You don't have a submission that is not late.

Question 1. Word Bank Matching (1 point each, 16 points total)

For each statement below, input the letter of the term that is *best* described. Note that you can click each word (cell) to mark it off. Each word is used at most once.

A. — Adapter Design Pattern	B. — Bug Bounties	C. — Concurrency Bug	D. — Delegation
E. — Delta Debugging	F. — Factory Method Pattern	G. — Functional Requirement	H. — Mutation Testing
I. — Observer Design Pattern	J. — Perverse Incentive	K. — Profiling	L. — Quality Requirement
M. — Requirements Elicitation	N. — Risk Assessment	O. — Singleton Design Pattern	P. — Stakeholders
Q. — Strong Conflict	R. — Traceability	S. — Validation	T. — Verification
U. — Weak Conflict			

Q1.1: P

Robbie wants to ensure that their new business idea for Golddex is viable, so they discuss with people like customers, regulatory bodies, and the relevant leadership at Golddex

Q1.2: H

Carey wants to ensure their collection of test cases is strong enough to get full points on the Autograder. So, they purposely make small changes to the source code of their program to see if the test cases can detect there is a change.

Q1.3: S

After eliciting requirements from all the relevant stakeholders, Kamilla - the product manager at Sunnamplex - wants to ensure that the requirements are complete and accurately reflect the customers' needs. So, they conduct interviews and create prototypes to ensure their requirements are correct.

Q1.4: F

Jan creates a C++ class 'Animal', with subclasses 'Dog' and 'Cat'. They have a function called `MakeAnimal` that can construct 'Animal' objects (which are either 'Dog' or 'Cat'), though the actual subclass of the returned object is not revealed in the function signature of `MakeAnimal`.

Q1.5: T

Ollie is finishing an EECS project, but wants to *formally* ensure that their software is correct and satisfies the specification. Specifically, they want to prove the absence of a particular class of bugs. So, Ollie uses *this* technique (that is based on rigorous mathematical proofs) to prove correctness of their software.

Q1.6: Q

Bioplex is releasing a new video game for mobile devices. As they do requirements elicitation, one stakeholder says 'players should not be able to see other users' match history', while another stakeholder says 'it should be possible for players to see other users' match history'.

Q1.7: L

Kan-code, a cloud-hosting solution for businesses, has the following goal - 'the platform should be secure and easy-to-use'.

Q1.8: N

Toughzap, a startup creating an AI-powered financial advisor, identifies potential issues they may face in the future, such as data breaches or incorrect predictions. For each issue, they develop a mitigation strategy.

Q1.9: **U**

Openlane is creating an eBook mobile application. One stakeholder provides the requirement that 'users' books should expire from their library after 60 days of first opening the book'. Another stakeholder, however, says 'books should not expire; users should be able to read any books they've purchased in the past'

Q1.10: **C**

Jessie notices an issue in their multi-threaded program where two threads are accessing the same variable, and one thread is writing. They use CHESS to detect this.

Q1.11: **J**

Statholdings wants their engineers to complete tasks quicker, so they give performance bonuses based on each employee's 'lines of code added' in hopes of getting them to code more. However, engineers start writing unreadable and duplicate code to boost their metric.

Q1.12: **O**

Domzoom is building an application that requires a global object used by multiple classes. They ensure that there is only one instance of this object across all the classes when the application is running.

Q1.13: **I**

Kris is working on a personal project that allows anyone that accesses their website to view the real-time price of a given stock. When a user enters a stock ticker, they are 'subscribed' to any changes for that stock. If the stock price changes, the new price is published to all users subscribed to it.

Q1.14: **M**

Donware is creating a tool for customers to track all their current shipping orders. In order to create a useful tool for their customers, they *gather input* from their customers and other stakeholders to better understand their needs and preferences.

Q1.15: **D**

Paula is creating a custom sorting function, which relies on a comparator function they wrote. The comparator function enables code reuse.

