Identifying Learning Strategies in Princeton University Course Evaluations

Sophie Goldman
Adviser: Christiane Fellbaum

Abstract

Learning strategies are sets of skills and approaches intended to improve student engagement and success in academic coursework. Examples of these strategies include methods for learning effectively from course material, preparing for exams, and improving concentration. At Princeton University, the McGraw Center for Teaching and Learning promotes student engagement with learning strategies through workshops and individualized programming. However, there is no existing work quantifying the extent to which Princeton students engage with learning strategies. This paper identifies the prevalence of five categories of learning strategies mentioned by Princeton University undergraduate students in 137,128 undergraduate course evaluations from the Fall 2014 through the Spring 2021 semesters. A logistic regression model is used to determine statistically significant relationships between strategy mentions and course level, size, and discipline. Results can be used to better support the undergraduate academic experience at Princeton by supplementing the McGraw Center's learning strategies programming and other campus resources.

1 Introduction

"Learning strategies" have been discussed in education research settings since the 1960s and are a continuing topic of research [1]. Learning strategies emphasize the *metacognitive process* —"thinking about [one's] own thinking and learning processes." [1] Instruments such as the Learning and Study Strategies Inventory (LASSI) have been developed to measure student awareness and use of learning strategies, including categories such as "Information Processing," "Test Taking Strategies,", and "Time Management." [2] Positive correlations have been found between LASSI measurements and academic outcomes, including GPA, test scores, and persistence, showing the value of learning strategies in students' academic experience [3].

Princeton University's McGraw Center for Teaching and Learning draws on learning strategies research and promotes the use of learning strategies through small workshops and individualized programming. The Learning Strategy Consultation program is the main source of learning strategy programming at the McGraw Center [4]. Through the program, juniors and seniors host individualized consultations with students to collaboratively determine effective learning strategies for students to implement in their work. These strategies equip students to address the "expectations and demands" of Princeton's academic curriculum, which often differ from high school environments and require students to adapt to new ways of learning [5]. Additionally, Principedia, a student-run resource supported by the McGraw Center, publishes analyses of Princeton's courses and promotes student discussion of approaches to learning at Princeton [6]. Students can make use of Principedia to gain specific insight into the structure and demands of their courses. My work as a learning consultant and on Principedia, which both address student learning on an individual level, motivated me to quantify the extent to which the undergraduate student body at Princeton as a whole engages with different types of learning strategies. The results of this work could support the McGraw Center by providing more specific data about how undergraduate students express their approaches to academics, which in turn can benefit students who use the McGraw Center's resources.

While research on learning strategies often emphasizes specific cognitive skills used in the learning process, this paper will examine learning strategies more broadly, aiming to identify overall categories of strategies that are mentioned by students. The goal of this project is to identify the extent to which students discuss five learning strategies —exam preparation, problem solving, reading strategies, resource usage, and time management —by using written responses from Princeton course evaluations. Course evaluations are a measurement of students' experiences in courses, especially with regard to student learning. Evaluations inform professors and university administration of student perspectives on courses and the learning experience, and allow students to learn from their peers about courses firsthand. At Princeton, students submit ratings and written responses to questions about a course's lectures, readings, and other course components. At the end of an evaluation, students are asked "What advice would you give to another student considering taking this course?" [7] The responses to this question are published internally for undergraduate students to read. While some of the provided advice emphasizes course quality, meant to assist students who are in the process of selecting courses for an upcoming semester, other responses are more practically oriented and serve as an opportunity for students to reflect on their learning experience in a course. I consider these responses as evidence of student engagement with learning strategies at Princeton. For example, a student might write "I suggest forming a study group with people. If I had not done so I would have had a lot more trouble understanding the material,"[1], or "Write summaries of every reading, just a sentence or two. It will help you SO much when you have to go back and reread things for the papers."[2]

Using the database of undergraduate course evaluations from Fall 2014 to Spring 2021, this project will address the following research questions and corresponding hypotheses:

Research Question 1: To what extent do Princeton students discuss learning strategies in course evaluations?

Hypothesis 1: The five learning strategies of interest —exam preparation, problem solving, reading strategies, resource usage, and time management —will be mentioned at different rates in the course evaluations, as the strategies are unlikely to be equally useful in every course. The null hypothesis states that there is no difference in the proportion of evaluations mentioning each of the learning strategies.

Hypothesis 1.1: Time management will be the most commonly discussed learning strategy. This strategy is most broadly applicable across all types of courses and is a necessary skill for students to establish, even outside of academic coursework. Furthermore, Princeton's 12-week semester operates at a rapid pace, potentially making the need for time management more apparent to all

[1] Evaluation for CHM 201: General Chemistry I, Fall 2014
[2] Evaluation for HIS 280: Approaches to American History, Spring 2015

students than other strategies, such as resource usage, and therefore more likely to be mentioned in evaluations. Exam preparation, problem solving, and reading strategies will be mentioned less frequently in the course evaluations, as they make use of skills that are more specific to particular academic disciplines. The null hypothesis states that another strategy is mentioned more often than time management.

Research Question 2: How does engagement with learning strategies depend on the course in which those strategies are used? In particular, how do course level, size, and discipline affect students' engagement with learning strategies?

Hypothesis 2: There will be significant relationships between the course level, size, and discipline and all five learning strategies. The null hypothesis states that there is no relationship between any of these characteristics and the mention of the strategies.

Hypothesis 2.1: Introductory 100-level and 200-level courses will have higher rates of learning strategy mentions compared to upper-level courses. These courses tend to offer a wider range of course support than upper-level courses, as material in these courses often serve as foundations for more advanced coursework. Furthermore, when students enter Princeton, they may not be familiar with the academic strategies necessary for success in college coursework. Campus resources such as the McGraw Center serve all students, but place a particular emphasis on first year and sophomore students, who are more likely to be in introductory courses. This made lead these students to discuss learning strategies more often in evaluations. Finally, students in upper-level courses may not feel as compelled to discuss learning strategies explicitly in their evaluations. It may be assumed by these evaluators that future students will already be aware of the strategies needed, or these evaluators may be less conscious of strategies they have used in the course as these strategies have become more familiar over time. The null hypothesis states that introductory courses do not have higher rates of learning strategy mentions compared to upper level courses.

Hypothesis 2.2: Courses with more students will have higher rates of learning strategy mentions compared to smaller courses. Like introductory courses, larger courses tend to have more avenues for student support and a larger staff to engage students. This may encourage a greater use of learning strategies that benefit from course-specific support. Additionally, many larger courses are expected to serve as prerequisite courses for departments and are offered repeatedly over a period of years. This could encourage students to give more specific advice related to academic strategies, as a larger body of students could potentially benefit from this advice. The null hypothesis states that larger courses do not have higher rates of learning strategy mentions compared to smaller courses.

Hypothesis 2.3: Courses from STEM disciplines will have higher rates of learning strategy mentions compared to humanities and social science courses. These courses are often structured around different disciplinary approaches to teaching and learning. STEM courses tend to make use of weekly assignments and examinations to test student understanding, while humanities and social sciences courses tend to emphasize readings and assign papers or projects to assess students. Though there is overlap in course structure across discipline, generally STEM course structure allows for students to receive more incremental feedback on their performance throughout the semester through assignment grades. This might provide these students with more opportunities to actively consider and adjust their learning strategy usage throughout the semester prior to evaluating the course. Additionally, the McGraw Center tutoring program focuses almost entirely on supporting introductory STEM courses [8]. This connection with McGraw might encourage more STEM students to engage with strategies more explicitly, even after enrollment in introductory courses. The null hypothesis states that STEM courses do not have higher rates of learning strategy mentions compared to other disciplines.

2 Problem Background and Related Work

This section will begin by examining existing research on identifying student engagement with learning strategies. Next, it will review research that analyzes course evaluations and other student evaluative writing, including in Princeton specific contexts. Finally, it will discuss related work linking course evaluation data to learning strategies.

2.1 Prior Research on Learning Strategy Usage

Research on understanding student engagement with learning strategies often relies on self-report questionnaires, though some research has taken a more in-depth approach through studying focus groups of recruited students [9]. Prior research has identified significant relationships between the use of learning strategies and the value of the learning experience for students [10]. Other work has linked the use of learning strategies to types of course assignments and assessments used in online courses [11]. Simesek and Balaban identified differences in learning strategy usage by college seniors from five disciplines using a self-report questionnaire, finding positive correlations between learning strategy usage and student success [12]. As a whole, learning strategies research is qualitatively focused, characterized by a small sample size (limited to a single course, semester, or pre-identified group of students) and consists of explicit investigation of student engagement with learning strategies. Additionally, this research tends to focus on more specific categories of learning strategies, such as rehearsal (methods of memorizing and encoding information), elaboration (methods of solidifying and deepening understanding of information), and organization (methods of categorizing and sorting information) [11, 12]. This level of detailed understanding results from the use of self-report questionnaires. Unlike these sources, this paper will make use of course evaluations, which can serve as an implicit measurement of learning strategy engagement, and will take a quantitative approach driven by natural language processing tools to allow for a broader overview of learning strategy engagement by a university's full undergraduate community.

2.2 Natural Language Processing Using Course Evaluations and Other Student Writing

Course evaluations are used in many academic contexts to understand student experiences and assess teaching. Natural language processing approaches are valuable for analyzing text responses from these evaluations in order to quantify student responses and better understand patterns and themes. When using course evaluation corpora, this can include linking written responses to quantitative evaluation ratings as well as identifying evaluation sentiment. Sliusarenko et al. used text mining and factor analysis to determine the most common topics in a single engineering course's positive and negative text evaluations over multiple semesters. They used a logistic regression to identify the relationship between positive and negative topics and the numerical ratings students provided, finding significant relationships between a number of topics and ratings [13]. Additional work focused on classifying evaluation sentiment has been done using a much larger corpus of 66,000 evaluations from massive open online courses to test a variety of machine learning models, with a reported 95.80% accuracy rate when using deep learning architecture combined with word embeddings [14]. These papers demonstrate the continuously developing potential of NLP for the analysis of course evaluations.

Additional papers that have taken an NLP based approach to student metacognitive writing can also provide important background that more closely relates to the goals of this paper. Kovanović et al. analyzed student academic reflections on personal academic work from four courses, with the goal of categorizing the reflections into three levels (making observations, identifying motive/reason, or setting goals for future). This approach classified reflections using both the most frequent n-grams and features from the Linguistic Inquiry and Word Count (LIWC) and Coh-Metrix tools, which both identify psychological and cognitive elements in text data [15]. LIWC implements a dictionary based approach to identify the percentages of words in a text that can be categorized under various psychological themes [16].

Finally, Princeton University senior theses have previously used Princeton course evaluation responses as a source of data. However, no projects have analyzed these course evaluations through the lens of learning strategies. Certain theses have emphasized the emotional components of course evaluations through sentiment and affect analysis [17, 18]. Others have used evaluations to calculate course difficulty [19] or to identify factors that affect the content of evaluations, such as the professor's gender or the course discipline [20]. The scope of these papers has ranged from a single department's evaluation data from 2008-2017 [18] to evaluation data from all courses from a single semester [17]. These papers obtained their data through scraping the Registrar's course evaluation site. This paper will use a broader range of evaluation data from the Princeton Courses website, which stores evaluation data that is no longer accessible through the Registrar's site, as well as the most recent evaluation data through Spring 2021. Furthermore, this paper will consider all undergraduate courses in the selected time period, rather than examining only one department or a subset of courses.

2.3 Using Course Evaluations to Understand Learning Strategies

Course evaluations have been previously cited as a potential source of knowledge on learning strategies [21]. However, minimal research has been found to use an evaluation-based approach to understanding learning strategies. Prosser and Trigwell linked numerical ratings in student course evaluations to students' approaches to learning, as measured through a questionnaire directly aimed at identifying approaches to learning, but did so with a sample of only eleven courses in both of their studies [22, 23]. Entwistle and Tait undertook a similar analysis to identify the relationship between evaluations and approaches to learning using a questionnaire distributed directly to students [24]. In these research settings, student evaluations of courses were collected at a small scale for the purposes of studying the relationship to learning approaches. In contrast, this paper will examine existing student responses, which were not collected with the explicit goal of measuring learning strategy usage. By using a much larger sample of evaluations, spanning the full university curriculum over many semesters, this paper will support more general conclusions about the role of learning strategies in the Princeton undergraduate academic experience.

