SYMPOSIUM ON LEARNING ANALYTICS AT MICHIGAN

The Programming Performance Prophecies

Predicting Student Achievement in a First-Year Introductory Programming Course

Jeffrey Ringenberg
Marcial Lapp and T. Jeff Fleszar

SLAM
22 February 2012

MichiganEngineering

Engineering at U of M

All incoming engineering students enroll in a programming course during year 1.

ENGR 101

Matlab / C++ programming for everyone.

ENGR 151

Advanced students with an adeptness for programming.

Motivation for Prediction Exam

- Introductory programming courses at many universities typically include students with a wide variety of computing backgrounds.

- If these groups of students can be identified, students can be separated.

Previous Work

- Common factors that have been studied.
  - High school GPA, ACT/SAT scores, personality traits, mathematical aptitude, prior exposure to specific courses, etc…
  - Some correlations have been found to exist
  - Data can be difficult/work intensive to obtain.

Our Approach

- Measure students' ability through an online examination.
  - Low overhead
  - The exam is given before the students have received any formal instruction.
    - e.g. all incoming students take the exam during summer orientation.
- Questions contain combinations of algorithmic, math, and logic-based questions.

Testing Approach

Fall 2009 - Initial release of exam.
Class sample size: 350 students.

Stage I: Analysis and Question Refinement

Fall 2010 – Exam Administration
Class sample size: 450 students

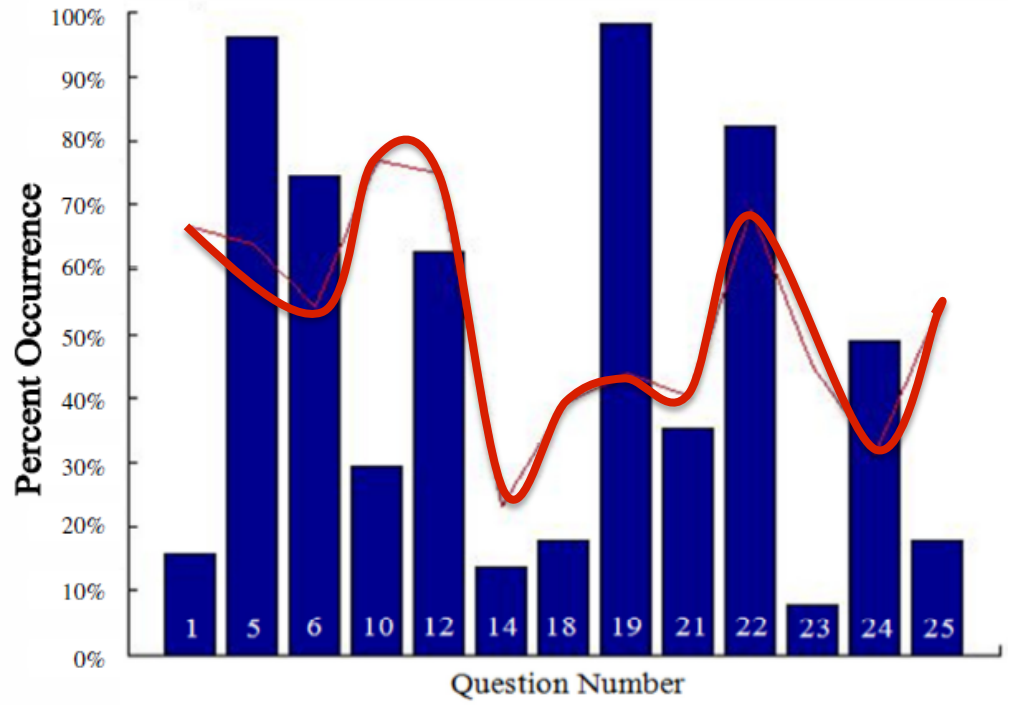
Stage II: Analysis

Correlation Analysis

- All questions that over 80% of the students answered correctly were ignored.
  - These questions were considered *ineffective.*
- Question sets, containing 7 questions each.
  - Each set of 7 questions was then correlated to overall course grades.
  - Count the number of times that each *effective* question appeared in an *effective* set was calculated as a percentage.

