courses at the college level.	1
Inspect sets or exhibits.	2
Manage documentation to ensure organization or accuracy.	3
Maintain current knowledge related to work activities.	4
Record information from meetings or other formal proceedings.	5
Coordinate student extracurricular activities.	6
Maintain student records.	7
Compile specialized bibliographies or lists of materials.	8
Teach online courses.	9
Document operational procedures.	10

Computer Science

Develop specifications for computer network operation.	1
Design software applications.	2
Develop models of information or communications systems.	3

Design integrated computer systems.	4
Analyze security of systems, network, or data.	5
Develop procedures for data entry or processing.	6
Design computer modeling or simulation programs.	7
Analyze design requirements for computer or electronics systems.	8
Implement security measures for computer or information systems.	9
Compile technical information or documentation.	10

Cosmetology

Apply solutions to hair for therapeutic or cosmetic purposes.	1
Apply cleansing or conditioning agents to client hair, scalp, or skin.	2
Assess skin or hair conditions.	3
Maintain professional knowledge or certifications.	4
Groom wigs or hairpieces.	5
Communicate with clients about products, procedures, and policies.	6
Establish standards for products, processes, or procedures.	7
Trim client hair.	8
Evaluate personnel practices to ensure adherence to regulations.	9
Establish work standards.	10

Criminal Justice

Interview people to gather information about criminal activities.	1
Research relevant legal materials to aid decision making.	2
Investigate crimes committed within organizations.	3
Analyze forensic evidence to solve crimes.	4
Process forensic or legal evidence in accordance with procedures.	5
Examine records or other types of data to investigate criminal activities.	6
Apprehend criminal suspects.	7
Identify implications for cases from legal precedents or other legal information.	8
Interview witnesses, suspects, or claimants.	9
Examine crime scenes to obtain evidence.	10

Culinary Arts

Estimate supplies, ingredients, or staff requirements for food preparation activities.	1
Evaluate quality of food ingredients or prepared foods.	2
Prepare foods for cooking or serving.	3
Present food or beverage information or menus to customers.	4
Determine food production methods.	5
Clean food preparation areas, facilities, or equipment.	6
Communicate dining or order details to kitchen personnel.	7
Inspect food products.	8
Coordinate activities of food service staff.	9
Create new recipes or food presentations.	10

Dance

Choreograph dances.	1
Perform dances.	2
Train others on performance techniques.	3
Communicate with others to coordinate material handling or movement.	4
Coordinate training activities.	5
Evaluate performance of educational staff.	6
Evaluate program effectiveness.	7
Coordinate musical rehearsals or performances.	8
Create musical compositions, arrangements or scores.	9
Schedule activities or facility use.	10

Dentistry

Diagnose dental conditions.	1
Treat dental problems or diseases.	2
Examine mouth, teeth, gums, or related facial structures.	3
Adjust dental devices or appliances to ensure fit.	4
Inspect medical or dental assistive devices.	5
Examine medical instruments or equipment to ensure proper operation.	6
Confer with other professionals to plan patient care.	7
Assist patients with hygiene or daily living activities.	8
Supervise patient care personnel.	9
Teach medical procedures or medical equipment use to patients.	10

Earth Sciences

Analyze geological samples.	1
Analyze geological or geographical data.	2
Collect geographical or geological field data.	3
Collect geological samples.	4
Survey land or bodies of water to measure or determine features.	5
Research hydrologic features or processes.	6
Analyze environmental data.	7
Inspect condition of natural environments.	8
Locate natural resources using geospatial or other environmental data.	9
Collect environmental data or samples.	10

Economics

Apply mathematical models of financial or business conditions.	1
Analyze market conditions or trends.	2
Determine pricing or monetary policies.	3
Develop financial analysis methods.	4
Analyze financial information.	5
Analyze market research data.	6
Analyze business or financial data.	7
Analyze industry trends.	8
Implement financial decisions.	9
Evaluate potential of products, technologies, or resources.	10

Education

Evaluate performance of educational staff.	1
Assist other educational professionals with projects or research.	2
Discuss child development and behavior with parents or guardians.	3
Evaluate program effectiveness.	4
Coordinate student extracurricular activities.	5
Establish organizational guidelines or policies.	6
Conduct employee training programs.	7
Supervise school or student activities.	8
Establish work standards.	9
Develop daily schedules for children or families.	10

Engineering

Research advanced engineering designs or applications.	1
Analyze design requirements for computer or electronics systems.	2
Document technical design details.	3
Design electrical equipment or systems.	4
Design industrial equipment.	5
Design civil structures or systems.	6
Evaluate technical data to determine effect on designs or plans.	7

Design energy production or management equipment or systems.	8
Design structures or facilities.	9
Design electromechanical equipment or systems.	10

English Literature

Manage documentation to ensure organization or accuracy.	1
Document design or operational test results.	2
Issue warnings or citations.	3
Record information from meetings or other formal proceedings.	4
Document technical specifications or requirements.	5
Assist other educational professionals with projects or research.	6
Teach humanities courses at the college level.	7
Prepare documentation for contracts, transactions, or regulatory compliance.	8
Document operational procedures.	9
Send information, materials or documentation.	10

Film and Photography

Develop artistic or design concepts for decoration, exhibition, or commercial purposes.	1
Review art or design materials.	2
Discuss production content and progress with others.	3
Construct distinctive physical objects for artistic, functional, or commercial purposes.	4
Draw detailed or technical illustrations.	5
Prepare film for distribution or use.	6
Collaborate with others to develop or refine designs.	7
Prepare production storyboards.	8
Review production information to determine costume or makeup requirements.	9
Design video game features or details.	10

Fine Arts

Develop artistic or design concepts for decoration, exhibition, or commercial purposes.	1
Review art or design materials.	2
Coordinate artistic activities.	3
Construct distinctive physical objects for artistic, functional, or commercial purposes.	4
Draw detailed or technical illustrations.	5
Create diagrams or blueprints for workpieces or products.	6
Collaborate with others to develop or refine designs.	7
Design layout of art or product exhibits, displays, or promotional materials.	8
Create graphical representations of structures or landscapes.	9
Draw guide lines or markings on materials or workpieces using patterns or other references.	10

Fitness and Leisure

Advise athletes, coaches, or trainers on exercise regimens, nutrition, or equipment use.	1
Coordinate training activities.	2
Schedule activities or facility use.	3
Evaluate program effectiveness.	4
Supervise trainees.	5
Conduct health or safety training programs.	6
Train others on performance techniques.	7
Teach health management classes.	8
Conduct employee training programs.	9
Evaluate performance of educational staff.	10

French

Attend events to develop professional knowledge.	1
Assist other educational professionals with projects or research.	2
Record information from meetings or other formal proceedings.	3

Teach online courses.	4
Accompany individuals or groups to activities.	5
Maintain current knowledge related to work activities.	6
Supervise trainees.	7
Communicate with others to coordinate material handling or movement.	8
Liaise between departments or other groups to improve function or communication.	9
Teach humanities courses at the college level.	10

Geography

Survey land or bodies of water to measure or determine features.	1
Analyze geological or geographical data.	2
Create maps.	3
Locate natural resources using geospatial or other environmental data.	4
Collect geographical or geological field data.	5
Analyze environmental data.	6
Create graphical representations of structures or landscapes.	7
Prepare maps.	8
Gather physical survey data.	9
Research hydrologic features or processes.	10