Q1.16: **K**

Frankie wants to find out the specific functions that are causing a longer-than-expected runtime in their application. So, they use *this technique* which tells them which functions take a long time to run.

Question 2. Delta Debugging (20 points)

Consider the scenarios below. For each scenario, specify if Delta Debugging can be used to solve the problem. If it can, select the answer choice marked "useful". If it cannot be used to solve the problem, select an assumption which may be violated (if multiple assumptions are violated, you can choose any one of them).

(a-i) (3 points) Suppose we have an online order form open for UMich undergraduate students with an offer for free t-shirts, but we want to make sure that we only count one submission per student. Each form includes a student's UMID. Using a list of these UMIDs, we hope to find the smallest subset that contains all duplicate UMIDs, such that we can remove the corresponding duplicate submissions. Interesting(x) returns true if the set contains a duplicate UMID.

○ A) Useful
○ B) Not monotonous
○ C) Not unambiguous
○ D) Not consistent

ANSWER: C

(a-ii) (3 points) We have a large bug database for a search engine that contains hundreds of bug reports. Each report consists of many lines of information, some of which may not be relevant to the bug. We want to simplify these bug reports by eliminating all lines that do not produce the bug, and keep only those that cause the bug to surface. Interesting(x) returns true on a subset of lines if including the line produces the bug.

○ A) Useful

○ B) Not monotonous
○ C) Not unambiguous
○ D) Not consistent

ANSWER: A

(a-iii) (3 points) We have a website consisting entirely of HTML files. The site currently crashes under certain conditions, and we want to know which subset of files cause the browser to crash. Interesting(x) returns true on the subset of files x if x causes a crash.

○ A) Useful
○ B) Not monotonous
○ C) Not unambiguous
○ D) Not consistent

ANSWER: A

Consider the following algorithm, Tribug Debugging, which functions in a very similar way as the Delta Debugging algorithm covered in the lecture and shares the same interface. That is, Tribug Debugging also takes two inputs that meet the following two preconditions: (1) P which is a list that is alone not interesting and (2) C which is a list such that the union of P and C is interesting.

It returns smallest, which represents a smallest subset of C such that the union of P and smallest is interesting. The key difference between Tribug Debugging and Delta Debugging is that, rather than splitting C in half as in Delta Debugging, Tribug Debugging partitions C into thirds. The following tribug function is an implementation of this algorithm.

```
1  def split(list):
2      third = math.floor(len(list) / 3)
3      // Note that C is a list, so order matters
4      return list[:third], list[third:third*2], list[third*2:]
5
6  def simple_tribug(P, C):
7      for element in C:
8          if interesting(P + element): return element
9      return C
10
11 def tribug(P, C):
12     if len(C) < 3: return simple_tribug(P, C)
13
14     C1, C2, C3 = split(C)
15     if (interesting(P + C1)): return tribug(P, C1)
16     if (interesting(P + C2)): return tribug(P, C2)
17     if (interesting(P + C3)): return tribug(P, C3)
18
19     smallest = union(
20             tribug(P + C1, C2 + C3),
21             tribug(P + C2, C1 + C3),
22             tribug(P + C3, C1 + C2)
23             )
24
25     return smallest
26
```

Consider an interesting(x) function that returns True if x contains the elements 3 and 5. How many calls to interesting are made by tribug on each of the following inputs?

(b-i) (2 points) P = [], C = [3, 0, 5]

> Your answer here.

ANSWER: 8

(b-ii) (2 points) P = [], C = [5, 4, 3]

> Your answer here.

ANSWER: 8

(b-iii) (2 points) P = [], C = [5, 2, 8, 3]

> Your answer here.

