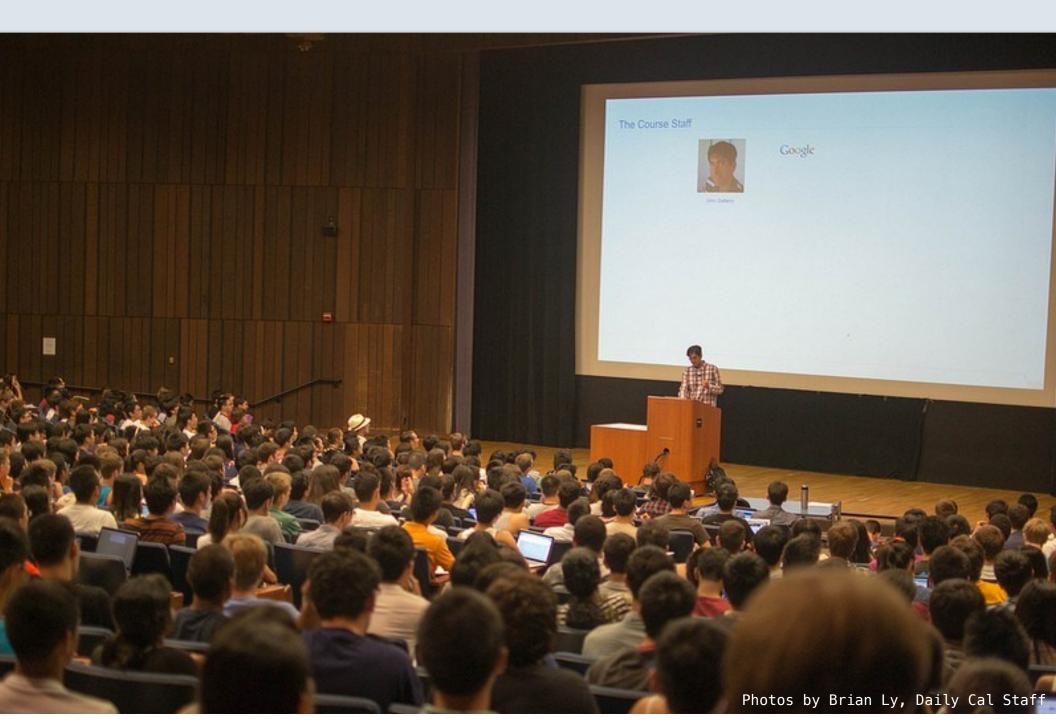
# Teaching Composition Quality at Scale Human Judgment in the Age of Autograders

#### John DeNero & Stephen Martinis

Google & Berkeley



#### Computer Science Courses are Growing



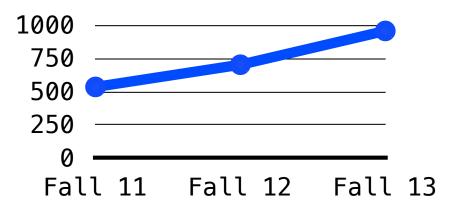
#### Computer Science Courses are Growing



#### Computer Science Courses are Growing



End-of-semester enrollment
in CS 61A (CS1 @ Berkeley)





"Programs must be written for people to read, and only incidentally for machines to execute." - Structure and Interpretation of Computer Programs

• Guidelines for effective program composition

- Guidelines for effective program composition
- Useful examples

- Guidelines for effective program composition
- Useful examples
- Incentives

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

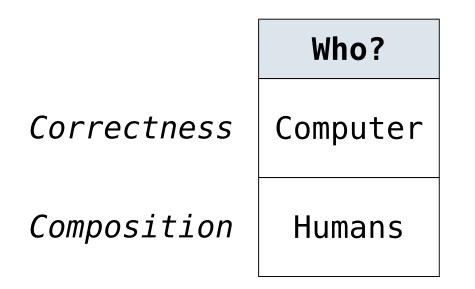
"Programs must be written for people to read, and only incidentally for machines to execute." - Structure and Interpretation of Computer Programs

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

Correctness

Composition

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback



- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

	Who?	How?
Correctness	Computer	Script
Composition	Humans	Code Reviews

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

	Who?	How?	Why?
Correctness	Computer	Script	78/90
Composition	Humans	Code Reviews	12/90

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

	Who?	How?	Why?	When?
Correctness	Computer	Script	78/90	Instant
Composition	Humans	Code Reviews	12/90	~2 weeks

"Programs must be written for people to read, and only incidentally for machines to execute." - Structure and Interpretation of Computer Programs

1

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

	Who?	How?	Why?	When?
Correctness	Computer	Script	78/90	Instant
Composition	Humans	Code Reviews	12/90	~2 weeks

"Programs must be written for people to read, and only incidentally for machines to execute." - Structure and Interpretation of Computer Programs

1

2

- Guidelines for effective program composition
- Useful examples
- Incentives
- Feedback

	Who?	How?	Why?	When?
Correctness	Computer	Script	78/90	Instant
Composition	Humans	Code Reviews	12/90	~2 weeks

# Guidelines & Useful Examples

• Demonstrations of programming language features often lack enough context for names to matter.