German

Attend events to develop professional knowledge.	1
Record information from meetings or other formal proceedings.	2
Liaise between departments or other groups to improve function or communication.	3
Accompany individuals or groups to activities.	4
Present business-related information to audiences.	5
Maintain current knowledge related to work activities.	6
Assist other educational professionals with projects or research.	7
Manage documentation to ensure organization or accuracy.	8
Teach humanities courses at the college level.	9
Supervise trainees.	10

Health Technician

Examine medical instruments or equipment to ensure proper operation.	1
Confer with other professionals to plan patient care.	2
Interact with patients to build rapport or provide emotional support.	3
Monitor patients following surgeries or other treatments.	4
Instruct patients in the use of assistive equipment.	5
Supervise patient care personnel.	6
Supervise technical medical personnel.	7
Collect biological specimens from patients.	8
Diagnose medical conditions.	9
Teach medical procedures or medical equipment use to patients.	10

Hebrew

Supervise trainees.	1
Teach online courses.	2
Manage documentation to ensure organization or accuracy.	3
Prepare documentation for contracts, transactions, or regulatory compliance.	4
Document technical specifications or requirements.	5
Record information from meetings or other formal proceedings.	6
Maintain current knowledge related to work activities.	7
Prepare information or documentation related to legal or regulatory matters.	8
Conduct employee training programs.	9
Maintain student records.	10

History

Teach humanities courses at the college level.	1
Interview people to obtain information about actions or status of individuals.	2
Evaluate characteristics of archival or historical objects.	3
Manage documentation to ensure organization or accuracy.	4
Record information from meetings or other formal proceedings.	5
Monitor current trends.	6
Conduct anthropological or archaeological research.	7
Teach social science courses at the college level.	8
Stay informed about current developments in field of specialization.	9
Send information, materials or documentation.	10

Japanese

Attend events to develop professional knowledge.	1
Teach online courses.	2
Supervise trainees.	3
Present business-related information to audiences.	4
Communicate with others to coordinate material handling or movement.	5
Accompany individuals or groups to activities.	6
Maintain current knowledge related to work activities.	7
Conduct amusement or gaming activities.	8
Maintain professional knowledge or certifications.	9
Assist other educational professionals with projects or research.	10

Journalism

Interview others for news or entertainment purposes.	1
Present information to the public.	2
Manage content of broadcasts or presentations.	3
Interview people to obtain information about actions or status of individuals.	4
Write articles, books or other original materials in area of expertise.	5
Discuss production content and progress with others.	6
Document events or evidence, using photographic or audiovisual equipment.	7
Provide information to coworkers.	8
Provide information to the general public.	9
Coordinate reporting or editing activities.	10

Law

Research relevant legal materials to aid decision making.	1
Identify implications for cases from legal precedents or other legal information.	2
Prepare written decisions for legal proceedings.	3
Make decisions in legal cases.	4
Prepare information or documentation related to legal or regulatory matters.	5
Represent the interests of clients in legal proceedings.	6
Rule on admissibility of legal proceedings.	7
Prepare documentation of legal proceedings.	8
Update knowledge of legal or regulatory environments.	9
Document information related to legal proceedings.	10

Library Science

Help patrons use library or archival resources.	1
Organize informational materials.	2
Compile specialized bibliographies or lists of materials.	3
Develop policies or procedures for archives, museums or libraries.	4
Develop library or archival databases.	5
Manage documentation to ensure organization or accuracy.	6
Distribute instructional or library materials.	7

Identify information technology project resource requirements.	8
Provide information to coworkers.	9
Present information to the public.	10

Linguistics

Send information, materials or documentation.	1
Manage documentation to ensure organization or accuracy.	2
Document technical specifications or requirements.	3
Document design or operational test results.	4
Interview people to obtain information about actions or status of individuals.	5
Develop procedures for data entry or processing.	6
Develop models of information or communications systems.	7
Record information from meetings or other formal proceedings.	8
Stay informed about current developments in field of specialization.	9
Analyze market research data.	10

Marketing

Collaborate with others to develop or implement marketing strategies.	1
Perform marketing activities.	2
Analyze market research data.	3
Develop business or market strategies.	4
Collaborate with others in marketing activities.	5
Train customers in the use of products.	6
Deliver promotional presentations to current or prospective customers.	7
Develop promotional materials.	8
Analyze market or customer related data.	9
Create marketing materials.	10

Mathematics

Apply mathematical models of financial or business conditions.	1
Develop financial analysis methods.	2
Determine operational methods.	3
Review blueprints or other instructions to determine operational methods or sequences.	4
Analyze data to determine project feasibility.	5
Analyze data to assess operational or project effectiveness.	6
Calculate data to inform organizational operations.	7
Develop procedures for data entry or processing.	8
Analyze data to inform personnel decisions.	9
Analyze business or financial data.	10

Mechanic / Repair Tech

Determine types of equipment, tools, or materials needed for jobs.	1
Conduct validation tests of equipment or processes.	2
Select tools, equipment, or technologies for use in operations or projects.	3
Test mechanical equipment to ensure proper functioning.	4
Inspect equipment to ensure safety or proper functioning.	5
Inspect safety equipment to ensure proper functioning.	6
Monitor work areas or procedures to ensure compliance with safety procedures.	7
Evaluate characteristics of equipment or systems.	8
Develop safety standards, policies, or procedures.	9
Inspect facilities or equipment to ensure specifications are met.	10

Media / Communications

Present business-related information to audiences.	1
Present information to the public.	2
Liaise between departments or other groups to improve function or communication.	3

Manage content of broadcasts or presentations.	4
Provide information to coworkers.	5
Communicate organizational information to customers or other stakeholders.	6
Deliver promotional presentations to current or prospective customers.	7
Conduct surveys in organizations.	8
Establish organizational guidelines or policies.	9
Establish interpersonal business relationships to facilitate work activities.	10

Medicine

Analyze health-related data.	1
Advise patients on healthcare system processes.	2
Collaborate with healthcare professionals to plan or provide treatment.	3
Develop patient therapy programs.	4
Confer with other professionals to plan patient care.	5
Interact with patients to build rapport or provide emotional support.	6
Establish standards for medical care.	7
Establish nursing policies or standards.	8
Assess patient work, living, or social environments.	9
Implement therapeutic programs to improve patient functioning.	10

Military Science

Recruit personnel.	1
Supervise trainees.	2
Confer with others to conduct or arrange operational activities.	3
Establish operational policies.	4
Direct activities of subordinates.	5
Coordinate personnel recruitment activities.	6
Coordinate training activities.	7
Resolve personnel problems.	8
Liaise between departments or other groups to improve function or communication.	9
Evaluate effectiveness of personnel policies or practices.	10

Music

Create musical compositions, arrangements or scores.	1
Coordinate musical rehearsals or performances.	2
Train others on performance techniques.	3
Perform for recordings.	4
Perform music for the public.	5
Adjust tuning or functioning of musical instruments.	6
Audition or interview potential performers or staff members.	7
Choreograph dances.	8
Evaluate performance of educational staff.	9
Develop artistic or design concepts for decoration, exhibition, or commercial purposes.	10

Nursing

Establish nursing policies or standards.	1
Confer with other professionals to plan patient care.	2
Collaborate with healthcare professionals to plan or provide treatment.	3
Interact with patients to build rapport or provide emotional support.	4
Implement therapeutic programs to improve patient functioning.	5
Develop patient therapy programs.	6
Collaborate with other professionals to assess client needs or plan treatments.	7
Advise patients on healthcare system processes.	8
Assess patient work, living, or social environments.	9
Supervise patient care personnel.	10