Correlation Analysis (2)

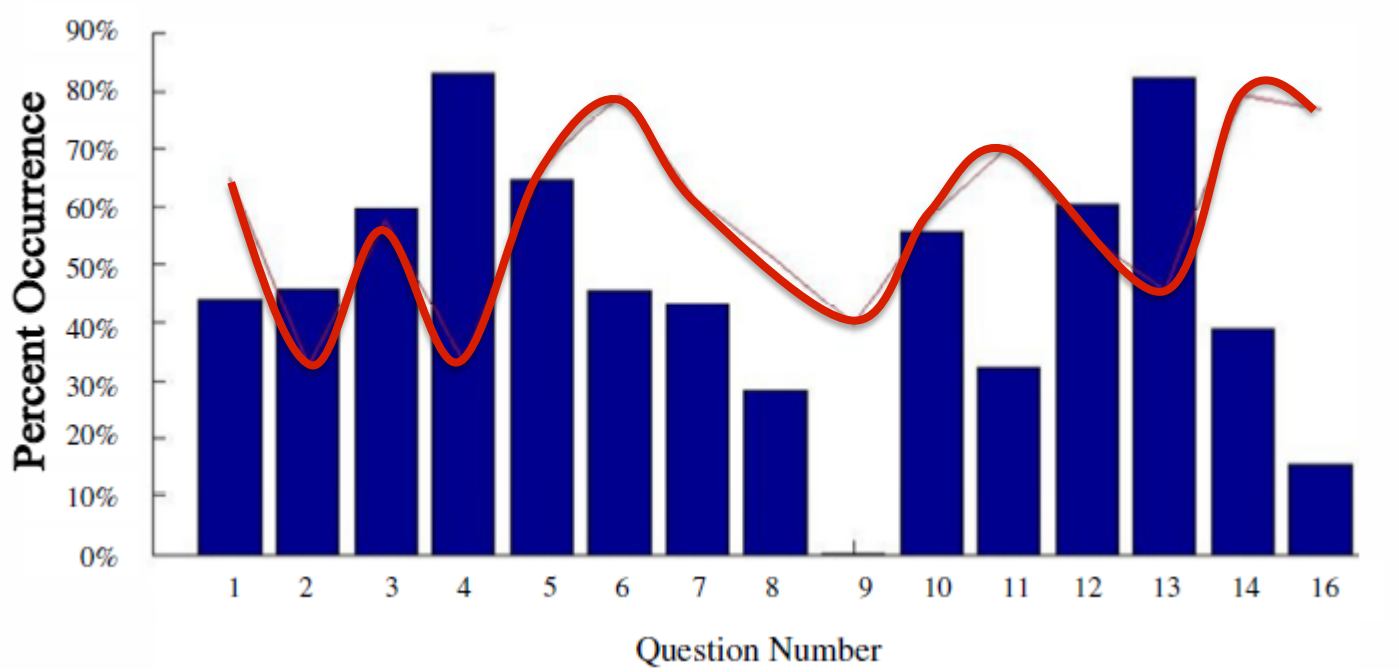
- High bar indicates that a question appeared in many of the predictive/effective sets.



Fall 2009 Data Analysis

Correlation Analysis (3)

- Fall 2010 – Correlation Results



Data Trending

- Correlation of the predictive exam to the overall course grade *increased* from 0.33 to 0.39.

Exam 1	0.38
Exam 2	0.40
Exam 3	0.37
Exam 4	0.35
Overall Course Grade	0.33

Fall 2009
Regular Course (ENGR 101)

Exam 1	0.36
Exam 2	0.42
Exam 3	0.36
Overall Course Grade	0.39

Fall 2010
Regular Course (ENGR 101)

```
* Import the Excel file into SPSS .

GET DATA /TYPE=XLSX
  /FILE='\\afs\umich.edu\user\j\r\jringenb\Download\F10_ROC_Data.xlsx'
  /SHEET=name 'ROC_DATA'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.

SAVE OUTFILE='\\afs\umich.edu\user\j\r\jringenb\Download\F10_SPSS_savefile.sav
'
  /COMPRESSED.

* Run ROC analyses .

DATASET ACTIVATE DataSet1.
ROC ROC_Survey BY ROC_State (1)
  /PLOT=CURVE(REFERENCE)
  /PRINT=SE COORDINATES
  /CRITERIA=CUTOFF(INCLUDE) TESTPOS(LARGE) DISTRIBUTION(FREE) CI(95)
  /MISSING=EXCLUDE.
```