- Demonstrations of programming language features often lack enough context for names to matter.
  - My first non-trivial lecture example: mul(add(2, mul(4, 6)), add(3, 5))

- Demonstrations of programming language features often lack enough context for names to matter.
  - My first non-trivial lecture example: mul(add(2, mul(4, 6)), add(3, 5))
  - Placeholders: foo, bar, baz, my\_x, helper, etc.

- Demonstrations of programming language features often lack enough context for names to matter.
  - My first non-trivial lecture example: mul(add(2, mul(4, 6)), add(3, 5))
  - Placeholders: foo, bar, baz, my\_x, helper, etc.
- Some students claim that a program composed of many small functions is hard for them to follow.

- Demonstrations of programming language features often lack enough context for names to matter.
  - My first non-trivial lecture example: mul(add(2, mul(4, 6)), add(3, 5))
  - Placeholders: foo, bar, baz, my\_x, helper, etc.
- Some students claim that a program composed of many small functions is hard for them to follow.
  - Tracing execution is easier without abstraction.

- Demonstrations of programming language features often lack enough context for names to matter.
  - My first non-trivial lecture example: mul(add(2, mul(4, 6)), add(3, 5))
  - Placeholders: foo, bar, baz, my\_x, helper, etc.
- Some students claim that a program composed of many small functions is hard for them to follow.
  - Tracing execution is easier without abstraction.
- Many composition issues only arise in programs of sufficient complexity.

- Demonstrations of programming language features often lack enough context for names to matter.
  - My first non-trivial lecture example: mul(add(2, mul(4, 6)), add(3, 5))
  - Placeholders: foo, bar, baz, my\_x, helper, etc.
- Some students claim that a program composed of many small functions is hard for them to follow.
  - Tracing execution is easier without abstraction.
- Many composition issues only arise in programs of sufficient complexity.
- Our solution: Emphasize composition in projects.

# Focus on Human Judgment

#### Focus on Human Judgment

Focus on Human Judgment

Composition is about communication among people.

Code Style Rules of Thumb

Effective Communication

**Rules of Thumb** Code Style sum\_of(x, y) A function body Why is each SumOf(x,y) most k lines. introduced?

vs should have at variable name

#### Effective Communication

Code Style	Rules of Thumb	Effective Communication
<pre>sum_of(x, y)</pre>	A function body should have at most k lines.	Why is each variable name introduced?
Scriptable		Humans required

Code Style	Rules of Thumb	Effective Communication
<pre>sum_of(x, y)</pre>	A function body should have at most k lines.	Why is each variable name introduced?
Scriptable		Humans required
<b>Objective</b>		Subjective

Code Style	Rules of Thumb	Effective Communication
<pre>sum_of(x, y)</pre>	A function body should have at most k lines.	Why is each variable name introduced?
Scriptable		Humans required
<b>Objective</b>		Subjective
Distracting		Essential

#### Composition Guidelines Given to Students

#### Composition Guidelines Given to Students

Names Choose names that evoke the purpose or meaning of the values to which they refer.

#### Composition Guidelines Given to Students

- Names Choose names that evoke the purpose or meaning of the values to which they refer.
- **Functions** Define functions with a clear role in the program that abstract the implementation of a behavior.

### Composition Guidelines Given to Students

- Names Choose names that evoke the purpose or meaning of the values to which they refer.
- **Functions** Define functions with a clear role in the program that abstract the implementation of a behavior.
- Purpose Remove all unused and unnecessary code and comments, leaving behind only code with a purpose.

### Composition Guidelines Given to Students

- Names Choose names that evoke the purpose or meaning of the values to which they refer.
- **Functions** Define functions with a clear role in the program that abstract the implementation of a behavior.
- Purpose Remove all unused and unnecessary code and comments, leaving behind only code with a purpose.
- **Brevity** Prefer concise implementations to convoluted ones, without explicitly minimizing the line count.