Nutrition

Monitor nutrition related activities of individuals or groups.	1
Manage preparation of special meals or diets.	2
Evaluate quality of food ingredients or prepared foods.	3
Determine food production methods.	4
Estimate supplies, ingredients, or staff requirements for food preparation activities.	5
Analyze health-related data.	6
Prepare foods or meals.	7
Research methods to improve food products.	8
Prepare foods for cooking or serving.	9
Inspect food products.	10

Philosophy

Identify implications for cases from legal precedents or other legal information.	1
Resolve interpersonal conflicts.	2
Research relevant legal materials to aid decision making.	3
Mediate disputes.	4
Review laws or regulations to maintain professional knowledge.	5
Maintain knowledge of laws or regulations.	6
Maintain current knowledge related to work activities.	7
Teach humanities courses at the college level.	8
Manage documentation to ensure organization or accuracy.	9
Issue warnings or citations.	10

Physics

Measure physical or chemical properties of materials or objects.	1
Create physical models or prototypes.	2
Design electromechanical equipment or systems.	3
Create electrical schematics.	4
Design electrical equipment or systems.	5
Research new technologies.	6
Research advanced engineering designs or applications.	7
Calibrate scientific or technical equipment.	8
Send information, materials or documentation.	9
Review blueprints or other instructions to determine operational methods or sequences.	10

Political Science

Advise others on matters of public policy.	1
Teach social science courses at the college level.	2
Analyze impact of legal or regulatory changes.	3
Research relevant legal materials to aid decision making.	4
Research issues related to the environment or sustainable business practices.	5
Forecast economic, political, or social trends.	6
Identify implications for cases from legal precedents or other legal information.	7
Appraise environmental impact of regulations or policies.	8
Interview people to obtain information about actions or status of individuals.	9
Mediate disputes.	10

Psychology

Treat patients using psychological therapies.	1
Develop patient therapy programs.	2
Discuss child development and behavior with parents or guardians.	3
Conduct surveys in organizations.	4
Encourage patients or clients to develop life skills.	5
Analyze data to inform personnel decisions.	6
Counsel clients regarding interpersonal issues.	7

Intervene in crisis situations to assist clients.	8
Analyze health-related data.	9
Evaluate potential problems in home or work environments of clients.	10

Public Safety

Develop emergency response plans or procedures.	1
Develop fire safety or prevention programs or plans.	2
Advise others on management of emergencies or hazardous situations or materials.	3
Communicate safety or hazard information to others.	4
Inspect facilities to ensure compliance with fire regulations.	5
Develop safety standards, policies, or procedures.	6
Develop emergency procedures.	7
Monitor work areas or procedures to ensure compliance with safety procedures.	8
Maintain contingency plans for disaster recovery.	9
Direct fire fighting or prevention activities.	10

Religion

Develop promotional strategies for religious organizations.	1
Teach humanities courses at the college level.	2
Conduct anthropological or archaeological research.	3
Manage documentation to ensure organization or accuracy.	4
Send information, materials or documentation.	5
Teach online courses.	6
Record information from meetings or other formal proceedings.	7
Assist other educational professionals with projects or research.	8
Establish work standards.	9
Interview people to obtain information about actions or status of individuals.	10

Sign Language

Assist individuals with special needs.	1
Test patient hearing.	2
Communicate with others to coordinate material handling or movement.	3
Instruct patients in the use of assistive equipment.	4
Operate audio-visual equipment.	5
Evaluate performance of educational staff.	6
Accompany individuals or groups to activities.	7
Evaluate program effectiveness.	8
Assist other educational professionals with projects or research.	9
Liaise between departments or other groups to improve function or communication.	10

Social Work

Train staff members in social services skills.	1
Develop patient therapy programs.	2
Maintain professional social services knowledge.	3
Intervene in crisis situations to assist clients.	4
Collaborate with other professionals to assess client needs or plan treatments.	5
Help clients get needed services or resources.	6
Implement therapeutic programs to improve patient functioning.	7
Refer individuals to educational or work programs.	8
Encourage patients or clients to develop life skills.	9
Evaluate program effectiveness.	10

Sociology

Conduct surveys in organizations.	1
Teach social science courses at the college level.	2
Analyze market research data.	3

Advise others on social or educational issues.	4
Interview people to obtain information about actions or status of individuals.	5
Research issues related to the environment or sustainable business practices.	6
Conduct opinion surveys or needs assessments.	7
Analyze jobs using observation, survey, or interview techniques.	8
Evaluate potential problems in home or work environments of clients.	9
Analyze health-related data.	10

Spanish

Assist other educational professionals with projects or research.	1
Evaluate program effectiveness.	2
Evaluate performance of educational staff.	3
Attend events to develop professional knowledge.	4
Accompany individuals or groups to activities.	5
Conduct eligibility or selection interviews.	6
Maintain current knowledge related to work activities.	7
Record information from meetings or other formal proceedings.	8
Supervise trainees.	9
Advise others on social or educational issues.	10

Theatre Arts

Direct productions or performances.	1
Audition or interview potential performers or staff members.	2
Review production information to determine costume or makeup requirements.	3
Entertain public with comedic or dramatic performances.	4
Train others on performance techniques.	5
Coordinate musical rehearsals or performances.	6
Prepare production storyboards.	7
Develop artistic or design concepts for decoration, exhibition, or commercial purposes.	8
Choreograph dances.	9
Discuss production content and progress with others.	10

Theology

Develop promotional strategies for religious organizations.	1
Manage outreach activities.	2
Provide counsel, comfort, or encouragement to individuals or families.	3
Encourage patients or clients to develop life skills.	4
Establish work standards.	5
Manage documentation to ensure organization or accuracy.	6
Establish organizational guidelines or policies.	7
Supervise trainees.	8
Intervene in crisis situations to assist clients.	9
Teach humanities courses at the college level.	10

Transportation

Plan flight operations.	1
Coordinate flight control or management activities.	2
Communicate safety or hazard information to others.	3
Develop safety standards, policies, or procedures.	4
Operate transportation equipment to demonstrate function or malfunction.	5
Pilot aircraft.	6
Confer with other personnel to resolve design or operational problems.	7
Monitor work areas or procedures to ensure compliance with safety procedures.	8
Conduct validation tests of equipment or processes.	9
Inspect safety equipment to ensure proper functioning.	10

Veterinary Medicine	
Examine animals to detect illness, injury or other problems.	1
Treat animal injuries or illnesses.	2
Prepare biological specimens for laboratory analysis.	3
Perform animal breeding procedures.	4
Monitor animal behavior or condition.	5
Research livestock management methods.	6
Prepare biological samples for testing or analysis.	7
Collect biological specimens from patients.	8
Provide care for animals.	9
Analyze laboratory specimens to detect abnormalities or other problems.	10
Women's Studies	
Teach social science courses at the college level.	1
Advise others on social or educational issues.	2
Teach humanities courses at the college level.	3
Assist other educational professionals with projects or research.	4
Conduct surveys in organizations.	5
Research issues related to the environment or sustainable business practices.	6
Establish work standards.	7
Interview people to obtain information about actions or status of individuals.	8
Advocate for individual or community needs.	9
Send information, materials or documentation.	10

Table 7: **Top 10 inferred workplace activities with the highest DWA scores per FOS.** We mask the frequently occurring skills based on their prevalence. This was implemented by evaluating the top 100 DWAs for each FOS and masking those that were commonly observed. Note that masked DWAs remain accessible in the published dataset.