You have decided to test out the Tribug algorithm from 2b) by running it on an input and adding print statements to see the values of P, C, and which subset of C is interesting at each iteration. Adding your print statements has created the code in the box below.

```
1
2  def split(list):
3      third = math.floor(len(list) / 3)
4      return list[:third], list[third:third*2], list[third*2:]
5
6  def simple_tribug(P, C):
7      for element in C:
8          if interesting(P + element): return element
9      return C
10
11 def tribug(P, C):
12     print("========= Top of Tribug =========") // PRINT
13     print("P is:" + P + " and C is: " + C) // PRINT
14
15     if len(C) < 3: return simple_tribug(P, C)
16
17     C1, C2, C3 = split(C)
18
19     if (interesting(P + C1)):
20         print("P U C1 is interesting") // PRINT
21         return tribug(P, C1)
22     if (interesting(P + C2)):
23         print("P U C2 is interesting") // PRINT
24         return tribug(P, C2)
25     if (interesting(P + C3)):
26         print("P U C3 is interesting") // PRINT
27         return tribug(P, C3)
28
29     smallest = union(
30             tribug(P + C1, C2 + C3),
31             tribug(P + C2, C1 + C3),
32             tribug(P + C3, C1 + C2)
33             )
34
35     return smallest
36
37 def main():
38     result = tribug([], [7, 1, 3, 0])
39
40     print("========= Main =========")
41     print("result is: " + result )
42
43
```

Below is the resulting log from running the main method on the input [7, 1, 3, 0]. The `interesting(x)` function here returns true if x contains the element 3.

```
1
2  =============== Top of Tribug ===============
3  P is [] and C is [7, 1, 3, 0]
4  _____2ci_____
5
6  =============== Top of Tribug ===============
7  P is ___2cii___ and C is ___2cii___
8  _____2ciii_____
9
10 =============== Main ===============
11 The result is: [3]
12
13
```

Given the code and the log above, answer the following questions. Note that some of the questions may not be specific to the code/log above and may test your understanding of the Delta Debugging algorithm in a general context.

(c-i) (1 points) Which of the followings should be filled in 2ci? (line 4 of the log output)

○ A) P U C1 is interesting
○ B) P U C2 is interesting
○ C) P U C3 is interesting
○ D) None (none of the above print statements are printed)

ANSWER: C

(c-ii) (1 points) Which of the followings should be filled in two blanks of 2cii? (line 7 of the log output)

○ A) [], [3, 0]
○ B) [3], [0]
○ C) [7, 1], [3, 0]

ANSWER: A

(c-iii) (1 points) Which of the followings should be filled in 2ciii? (line 8 of the log output)

○ A) P U C1 is interesting
○ B) P U C2 is interesting
○ C) P U C3 is interesting
○ D) None (none of the above print statements are printed)

ANSWER: D

(c-iv) (1 points) Unlike Delta Debugging, Tribug Debugging does not require that the Interesting(x) function be consistent on inputs x.

○ True
○ False

ANSWER: False

(c-v) (1 points) In the Delta Debugging algorithm as shown in the lecture, interference occurs when there is an error in the running of the Interesting(x) function, which interferes with the results.

○ True
○ False

ANSWER: False

Question 3. Short Answer (4 points each, 24 points)

(a) (4 points)

Your friend implemented a new automated program repair tool, GenMutantProg. In particular, GenMutantProg generates candidate higher-order mutants — from the original, buggy program using a new set of mutation operators designed by your friend — in hope that at least one of the candidate mutants will fix the bug from the original program and pass the test suite.

Explain what **"higher-order"** mutants mean in one sentence? Please also list **one advantage** and **one disadvantage** of generating higher-order mutants, compared to generating any mutants in general. Please use at most two sentences for the advantage and at most two sentences for the disadvantage.

```
Your answer here.
```

ANSWER:

Answer:

- Explanation of what high-order mutants mean. (1pt): High-order mutants change multiple operators at the same time from the original program.
- Advantage of high-order mutants in this case. (1.5pt): high-order mutants are more likely to fix the bug in the original program
- Disadvantage of high-order mutants in this case. (1.5pt): Even though high-order mutants might fix the bug in the original program, it is likely that at the same time one (or more) of the mutants introduces more bugs.