3 Approach

This paper will examine categories of learning strategies that are most clearly recognizable in the course evaluations. These categories are more commonly discussed in the evaluation responses as compared to particular theoretical learning strategies, such as those described in Section 2.1. This section will describe the overall approach used to identify student engagement with five types of learning strategies: exam preparation, problem solving, reading strategies, resource usage, and time management.

3.1 Selecting Categories of Learning Strategies

The McGraw Center lists the following when describing potential topics that could be discussed in a Learning Strategies Consultation [4]:

- Time management and planning
- Motivation and overcoming procrastination
- Managing large projects (i.e. JPs and Theses)
- Engaged, efficient reading and learning from text
- Tackling P-sets and problem-solving strategies
- Effective note-taking
- Exam preparation and studying
- Mastering large amounts of information
- Learning and succeeding in specific courses
- Making the most of office hours and other resources

This list was used to select the five learning strategy categories that are identified in the course evaluations. The chosen strategies were selected due to their assumed applicability to a variety of courses. Additionally, after inspecting the course evaluations manually, these strategies appeared to be the most easily identifiable in the course evaluation responses. The chosen categories are defined below, along with examples of ways students might implement each strategy.

- Exam Preparation: approaches to studying in preparation for exams

  Examples: practicing problems, making use of lecture learning goals, memorizing course content

- Problem Solving: approaches to completing problem sets and other assignments based in student-led problem solving

  Examples: learning from course examples, applying conceptual material from lecture

- Reading Strategies: approaches to making use of written course material, particularly textbooks, papers, and other assigned readings

  Examples: summarizing readings, annotating passages, reading in advance of lecture

- Resource Usage: approaches to making use of interpersonal resources, such as professors, preceptors, peers, and tutors

  Examples: attending office hours, collaborating with peers, attending McGraw Center tutoring

- Time Management: approaches to planning, scheduling, and balancing course workload

  Examples: starting assignments early, prioritizing certain courses over others, keeping up with course material on a weekly basis

The selected learning strategies do not represent entirely distinct approaches to learning. Figure 1 shows one representation of the overlap between the five strategies. All strategies share some attributes in common with at least another strategy. For example, a student using problem solving strategies may incorporate resource usage and time management strategies in order to receive help on assignments and to complete assignments on time. Exam preparation strategies may make use of problem sets as an important study tool. Similarly, reading strategies may be necessary as an element of effective exam preparation. There are certainly many other ways in which these strategies can be connected, and in practice students are likely to use multiple strategies simultaneously in one course. Additionally, two of the strategies, resource usage and time management, tend to be more broadly applicable to a variety of courses. Exam preparation is relevant in courses across many disciplines, but tends to be emphasized more often in quantitative courses. Problem solving strategies are most relevant in STEM contexts, while reading strategies apply most often in humanities and social science courses. It is important to note these commonalities and differences between the selected learning strategies, as they provide important context for the design of the learning strategy dictionaries described in the next section.

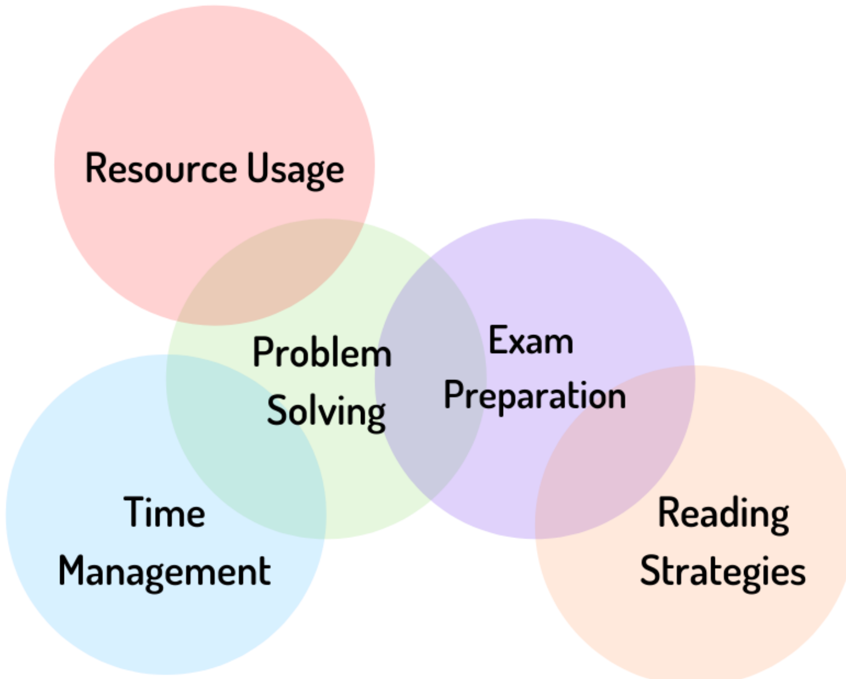

Figure 1: Overlapping Relationships Between Learning Strategies.

3.2 Identifying Learning Strategy Mentions

The course evaluation responses provide evidence for the extent to which students mention each of the five selected learning strategy categories. Since students may use learning strategies in courses but not describe them in course evaluations, the responses do not capture all instances where learning strategies were used. However, it is assumed that many instances of learning strategy usage can be represented through learning strategy mentions. This paper uses a customized dictionary based approach similar to that used by the LIWC tool, and undertakes a similar classification task as Kovanović et al., using evaluation data instead of student academic reflections to identify student metacognitive processes in an academic setting. The dictionaries of tokens relevant to each strategy were used to calculate the number of evaluations that mention tokens associated with a particular strategy's dictionary. The learning strategy dictionaries were compiled through the identification of common terminology used by the McGraw Center to discuss the selected learning strategies in its online resources, workshops, and learning consultation materials. Additionally, after manually identifying course evaluations that were indicative of learning strategies, new tokens that appeared to be relevant to those strategies were added to the learning strategies dictionaries. Each evaluation was labeled with a binary value for each strategy representing if a token in the evaluation was found in the relevant strategy's dictionary.

3.3 Relationship between Strategy Mentions and Course Characteristics

In order to identify the relationship between strategy mentions and course characteristics, a logistic regression was used to predict whether an evaluation would be labeled with a 1 or a 0 for each learning strategy. The predictors included three categorical variables representing the course level, size and discipline. For course level, each course was labeled based on the first digit of its course code number. These values range from 1 through 4, corresponding to the four levels of undergraduate courses. Courses were divided into four categories by size: 1-20 students, 21-50 students, 51-100 students, and 101+ students. This division represents seminars, and small, medium, and large lectures, respectively. The course disciplines were divided into Engineering, Humanities, Natural Sciences and Social Sciences. The coefficients of the logistic regression are used to identify the relationship between the course characteristics and the learning strategy mentions.

4 Implementation

The flowchart in Figure 2 depicts the implementation process. Each step of the implementation process is described in the following sections.

4.1 Data Collection

Data was obtained from the Princeton Courses website, a platform run by the Undergraduate Student Government that maintains a record of course evaluation data from the University Office of the Registrar for the Fall 2014 through Spring 2021 semesters [25].

The initial dataset included 148,958 evaluations. Each evaluation is associated with a course ID, which identifies the evaluation's course and the semester in which that course was offered.

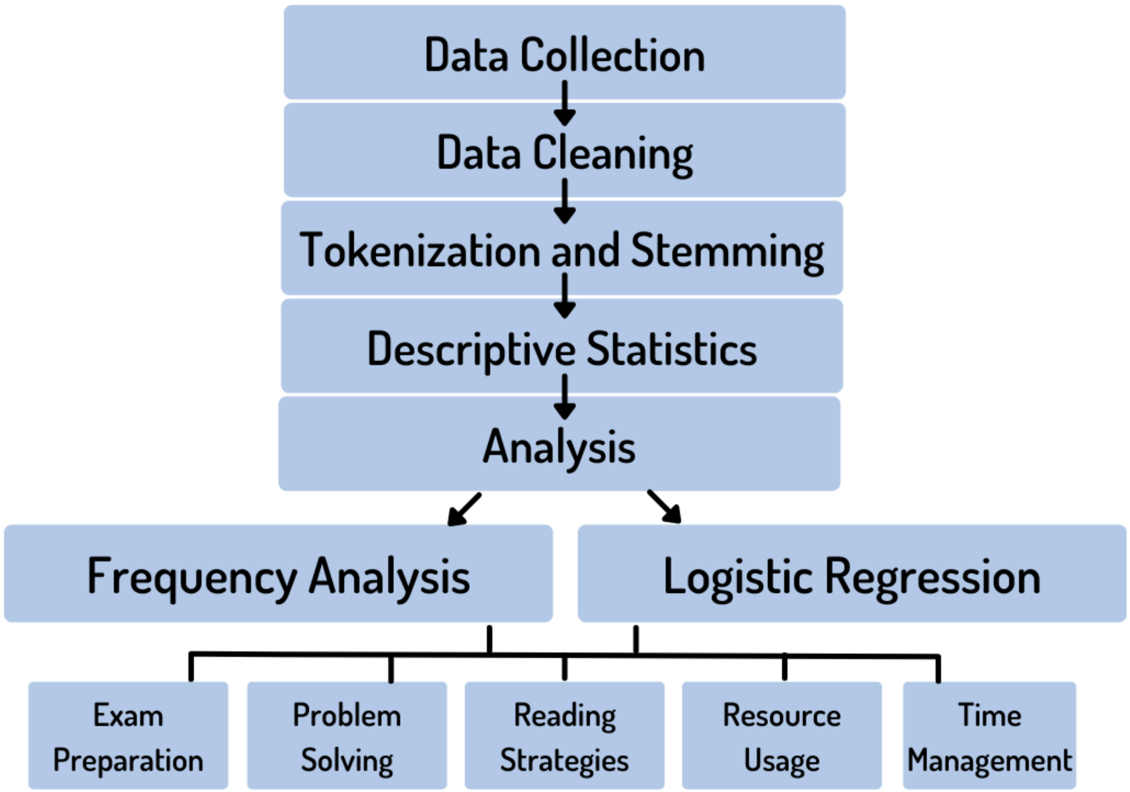

Figure 2: Implementation Flowchart.

This data was merged with Princeton Courses data containing course specific details, including the department and catalog number, section-level enrollment, and whether the course was labeled as an undergraduate or graduate course. Other provided data, such as the list of cross-listed course codes associated with the course, were not used.

The provided data on a course's department and catalog number can be combined to determine the course code. Course codes are used to identify a course's primary department or program and the course's level, using a three letter abbreviation for the department and a three digit number ranging from 100 to 499 for undergraduate courses. Some course codes correspond to departments that offer concentrations, while other correspond to certificate programs or more specific topics, such as foreign languages. While not all course codes directly refer to a department or program, this paper will use the terms department and program interchangeably to refer to the course code abbreviations.

In order to determine the course enrollment, individual section enrollments were added together.

This required parsing of the section-specific data for each course, such that only enrollments from primary sections of a course (lectures, seminars, studios, etc.) were counted, rather than including precepts and secondary sections that contain the same students.

Additionally, departments were manually labeled with their primary discipline. The disciplines were identified using the University's "Areas of Study" web page, which lists the following disciplines: "Engineering," "Humanities", "Natural Sciences," and "Social Sciences".[3] The pages for each discipline list associated departments and programs. Some departments are listed under multiple categories due to their interdisciplinary nature. For the purposes of this project, one discipline was chosen for each course code, using the University's categorization as a guide. A list of the course codes, corresponding department/program names, and labeled disciplines can be found in Table A1 in the Appendix.

