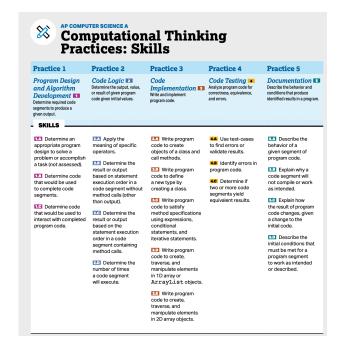


In my experience with high school curriculum, I have found a wide variation in the rigor of course material. This project seeks to develop a tool for evaluating the rigor of a curriculum, by measuring its alignment to the College Board's respective AP Course. This project focuses on the College Board's AP Computer Science A course, which covers a first year Java and Object Orientied Design course.

For this course, the College Board defines a set of "Computational Thinking Practices" (skills) and content that will be assessed on a year-end summative assessment to determine student's mastery of the course.

There are 5 main Computational Thinking Practices identified by the College Board, which it then breaks down into subskills:



In addition, the College Board defines a set of "Essential Knowledge" (the content) to be assessed in the course, which it organizes under 5 "Big Ideas." For example, the content for a lesson on iteration is:

## **ENDURING UNDERSTANDING**

CON-2

Programmers incorporate iteration and selection into code as a way of providing instructions for the computer to process each of the many possible input values.

## **LEARNING OBJECTIVE**

### CON-2.C

Represent iterative processes using a while loop.

## **ESSENTIAL KNOWLEDGE**

### CON-2.C.1

Iteration statements change the flow of control by repeating a set of statements zero or more times until a condition is met.

#### ON-2.C.2

In loops, the Boolean expression is evaluated before each iteration of the loop body, including the first. When the expression evaluates to true, the loop body is executed. This continues until the expression evaluates to false, whereupon the iteration ceases.

#### CON-2.C.3

A loop is an infinite loop when the Boolean expression always evaluates to true.

### CON-2.C.4

If the Boolean expression evaluates to false initially, the loop body is not executed at all.

## CON 2 C F

Executing a return statement inside an iteration statement will halt the loop and exit the method or constructor.

continued on next page

Every question on the College Board's end-of-course summative exam is aligned to a particular computational thinking skill and essential knowledge. As a note, some school networks have found the College Board's standards to be very complete, and "backwards plan" their middle school and pre-AP high school courses to prepare students for the AP level work.

As a first step, this project will focus on the assessment questions used in a particular curriculum, and measure how well they align to the College Board's Computational Thinking Practice and Curriculum Framework. (As a note, AP classes in most subjects have an analagous set of thinking practices and framework standards, so one day, this work may be generalized to assess curriculums in other subject areas.)

## Two questions to assess are:

- Can a TF-IDF vectorization of College Board question prompt with a Logistic Regressor successfully classify an assessment question by Computational Thinking Practice?
- 2. If ChatGPT is supplied only with the College Board Framework for Computational Thinking, can it successfully identify the particular thinking practice being assessed by a question prompt?

## **Initial Conclusions:**

 When just classifying between two different computational thinking practices, both the Logistic Regression and ChatGPT classify with 100% accuracy

## **Next Steps:**

1. Expand this to include all 15

- computational thinking practices.
  Compare accuracy of Logistic
  Regression and ChatGPT.
- 2. Determine whether the classifier can also identify the "Essential Knowledge" assessed by the question, not just the computational skill.
- 3. Attempt to generalize the classifiers to classify non-assessment questions such as lecture material, lab questions, and homework problems.
- Create a visualization that shows the distribution of thinking skills and content assessed over the course of the curriculum.

# Classifying Questions Using Logistic Regression

As first step, this section will try to classify prompts as assessing one of these two AP Computational Thinking Practices (CTP):

 CTP 2.A: Apply the meaning of specific operators. For example: Consider the following code segment.

```
int x = 7;
int y = 3;
if ((x < 10) && (y < 0))

System.out.println(""Value
is: "" + x * y);
else

System.out.println(""Value
is: "" + x / y)</pre>
```

What is printed as a result of executing the code segment?

2. CTP 2.B: Determine the results or output based on statement execution order in a code segment without method calls (except for output). For example:

Consider the following code segment.

```
int[] arr = {7, 2, 5, 3,
0, 10};
for (int k = 0; k <
arr.length - 1; k++) {
  if (arr[k] > arr[k + 1])
    System.out.print(k +
"" "" + arr[k] + "" "");
}
```

What will be printed as a result of executing the code segment?

## **Preprocessing the Data**

In this first step, we will:

1. Read in example prompts for each

category.

Preprocess the data: tokenize, lemmatize, and look at the token frequency distribution by question category.

# Testing the Logistic Regression Classifier on the Questions

Here, we build a pipeline that does the following:

- Preprocess Text: tokenize and lemmatize the text, vectorize using TF-IDF, restricting to 10 features.
- 2. Train and test a logistic regression classifier. Evaluate the model appropriately.

# **Build and Test a ChatGPT Classifier**

This classifier asks ChatGPT to determine the Computational Thinking Skill being assessed in the problem.

- Create a function call to ask
   ChatGPT for the classification
- 2. Test ChatGPT on data set and evaluate classification.

## Summary

When just looking at two categories of questions, both the Logistic Regression and ChatGPT classifiers classify the computational thinking practice with 100% accuracy. This should be expanded to incorporate classify questions from all 15 thinking practices. In addition, it should be trained to classify the content assessed, not just the skill.