Student ID: _____

Software Metrics and Models
Comprehensive Examination
Spring 2011

I. (10 points)
Indicate to the left of each statement whether the statement is True (T) or False (F).
.

_____1. A development team in crisis should postpone collecting metrics until the problems have been solved.

_____2. It has been reported that the value of the McCabe Cyclomatic Complexity Metric for a given software component correlates well with the number of defects discovered in the code.

_____3. It is an accepted fact that an organization (1) plan for at least a year before starting a metrics program and (2) not pressure anyone who resists reporting metrics.

_____4. Barry Boehm claims that finding and fixing a software problem after delivery of the software is 100 times more expensive than finding and fixing it during requirements and early design phases.

_____5. Boehm also claims that for every dollar spent on the development of a software system, 10 cents will be spent on maintenance.

_____6. We say that a statement involving measurement is meaningful if its truth value is invariant under transformations of allowable scales.

_____7. The median is an appropriate statistic to use in analyzing ordinal scale data.

_____8. The sample mean is an appropriate statistic to use in analyzing ratio scale data.

_____9. To determine a preliminary estimate of the size of a software product in lines of code, it is necessary to be able to envision at least one design approach, even if it is not the design selected later in the project.

_____10. The estimated number of lines of code in a software product is the only known basis for estimating development cost and schedule.

II. (10 points)
The following statements have omitted words or phrases which may be filled in from the list below. Certain words from the list may not be used at all; others may be used more than once. **FILL IN BY NUMBER ONLY.**

1. The total cost of software development is composed of the cost of performance of the specific work plus the Cost of (Poor) Quality that is composed of

   (1) _____: Unnecessary repetition of any process step due to error.

   (2) _____: Process steps needed to evaluate the product to determine if there are any errors.

   (3) _____: Process steps to prevent (or reduce) errors from getting into the product.

2. Identify the scale type that is best described by the following characteristics:

   (1) _____ Scale: Addition and subtraction are acceptable, but multiplication and division are not.

   (2) _____ Scale: The empirical relation system consists only of different classes for which there is no notion of ordering.

   (3) _____ Scale: The numbers can be ranked, but addition, subtraction and other arithmetic operations have no meaning.

   (4) _____ Scale: The measurement mapping must start at zero and increase at equal intervals, known as units.

3. In the statement: "x is taller than y", the phrase "taller than" is a(n) _____ for height.

4. A generally true measurement statement is that a metric should:

   (1) measure the _____ things, and

   (2) deliver more _____ than it costs.

Selection List

1. accuracy	10. detailed design	19. prevention
2. appraisal	11. empirical relation	20. processes
3. attributes	12. heuristics	21. products
4. code and unit test	13. interval	22. ratio
5. cohesion	14. measure	23. resources
6. consistent	15. measurement	24. rework
7. correctness	16. nominal	25. right
8. costs	17. ordinal	26. value
9. coupling	18. precision	27. yardstick

III. (25 points)

Consider the following Java code:

```
public static void selectionsort(int[ ] data, int first, int n)
{
   int i, j;                  // Loop control variables
   int big;                   // Index of largest value in data[first]...data[first+i]
   int temp;                  // Used during the swapping of two array values
   i = n – 1;
   while( i > 0)
   {
      // Calculate big as the index of the largest value in data[first]...data[first+i]:
      big = first;
      j = first+1;
      while( j <= first + i )
          {   if (data[big] < data[j])
                    big = j;
              j++;
          };
      // swap data[first+i] with data[big]:
      temp = data[first+i];
      data[first+i] = data[big];
      data[big] = temp;

      i--;
   }
}
```

On the following page:

III.A. Create a flowgraph of this selection sort (15 points)

III.B. Using any technique that you have learned, calculate the value of the McCabe Cyclomatic Complexity Metric for this algorithm? (5 points)

III.C. For this algorithm, express the McCabe Metric in terms of the number of nodes and the number of edges of the flowgraph. (5 points)

III. continued

IV. (15 points)

For the selection sort in problem III, count the number of lines of code under the following conditions.

1. Count the number of physical lines (including blank lines). (3 points)

2. Count the number of physical lines except for blank lines and comments. (3 points)

3. Count the number of executable statement. (3 points)

4. In using line of code counts as a basis for prediction development cost and schedule, which of the above three methods would you use. Explain your choice. (6 points)

V. (5 points)

Identify the following variables as N for Nominal, O for Ordinal, I for Interval or R for ratio scale.

1. _____: Categorization of software components by Programming Language
2. _____: Identification of defect impact as Low, Medium or High
3. _____: Labor hours charged to system test phase
4. _____: The number of calendar months spent in design
5. _____: Number of defects found during an inspection

VI. (10 points)

The COCOMO Cost Estimation Model is based on the following mathematical model: $y = ax^b$

1. What does x represent, i.e., what is the input to the model?

2. What does y represent, i.e., what is the output of the model?

3. The parameters a and b were originally calculated by a regression analysis performed on which one of the following pairs?
   i. x and y
   ii. x and log(y)
   iii. log(x) and y
   iv. log(x) and log(y)

4. Should a somewhat immature company begin immediately using COCOMO or a similar model to estimate cost and schedule or should the company wait until it is more mature? Answer yes or no and justify your answer.

VII. (10 points)

Suppose a post delivery analysis of a recent project shows that the origins of the 125 defects identified by reviews and testing are as follows:

Defect Origin	Requirements Specification Phase	Top-Level Design Phase	Detailed Design Phase	Coding Phase	Total
Total	19	13	31	62	125

Produce a Pareto chart of the number of defects found by lifecycle phase (ignoring total defects column). What information is conveyed by the chart?

VIII. (15 points)

Suppose you own a small software company whose product, a patient information system, has been sold to several hundred physicians' groups. Although the fielded software is very buggy, your customers are reasonably happy because your developers give them very quick turn-around in fixing the defects as they are reported. Your problem is that the developers are supposed to be working on the next generation of the product, but are making no progress because of the daily customer support that they have to do. You decide to tackle the problem using the Goal-Question-Metric approach. Your goal is "to devote at least 60% of the development staff to the next generation of the product, while continuing to provide the same level of support to our current customers." Identify 3 questions that must be answered to know if progress is being made toward your goal. Finally, identify one or more metrics per question that will supply the necessary information to answer the question. Note: this problem does not have a unique answer. Use the next page if you need more space.

VIII. continued