4.2 Data Cleaning

As the focus of this paper is on undergraduate students, only evaluations for undergraduate courses were included. All evaluations for courses labeled as graduate courses were removed, leaving 139,558 evaluations. Additionally, the provided data included evaluations from Summer 2021. These evaluations were removed since it is not common for Princeton to offer summer courses on campus and these evaluations do not necessarily reflect the typical learning experiences of the undergraduate student body during the academic year. There were also a small number of courses with irregular course codes. Any courses with course code numbers outside of the range 100-499 were removed. After this stage, 139,140 evaluations remained.

Initially, evaluations were provided for 95 departments and programs. The 20 smallest programs (measured by the number of evaluations across all semesters) were dropped, as each had around 100 evaluations or fewer. These course codes represented foreign language courses with very low enrollments and small certificate programs. Since certificate programs are more often cross-listed and this paper only considers the primary course code for each course, eliminating these evaluations

[3] https://www.princeton.edu/academics/areas-of-study

does not eliminate courses associated with those certificates entirely. In total, 138,080 evaluations for the top 75 departments and programs were left.

The provided dataset also included some duplicate evaluations across different offerings of a single course. This is a result of Princeton Courses' design, as in the event that evaluations are not available for a given semester's offering of a course, evaluations from a different semester will be shown to students. Therefore, it was necessary to remove these evaluations. Duplicate evaluations were identified through two approaches. First, any evaluations associated with a course that had an enrollment of 0 (meaning that the course was cancelled in that semester) were removed. Secondly, evaluations associated with a course with over a 100% response rate were removed, as having more evaluations than enrolled students implied that the evaluations were related to a different semester. This assumption was confirmed through manually inspecting the relevant courses and comparing the published evaluations between semesters.

This data cleaning process resulted in a final dataset of 137,128 evaluations.

4.3 Tokenization and Stemming

The Python Natural Language Toolkit [26] was used to tokenize and stem all evaluation text. A modified version of the default English stopwords was used, as certain stopwords were determined to be relevant to potential learning strategies. The tokens that were not included in the stopwords list were 'up,' 'before,' 'during,' and 'after' as these tokens were thought to be potentially relevant to the time management strategy. A list of all stopwords is included at the end of the Appendix. All punctuation was removed from the course evaluations and tokens were converted to lowercase.

By removing extraneous characters and stopwords, it was possible to identify the most frequent meaningful words in the dataset. This also made the process of labeling evaluations for mentions of each learning strategy more efficient by reducing the size of the dataset.

A Porter stemmer was used to stem all tokens. Stemming reduces words with the same root to a single form. For example, 'collaborate' and 'collaboration' were both reduced to 'collabor,' and plural nouns were reduced to their singular form. By applying the same stemmer to the learning strategy dictionaries, it was possible to identify all instances of a particular stem in the evaluations without needing to specify all forms of the token in the dictionaries.

4.4 Descriptive Statistics and Exploratory Data Analysis

Distribution of Evaluations

While the evaluation dataset contains 137,128 evaluations representing 2,708 courses, most evaluations are associated with a small percentage of all courses. The top 30 courses, measured by total number of evaluations (28,605 evaluations total), make up 20.9% of all evaluations. As seen in Figure 3, the courses with the most evaluations primarily come from introductory 100 and 200 level courses in the COS, ECO, MAT, PHY, CHM, and MOL departments, along with courses from LIN, POL, MAE, PSY, EEB, and SPA. In comparison, the bottom 2000 courses (26,825 evaluations total) make up 19.6% of all evaluations. This is due to two main factors: some courses are only offered in one semester, which limits the number of total evaluations possible, and others are naturally limited in size by their format, as in the case of a seminar or studio course. As such, it is expected that most of the evaluations would be dominated by a fairly small set of courses. The evaluations are also distributed unevenly across departments, with the top 3 departments (COS, ECO, and MAT) making up 19.6% of all evaluations. The full distribution of evaluations across departments can be seen in Table A1. This distribution is likely a result of the increased enrollment in these concentrations and the relevance of courses in these departments to a variety of concentrations.

In order to understand the proportion of a course's enrollment represented in the evaluations, the number of evaluations for a course in a given semester was divided by the course's enrollment. On average, 52% of students in a course submitted an evaluation, and the median response rate was 51%. Figure 4 shows that the response rate is distributed fairly symmetrically around the average, and most courses have a response rate above 30%. This indicates that evaluations tend to accurately measure the experiences of students in a course, though evaluations courses with lower response rates may be less reflective of student experiences due to potential non-response bias.

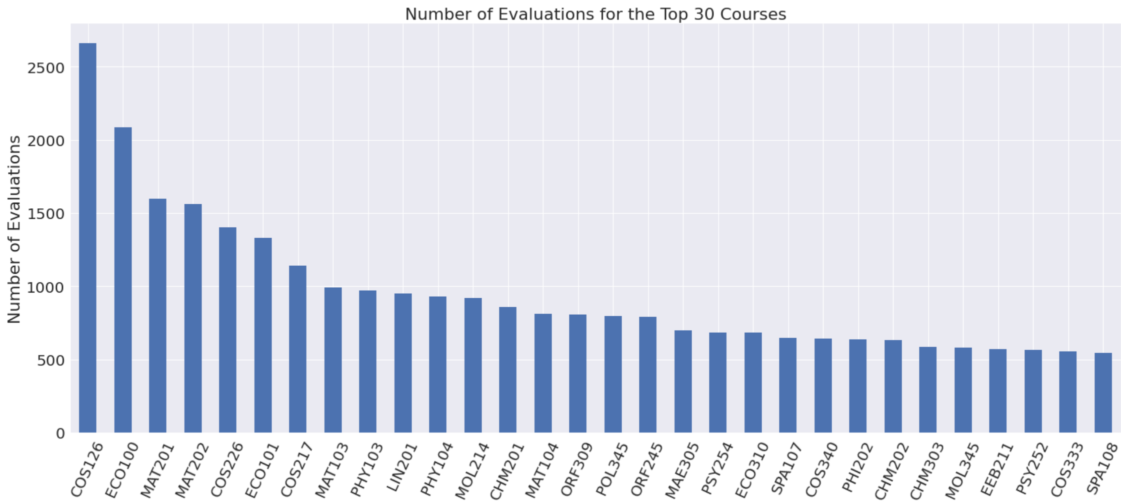

Figure 3: The 30 courses with the highest number of evaluations.

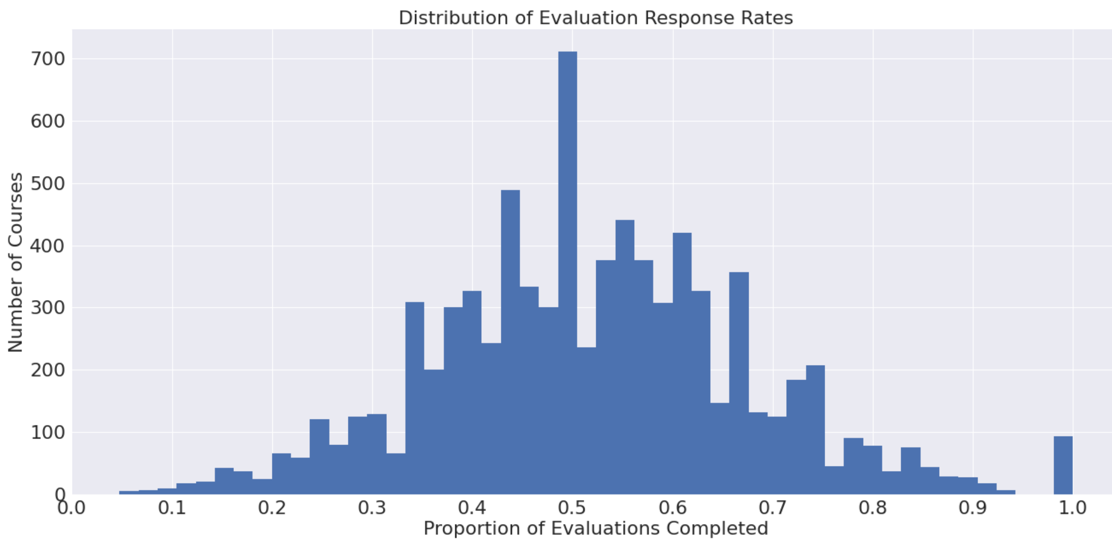

Figure 4: The distribution of response rates for course evaluations.

Descriptive Statistics of Evaluations and Comments

Statistic	Value
Total number of tokens	2,504,754
Unique tokens	30,342
Average number of tokens in evaluation	18
Average evaluations per department across all semesters (n = 75)	1,828
Median evaluations per department across all semesters (n = 75)	1,134
Average evaluations per course across all semesters (n = 2,708)	51

Median evaluations per course across all semesters (n = 2,708)	16

Table 1: Descriptive Statistics of Course Evaluation Data

Table 1 contains an overview of the evaluation comments. All statistics refer to data after stopwords are removed. The average evaluation is short (about a sentence in length after removing stopwords).

Exploratory Data Analysis

It is possible to identify the most frequent n-grams in the cleaned evaluations in order to obtain an initial indication of which tokens relevant to learning strategies are among those most frequently used. Figure 5 shows the top 60 trigrams after stopwords were removed. The most common trigram, 'go office hours' directly reflects concepts relevant to a resource usage learning strategy. Though most of the other trigrams discuss the quality of the course and student opinions about taking various courses, other relevant trigrams include 'keep up readings,' 'start assignments early,' and 'pay attention lecture.' These trigrams demonstrate the most common phrases used to reference learning strategies, though many more specific phrases are used in less frequent cases.

Alongside analyzing the tokens in the dataset by identifying trigram frequency, topic modeling can be used to map related tokens to a set of unlabeled topics. Using the scikit-learn package, a Latent Dirchelet Allocation topic model was produced on the evaluations, with 10 topics [27]. This provided a preliminary sense of any inherent categories of evaluations to confirm the project's goal of classifying evaluations by learning strategy. The list of the top 20 words in each of the 10 topics can be found in Figure A2. While some topics refer to broader course characteristics, such as the quality of the course (topic 6, topic 8) or subject matter (topic 1, topic 3), others refer more specifically to tokens that seem relevant to learning strategies. For example, topic 5 appears to be related to resource usage, with tokens such as 'office,' 'hour,' 'help,' 'ask,' 'professor,' and 'preceptor' included in this topic. Topic 4 appears to be related to exam preparation, mentioning tokens such as 'exam,' 'midterm,' 'final,' 'studi,' and 'practic.' The topic modeling results provide an