5 Community Detection and Clustering

We employ the Louvain community detection method [1] to cluster the most closely related academic majors. This process entailed averaging the scores for each DWA within identical majors, culminating in a unified vector representation for each field of study. This analysis discerned four distinct communities (see Figure 4). The two most extensive communities generally correspond to Non-STEM (highlighted in green) and STEM (in blue) disciplines. Notably, the smallest cluster, comprising Accounting, Economics, and Marketing, tends more towards Non-STEM fields, likely attributed to their social sciences orientation. The fourth community encapsulates fields such as Fitness, Education, and Nutrition.

Adopting an analogous approach with "Tasks" rather than DWAs, the results exhibit a significant variation (see Figure 5). The key divergence when employing "Tasks" is the formation of three communities instead of four. The group previously including Accounting, Economics, and Marketing amalgamated with the broader STEM category, indicating a shift in community dynamics based on the classification criterion used.

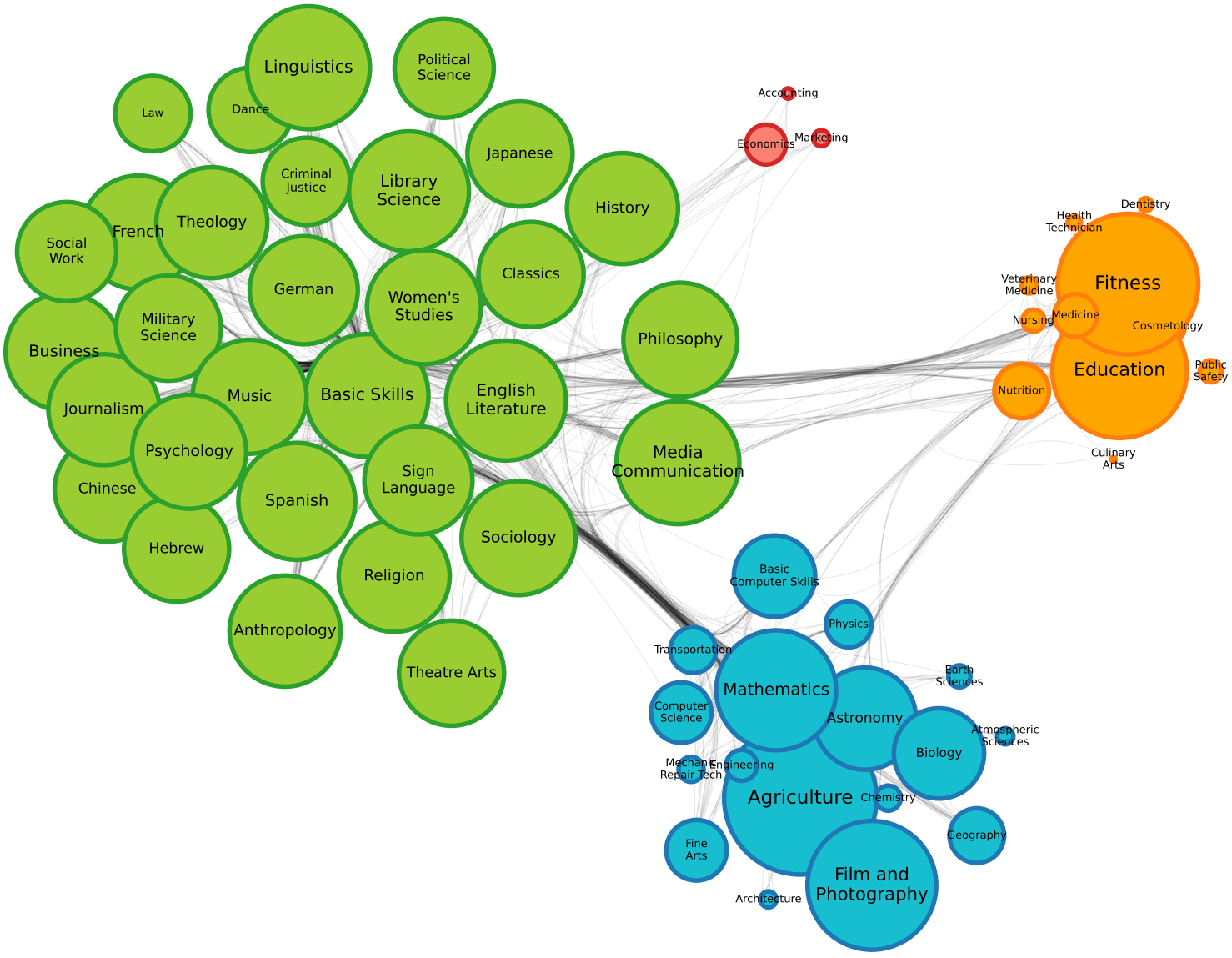


Figure 4: **The similarity of FOS according to their DWA profiles.** The graph represents the results of the Louvain community detection method [1] to identify the similar FOS. (*Note:* Node size represents the Degree Centrality).