ROC Curve

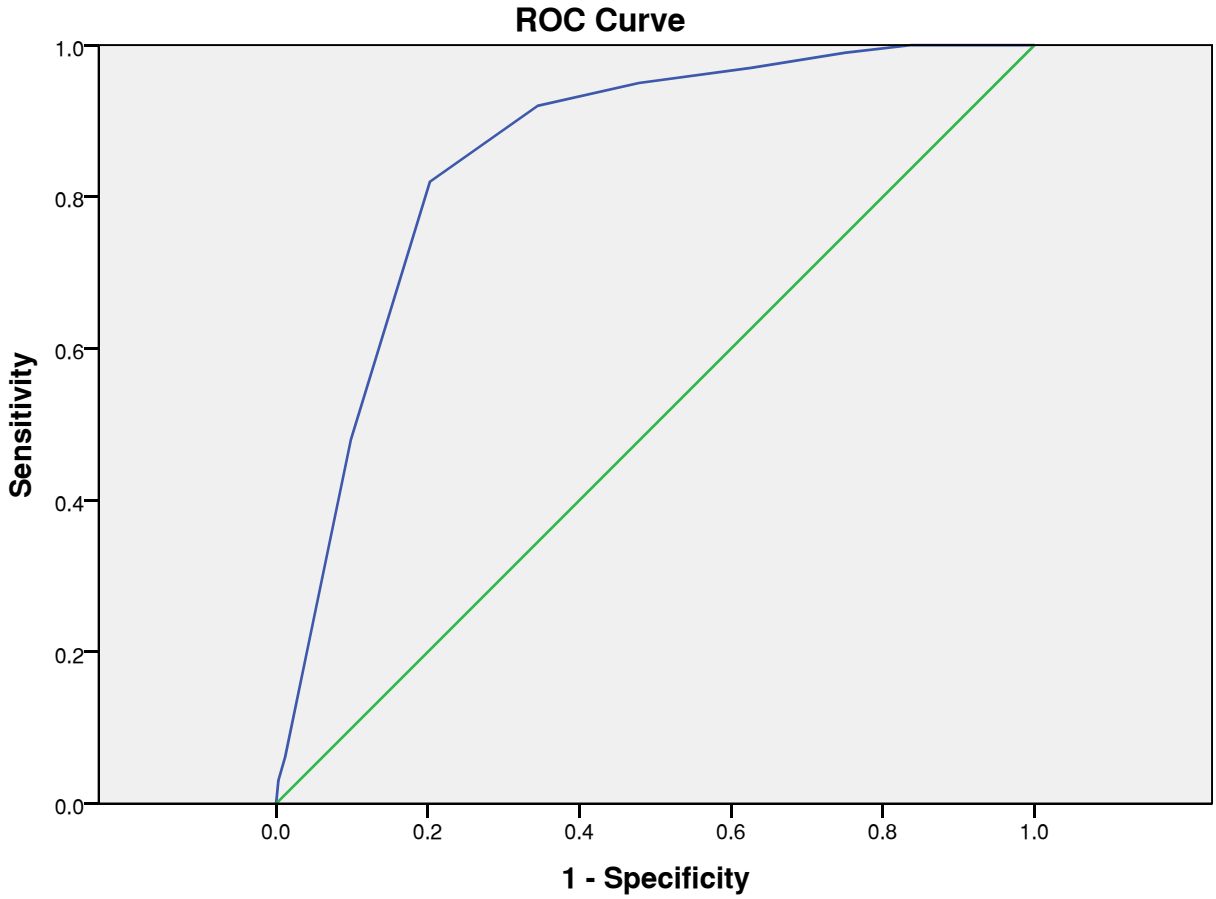
```
[DataSet1] \\afs\umich.edu\user\j\r\jringenb\Download\F10_SPSS_savefile.sav
```

Case Processing Summary

ROC_State	Valid N (listwise)
Positive[a]	100
Negative	345

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s):ROC_Survey

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.853	.020	.000	.814	.891

The test result variable(s): ROC_Survey has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):ROC_Survey