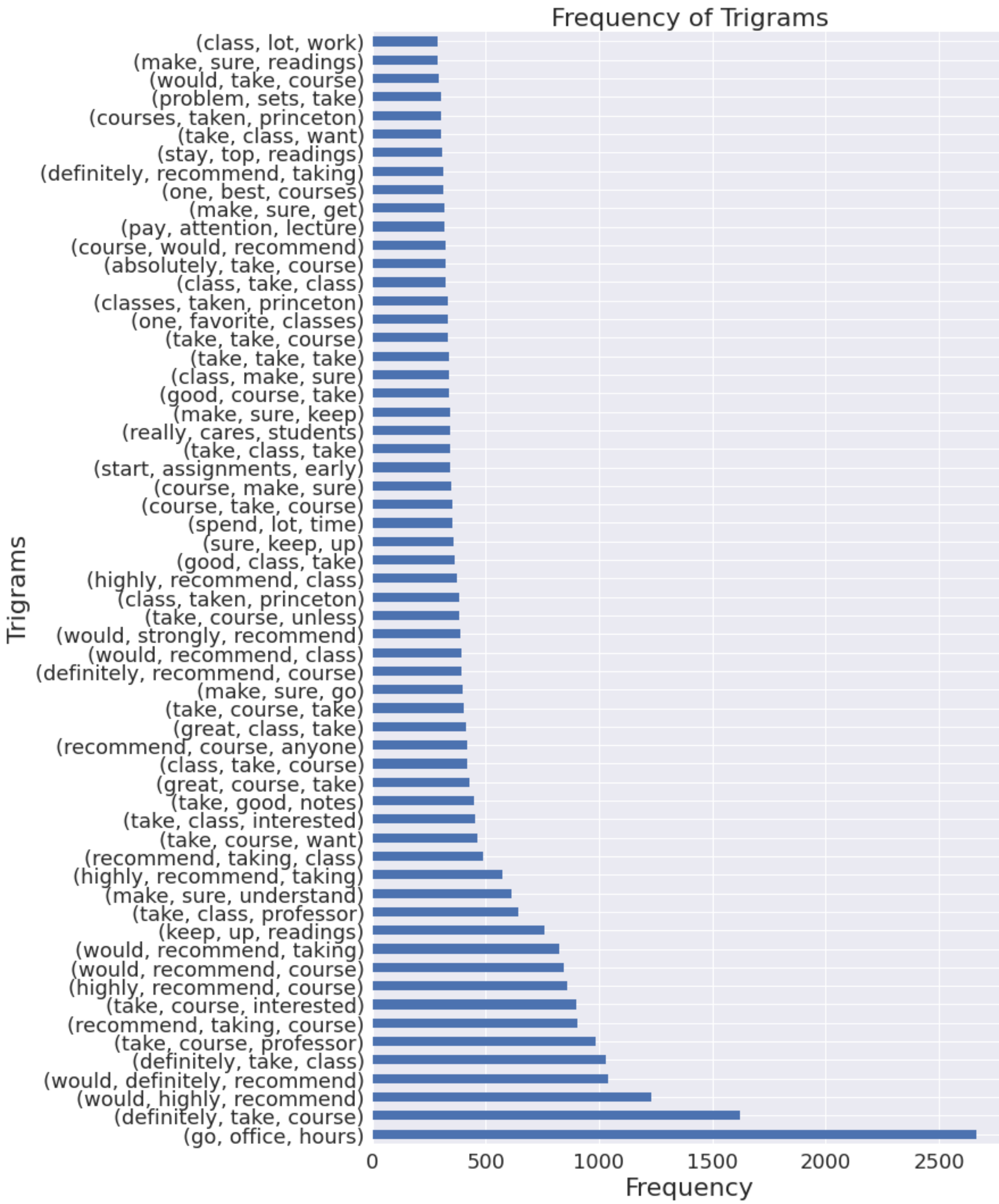

Figure 5: The top 60 trigrams in the course evaluation responses.

indication that there are coherent themes across the evaluations, including those related to learning strategies. This provides additional support for the selected approach of labeling evaluations using the learning strategies dictionaries.

4.5 Learning Strategy Dictionaries

The learning strategy dictionaries were designed to reflect terms that were most likely to be used in the context of their respective learning strategy. As described in Section 3.2, existing language used by the McGraw Center to discuss learning strategies, as well as language observed in the evaluations, were both drawn from when compiling the dictionaries. Additionally, labeled evaluations were examined to identify instances of false positives where tokens may be capturing evaluations unrelated to the labeled learning strategy, leading to an iterative editing process of the dictionaries.

Exam Preparation	Problem Solving	Reading Strategies	Resource Usage	Time Management
'assessment'	'calculate'	'annotate'	'classmate'	'balance'
'exam'	'homework'	'argument'	'collaborate'	'calendar'
'final'	'mcgraw'	'article'	'friend'	'cram'
'flashcards'	'problem'	'author'	'instructor'	'deadline'
'memorize'	'problem sets'	'book'	'lecturer'	'early'
'midterm'	'pset'	'chapter'	'mcgraw'	'hour'
'notes'	'solution'	'criticize'	'office hours'	'fast'
'outline'	'solve'	'essay'	'partner'	'keep up'
'quiz'		'highlight'	'preceptor'	'late'
'practice'		'papers'	'study buddy'	'manage'
'review'		'read'	'ta'	'plan'
'study'		'summarize'	'tas'	'planner'
'synthesize'		'text'	'tutor'	'prioritize'
'summarize'		'textbook'		'procrastinate'
'test'				'routine'
				'schedule'
				'time'
				'weekly'

Table 2: Learning Strategy Dictionaries

Designing the dictionaries included making subjective trade-offs between the breadth of the dictionaries, in order to account for as many potentially relevant tokens as possible, and their specificity, in order to produce more accurate results. An overview of some challenges and subsequent decisions made when building these dictionaries is provided in the following subsections.

Intentionally Excluded Tokens

Some tokens that are relevant to a particular learning strategy were at times excluded intentionally due to overlapping meaning. For example, 'assignments' was not included in the problem solving dictionary since it is often used to refer to written assignments, which do not make use of the same approaches. Instead, tokens such as ‘pset’ and ‘homework’ were used. These tokens refer more specifically to assignments with an emphasis on problem solving. Additionally, the token ‘organized’ was excluded, as while it might seem that this token would refer to student organization in connection to time management, inspection of the evaluations using the stem ‘organ’ revealed many evaluations reviewing the organization of the course itself, rather than students’ use of organization to succeed in the course. Finally, some tokens, such as ‘apply’ and ‘concept’ would seem to be very relevant to learning strategies, and are used often in more formal discussions of learning strategies. However, these terms are used more broadly in course evaluations, and including them would over-count evaluations that are not focused on the use of learning strategies.

Unintentionally Excluded Tokens

It is straightforward to determine when it is unnecessary to include a token in the dictionary, as it is possible to review the evaluations containing that particular token. However, it is more difficult to determine when a token is missing that should be included in the dictionary. It is possible that relevant tokens were excluded unintentionally from the dictionaries. This is somewhat addressed by the fact that certain tokens are more easily identifiable as being relevant to a particular learning strategy. These tokens tend to co-occur with other, more niche tokens. Ideally, the dictionaries contain most of the commonly used tokens so that the majority of relevant evaluations are captured, even if some evaluations are unintentionally excluded.

Tokens with Overlapping Meaning

Many of the included tokens hold multiple meanings. The token ‘hour’ can refer to a measurement of time, but is also used in the phrase ‘office hours.’ For the time management strategy, it was necessary to only count evaluations that contained the token ‘hour’ but not the token ‘office’. Though this might remove some evaluations that speak about both the measurement of time and the academic resource, not filtering out these evaluations would inflate the measurement of time management strategy mentions by counting all mentions of ‘office hours.’ Similarly, the token ‘review,’ used in the exam preparation strategy, could also be used to refer to course reviews. In this case, inspection of the evaluations indicated that 'review' was often used in discussions of exam preparation, so while some evaluations using this token may be falsely labeled, in most cases it can be assumed that this token is being used in an exam-relevant context.

Strategies with Overlapping Tokens

Some strategies might be referred to using the same token. For example, an important part of exam preparation or problem solving can be attending office hours. However, since 'office hours' is more directly related to the resource usage strategy, the token was only included in the resource usage dictionary. Since an evaluation only needs to contain one word from a dictionary in order to be labeled as mentioning that strategy, it is assumed that a more directly related word will be mentioned as well alongside a token that may overlap. This would cause an evaluation to be correctly labeled with multiple strategies. In other cases, tokens were included in multiple strategies. For example, 'summarize' is included in both the exam preparation and reading strategies dictionaries. Since summarizing is a skill applied in distinct ways to reading materials and course notes, this token is included in both categories to reflect the use of two separate strategies.

4.6 Frequency Analysis

After tokenizing and stemming each evaluation, all tokens in each evaluation were compared to the stemmed terms in each learning strategy dictionary until a match was found between an evaluation token and a dictionary token. This allowed all evaluations to be labeled with five binary values, one representing each strategy. Any evaluation containing at least one of the words in a particular learning strategy dictionary was labeled with a 1 for that learning strategy. If the evaluation contained none of the words in a particular dictionary, the evaluation was labeled with a 0. The proportions of all evaluations mentioning each strategy were calculated by finding the proportion of all evaluations labeled with a 1. The evaluations were then grouped by each course characteristic (level, size, and discipline) to similarly identify the proportions of learning strategy mentions within each category. For all statistical analysis in this paper used to determine if differences between proportions were significant, the Python package statsmodels was used [28].

4.7 Regression Analysis

The statsmodels package was also used to perform a logistic regression using a generalized linear model. The model used a logistic link function and a binomial distribution. The input for this model was the set of three categorical variables representing the course level, size, and discipline. The general form for the logistic regression is as follows, where X represents the matrix of independent variables, β is the vector of coefficients, and Y is the binary label for a given learning strategy. This equation gives the probability that an evaluation with course characteristics with be assigned the label 1.

$$\mathrm{P}(Y = 1|X) = \frac{1}{1 + \exp(-\beta X)} \tag{1}$$

The logistic regression output yields a set of coefficients corresponding to each of the categorical variables. Raising the exponential function to the power of each coefficient yields the odds ratio. The odds ratio describes how much more likely it is that the output will occur for a given variable compared to the baseline level in that category [29]. An odds ratio greater than 1 indicates a higher likelihood than the baseline, while an odds ratio less than 1 indicates a lower likelihood than the baseline. An odds ratio of 1 indicates no change in likelihood compared to the baseline.

5 Results and Evaluation

5.1 Overall Strategy Frequency

The first results address the research question regarding the overall frequency of each learning strategy across all evaluations.

Certain learning strategies are mentioned more frequently than others across the full evaluation dataset. Resource usage and time management are most prevalent, potentially due to their wider applicability across a wide range of course types. As shown in Figure 6, 27.79% of evaluations mention at least one of the tokens in the resource usage dictionary. In comparison, time management is mentioned in 24.34% of evaluations. Reading strategies, exam preparation, and problem solving

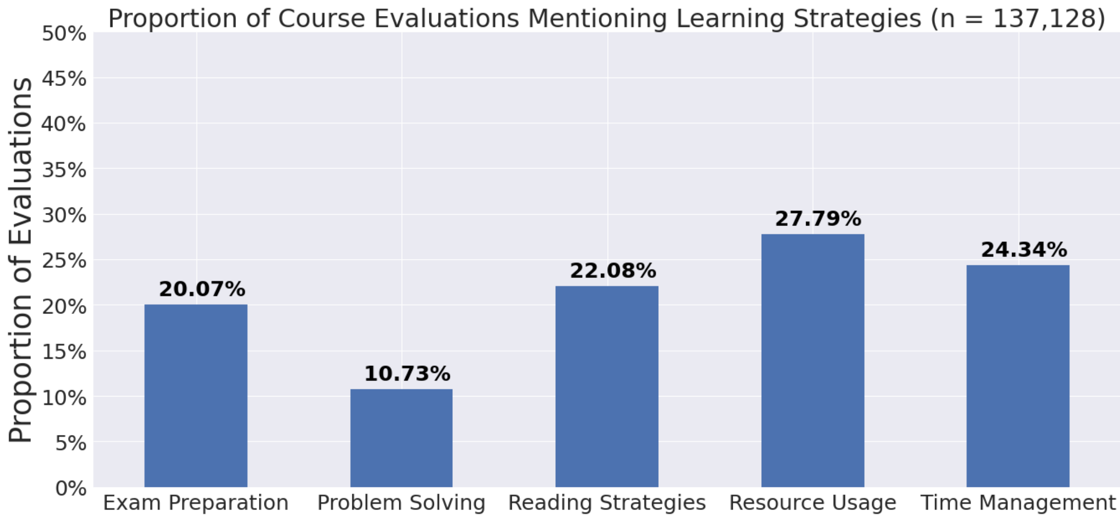

Figure 6: The proportion of evaluations mentioning each of the five selected learning strategies.

are mentioned in 22.08%, 20.07%, and 10.73% of the evaluations. In addition, it was determined that 57.48% of all evaluations were labeled with at least one learning strategy.

Since the proportions in Figure 6 are calculated from the same population and are not independent, they cannot be compared using a chi-square test. Instead, McNemar's test [30] can be used to test statistical differences between the proportions. Table 3 contains the p-values for all pairs of strategies. Any p-value smaller than the significance level of $p < .05$ is bolded in this table and in all following tables.