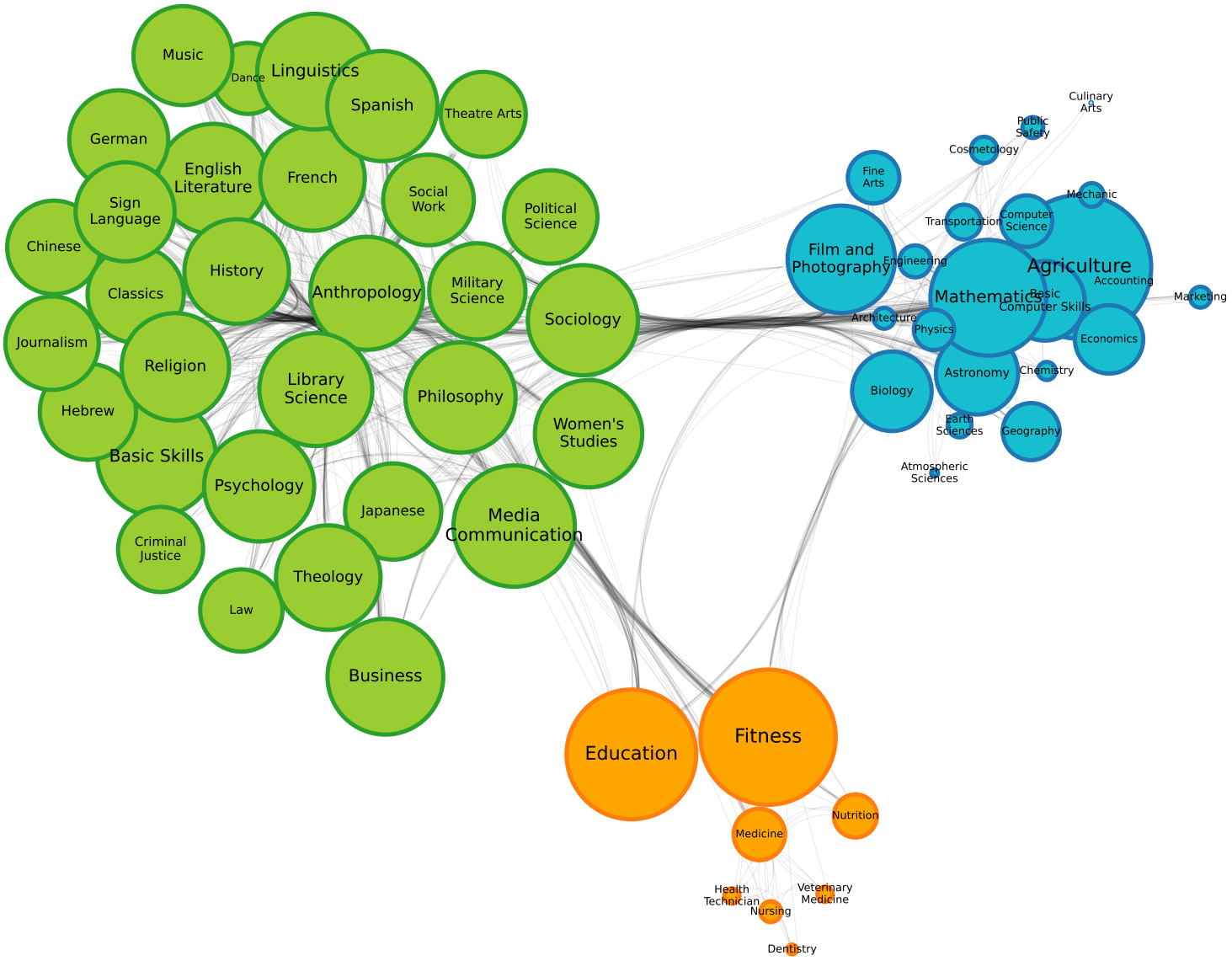


Figure 5: **The similarity of FOS according to their Task profiles.** The graph represents the results of the Louvain community detection method [1] to identify the similar FOS. (*Note:* Node size represents the Degree Centrality)

Moreover, we employ agglomerative hierarchical clustering technique [2] on the DWA vector representations of academic majors, aiming to elucidate their hierarchical relationships. Hierarchical clustering generates a nested sequence of clusters, allowing for an in-depth exploration of clusters at varying levels of granularity without predefining a specific number of categories (in this context, groups of majors). In this framework, FOS are deemed similar if they share akin work activities (see Figure 6 for DWAs and Figure 7 for Tasks).

The resulting dendrogram illustrates that majors cluster in a manner consistent with expectations. For instance, in case of DWAs (see Figure 6), Marketing and Accounting emerge as closely related, akin to the pairing of Linguistics and History. Progressing through the hierarchy, these clusters show increasing similarity to other fields, such as Anthropology. Notably, just prior to the final clustering step, which amalgamates all majors (indicated in blue), two predominant clusters are discernible: one representing STEM majors (in green) and the other Non-STEM majors (in orange).

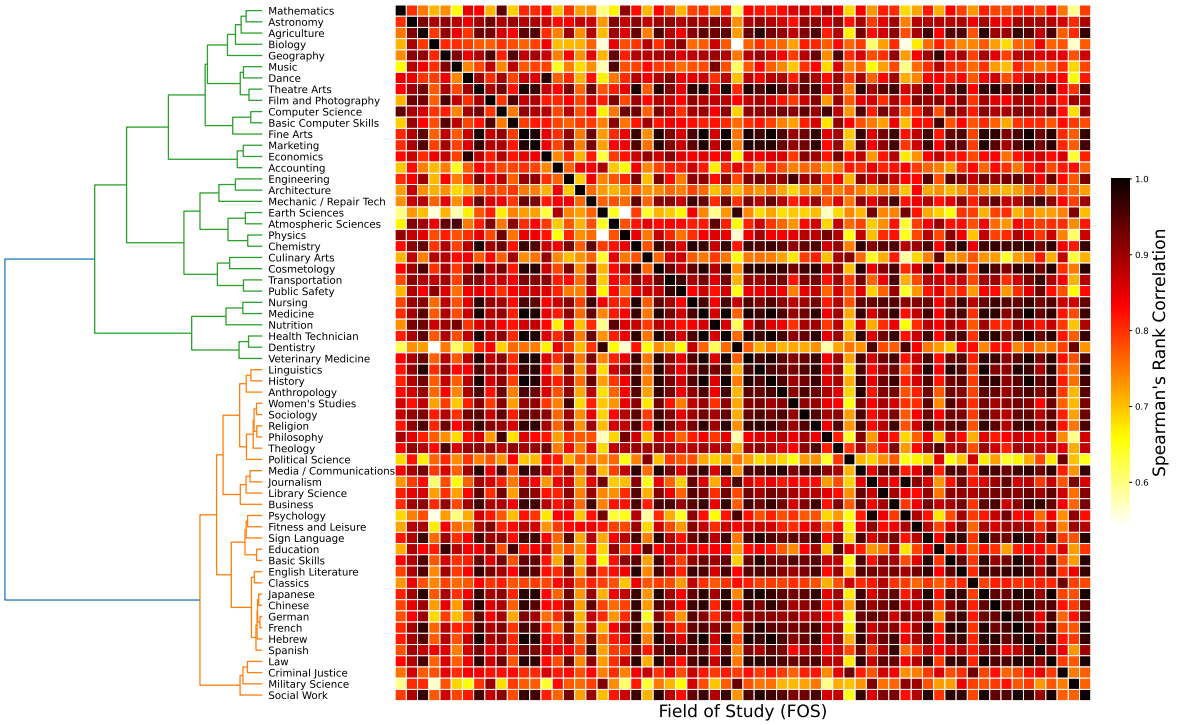


Figure 6: **The similarity of FOS according to their DWA profiles.** The heatmap shows the Spearman's rank correlation between the fields of study and the dendrogram represents the results of hierarchically clustering similar FOS.

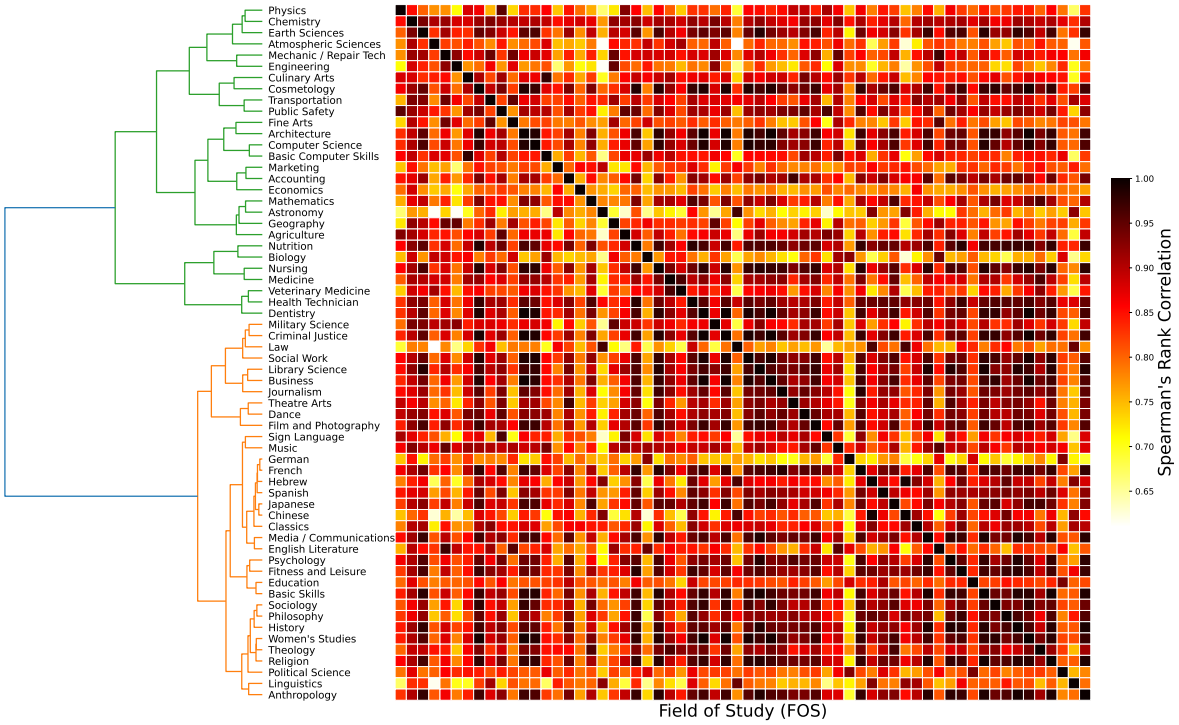


Figure 7: **The similarity of FOS according to their Task profiles.** The heatmap shows the Spearman's rank correlation between the fields of study and the dendrogram represents the results of hierarchically clustering similar FOS.