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.00	1.000	1.000
1.50	1.000	.997
2.50	1.000	.991
3.50	1.000	.977
4.50	1.000	.957
5.50	1.000	.916
6.50	1.000	.838
7.50	.990	.751
8.50	.970	.626
9.50	.950	.478
10.50	.920	.345
11.50	.820	.203
12.50	.480	.099
13.50	.200	.041
14.50	.060	.012
15.50	.030	.003
17.00	.000	.000

The test result variable(s): ROC_Survey has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

```
* Import the Excel file into SPSS .

GET DATA /TYPE=XLSX
  /FILE='\\afs\umich.edu\user\j\r\jringenb\Download\F11_Results.xlsx'
  /SHEET=name 'Summary'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.

SAVE OUTFILE='\\afs\umich.edu\user\j\r\jringenb\Download\F11_SPSS_savefile.sav
'
  /COMPRESSED.

* Run ROC analyses .

DATASET ACTIVATE DataSet1.
ROC Survey BY State (1)
  /PLOT=CURVE(REFERENCE)
  /PRINT=SE COORDINATES
  /CRITERIA=CUTOFF(INCLUDE) TESTPOS(LARGE) DISTRIBUTION(FREE) CI(95)
  /MISSING=EXCLUDE.
```

ROC Curve

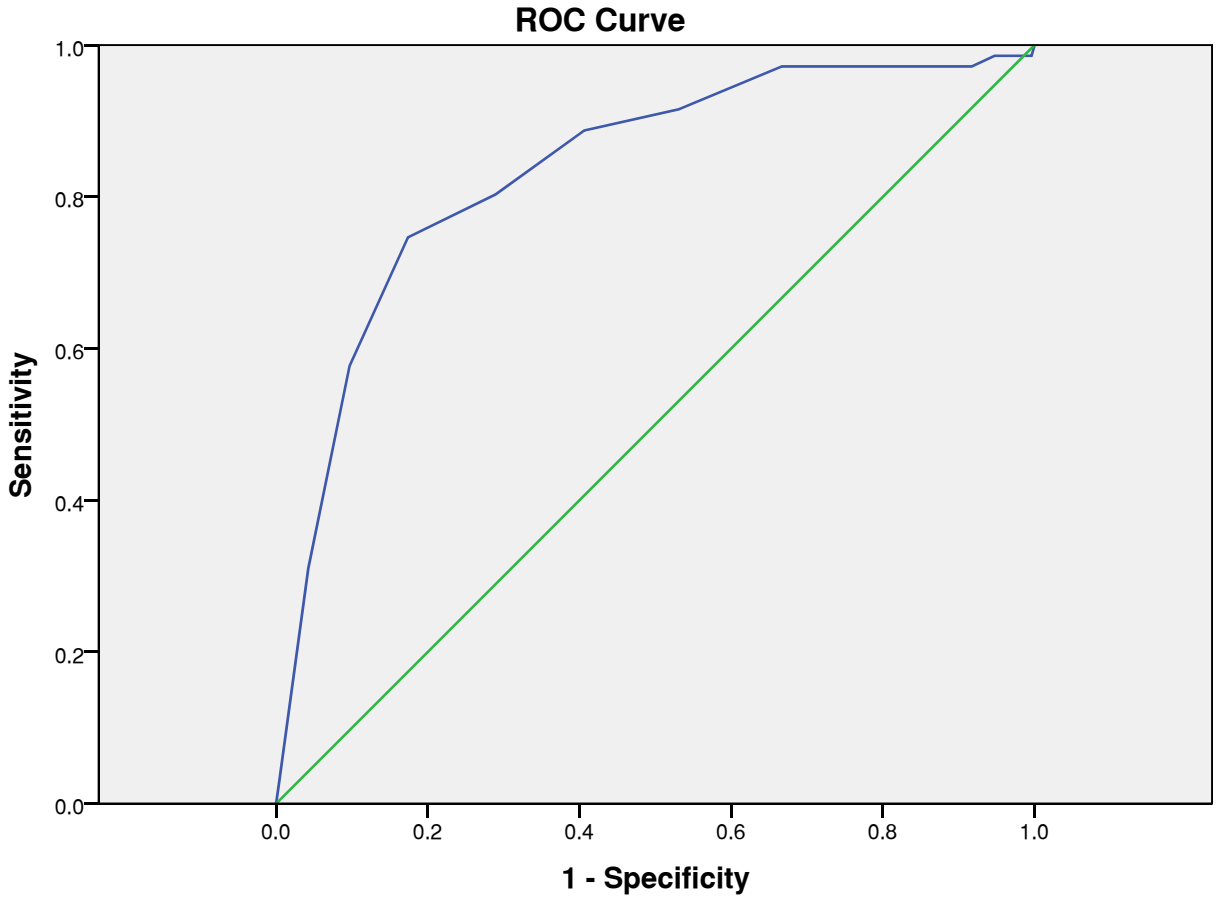
```
[DataSet1] \\afs\umich.edu\user\j\r\jringenb\Download\F11_SPSS_savefile.sav
```