Learning Strategy	Problem Solving	Reading Strategies	Resource Usage	Time Management
Exam Preparation	**< .0001**[4]	**4.84 $\times 10^{-45}$**	**< .0001**	**< .0001**
Problem Solving	-	**< .0001**	**< .0001**	**< .0001**
Reading Strategies	-	-	**< .0001**	**2.87 $\times 10^{-52}$**
Resource Usage	-	-	-	**2.05 $\times 10^{-109}$**

Table 3: p-values for McNemar's test on overall learning strategy proportions

[4]Used when p-value is reported by statsmodels as 0.0 and cannot be reported exactly.

5.2 Relationships Between Strategy Mentions and Course Characteristics

The next section will describe findings related to more detailed relationships between learning strategy mentions and the following course characteristics: course level, size, and discipline. In this section, comparisons will be made between levels of a given course characteristic for each learning strategy. A chi-square test of homogeneity is used to determine statistical significance, as the evaluations used now only appear in one level. Again, the level of significance used is $p < .05$, which indicates that there is a statistically significant difference between the proportion of evaluations mentioning a learning strategy between two levels of the course characteristic. Tables A2 through A4 show the exact proportions for each course characteristic that are represented in the following figures.

5.3 Course Level

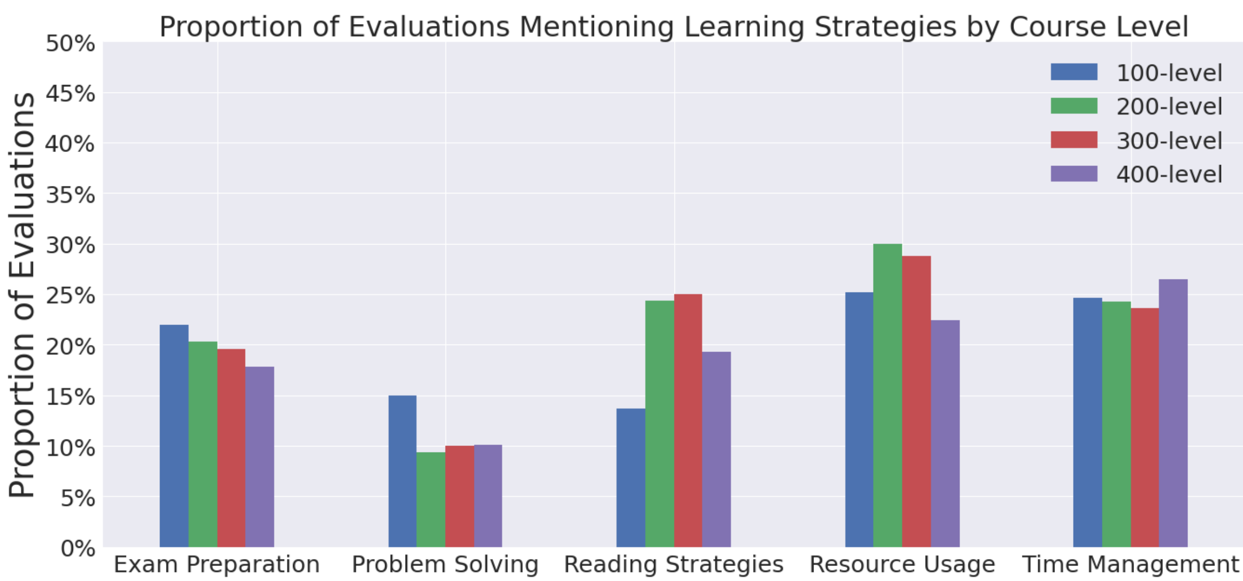

Figure 7: The proportion of evaluations mentioning each of the five selected learning strategies grouped by course level.

As shown in Figure 7, only one of the learning strategies, exam preparation, demonstrates a consistent negative relationship between course level and learning strategy mentions, with exam preparation being mentioned more often in lower level courses than upper level courses. All differences between each level for this strategy are statistically significant, as shown in Table 4. The problem solving strategy also shows a increased prevalence for the 100-level courses compared all other levels, though the same cannot be said for 200-level courses compared to upper level courses. Reading strategies and resource usage do not exhibit a clear trend in the proportions of evaluations mentioning these strategies, though the differences are statistically significant as shown in Tables 6 through 7. Finally, the results for the time management strategy show an increased proportion of mentions in 400-level courses compared to all other levels, though there is also a small decrease when comparing 100-level or 200-level courses to 300-level courses. These differences are also significant as shown in Table 8.

Course Level	100	200	300	400
100	-	**.0000005**	$\mathbf{2.70 \times 10^{-14}}$	$\mathbf{6.61 \times 10^{-23}}$
200	-	-	**.003**	$\mathbf{2.89 \times 10^{-11}}$
300	-	-	-	**0.000003**

Table 4: p-values for chi-square test on exam preparation proportions by course level

Course Level	100	200	300	400
100	-	$\mathbf{1.67 \times 10^{-113}}$	$\mathbf{1.48 \times 10^{-91}}$	$\mathbf{2.53 \times 10^{-46}}$
200	-	-	**.002**	**.015**
300	-	-	-	.764

Table 5: p-values for chi-square test on problem solving proportions by course level

Course Level	100	200	300	400
100	-	$\mathbf{8.23 \times 10^{-252}}$	$\mathbf{7.67 \times 10^{-284}}$	$\mathbf{2.84 \times 10^{-51}}$
200	-	-	**.020**	$\mathbf{3.52 \times 10^{-38}}$
300	-	-	-	$\mathbf{4.41 \times 10^{-48}}$

Table 6: p-values for chi-square test on reading strategies proportions by course level

Course Level	100	200	300	400
100	-	$\mathbf{1.94 \times 10^{-41}}$	$\mathbf{1.71 \times 10^{-24}}$	$\mathbf{2.70 \times 10^{-10}}$
200	-	-	**.00003**	$\mathbf{8.03 \times 10^{-72}}$
300	-	-	-	$\mathbf{1.11 \times 10^{-52}}$

Table 7: p-values for chi-square test on resource usage proportions by course level

Course Level	100	200	300	400
100	-	.263	**.001**	**.00006**
200	-	-	**.013**	**.00000008**
300	-	-	-	$\mathbf{6.48 \times 10^{-13}}$

Table 8: p-values for chi-square test on time management proportions by course level

5.4 Course Size

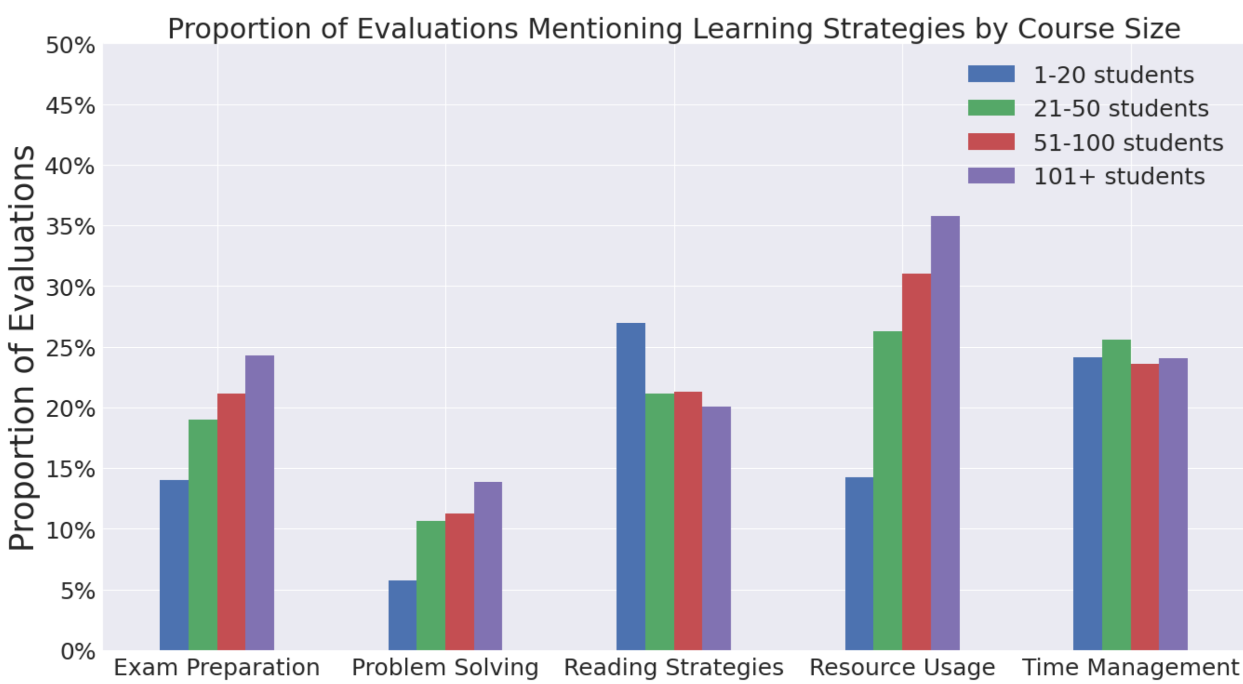

Figure 8: The proportion of evaluations mentioning each of the five selected learning strategies grouped by course size.

Figure 8 shows the frequency of each learning strategy among the four categories of course size. Three of the strategies demonstrated a consistent positive relationship: as course size increased, exam preparation, problem solving, and resource usage strategies were mentioned more often. All differences in proportions are statistically significant, as shown in Tables 9, 10, and 12 respectively. The prevalence of reading strategies mentions decreased as course size increased, with statistical differences between all levels except between 21-50 students and 51-100 students as shown in Table 11. Finally, Table 13 shows no significant differences between course sizes for time management mentions except between courses in the 21-50 student level and all other levels. Courses with 21-50 students have a slightly higher proportion of evaluations mentioning time management.

Course Level	1-20	21-50	51-100	101+
1-20	-	$\mathbf{3.60 \times 10^{-61}}$	$\mathbf{6.29 \times 10^{-117}}$	$\mathbf{5.27 \times 10^{-249}}$
21-50	-	-	$\mathbf{5.78 \times 10^{-12}}$	$\mathbf{3.27 \times 10^{-68}}$
51-100	-	-	-	$\mathbf{7.44 \times 10^{-24}}$

Table 9: p-values for chi-square test on exam preparation proportions by course size

Course Level	1-20	21-50	51-100	101+
1-20	-	$\mathbf{3.32 \times 10^{-109}}$	$\mathbf{8.57 \times 10^{-132}}$	$\mathbf{7.31 \times 10^{-270}}$
21-50	-	-	**.012**	$\mathbf{4.76 \times 10^{-41}}$
51-100	-	-	-	$\mathbf{1.33 \times 10^{-26}}$

Table 10: p-values for chi-square test on problem solving proportions by course size

Course Level	1-20	21-50	51-100	101+
1-20	-	$\mathbf{2.74 \times 10^{-64}}$	$\mathbf{6.74 \times 10^{-61}}$	$\mathbf{3.60 \times 10^{-105}}$
21-50	-	-	.730	**.0001**
51-100	-	-	-	**.00003**

Table 11: p-values for chi-square test on reading strategies proportions by course size

Course Level	1-20	21-50	51-100	101+
1-20	-	$\mathbf{5.09 \times 10^{-299}}$	**< .0001**	**< .0001**
21-50	-	-	$\mathbf{1.99 \times 10^{-40}}$	$\mathbf{1.82 \times 10^{-172}}$
51-100	-	-	-	$\mathbf{3.41 \times 10^{-43}}$

Table 12: p-values for chi-square test on resource usage proportions by course size

Course Level	1-20	21-50	51-100	101+
1-20	-	**.00002**	.146	.809
21-50	-	-	**.000000005**	**.0000009**
51-100	-	-	-	.179

Table 13: p-values for chi-square test on time management proportions by course size