Fill-in-the-blank starter code and a full test suite

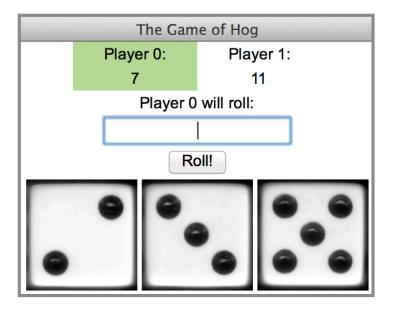
#### Fill-in-the-blank starter code and a full test suite

The Game of Hog					
	Player 0: Player 1:			/er 1:	
	7		11		
Player 0 will roll:					
Roll!					
	0	•		•••	

Fill-in-the-blank starter code and a full test suite Composition can be taught by example in context.

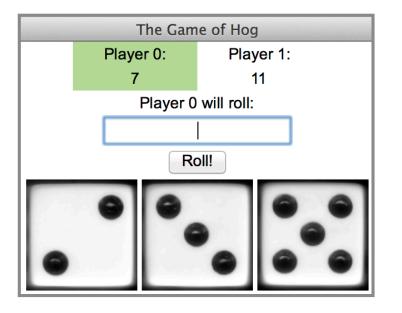
The Game of Hog					
	Playe	er 0:	Player 1:		
	7		1	11	
Player 0 will roll:					
Roll!					
	0	•		••••	

Fill-in-the-blank starter code and a full test suite Composition can be taught by example in context.



```
def make_averaged(fn, num_samples=1000):
    """Return a function that returns the average
    return value of FN called NUM_SAMPLES times.
    >>> dice = make_test_dice(3, 1, 5, 6)
    >>> averaged_dice = make_averaged(dice, 1000)
    >>> averaged_dice()
    3.75
    """
    "*** YOUR CODE HERE ***"
```

Fill-in-the-blank starter code and a full test suite Composition can be taught by example in context.



```
def make_averaged(fn, num_samples=1000):
    """Return a function that returns the average
    return value of FN called NUM_SAMPLES times.
    >>> dice = make_test_dice(3, 1, 5, 6)
    >>> averaged_dice = make_averaged(dice, 1000)
    >>> averaged_dice()
    3.75
    """
    "*** YOUR CODE HERE ***"
```

Shameless advertisement: All four course projects
are freely available with comprehensive autograders.
(See one in the Nifty Assignments session tomorrow.)

# Online Composition Feedback System

 Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)

- Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)
- Summary statements and comments on specific lines. (Heaney & Daly, 2004)

- Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)
- Summary statements and comments on specific lines. (Heaney & Daly, 2004)
- Text comments inlined with the submission source code. (MacWilliam & Malan, 2013)

```
def find_centroid(polygon):
    def get_x(p):
        return latitude(p)
    def get_y(p):
        return longitude(p)
```

- Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)
- Summary statements and comments on specific lines. (Heaney & Daly, 2004)
- Text comments inlined with the submission source code. (MacWilliam & Malan, 2013)

```
fsum = 0;
def f( p1,p2):
    return get_x(p1)*get_y(p2)-get_x(p2)*get_y(p1)
area = ...
if(area == 0):
    return (latitude(polygon[0]),longitude(polygon[0]),0)
```

```
def find_centroid(polygon):
    def get_x(p):
        return latitude(p)
    def get_y(p):
        return longitude(p)
```

```
To give a new name to a function,
you can use assignment:
get_y = longitude
```

- Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)
- Summary statements and comments on specific lines. (Heaney & Daly, 2004)
- Text comments inlined with the submission source code. (MacWilliam & Malan, 2013)

```
fsum = 0;
def f( p1,p2):
    return get_x(p1)*get_y(p2)-get_x(p2)*get_y(p1)
area = ....
if(area == 0):
    return (latitude(polygon[0]),longitude(polygon[0]),0)
```

```
def find_centroid(polygon):
    def get_x(p):
        return latitude(p)
    def get_y(p):
        return longitude(p)
```

```
To give a new name to a function,
you can use assignment:
get_y = longitude
```

- Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)
- Summary statements and comments on specific lines. (Heaney & Daly, 2004)
- Text comments inlined with the submission source code. (MacWilliam & Malan, 2013)

```
fsum = 0;
def f( p1,p2):
    return get_x(p1)*get_y(p2)-get_x(p2)*get_y(p1)
area = ....
if(area == 0):
    return (latitude(polygon[0]),longitude(polygon[0]),0)
```

polygon[0] is repeated; perhaps give it the name "first" instead.