6 Gender, Education, and Skills

To make a more direct comparison with the OECD report, we quantify these differences at the O*NET Ability level using our `DWA2Ability` trained models (see Section "Mapping O*NET DWAs to Abilities (`DWA2Ability`)"). In Table 8, we compare the percentage of the difference between female students' abilities and male students' abilities

Ability	**Relative Difference % (Women - Men)/Men**	**Ability**	**Relative Difference % (Women - Men)/Men**
Visualization	-1.39%	Written Expression	0.80%
Depth Perception	-1.36%	Information Ordering	0.84%
Rate Control	-1.17%	Category Flexibility	0.86%
Manual Dexterity	-1.16%	Speed of Closure	0.87%
Arm-Hand Steadiness	-1.06%	Originality	0.90%
Visual Color Discrimination	-1.04%	Selective Attention	0.94%
Control Precision	-0.87%	Trunk Strength	0.97%
Wrist-Finger Speed	-0.80%	Speech Clarity	1.01%
Reaction Time	-0.73%	Speech Recognition	1.03%
Multilimb Coordination	-0.68%	Deductive Reasoning	1.04%
Finger Dexterity	-0.62%	Gross Body Coordination	1.10%
Perceptual Speed	-0.45%	Time Sharing	1.17%
Spatial Orientation	-0.37%	Problem Sensitivity	1.30%
Number Facility	-0.34%	Stamina	1.39%
Mathematical Reasoning	-0.22%	Static Strength	1.42%
Near Vision	-0.18%	Inductive Reasoning	1.51%
Far Vision	-0.10%	Memorization	1.64%
Glare Sensitivity	0.11%	Auditory Attention	1.67%
Written Comprehension	0.29%	Extent Flexibility	1.72%
Oral Comprehension	0.38%	Hearing Sensitivity	2.22%
Oral Expression	0.54%	Night Vision	2.68%
Response Orientation	0.54%	Sound Localization	3.46%
Fluency of Ideas	0.58%	Peripheral Vision	3.75%
Flexibility of Closure	0.63%	Explosive Strength	8.06%
Dynamic Strength	0.64%	Speed of Limb Movement	8.73%
Gross Body Equilibrium	0.78%	Dynamic Flexibility	24.97%

Table 8: **The inferred workplace abilities inferred from syllabi reveal key differences among women and men.** Inferred workplace abilities relative difference between Men and Women weighted by the number of degrees awarded according to IPEDS. Negative values suggest more male abilities, and positive values suggest more female abilities.

7 Taught Skill Distinctiveness and Labor Market Outcomes

We compare the distinctiveness of FOS's taught skills with earnings of 25- to 29-year-old bachelor's degree holders reported by the National Center for Education Statistics (NCES)[9] (see Table 9). We measure skill distinctiveness by noting various percentile skill RCA score within each FOS (see figures 8, 9, 10, 11, and 12); under this measure, FOS with higher diversity scores teach more distinctive skills compared to other FOS.

FOS	Salary	FOS	Salary
Accounting	60,000	Hebrew	N/A
Agriculture	45,380	History	45,090
Anthropology	45,230	Japanese	N/A
Architecture	52,170	Journalism	N/A
Astronomy	N/A	Law	N/A
Atmospheric Sciences	N/A	Library Science	N/A
Basic Computer Skills	N/A	Linguistics	45,370
Basic Skills	N/A	Marketing	52,160
Biology	50,450	Mathematics	54,560
Business	50,260	Mechanic / Repair Tech	N/A
Chemistry	77,705	Media / Communications	45,570
Chinese	N/A	Medicine	N/A
Classics	N/A	Military Science	N/A
Computer Science	70,140	Music	N/A
Cosmetology	N/A	Nursing	58,690
Criminal Justice	41,810	Nutrition	N/A
Culinary Arts	N/A	Philosophy	48,840
Dance	N/A	Physics	47,570
Dentistry	N/A	Political Science	50,640
Earth Sciences	N/A	Psychology	41,420
Economics	64,860	Public Safety	49,850
Education	41,510	Religion	N/A
Engineering	68,860	Sign Language	N/A
English Literature	44,640	Social Work	40,480
Film and Photography	N/A	Sociology	43,140
Fine Arts	40,500	Spanish	N/A
Fitness and Leisure	45,500	Theatre Arts	N/A
French	N/A	Theology	35,230
Geography	45,920	Transportation	59,920
German	N/A	Veterinary Medicine	N/A
Health Technician	50,270	Women's Studies	N/A

Table 9: Median annual earnings (salary) obtained from NCES.

[9] https://nces.ed.gov/programs/digest/d19/tables/dt19_505.10.asp (accessed on 12th February 2024)