5.5 Course Discipline

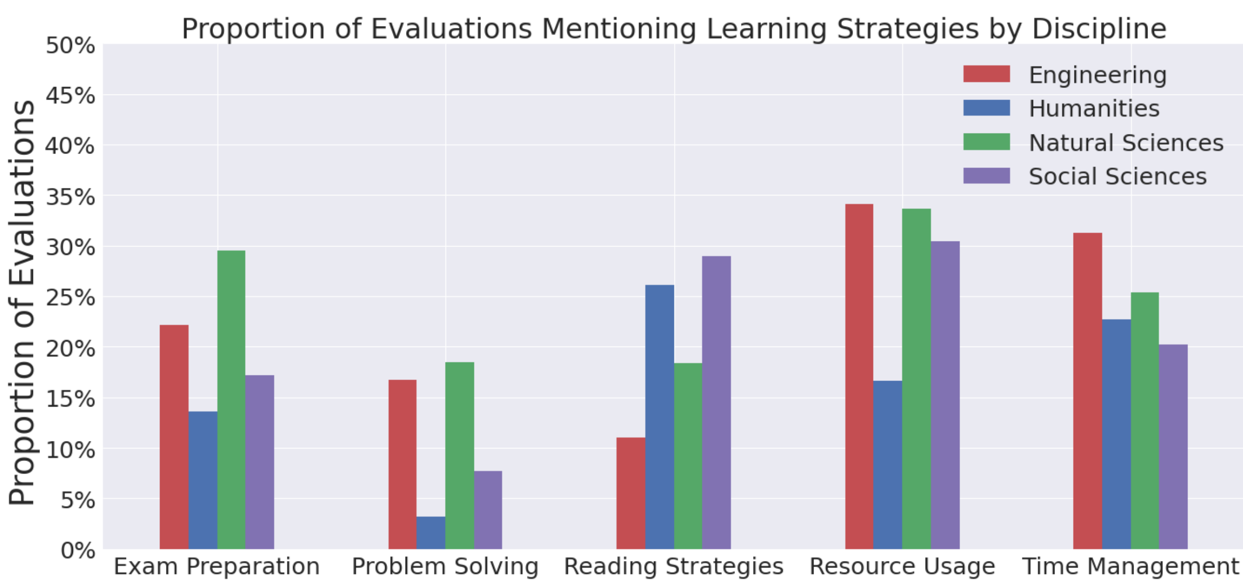

Figure 9: The proportion of evaluations mentioning each of the five selected learning strategies grouped by course discipline.

Figure 9 shows the frequency of learning strategy mentions in each of the four course disciplines.

All strategies except reading strategies are mentioned more frequently in evaluation for courses in the two STEM disciplines, Engineering and Natural Sciences, than the evaluation for Humanities or Social Sciences courses. Tables 14 through 18 show the statistical significance of these results for each strategy. All differences between STEM and non-STEM disciplines and between Humanities and Social Sciences are significant. Differences between Engineering and Natural Sciences are significant for all strategies except resource usage.

Course Discipline	Engineering	Humanities	Natural Sciences	Social Sciences
Engineering	-	**1.83×10^{-184}**	**1.99×10^{-92}**	**1.38×10^{-55}**
Humanities	-	-	**$< .0001$**	**1.66×10^{-42}**
Natural Sciences	-	-	-	**$< .0001$**

Table 14: p-values for chi-square test on exam preparation proportions by course discipline

Course Discipline	Engineering	Humanities	Natural Sciences	Social Sciences
Engineering	-	**$< .0001$**	**.00000001**	**9.75×10^{-265}**
Humanities	-	-	**$< .0001$**	**8.76×10^{-172}**
Natural Sciences	-	-	-	**$< .0001$**

Table 15: p-values for chi-square test on problem solving proportions by course discipline

Course Discipline	Engineering	Humanities	Natural Sciences	Social Sciences
Engineering	-	**$< .0001$**	**3.05×10^{-136}**	**$< .0001$**
Humanities	-	-	**2.83×10^{-140}**	**2.92×10^{-18}**
Natural Sciences	-	-	-	**5.80×10^{-234}**

Table 16: p-values for chi-square test on reading strategies proportions by course discipline

Course Discipline	Engineering	Humanities	Natural Sciences	Social Sciences
Engineering	-	**< .0001**	.260	$\mathbf{1.05 \times 10^{-22}}$
Humanities	-	-	**< .0001**	**< .0001**
Natural Sciences	-	-	-	$\mathbf{3.00 \times 10^{-20}}$

Table 17: ***p*-values for chi-square test on resource usage proportions by course discipline**

Course Discipline	Engineering	Humanities	Natural Sciences	Social Sciences
Engineering	-	$\mathbf{3.84 \times 10^{-137}}$	$\mathbf{3.65 \times 10^{-58}}$	$\mathbf{4.54 \times 10^{-217}}$
Humanities	-	-	$\mathbf{1.41 \times 10^{-17}}$	$\mathbf{5.50 \times 10^{-16}}$
Natural Sciences	-	-	-	$\mathbf{2.43 \times 10^{-57}}$

Table 18: ***p*-values for chi-square test on time management proportions by course discipline**

5.6 Results of Logistic Regression

The logistic regression was performed using the following formula, where each categorical variable takes on one of the four levels corresponding to each course characteristic:

strategy $\sim$ course_level + course_size + course_discipline

Tables 19 through 23 show the coefficients obtained for each strategy's regression, the calculated odds ratio, and the p-value for the coefficient. The variable names are bolded for significant results with $p < .05$. The regression results quantify the nature of the relationships observed in Figures 7 through 9. The baseline categories for course level, course size, and course discipline were chosen to be 100-level, 1-20 students, and Engineering courses, respectively. All odds ratios reflect the likelihood of strategy mentions for a given level relative to these baselines. The results indicate significant relationships between all learning strategies and the three course characteristics.

Variable	Coefficient (Standard Error)	Odds Ratio	p-value
200-level	-0.092 (0.019)	0.912	.000002
300-level	0.021 (0.02)	1.022	.297
400-level	-0.095 (0.029)	0.909	.001
21-50 students	0.242 (0.023)	1.274	4.69×10^{-26}
51-100 students	0.346 (0.024)	1.413	5.79×10^{-49}
101+ students	0.431 (0.024)	1.539	2.06×10^{-74}
Humanities	-0.477 (0.023)	0.621	3.69×10^{-99}
Natural Sciences	0.39 (0.02)	1.477	6.09×10^{-87}
Social Sciences	-0.315 (0.021)	0.729	9.43×10^{-51}

Table 19: Logistic regression results for exam preparation strategy

Variable	Coefficient (Standard Error)	Odds Ratio	p-value
200-level	-0.593 (0.025)	0.552	2.26×10^{-129}
300-level	-0.385 (0.026)	0.681	9.27×10^{-51}
400-level	-0.59 (0.036)	0.554	3.88×10^{-59}
21-50 students	0.321 (0.032)	1.379	3.49×10^{-23}
51-100 students	0.235 (0.033)	1.265	1.92×10^{-12}
101+ students	0.301 (0.033)	1.351	1.18×10^{-19}
Humanities	-1.783 (0.035)	0.168	$< .0001$
Natural Sciences	0.104 (0.023)	1.109	.000004
Social Sciences	-0.873 (0.027)	0.418	3.23×10^{-236}

Table 20: Logistic regression results for problem solving strategy

Variable	Coefficient (Standard Error)	Odds Ratio	p-value
200-level	0.717 (0.022)	2.048	1.00×10^{-243}
300-level	0.672 (0.022)	1.958	7.51×10^{-199}
400-level	0.582 (0.03)	1.79	1.09×10^{-81}
21-50 students	-0.145 (0.02)	0.865	4.03×10^{-13}
51-100 students	-0.072 (0.021)	0.931	.0006
101+ students	-0.081 (0.021)	0.922	.0001
Humanities	1.053 (0.024)	2.866	< .0001
Natural Sciences	0.601 (0.025)	1.824	5.40×10^{-130}
Social Sciences	1.145 (0.023)	3.141	< .0001

Table 21: Logistic regression results for reading strategies strategy

Variable	Coefficient (Standard Error)	Odds Ratio	p-value
200-level	0.286 (0.018)	1.332	1.27×10^{-56}
300-level	0.365 (0.019)	1.441	1.86×10^{-82}
400-level	0.126 (0.027)	1.134	.000003
21-50 students	0.644 (0.022)	1.903	5.48×10^{-194}
51-100 students	0.846 (0.022)	2.33	< .0001
101+ students	1.005 (0.022)	2.732	< .0001
Humanities	-0.695 (0.02)	0.499	1.22×10^{-258}
Natural Sciences	-0.018 (0.018)	0.982	.316
Social Sciences	-0.163 (0.018)	0.849	1.88×10^{-19}

Table 22: Logistic regression results for resource usage strategy

Variable	Coefficient (Standard Error)	Odds Ratio	p-value
200-level	-0.031 (0.018)	0.97	.091
300-level	-0.061 (0.019)	0.94	0.001
400-level	-0.089 (0.026)	0.915	.0006
21-50 students	-0.016 (0.019)	0.985	0.424
51-100 students	-0.145 (0.021)	0.865	2.51×10^{-12}
101+ students	-0.145 (0.021)	0.865	4.32×10^{-12}
Humanities	-0.501 (0.02)	0.606	1.17×10^{-142}
Natural Sciences	-0.307 (0.019)	0.736	1.23×10^{-59}
Social Sciences	-0.593 (0.019)	0.553	4.44×10^{-208}

Table 23: Logistic regression results for time management strategy

6 Discussion

The goal of this paper was to quantify the prevalence of five selected learning strategies in course evaluation responses and the relationship between learning strategy mentions and course characteristics. This discussion section will compare the results to the initial hypotheses and reflect on the implications of these findings.

Hypotheses 1 expressed that there would be significant differences between the proportions of learning strategy mentions. As $p < .05$ for all values in Table 3, we can reject the null hypothesis for Hypothesis 1 and conclude that all of the learning strategies are mentioned with statistically significant differing frequencies. These results indicate that certain learning strategies have a more significant impact on students across all courses.

Hypothesis 1.1 stated that the most prevalent strategy would be time management. This was not the case, as resource usage was mentioned more frequently, with a statistically significant difference between resource usage and the next more common strategy, time management. Though there is only a 3.45 percentage point decrease between the two strategies, the overall popularity of resource usage was not anticipated. These results may be a result of the overwhelming number of evaluations that mention ‘office hours’, which was included in the most frequent trigram as shown in Figure 5. Exam preparation and reading strategies were both mentioned at similar, but statistically different rates. Problem solving, however, was mentioned the least often. This may be a result of the overall impact of problem solving on a student’s experience in a course. While problem sets and similar assignments tend to be a constant element of a course’s workload, they are not weighted as significantly as exams or larger assignments requiring reading strategies, such as papers. This might also be a result of the smaller dictionary used for this strategy, whereas other dictionaries used a wider range of tokens that may capture broader concepts, and therefore a larger proportion of all evaluations. The overall magnitude of the learning strategy frequencies, ranging from 10.3% to 27.79%, provides further evidence of the value of course evaluations in providing practical information to students.

Hypothesis 2 stated that there would be a statistically significant relationship between course characteristics and learning strategy mentions. The p-values for the logistic regression coefficients in Tables 19 through 23 are less than .05 for almost all variables. We can reject the null hypothesis and conclude that there is a statistically significant relationship between each course characteristic and learning strategy mentions, though not at every level of each category. For Hypotheses 2.1 through 2.3, the statistical significance holds, though the relationships determined do not all align with these hypotheses.