Nice work. Keep looking for more ways to eliminate repeated expressions by introducing names.

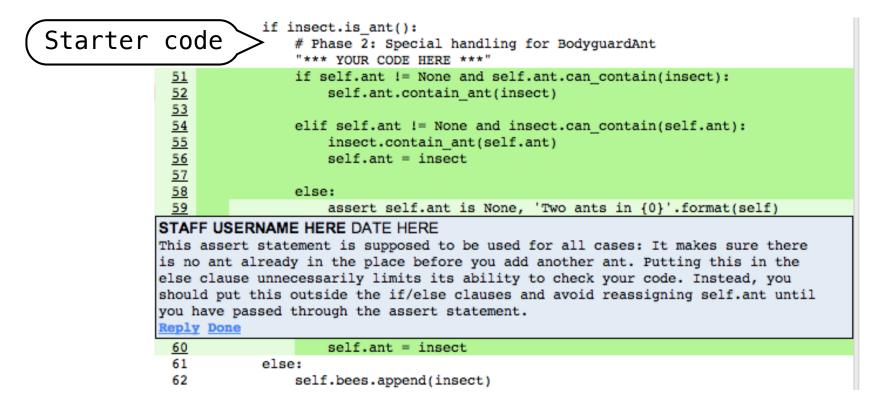
def find\_centroid(polygon):
 def get\_x(p):
 return latitude(p)
 def get\_y(p):
 return longitude(p)

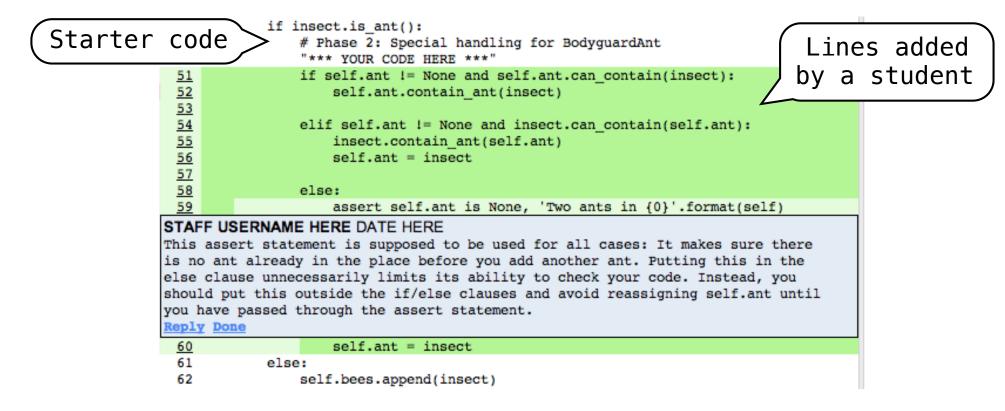
To give a new name to a function, you can use assignment: get\_y = longitude

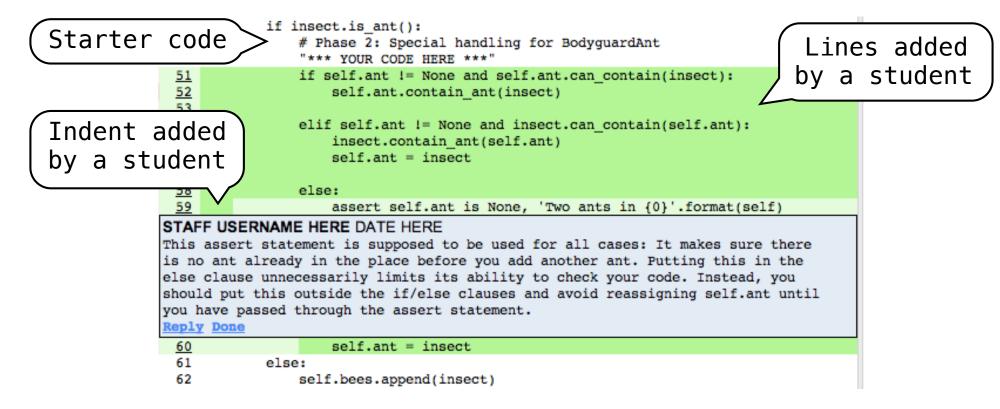
- Assignments are managed and graded with an online tool. (Jones & Jamieson, 1997; Jackson & Usher, 1997)
- Summary statements and comments on specific lines. (Heaney & Daly, 2004)
- Text comments inlined with the submission source code. (MacWilliam & Malan, 2013)