(b) (4 points)

Your team maintains a codebase and your manager proposes to apply readability metrics to each pull request with the goal of making sure the codebase contains highly readable code. Specifically, the plan is to calculate the readability score for the code in each pull request, and only those pull requests with a high readability score can be merged into the codebase.

Is this a good plan? Would this plan go wrong in any way? Please list **one possibility that this plan could go wrong**. Use at most two sentences in your answer.

[Your answer here.]

ANSWER:

Answer: (non-exhaustive, but typically it should mention it's not a good plan):

- It can lead to perverse incentives: It may be true that existing code with a few more blank lines is more readable. So what if we just insert a blank line between every line of code?

(c) (4 points)

Suppose you maintain a very large, currently closed-source project that is very popular among many users. At some point, you received too many feature requests and bug reports that you could not handle all by yourself alone. Now, you want to turn this project into an open-source one, with the goal of using contributors online to resolve issues (including implementing new features and/or fixing bugs).

However, one concern you have is that "random contributors" with different backgrounds may end up creating low-quality code (such as code that does not function properly, code that functions but has low readability, etc.). You don't want to allow online contributors to add such low-quality code into your project's codebase.

What is one approach that you can use to make sure no (or very little) low-quality code can be inserted into the codebase? Limit your answer to at most two sentences.

[Your answer here.]

ANSWER:

Answer: (non-exhaustive):

- Code review
- Automated unit tests/regression tests/style checks

(d) (4 points)

Imagine you are working on the autopilot feature for an airplane manufacturer. Once implemented, this feature will be deployed on all airplanes.

Please list **one informal quality requirement** and **one verifiable quality requirement** for this task.

[Your answer here.]

ANSWER:

Answer: (non-exhaustive):

- Informal quality requirements: The feature should run reliably/accurately/efficiently
- Verifiable quality requirements: The answer that receives full credits will include verifiable statistics (typically a number). For example: in 1000 runs, it should correctly identify the optimal path in 999 runs. Another example: it should pass all 70 unit tests.

(e) (4 points)

WebRobot and FlashFill are both program synthesis techniques that generate programs from a domain-specific language (or DSL). **What is a DSL and how is it different from Python?** Use at most two sentences in your answer.

Also, how is WebRobot different from FlashFill in terms of the underlying synthesis approach? Please use at most two sentences for synthesis approach.

Your answer here.

ANSWER:

Answer:

- DSL vs Python (2pt):
  - DSL is specialized to a particular application domain whereas the applicability of Python is not constrained to a certain domain.
- Specification (2pt)
  - WebRobot is based on specification of execution traces, which makes it PBD (Programming-by-Demonstration).
  - FlashFill is based on specification of input-output (I/O) examples, which makes it PBE (Programming-by-example).

(f) (4 points)

You have developed a tool, ErrorFixer481, that automatically fixes "Last-Mile Errors". In particular, ErrorFixer481 is based on a neural approach.

From the guest lecture presented by Jose Cambronero, what are Last-Mile Errors? What is one disadvantage of using a neural approach? Please limit your answer to 4 sentences.

Your answer here.

ANSWER:

Answer

- Definition of last-mile errors: errors that require few edits to fix (2pts)
- One disadvantage of the neural approach (non-exhaustive)(2pts): need new data and retraining for new domain

---

Question 4. Design Patterns (20 points)

For each code snippet below...
- Which design pattern does this code employ? [2 pts]
- What is one potential pro of employing this design pattern? Please limit your answer to one sentence. [1 pt]
- What is one potential con of employing this design pattern? Please limit your answer to one sentence. [1 pt]