Hypothesis 2.1 stated that introductory (100 and 200 level) courses would have higher rates of learning strategy mentions compared to upper level courses. This was the case for some, but not all learning strategies, meaning the hypothesis was not fully correct. Time management demonstrated the clearest negative relationship between course level and strategy mentions. However, the relationship, though statistically significant, was very small. Compared to 100-level courses, 300-level and 400-level courses are 6% and 8.5% less likely to mention time management. This may reflect time management’s broad applicability, as has been discussed previously. Compared to 100-level courses, reading strategies are 104%, 95.8%, and 79% more likely to be mentioned in evaluations for 200-level, 300-level, and 400-level courses. It is unclear why reading strategies are much more common in 200-level courses compared to 100-level courses. The slight decrease from 200-level courses to higher level courses indicates some support for the hypothesis, and it may be that reading strategies are not expected to be used as intensely in 100-level courses, and are only developed in 200-level courses and above. On the other hand, the resource usage strategy is mentioned with increasing likelihood for 200-level and 300-level courses, and more frequently for 400-level courses compared to 100-level courses. This disagreement with the hypothesis may reflect students' initial unfamiliarity with the availability and importance of seeking out resources and help in college level coursework, which is gained over time as students progress to upper level courses and eventually becomes a more routine part of students' learning approaches.

Hypothesis 2.2 stated that large courses would have higher rates of learning strategy mentions. Once again, this was the case for some, but not all of the learning strategies. The hypothesis was most clearly supported for the exam preparation and resource usage strategies, as shown by the consistently increasing odds ratios in Tables 19 and 22 as course size increases. Problem solving strategies were 38%, 27%, and 35% more likely to be mentioned compared to the baseline, though the small magnitude of these differences indicates the overall similarity of problem solving strategy mentions in lectures of different sizes. However, time management strategy mentions were 13.5% less likely to appear in evaluations for larger courses with 50-100 or over 100 more students. It is possible that larger courses benefit less from these strategies, or that students choose not to discuss time management, even if it may play a role in their approach to the course. Reading strategies were more likely to be used in the smallest sized course compared to all other courses, though again the likelihood compared to the baseline was at most 13.5% less for any of the course sizes. This may be a result of the prevalence of discussion-based seminars in the baseline enrollment category, which often rely on close reading of existing texts. As a whole, overall patterns reflect a positive relationship between increasing course size and increasing strategy mentions.

Hypothesis 2.3 stated that courses from STEM disciplines would have higher rates of learning strategy mentions compared to humanities and social science courses. This was true for all strategies except reading strategies, supporting the hypothesis. Exam preparation and problem solving strategies were more likely to be mentioned in natural sciences course evaluations compared to engineering evaluations, and less like to be mentioned in humanities and social science evaluations compared to engineering evaluations. This aligns with the known methods of assessment used in these courses. In contrast, reading strategies were over three times as likely to be mentioned in social science courses, and nearly three times as likely to be mentioned in humanities courses. Since humanities and social science courses are known to have a larger emphasis on close reading and responding to texts, this particular result is not surprising.

It is more interesting to note the relationship between course discipline and resource usage mentions, as well as for time management mentions. As previously stated, these strategies would appear to be relevant across disciplines, though these results were expected according to Hypothesis 2.3. The logistic regression identifies that humanities and social sciences course evaluations are respectively 50% and 15% less likely to mention resource usage compared to engineering courses. No conclusions can be drawn about the difference in resource usage mentions between engineering and natural science courses, as no statistically significant difference was confirmed as indicated in Table 17. These results may indicate different motivations held by students when making use of course resources, such as office hours. In STEM courses, office hours tend to be used for clarification of assignments and for explaining relevant course material, which are both directly related to student success in a course. On the other hand, office hours in other disciplines may be more focused on open discussion of course topics, and not play a direct impact on a student's course performance. This might lead students in humanities and social science courses to reference office hours or other resources less often when being asked to provide advice. Additionally, as previously mentioned, the McGraw Center tutoring program focuses almost entirely on STEM courses, providing STEM students with a greater variety of resources than students in other courses.

When examining the time management strategy, we can observe significant results when comparing across disciplines: humanities, natural sciences, and social sciences courses were 26%, 39%, and 45% less likely to mention time management compared to engineering courses. This may reflect the different course and assignment schedules used in these disciplines, where assignments are due on a weekly or bi-weekly basis and are cumulative, and therefore require more active management throughout the semester. Humanities courses may demonstrate a higher likelihood of time management strategy mentions compared to natural or social science courses due to a higher reading load, which requires additional planning to manage.

As a whole, disciplinary differences in learning strategy mentions are evident in these results, even when considering broadly valuable strategies that do not reflect discipline-specific skills. It should be noted that these results do not necessarily reflect a difference in the rates at which students use these strategies in their coursework. Instead, they represent a measurement of the ways in which students from different disciplines discuss their learning strategy usage. These results might indicate that students from separate disciplines actively prioritize these strategies to different degrees. Students may also use differing terminology to discuss the strategies across disciplines that were not fully captured by the learning strategy dictionaries.

These results can be used to support multiple audiences on Princeton's campus: students, who are the direct readers of these evaluation responses, as well the members of students' "learning support network" [31], including faculty, course staff, and the McGraw Center. Since results show that learning strategies are mentioned with meaningful frequency in the evaluations, a labeling system similar to the one designed in this paper could be applied to existing course evaluations and presented to students. This would provide an organized method for students to find targeted advice from the evaluations. These students could be encouraged to use the evaluations as a resource while enrolled in their courses and implement course-specific, applicable strategies.

Faculty and course staff could use these results in order to gain more specific insight into the student experience in courses. While students submit additional evaluation responses directly to course staff, the responses examined in this paper function as an important indicator of how students approached a particular course. Faculty can use this measurement of learning strategy mentions to identify the ways in which student approaches are positively or negatively supporting the fulfillment of a course's goals.

Finally, as mentioned at the start of this paper, the McGraw Center is a crucial resource on campus that supports student engagement with learning strategies. By applying these results to its work, the McGraw Center Learning Consultation program could produce more targeted support for students by reflecting on the types of courses included in a student's schedule and the strategies most often used in those courses. In addition, this work highlights strategies that are potentially under-utilized by students, allowing for further investigation into the ways McGraw Center programming could promote certain strategy usage.

7 Summary

7.1 Conclusion

This paper implemented a dictionary-based natural language processing approach in order to quantify the prevalence of five popular learning strategy categories in 137,128 Princeton undergraduate course evaluations. Prior research has demonstrated the importance of learning strategies for student success and the potential of analyzing course evaluations through computational techniques. This paper completed a large-scale analysis of the Princeton undergraduate course evaluations, enabling a new understanding of how the student body as a whole approaches learning at Princeton. Resource usage and time management were found to be the most frequently mentioned strategies, followed in decreasing order by reading strategies, exam preparation, and problem solving. The course level, size, and discipline were all shown to have a statistically significant relationship with mentions of each of the five learning strategies. Quantifying the prevalence of these strategies can provide clearer insight into the extent to which students engage with these strategies across the Princeton curriculum. Results can be used to supplement existing practices of individualized learning strategy engagement at the student, faculty, and institutional levels.

7.2 Limitations

Though the learning strategies dictionaries appear to be an effective tool for measuring learning strategy usage, there is not a direct relationship between mentions of individual tokens and the discussion of explicit learning strategies. For example, students may provide feedback regarding the difficulty of exams or readings without discussing strategies for approaching these course components. Student discussion of potential areas for learning strategy usage was still included in the count of learning strategy mentions, since this discussion can still provide valuable information to students intending to develop learning strategies based on the evaluations. However, due to this limitation, these results only represent a high-level overview of learning strategy engagement.

An additional limitation of this paper is the assumption of the learning strategy dictionaries' accuracy. Given more time, a machine learning model could be trained on manually labeled evaluations and used to predict the presence of each learning strategy in the evaluations. This model's accuracy could be evaluated, and then the proportion of evaluations predicted by the model as mentioning each learning strategy could be calculated. Comparing the model's proportions to the results of this paper could allow for a validation of the learning strategies dictionary approach.

7.3 Further Work

Further work could examine the department-level differences in learning strategy mentions. The results for this work is available, but was excluded from this paper in order to focus on the relationship between course discipline and learning strategy usage. This work could be used to compare departments within disciplines, as well as to identify the prevalence of each learning strategy at the department level. Further work could also explore common themes present in the course evaluations within each learning strategy. This paper's approach could be supplemented by an analysis of which tokens from the learning strategies dictionaries appear most often in the labeled evaluations. This would provide more specific insight into the breadth and depth of learning strategies used by students. Finally, further work could compare ways in which Princeton students might conceptualize or implement learning strategies in light of the differences seen between types of courses. For example, students in STEM and humanities courses could be surveyed regarding their approaches to time management and resource usage to better understand how to engage students in these disciplines in learning strategies.

Learning strategies will continue to play an influential role in the undergraduate academic experience at Princeton. This paper, along with future work, can play a role in providing students and campus resources with a better understanding of these strategies, supporting student success across the university.

8 Acknowledgements

Thank you to Professor Fellbaum for her guidance throughout the semester. It has been a valuable and enjoyable learning experience to work with her on this paper as a member of the NLP Independent Work seminar. Thank you as well to seminar TAs Alan Ding and Pranay Manocha, and all members of the seminar, for their support and feedback.

Additionally, I greatly appreciate the support of Nic Voge, Senior Associate Director at the McGraw Center, who has been an influential mentor during my work in the Learning Strategies Consultation program. I am grateful to him and my fellow learning consultants for our engaging discussions about the role of learning strategies at Princeton.

Thank you as well to Charles An for providing me with access to the Princeton Courses course evaluation data and to Oscar Torres-Reyna in Data & Statistical Services for helping me interpret the statistical results included in this paper.

9 Honor Code

This paper represents my own work in accordance with University regulations.

/s/ Sophie Goldman

References

[1] B. L. McCombs, "Historical Review of Learning Strategies Research: Strategies for the Whole Learner—A Tribute to Claire Ellen Weinstein and Early Researchers of This Topic," *Frontiers in Education*, vol. 2, p. 6, 2017. [Online]. Available: https://www.frontiersin.org/article/10.3389/feduc.2017.00006

[2] "LASSI 3rd Edition." [Online]. Available: https://www.hhpublishing.com/ap/_assessments/LASSI-3rd-Edition.html

[3] C. J. Fong, M. R. Krou, K. Johnston-Ashton, M. A. Hoff, S. Lin, and C. Gonzales, "LASSI's great adventure: A meta-analysis of the Learning and Study Strategies Inventory and academic outcomes," *Educational Research Review*, vol. 34, p. 100407, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1747938X21000300

[4] "Learning Strategies Consultations." [Online]. Available: https://mcgraw.princeton.edu/undergraduates/learning-strategies-consultations-undergraduates