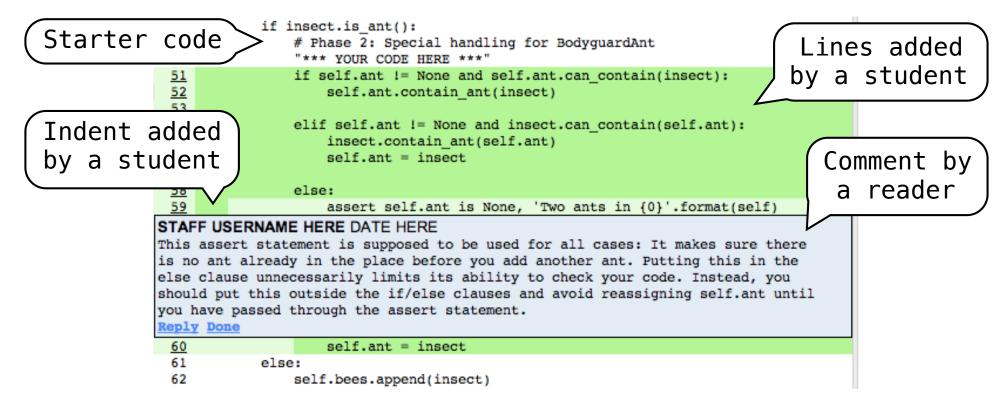
```
fsum = 0;
def f( p1,p2):
    return get_x(p1)*get_y(p2)-get_x(p2)*get_y(p1)
    area = ...
    if(area == 0):
        return (latitude(polygon[0]),longitude(polygon[0]),0)
polygon[0] is repeated; perhaps give it the name "first" instead.
```

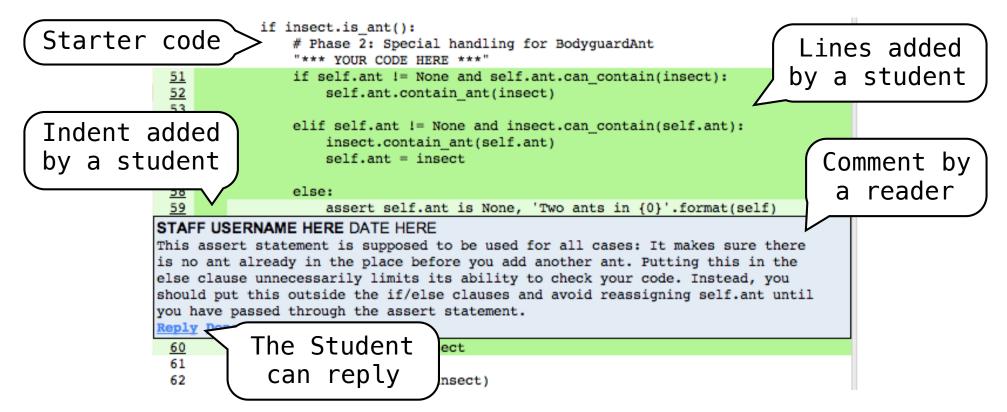
48 49 50	# Phase 2: Special handling for BodyguardAnt				
51	if self.ant != None and self.ant.can contain(insect):				
	self.ant.contain_ant(insect)				
52 53 54 55 56 57 58					
54	elif self.ant != None and insect.can_contain(self.ant):				
55	insect.contain_ant(self.ant)				
56	self.ant = insect				
57					
	else:				
<u>59</u>	assert self.ant is None, 'Two ants in {0}'.format(self)				
STAFF USERNAME HERE DATE HERE					
This	This assert statement is supposed to be used for all cases: It makes sure there				
is no	is no ant already in the place before you add another ant. Putting this in the				
else clause unnecessarily limits its ability to check your code. Instead, you					
should put this outside the if/else clauses and avoid reassigning self.ant until					
you have passed through the assert statement.					
Reply	y Done				
60	self.ant = insect				
61	else:				
62	self.bees.append(insect)				