(ai) (2 points)

```cpp
1  void print_elements(Container<int>::traverse begin, Container<int>::traverse end) {
2      for (Container<int>::traverse trav = begin; trav != end; ++trav) {
3          cout << *trav << " ";
4      }
5      cout << endl;
6  }
7
8
9  int main() {
10     ///VECTOR
11     vector<int> my_vec {1, 2, 3, 4};
12     print_elements(my_vec.begin(), my_vec.end());
13
14     //MAP
15     map<int,int> my_map;
16     my_map.insert(pair<int, int>(1, 1));
17     my_map.insert(pair<int, int>(2, 2));
18     my_map.insert(pair<int, int>(3, 3));
19     print_elements(my_map.begin(), my_map.end());
```

```
20
21      //SET
22      set<int> my_set;
23      my_set.insert(1);
24      my_set.insert(2);
25      my_set.insert(3);
26      print_elements(my_set.begin(), my_set.end());
27  }
28
```

○ A) Adapter Pattern
○ B) Composite Pattern
○ C) Proxy Pattern
○ D) Named Constructor Pattern
○ E) Factory Pattern
○ F) Singleton Pattern
○ G) Iterator Pattern
○ H) Observer Pattern
○ I) Template Pattern
ANSWER: Iterator Pattern

(aii) (2 points) Pros/Cons

> Your answer here.

ANSWER: Design Pattern: Iterator Pattern Pro: The interface for traversing containers is uniform so we don't have to remember different traversal methods based on the data structure or container. Con: Having a lot of seemingly similar iterators can cause confusion as to which iterator object is associated with which data structure in the program. A correct answer will list a pro/con of the iterator pattern, not of iterators in general. For example, iterators reading being read only is a property of iterators, not a property of the iterator pattern. -1 for too brief or incorrect.

(bi) (2 points)
```
1  struct HashMap {
2      void Search() {}
3      void Delete() {}
4      void Insert() {}
5      ...
6      struct Vector {
7          int size() {}
8          void push_back() {}
9          void pop_back() {}
10         void resize() {}
11         ...
12         struct Array {
13             ...
14         };
15     };
16 };
17
```

○ A) Adapter Pattern
○ B) Composite Pattern
○ C) Proxy Pattern
○ D) Named Constructor Pattern
○ E) Factory Pattern
○ F) Singleton Pattern
○ G) Iterator Pattern
○ H) Observer Pattern
○ I) Template Pattern
ANSWER: Adapter Pattern

(bii) (2 points) Pros/Cons

> Your answer here.

(c) (3 points) Briefly explain 2 creational design patterns from lecture in your own words and give a brief example of one being better than the other. Please limit your answer to a maximum of 4 sentences (ex. 2 per design pattern)

> Your answer here.

ANSWER: Named Constructor Idiom: Here, the programmer makes the class's constructor private or protected, while making a public creation method that calls that constructor. This could be used to ensure that the object is always created with the new keyword. Factory Method Pattern: This pattern separates the code that creates these objects from the code that depends on its interface. This is usually done by creating factory classes that actually create the objects, so all the programmer has to do is select the correct factory class. If we are making a game and we want different enemies to spawn in each place based on difficulty, we can check the difficulty once to get the proper factory method rather than checking the difficulty every time we need to spawn an enemy. Singleton: The pattern helps ensure that there is only one instance of a class at runtime. Logging classes usually follow this pattern because its single instance is used repeatedly by many other classes. (1 pt for naming two creational design patterns, 1 pt for each correct explanation)

(d) (3 points) Suppose that a particular software engineering project spends 40% of its lifetime effort on implementation, 45% of its lifetime effort on design and documentation, and 15% of its lifetime effort on testing. You are considering a design that would (a) decrease the effort required for implementation by 20% (for example, if implementation previously took 10 hours, but that effort is increased by 35%, it would now take 13.5 hours); however, adopting this design would also (b) increase the effort required for design and documentation by 10% and (c) increase the effort required for testing by 10%. Assume the project originally required 100 hours to complete over its lifetime and calculate the new hours required by the project with your new proposed design. Should the new design be adopted? Explain your reasoning with no more than 3 sentences.