[5] "McGraw on a Page - Fall 2021." [Online]. Available: https://mcgraw.princeton.edu/node/4656
[6] "Principedia." [Online]. Available: https://mcgraw.princeton.edu/undergraduates/principedia
[7] Princeton University Office of the Registrar, "Course Evaluation Results." [Online]. Available: https://registrarapps.princeton.edu/course-evaluation
[8] "Group and Individual Tutoring." [Online]. Available: https://mcgraw.princeton.edu/undergraduates/group-and-individual-tutoring
[9] S. F. E. Rovers, R. E. Stalmeijer, J. J. G. van Merriënboer, H. H. C. M. Savelberg, and A. B. H. de Bruin, "How and Why Do Students Use Learning Strategies? A Mixed Methods Study on Learning Strategies and Desirable Difficulties With Effective Strategy Users," *Frontiers in Psychology*, vol. 9, p. 2501, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.02501
[10] S. Qureshi and R. Ullah, "Learning Experiences of Higher Education Students: Approaches to Learning as Measures of Quality of Learning Outcomes," *Bulletin of Education and Research*, vol. 36, no. 1, pp. 79–100, Jun. 2014, publisher: Institute of Education and Research. [Online]. Available: https://eric.ed.gov/?id=EJ1210435
[11] B. D. Arend, "Course assessment practices and student learning strategies in online courses," *Journal of Asynchronous Learning Networks*, vol. 11, no. 4, pp. 3–17, 2007, publisher: Citeseer.
[12] A. Simsek and J. Balaban, "Learning Strategies of Successful and Unsuccessful University Students," *Contemporary Educational Technology*, vol. 1, no. 1, Mar. 2010. [Online]. Available: https://www.cedtech.net/article/learning-strategies-of-successful-and-unsuccessful-university-students-5960
[13] T. Sliusarenko, L. H. Clemmensen, and B. Ersball, "Text mining in students' course evaluations," in *International Conference on Computer Supported Education. INSTICC Press*, 2013.
[14] A. ONAN, "Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach," *Computer Applications in Engineering Education*, vol. 29, no. 3, pp. 572–589, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.22253
[15] V. Kovanović, S. Joksimović, N. Mirriahi, E. Blaine, D. Gašević, G. Siemens, and S. Dawson, "Understand students' self-reflections through learning analytics," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. Sydney New South Wales Australia: ACM, Mar. 2018, pp. 389–398. [Online]. Available: https://dl.acm.org/doi/10.1145/3170358.3170374
[16] "LIWC." [Online]. Available: https://liwc.wpengine.com/how-it-works/
[17] C. Liu, "Understanding Academic Emotions at Princeton University," Senior Thesis, Princeton University, Princeton, NJ, 2018. [Online]. Available: http://arks.princeton.edu.ezproxy.princeton.edu/ark:/88435/dsp01pc289m804
[18] M. Hammel, "Analyzing Increasing Enrollments and Diversity in Princeton Computer Science," Senior Thesis, Princeton University, Princeton, NJ, 2018. [Online]. Available: http://arks.princeton.edu.ezproxy.princeton.edu/ark:/88435/dsp01qb98mj20j
[19] J. White, "Experimental Measures of Difficulty for Princeton Courses," Senior Thesis, Princeton University, Princeton, NJ, 2017. [Online]. Available: http://arks.princeton.edu/ark:/88435/dsp010k225d672
[20] C. Huang, "An Adversarial Fair Autoencoder for Debiased Representations of Data," Senior Thesis, Princeton University, Princeton, NJ, 2018. [Online]. Available: http://arks.princeton.edu.ezproxy.princeton.edu/ark:/88435/dsp01xw42nb64t
[21] A. Ho, "How do course evaluations improve teaching and learning?" Feb. 2019. [Online]. Available: https://scholar.harvard.edu/files/andrewho/files/how_do_course_evaluations_improve_teaching_and_learning_-_andrew_ho_-_linc_2019.pdf
[22] M. Prosser and K. Trigwell, "Student Evaluations of Teaching and Courses: Student Study Strategies as a Criterion of Validity," *Higher Education*, vol. 20, no. 2, pp. 135–142, 1990, publisher: Springer. [Online]. Available: http://www.jstor.org/stable/3447285
[23] ——, "Student evaluations of teaching and courses: Student learning approaches and outcomes as criteria of validity," *Contemporary Educational Psychology*, vol. 16, no. 3, pp. 293–301, Jul. 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0361476X9190029K
[24] N. Entwistle and H. Tait, "Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments," *Higher Education*, vol. 19, no. 2, pp. 169–194, Jun. 1990. [Online]. Available: https://doi.org/10.1007/BF00137106
[25] "Princeton Courses." [Online]. Available: https://www.princetoncourses.com/
[26] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc.", 2009.
[27] 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*, vol. 12, pp. 2825–2830, 2011.
[28] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.
[29] O. Torres-Reyna, "Logit, Probit and Multinomial Logit models in R." [Online]. Available: https://www.princeton.edu/~otorres/LogitR101.pdf

[30] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, Jun. 1947. [Online]. Available: https://doi.org/10.1007/BF02295996
[31] "Learning Support Network." [Online]. Available: https://mcgraw.princeton.edu/learning-support-network

10 Appendix

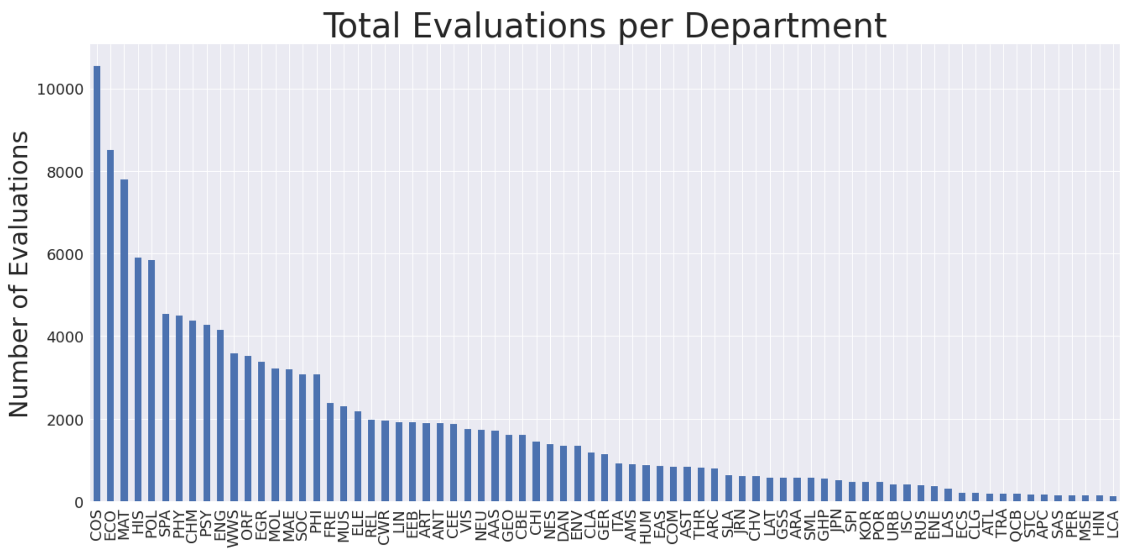

Figure A1: The number of evaluations for each of the 75 departments/programs.

Topic 0 words	Topic 1 words	Topic 2 words	Topic 3 words	Topic 4 words	Topic 5 words	Topic 6 words	Topic 7 words	Topic 8 words	Topic 9 words
read	take	class	interest	lectur	hour	class	problem	cours	take
lectur	cours	cours	cours	exam	go	cours	cours	professor	class
make	interest	semest	class	class	help	take	materi	class	lot
go	class	take	read	question	offic	recommend	set	take	learn
sure	major	like	take	studi	assign	professor	lectur	student	time
precept	would	first	histori	time	get	interest	good	great	cours
class	requir	much	great	make	lab	great	class	way	work
get	recommend	week	realli	attent	start	would	lot	help	fun
up	good	time	professor	final	earli	high	interest	learn	languag
take	scienc	get	lot	pay	class	definit	realli	realli	interest
good	need	math	topic	practic	take	one	pretti	write	definit
keep	learn	year	lectur	week	make	princeton	exam	think	want
cours	want	work	paper	note	realli	best	well	princeton	realli
stay	great	half	learn	ask	preceptor	amaz	work	engag	much
attend	unless	physic	discuss	problem	also	lectur	hard	one	put
easi	subject	took	like	go	professor	realli	learn	amaz	great
materi	definit	up	polit	midterm	question	engag	time	also	get
top	materi	think	good	take	ask	materi	difficult	incred	worth
tri	realli	would	much	befor	cours	super	understand	experi	assign
help	math	end	engag	read	understand	taken	great	care	prepar

Figure A2: Topic Model Results.

Course Code	Department/Program Name	Discipline
AAS	African American Studies	Social Sciences
AMS	American Studies	Social Sciences
ANT	Anthropology	Social Sciences
APC	Applied and Computational Math	Natural Sciences
ARA	Arabic	Humanities
ARC	Architecture	Humanities
ART	Art and Archaeology	Humanities
AST	Astrophysical Sciences	Natural Sciences
ATL	Atelier	Humanities
CBE	Chemical and Biological Engineering	Engineering
CEE	Civil and Environmental Engineering	Engineering
CHI	Chinese	Humanities
CHM	Chemistry	Natural Sciences
CHV	Center for Human Values	Humanities
CLA	Classics	Humanities
CLG	Classical Greek	Humanities
COM	Comparative Literature	Humanities
COS	Computer Science	Engineering
CWR	Creative Writing	Humanities
DAN	Dance	Humanities
EAS	East Asian Studies	Humanities
ECO	Economics	Social Sciences
ECS	European Cultural Studies	Social Sciences
EEB	Ecology and Evolutionary Biology	Natural Sciences
EGR	Engineering	Engineering

ELE	Electrical Engineering	Engineering
ENE	Energy Studies	Natural Sciences
ENG	English	Humanities
ENV	Environmental Studies	Natural Sciences
FRE	French	Humanities
GEO	Geosciences	Natural Sciences
GER	German	Humanities
GHP	Global Health and Health Policy	Social Sciences
GSS	Gender and Sexuality Studies	Social Sciences
HIN	Hindi	Humanities
HIS	History	Social Sciences
HUM	Humanistic Studies	Humanities
ISC	Integrated Science Curriculum	Natural Sciences
ITA	Italian	Humanities
JPN	Japanese	Humanities
JRN	Journalism	Humanities
KOR	Korean	Humanities
LAS	Latin American Studies	Social Sciences
LAT	Latin	Humanities
LCA	Lewis Center for the Arts	Humanities
LIN	Linguistics	Social Sciences
MAE	Mechanical and Aerospace Engineering	Engineering
MAT	Mathematics	Natural Sciences
MOL	Molecular Biology	Natural Sciences
MSE	Materials Science and Engineering	Engineering
MUS	Music	Humanities

NES	Near Eastern Studies	Humanities
NEU	Neuroscience	Natural Sciences
ORF	Operations Research and Financial Engineering	Engineering
PER	Persian	Humanities
PHI	Philosophy	Humanities
PHY	Physics	Natural Sciences
POL	Politics	Social Sciences
POR	Portuguese	Humanities
PSY	Psychology	Natural Sciences
QCB	Quantitative and Computational Biology	Natural Sciences
REL	Religion	Humanities
RUS	Russian	Humanities
SAS	South Asian Studies	Humanities
SLA	Slavic Languages and Literatures	Humanities
SML	Statistics and Machine Learning	Natural Sciences
SOC	Sociology	Social Sciences
SPA	Spanish	Humanities
SPI	School of Public and International Affairs	Social Sciences
STC	Science and Technology Council	Natural Sciences
THR	Theater	Humanities
TRA	Translation and Intercultural Communication	Social Sciences
URB	Urban Studies	Social Sciences
VIS	Visual Arts	Humanities
WWS	Woodrow Wilson School (now SPI)	Social Sciences

Table A1: Course Codes, Program Names, and Assigned Discipline.

Learning Strategy / Course Level	100	200	300	400
Exam Preparation	0.219	0.203	0.196	0.179
Problem Solving	0.150	0.094	0.100	0.101
Reading Strategies	0.137	0.244	0.250	0.193
Resource Usage	0.252	0.300	0.287	0.224
Time Management	0.247	0.243	0.236	0.264

Table A2: Course Level Proportions for Learning Strategy Mentions Corresponding to Figure 7

Learning Strategy / Discipline	1-20 students	21-50 students	51-100 students	101+ students
Exam Preparation	0.140	0.190	0.211	0.243
Problem Solving	0.057	0.106	0.112	0.139
Reading Strategies	0.270	0.212	0.213	0.200
Resource Usage	0.143	0.263	0.310	0.358
Time Management	0.241	0.256	0.236	0.240

Table A3: Course Size Proportions for Learning Strategy Mentions Corresponding to Figure 8

Learning Strategy / Course Size	Engineering	Humanities	Natural Sciences	Social Sciences
Exam Preparation	0.221	0.136	0.295	0.171
Problem Solving	0.167	0.032	0.185	0.077
Reading Strategies	0.111	0.261	0.184	0.290
Resource Usage	0.341	0.167	0.337	0.304
Time Management	0.313	0.227	0.253	0.203

Table A4: Course Discipline Proportions for Learning Strategy Mentions Corresponding to Figure 9

NLTK Stopwords

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of','at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'above', 'below', 'to', 'from', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where','why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't",'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan',"shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]