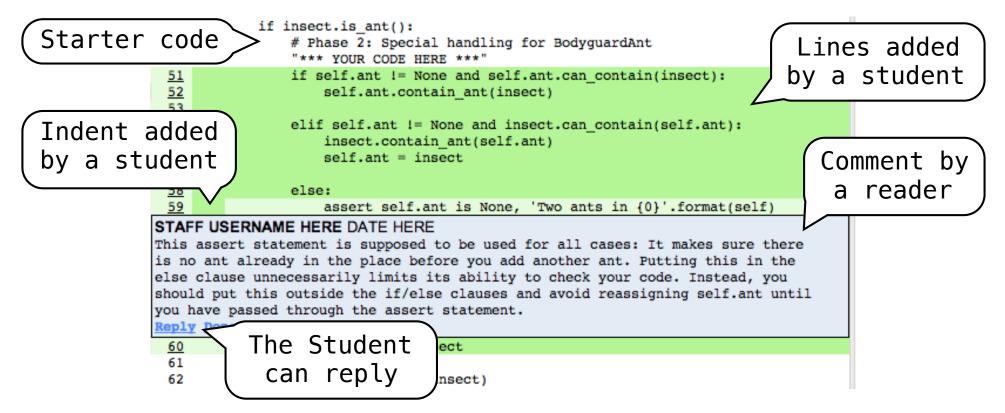




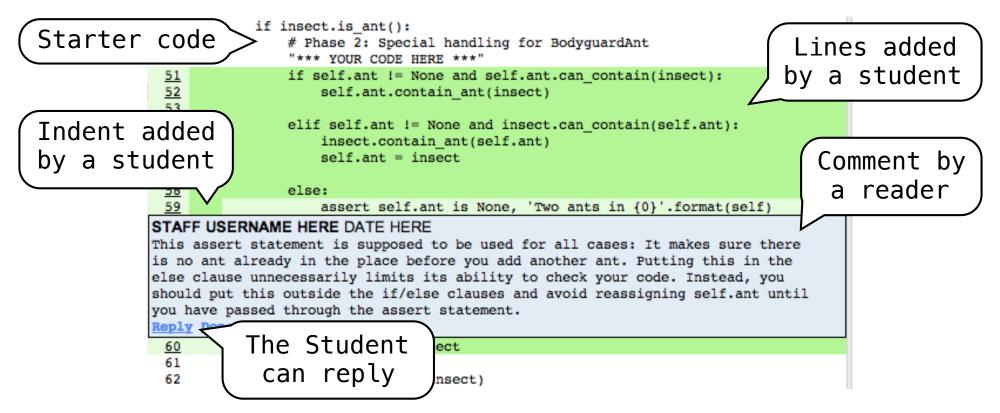




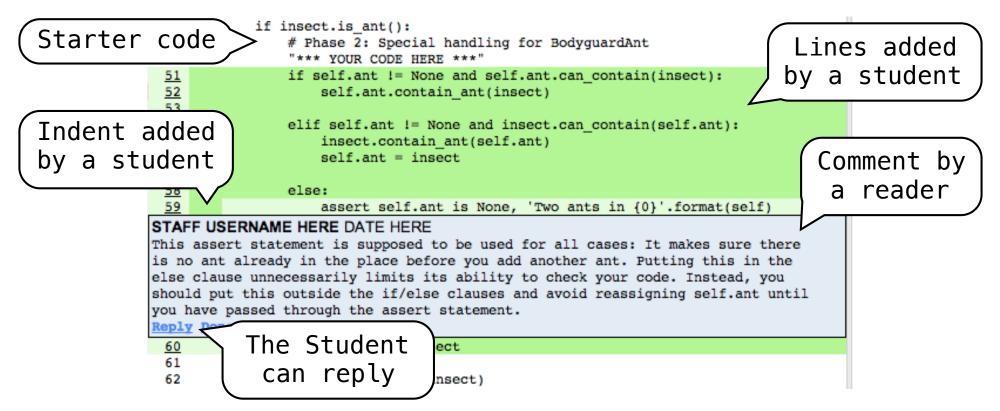
• Interface is adopted from the open source branch of Google's internal code review tool, Rietveld.



• Shared cache of comments among all graders.



- Shared cache of comments among all graders.
- Integration with existing version control systems.



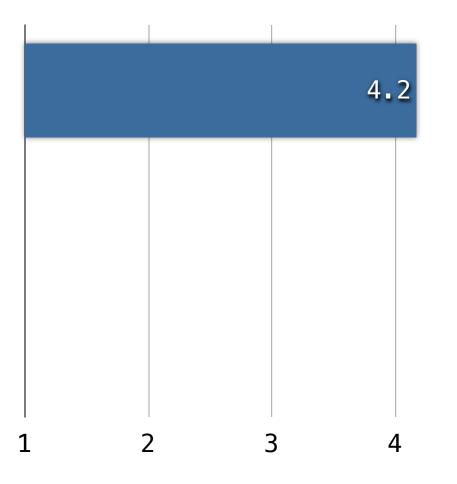
- Shared cache of comments among all graders.
- Integration with existing version control systems.
- Easy administration through Google App Engine.