> Your answer here.

ANSWER: Old Design: 100 hours New Design: 98 hours New design should be adopted. Overall, the lifetime hours are only 2% less. Even though less time is spent on implementation, we are focusing more on designing, documenting, and testing which are crucial for code's longevity and maintainability. (1 pt for correct calculation, 1 pt for choosing correct design, 1 pt for reasonable explanation)

(e) (3 points) Identify one anti-pattern, and explain why it is an anti-pattern. Briefly identify a better pattern to use. Use at most 4 sentences.

> Your answer here.

ANSWER: Answers will vary. One example, duplicating code, bad for maintainability and comprehension, prone to new bugs. Instead use reusable functions (1 pt for anti pattern, 1 point for why, 1 point for better solution)

(f) (3 points) Suppose you and your partner are working on a large project together. Your partner encourages you not to add any inline comments, as they say it makes the code look messy. Do you think your project would benefit from inline comments? If you aren't able to have inline comments, what is another strategy you could use to make sure your code is accessible/maintainable for new team members? Briefly explain how this strategy would be helpful. Use at most 4 sentences.

> Your answer here.

ANSWER: The code may or may not benefit from inline comments, depending on complexity and how many there are, a good argument about this is worth 1 point. 1 point for identifying alternative documentation (a really thorough README, git commit messages/issues/PRs, wiki type articles) and 1 point for explaining how that would help a new team member.

Question 5. Automated Program Repair (20 points)

Consider each of the following three scenarios in (a), (b), and (c). In each scenario, there is a subject program that fails some of the test cases from a given test suite. For each scenario, you can assume the test suite is very comprehensive and has very high quality, such that any program that passes this test suite is deemed correct.

Now you plan to use **Automated Program Repair (APR)** in each scenario to fix the bug(s) in each subject program. In particular, the APR tool you will use is based on mutation (which is similar to GenProg in the lecture). That is, the tool mutates the original subject program to generate mutants and checks if any of these generated mutants are correct by running each mutant on the test suite to see if it passes the test suite.

We will consider each of the three scenarios below.

(a) In the first scenario, you have a function that takes as input a string and returns another string. The desired behavior of a correct implementation is to return the reverse of the input string. For example, if the input string is "eecs", the desired output should be "scee". The test suite contains many tests that are used to check if the output is the reverse of the input. However, some of the tests are failing due to some bug(s) in the function. Above is the buggy `reserve_string` function.

```
1  def reverse_string(input_string):
2      reversed_string = ""
3      for i in range(0, len(input_string)):
4          reversed_string += input_string[len(input_string) - i + 1]
5      return reversed_string
6
7
```

(a-i) (2 points) Can you explain why it is wrong? Also show one input on which the program returns a wrong output. You can choose any input you like. Use at most four sentences in your answer.

> Your answer here.

ANSWER: The issue with this code is the index used to access the characters of the input string in the loop. It will cause an IndexError to be raised when trying to access the last character of the string.

(a-ii) (4 points) Recall that the APR tool is based on mutation. In particular, in each mutation step, the tool can only change an existing operator/constant to a different one. For example, it can change the "+" operator to "−", or it can change a constant "0" to "1", but it cannot do both in one single mutation step. Also suppose the tool can generate all first-order mutants, but cannot generate any higher-order mutants.

Can this APR tool fix the bug from the `reverse_string` program? Please justify your answer using at most four sentences.

> Your answer here.

ANSWER: The APR can fix this bug by changing the '+' operator to '−' operator. Example: `reversed_string += input_string[len(input_string) - i + 1]` -> `reversed_string += input_string[len(input_string) - i - 1]`

(b) The following class is based on the **Singleton Design Pattern**, where only one single instance of the ValueList class can be created. This ValueList instance stores a list of values and has functions that allow adding new values and fetching all existing values.