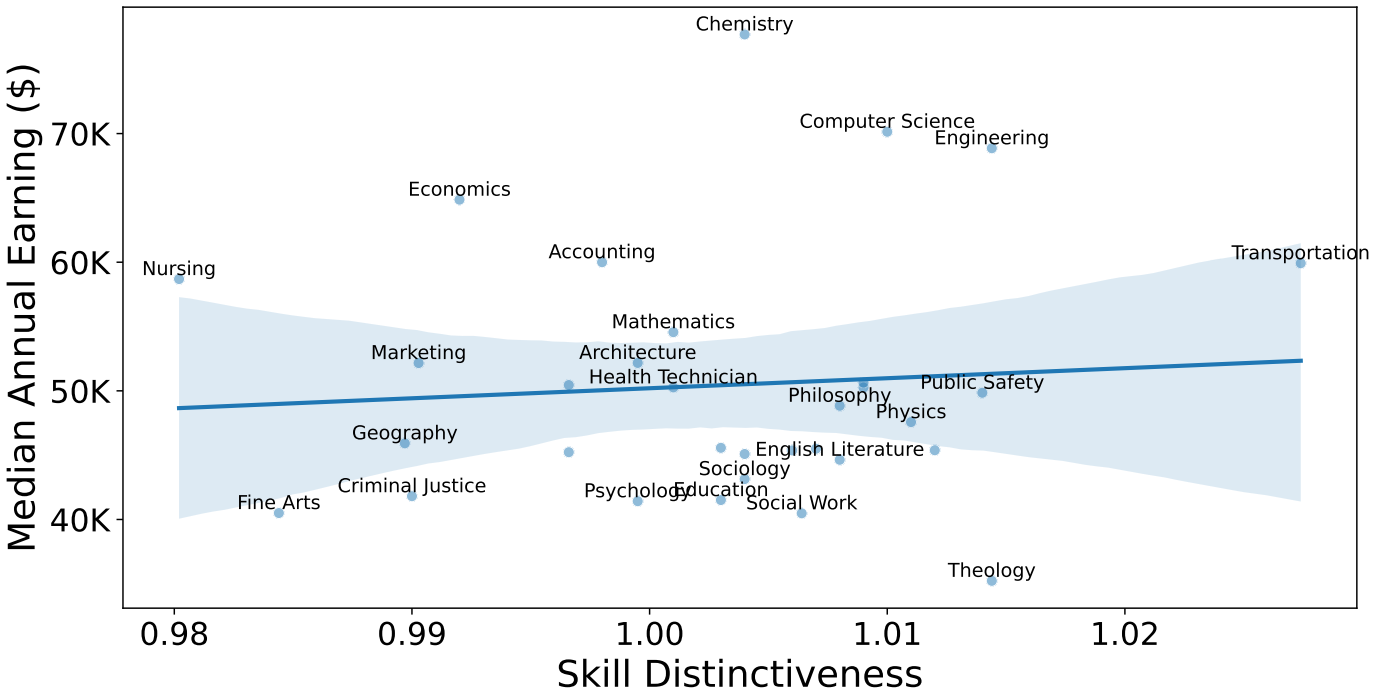


Figure 8: **Skill distinctiveness versus salary.** The x-axis shows the skill distinctiveness as RCA value at the 60^{th} percentile within each field of study and the y-axis shows the corresponding median annual earnings (salary) obtained from NCES. The line represents a regression line with $R^2 = 0.006$ (p-*value* $= 0.667$).

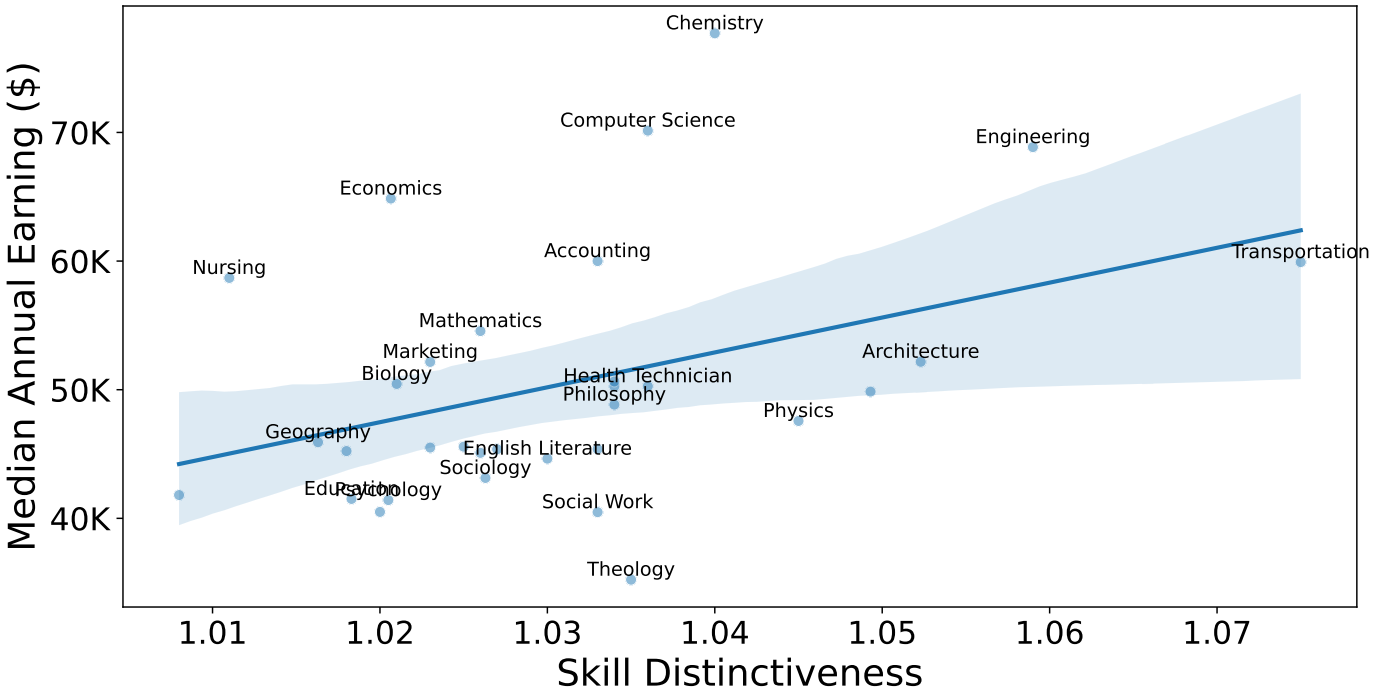


Figure 9: **Skill distinctiveness versus salary.** The x-axis shows the skill distinctiveness as RCA value at the 70^{th} percentile within each field of study and the y-axis shows the corresponding median annual earnings (salary) obtained from NCES. The line represents a regression line with $R^2 = 0.154$ (p-*value* $= 0.026$).

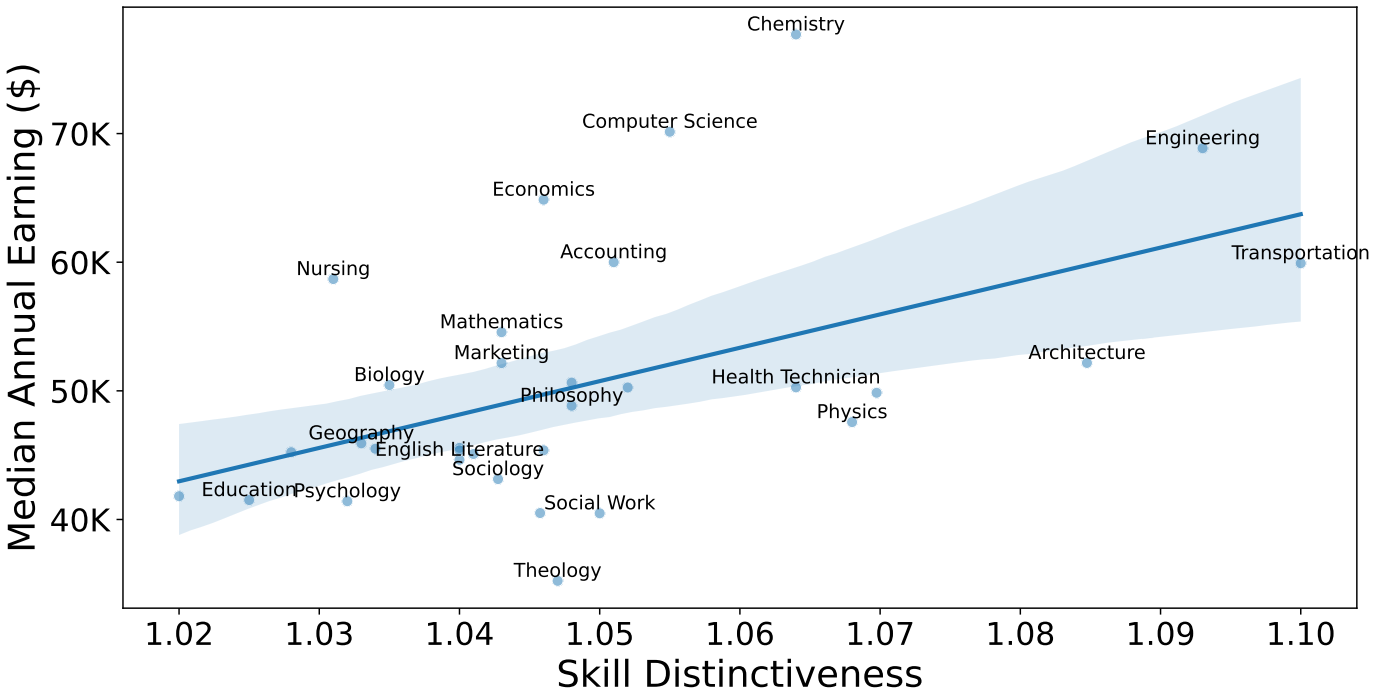


Figure 10: **Skill distinctiveness versus salary.** The x-axis shows the skill distinctiveness as RCA value at the 75^{th} percentile within each field of study and the y-axis shows the corresponding median annual earnings (salary) obtained from NCES. The line represents a regression line with $R^2 = 0.250$ (p-*value* $= 0.004$).