# Evaluations

Fall 2012 Average of 626 Responses (All questions rated on a scale from 1 to 5)

Fall 2012 Average of 626 Responses (All questions rated on a scale from 1 to 5)

How much did you read of the composition feedback that you received on your project submissions?

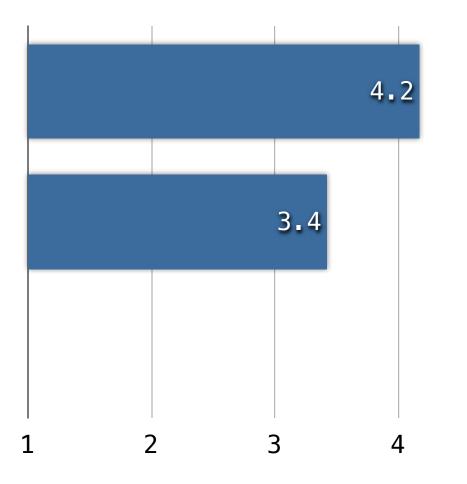


5

Fall 2012 Average of 626 Responses (All questions rated on a scale from 1 to 5)

How much did you read of the composition feedback that you received on your project submissions?

How helpful was the composition feedback you received?



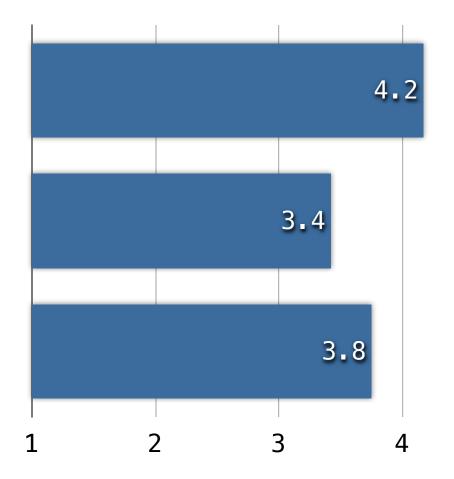
5

Fall 2012 Average of 626 Responses (All questions rated on a scale from 1 to 5)

How much did you read of the composition feedback that you received on your project submissions?

How helpful was the composition feedback you received?

How well did the composition feedback software system work for you?



### Blind Comparison of Student Submissions

### Blind Comparison of Student Submissions

Experiment Design

# Blind Comparison of Student Submissions

Experiment Design

• 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).

Experiment Design

- 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).
- 6 course graders (experts) judged implementations of a single function, find\_centroid, that didn't change significantly.

Experiment Design

- 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).
- 6 course graders (experts) judged implementations of a single function, find\_centroid, that didn't change significantly.

1

Average Responses by Graders (Scale from 1 to 3)

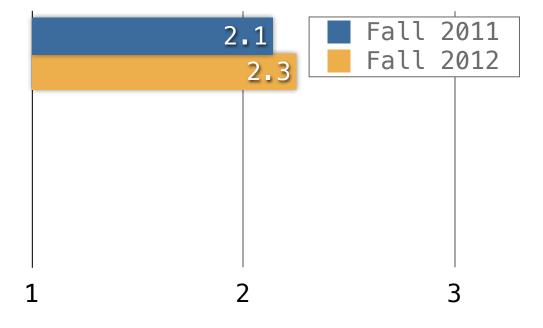
2 Fall 2011

Experiment Design

- 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).
- 6 course graders (experts) judged implementations of a single function, find\_centroid, that didn't change significantly.

### Average Responses by Graders (Scale from 1 to 3)

Do names convey the meaning & purpose of their values?



Experiment Design

- 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).
- 6 course graders (experts) judged implementations of a single function, find\_centroid, that didn't change significantly.

### Average Responses by Graders (Scale from 1 to 3)

Do names convey the meaning & 2.1 Fall 2011 purpose of their values? 2.3 Is the implementation concise? 2.2 1.8 2.2 1.8 2.2 1.8 2.2 3

### Experiment Design

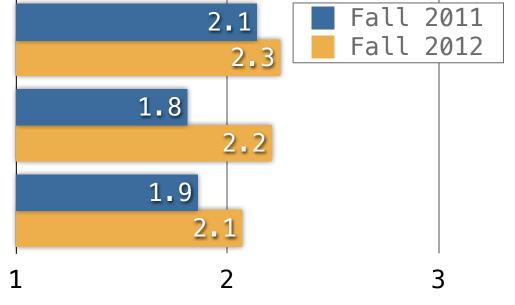
- 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).
- 6 course graders (experts) judged implementations of a single function, find\_centroid, that didn't change significantly.