Case Processing Summary

State	Valid N (listwise)
Positive[a]	71
Negative	495
Missing	154

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s):Survey

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.835	.026	.000	.783	.887

The test result variable(s): Survey has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Survey

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.00	1.000	1.000
1.00	.986	.996
2.50	.986	.990
3.50	.986	.978
4.50	.986	.947
5.50	.972	.917
6.50	.972	.840
7.50	.972	.764
8.50	.972	.667
9.50	.915	.531
10.50	.887	.406
11.50	.803	.289
12.50	.746	.174
13.50	.577	.097
14.50	.310	.042
15.50	.028	.004
17.00	.000	.000

The test result variable(s): Survey has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

		Actual Data			Projected Data						No Survey
		Survey < 11	11 <= Survey < 14	Survey >= 14	Count 1/7*	Count 1/6*	Count 1/5*	Count 1/4*	Count 1/3*	Count 1/2*	
Fall 2010**	ENGR 101	234	113	17	387	384	378	370	357	331	77
	ENGR 151	17	46	18	58	61	67	75	88	114	5
Fall 2010 Scaled	ENGR 101	312	151	23	***498***	***493***	***487***	477	460	428	103
	ENGR 151	17	46	18	***69***	***74***	***80***	90	107	139	7
Fall 2011	ENGR 101	294	156	52	***452***	***448***	***442***	433	419	390	111
	ENGR 151	8	19	37	***114***	***118***	***124***	133	147	176	12

* "Final Count 1/N" assumes the following students took 151: all students with (Survey >= 14) and only 1/N of students with (11 <= Survey < 14)

** Fall 2010 Data only includes results from 2 of the 3 sections of 101, so its initial counts will be lower by ~30% (see scaled data)

Note: Students that didn't take the survey are not counted in the course totals (see additional column)

	Median		Mean		STDEV	
	Fall 2010	Fall 2011	Fall 2010	Fall 2011	Fall 2010	Fall 2011
ENGR 101	9	10	9.33	9.7	2.67	2.99
ENGR 151	12	14	11.8	13	2.52	2.63

Appendix B - Questions Used for the Fall 2010 Predictive Exam

Question Number	Question Text	Possible Answers	Correct Answer
1	What does this list of instructions do? 1. Request a value for A 2. Request a value for B 3. Request a value for C 4. Assign the value of the equation (BxB-4xAxC) to D 5. Return D	a) It returns the discriminant for a quadratic equation Ax^2+Bx+C=0 b) It returns the solution to a quadratic equation Ax^2+Bx+C=0 c) It returns the equation Ax^2+Bx+C d) None of the above	a
2	What does this list of instructions do (assuming that the initial value for A is a positive integer)? 1. Request a value for A that is a multiple of 3 2. Assign A plus 1 to A 3. While A is not equal to 0, assign A minus 3 to A 4. Return A	a) It returns a multiple of A plus 1 b) It returns the smallest divisor of A c) It returns the value 0 d) It returns the value 1 e) The instructions will run forever	e
3	What does this list of instructions do (assuming that the initial values for A and B are positive integers and that instruction 5 will eventually execute)? 1. Request a value for A 2. Request a value for B 3. While B is greater than A, request a new value for B 4. Assign B divided by A to C 5. Return C	a) It returns a value less than or equal to 1 b) It returns 0 c) It returns a value greater than or equal to 1 d) It returns 1 e) None of the above	a
4	What does this list of instructions do (assuming that the initial values for X and Y are positive integers)? 1. Request a value for X 2. Request a value for Y 3. While X is greater than or equal to Y, assign X minus Y to X 4. Return X	a) It returns the value X divided by Y b) It returns the integer remainder of the initial value of X divided by the initial value of Y c) It returns the integer remainder of the initial value of Y divided by the initial value of X	b