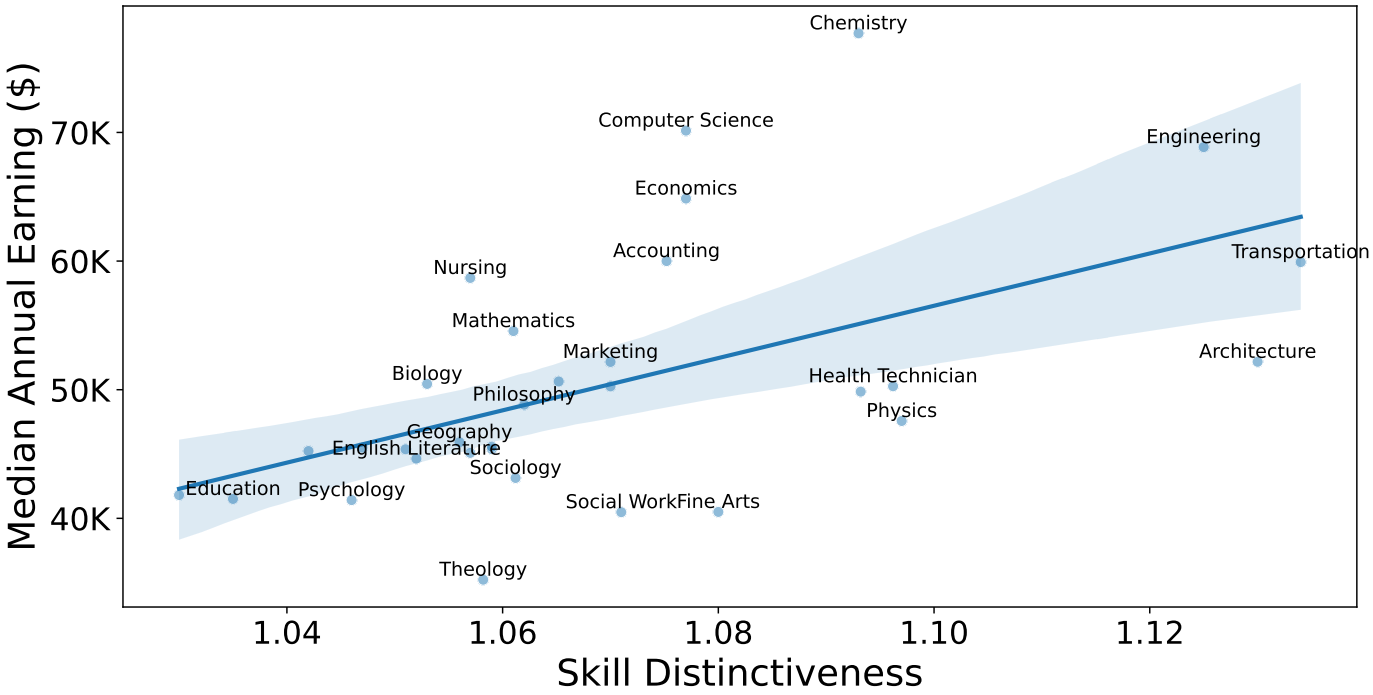


Figure 11: **Skill distinctiveness versus salary.** The x-axis shows the skill distinctiveness as RCA value at the 80^{th} percentile within each field of study and the y-axis shows the corresponding median annual earnings (salary) obtained from NCES. The line represents a regression line with $R^2 = 0.290$ (p-*value* $= 0.001$).

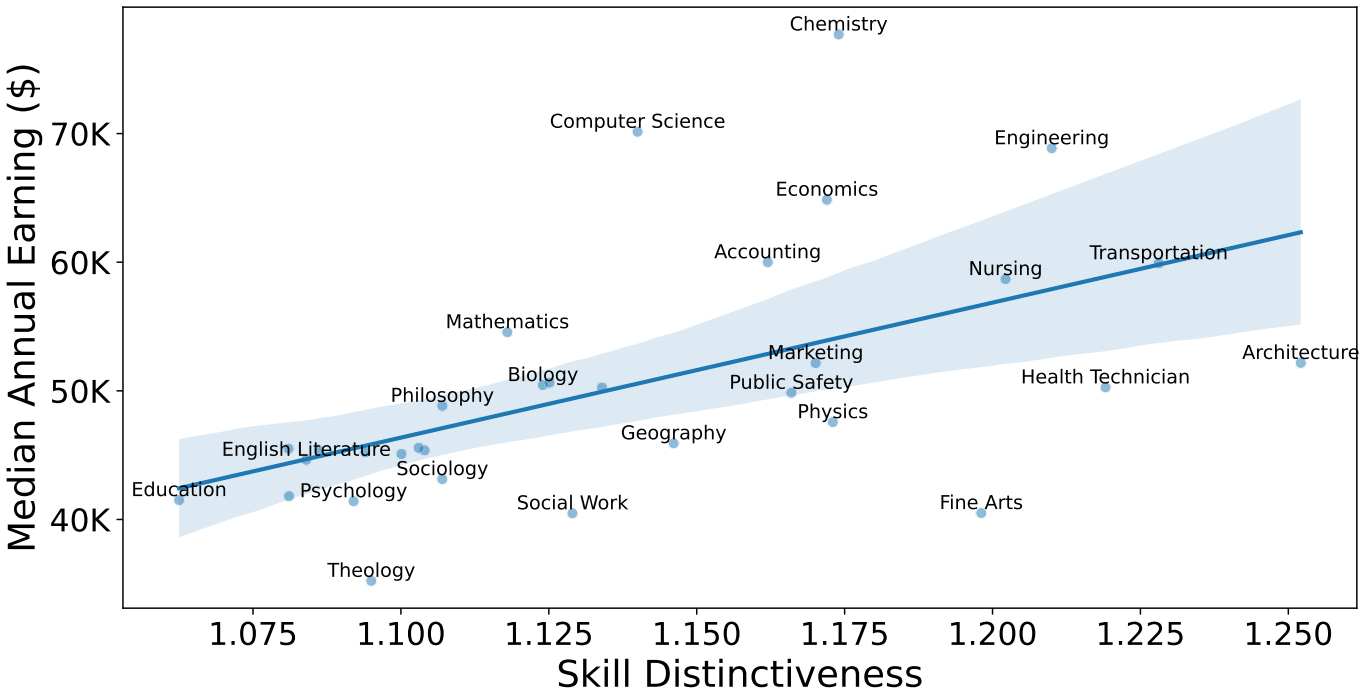


Figure 12: **Skill distinctiveness versus salary.** The x-axis shows the skill distinctiveness as RCA value at the 90^{th} percentile within each field of study and the y-axis shows the corresponding median annual earnings (salary) obtained from NCES. The line represents a regression line with $R^2 = 0.294$ (p-$value = 0.001$).

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[2] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[4] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*, 2020.