The test suite has many tests that check the functionality of all the functions in this class, but a number of them are failing now.

```
1  class ValueList:
2      __instance = None
3
4      @staticmethod
5      def get_instance():
6          if ValueList.__instance is None:
7              return Singleton.__instance
8          ValueList()
9
10     def __init__(self):
11         ValueList.__instance = self
12
13     def add_value(self, value):
14         self.values.append(value)
```

```
15
16      def get_all_values(self):
17          return self.values
18
19
```

(b-i) (2 points) Can you explain why the ValueList class implementation is wrong? A hint is that it has **no more than three** bugs. Please find as many as you can and explain them all. For each bug, please use no more than two sentences in your explanation.

<div>
Your answer here.
</div>

ANSWER: Bug1: If the ValueList's instance does not exist, we can't return it directly without creating a new one.
Bug2: The variable self.values is not declare/initialize before using.
Bug3: The Singleton at line 7 should be ValueList.
Student only need to find 2 out of 3 bugs to get full points.

(b-ii) (1 points) Let's still consider the APR tool from the previous scenario (for the `reverse_string` program). Would that tool be able to fix the program in this scenario?

<div>
Your answer here.
</div>

ANSWER: No, we have at least two bugs in this program that require at least two mutation steps (higher-order mutant), but the previous tool can only do one step (first order-mutants).

(b-iii) (3 points) Suppose we have a slightly improved version of the APR tool. It is still based on mutation but in each mutation step, it can not only change one operator/constant to another, but also can add an operator/constant, add a new line of code, or delete an existing line. In other words, in one mutation step, it can generate more mutants than in the tool in the previous scenario. Note that it can still make only one single change in one single mutation step (but now this change can involve addition or deletion of code).
A second improvement is that, instead of only considering all first-order mutants as in the previous scenario, now it also generates mutants that need two mutation steps. For example, it can generate a mutant that requires first adding a new line and then changing another operator in the original program.
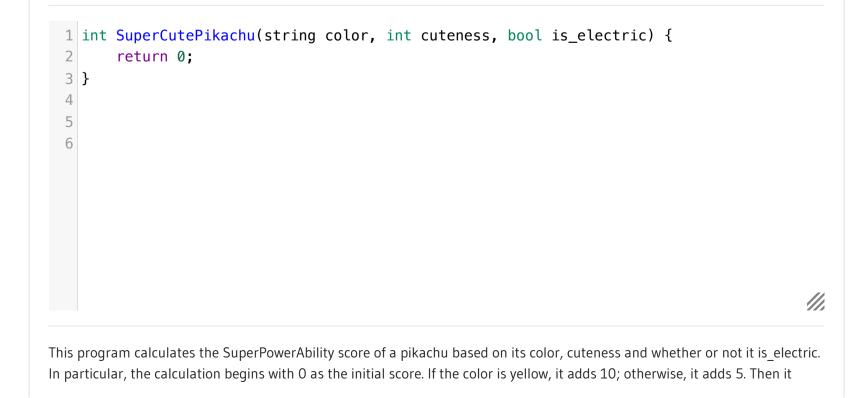Can this more powerful APR tool fix the program? Please justify your answer using no more than four sentences.

<div>
Your answer here.
</div>

ANSWER: Yes, because one can first change None to not None and then add the line self.values = [] that initializes the data field.
Depending on how many bugs the student found in (i), the answer may be 'yes' or 'no' since the improved APR tool can do at most two steps. We give points to both answer as long as it is fully justified.

(c) In this third scenario, consider the following program.

```
1  int SuperCutePikachu(string color, int cuteness, bool is_electric) {
2      return 0;
3  }
4
5
6
```

This program calculates the SuperPowerAbility score of a pikachu based on its color, cuteness and whether or not it is_electric. In particular, the calculation begins with 0 as the initial score. If the color is yellow, it adds 10; otherwise, it adds 5. Then it

multiplies the current score by cuteness. Finally, if the pikachu is_electric, it adds 20.