		d) It returns the value of the initial value of Y minus the initial value of X e) The instructions will run forever	
5	What does this list of instructions do (assuming that the initial values for A and B are positive integers)? 1. Request a value for A 2. Request a value for B 3. If A is less than or equal to B, return B multiplied by 3 4. Otherwise, return A multiplied by 3	a) It returns the maximum of A and B b) It returns the minimum of A and B c) It returns the maximum of A multiplied by 3 and B multiplied by 3 d) It returns the minimum of A multiplied by 3 and B multiplied by 3	c
6	What does this list of instructions do? 1. Request a value for A 2. Request a value for B 3. Assign the value (AxA) to C 4. Assign the value (C+BxB) to C 5. Assign the square root of C to C 6. Return C	a) It returns the value (C+BxB) b) It returns the value (AxA+BxB) c) It returns the square root of (BxB) d) It returns the square root of (AxA+BxB) e) None of the above	d
7	A man decides to buy a horse. He pays 60 dollars for the animal. After a year, the value of the horse has increased to 70 dollars and he decides to sell the horse. A few days later he regrets his decision to sell the horse, and he buys it again. Unfortunately, he has to pay 80 dollars to get it back, thereby paying 10 dollars more for the horse that he just sold! After another year of owning the horse, he finally decides to sell it for 90 dollars. What is the overall profit the man makes? (Please enter only whole numbers and no units or signs such as $$ or dollars.)	Free Response	20
8	A cube's exterior surfaces are painted with 3 different colors in such a way that the opposite sides of the cube are painted in the	a) 8 b) 16 c) 24	d

	same color. The cube is then cut into 64 smaller cubes of equal size. How many cubes have at least two of their sides painted in different colors?	d) 32 e) 56	
9	How many squares are there on a normal tic-tac-toe board assuming the board consists of 3 rows and 3 columns? (Please enter only whole numbers and no units such as squares.)	Free Response	14
10	Three people picked 65 apples altogether. At the first tree they each picked the same number of apples. At the second tree they each picked 3 times as many as they picked at the first tree. When they finished at the third tree, the group had 5 times as many apples as they had when they started at that tree. At the fourth tree the group picked just 5 apples. How many apples did each person pick at the first tree? (Please enter only whole numbers and no units such as apples.)	Free Response	1
11	At a party, everyone shook hands with everybody else. There were 66 handshakes. How many people were at the party? Hint: Consider how many handshakes occur with two people, three people, etc?	a) 8 b) 10 c) 12 d) 14	c
12 Question 6 from Fall '09	If $0 < s*t < 1$, then which of the following can be true?	a) $s < -1$ and $t > 0$ b) $s < -1$ and $t < -1$ c) $s > 1$ and $t < -1$ d) $s > -1$ and $t < -1$ e) $s > 1$ and $t > 1$	d
13 Based on Question 19 from Fall '09	Assuming a base 2 numbering system, 00110010 + 10001010 = ?	a) 10111010 b) 10111020 c) 00111200 d) 10011010 e) 10111100	e
14 Based on Question 5 from Fall '09	To make money as a painter, an artist must first charge X dollars for an original painting. The artist then charges (2X)/5 dollars for each of the first 10 copies of the original painting and finally (X/10) dollars for each additional copy. If the artist made $280 by selling the original painting along with 30 copies, what was the cost of the original painting?	a) $30 b) $35 c) $40 d) $45 e) $50	c

15 Question 22 from Fall '09	Assuming that only one of the four statements below is true, which one must it be? Statement 1: The number of false statements here is one Statement 2: The number of false statements here is two Statement 3: The number of false statements here is three Statement 4: The number of false statements here is four	a) Statement 4 b) Statement 3 c) Statement 2 d) Statement 1 e) None of the statements is true	b
16 Question 12 from Fall '09	A ranch has both horses and ponies. Exactly 5/6 of the ponies have horseshoes and exactly half of the ponies with horseshoes are Icelandic ponies. What is the minimum number of ponies on the ranch?	a) 6 b) 8 c) 10 d) 12 e) 18	d