For example, ("yellow", 25, true) would result in a SuperPowerAbility of (10*25)+20 = 270 Points! That's really powerful! ("blue",15,false) would result in a SuperPowerAbility of (5*15) = 75 Points. That's not very powerful.

In the program above, the developer created a placeholder for the implementation that returns the initial score 0, but they forgot to update this program. As a result, almost all test cases are currently failing.

(c-i) (2 points) In this scenario, we consider an even more powerful APR tool than the tools in two scenarios above. This new APR tool can generate higher-order mutants and it **does not limit** the number of mutation steps — this is the main difference from previous scenarios.. In each mutation step, it can change an existing operator/constant, change an existing line, add a new line, or delete an existing line (which is the same as in the second scenario). This new APR tool is clearly more powerful. Can this new tool repair the `SuperCutePikachu` program **in theory**? That is, mutate the program to one that can pass all tests in the test suite.

> Your answer here.

ANSWER: Yes, the tool can continuously "repair" the missing code until it works.

(c-ii) (3 points) Please justify your answer in (c-i).

In particular, if your answer is "Yes", please explain why and list one potential issue that the APR tool might run into **in practice** ("in practice" here means if you actually deploy the tool and run it on the program to repair it).

If your answer is "No", please explain why and propose a potential extension to the APR tool such that it can **in theory** repair the program.

Limit your entire answer to at most 4 sentences.

> Your answer here.

ANSWER: We accepted any reasonable answer for why. Here are a few: A high cost to develop the tool. Expensive compute time to run the tool. The tool takes too long.

(c-iii) (3 points) In addition to using a program repair technique, what is another **automated** approach that you could potentially use to "repair" the `SuperCutePikachu` program above? Please briefly explain this approach and justify why it could work in this scenario. Also compare and contrast this approach with the automated program repair approach. Please use no more than 6 sentences in your response.

> Your answer here.

ANSWER: One way they are different is that Program Synthesis focuses on generating code from nothing, whereas program repair is focused on mutating pre-existing code. Although the line gets blurred in the middle, this distinction is clear at the extremes.

Another way they are different is the search space. Program repair seeks and tests mutants of the preexisting code. In that way, the search space is mutated lines of the code that already exists. But the search space for automated synthesis is much larger. We are searching the entire grammar of the programming language for the solution. In this question, we asked students to correctly identify another automated technique (Program Synthesis) and asked them to differentiate between program repair and program synthesis.

The reason program synthesis could work here is: since we have the comprehensive test suite, it can serve as the specification. We can also define a domain-specific language (or we can use Python) as the search space. We can use the top-down search technique to search in this space to find a program that passes the test suite.

Question 6. Extra Credit (1 point each)

(*Feedback*) What was your favorite topic or activity during the course?

> [text box]

(*Feedback*) What do you think we should do more of next semester (or what is the thing you would most recommend that we change for future semesters)?

> [text box]

minutes remaining

Hide Time

Manual Save

(*Guest Lecture*) List one thing you learned from guest speaker José Cambronero of Microsoft that was not listed on an introductory summary slide or otherwise convince us that you paid careful attention during that lecture.

(*Guest Lecture*) List one thing you learned from guest speaker Roscoe Bartlett of Sandia National Laboratories that was not listed on an introductory summary slide or otherwise convince us that you paid careful attention during that lecture.

(*Optional Reading 1*) Identify a single optional reading that was assigned after Exam 1. Write two sentences about it that convince us you read it critically. (The most common student mistakes for these questions in Exam 1 were choosing a required reading instead of an optional reading or failing to "identify" or name the reading selected.)

(*Optional Reading 2*) Identify **another** optional reading that was assigned after Exam 1. Write two sentences about it that convince us you read it critically.