### Average Responses by Graders (Scale from 1 to 3)

Do names convey the meaning & purpose of their values?

Is the implementation concise?

Is the program composed well?



### Experiment Design

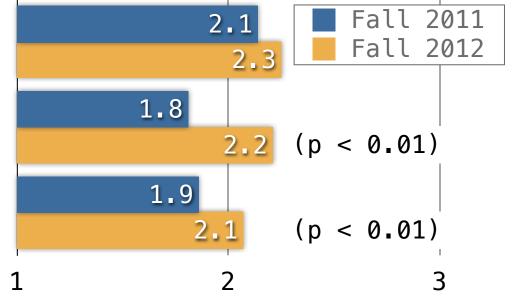
- 16 student submissions were sampled without replacement from Fall 2011 (before composition focus) and Fall 2012 (after).
- 6 course graders (experts) judged implementations of a single function, find\_centroid, that didn't change significantly.

### Average Responses by Graders (Scale from 1 to 3)

Do names convey the meaning & purpose of their values?

Is the implementation concise?

Is the program composed well?



## Another Shameless Advertisement

## Composing Programs: An Interactive Textbook

A product of public domain and open source content:

- A product of public domain and open source content:
- Derived from Structure and Interpretation of Computer Programs

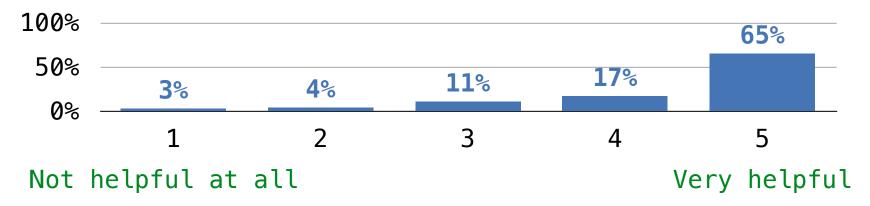
- A product of public domain and open source content:
- Derived from Structure and Interpretation of Computer Programs
- Examples diagrammed by the Online Python Tutor

- A product of public domain and open source content:
- Derived from Structure and Interpretation of Computer Programs
- Examples diagrammed by the Online Python Tutor

Demo: <u>http://composingprograms.com/pages/16-higher-order-functions.html#functions-as-arguments</u> Demo: <u>http://composingprograms.com/pages/23-sequences.html#recursive-lists</u>

- A product of public domain and open source content:
- Derived from Structure and Interpretation of Computer Programs
- Examples diagrammed by the Online Python Tutor

How helpful did you find the online tool for drawing environment diagrams in understanding course material?



Demo: <u>http://composingprograms.com/pages/16-higher-order-functions.html#functions-as-arguments</u> Demo: <u>http://composingprograms.com/pages/23-sequences.html#recursive-lists</u>

## Thanks!

Source Code: github.com/moowiz/ucb-codereview
Description: composingprograms.com/codereview.html

Interested in the Composing Programs textbook? Please fill out the survey on composingprograms.com!

## Extra

### End-of-Semester Survey Results

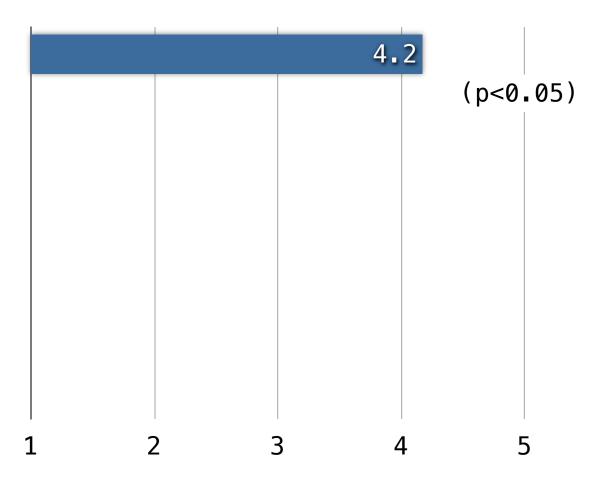
Fall 2012 Average Responses (Scale from 1 to 5)

(p<0.05)

## End-of-Semester Survey Results

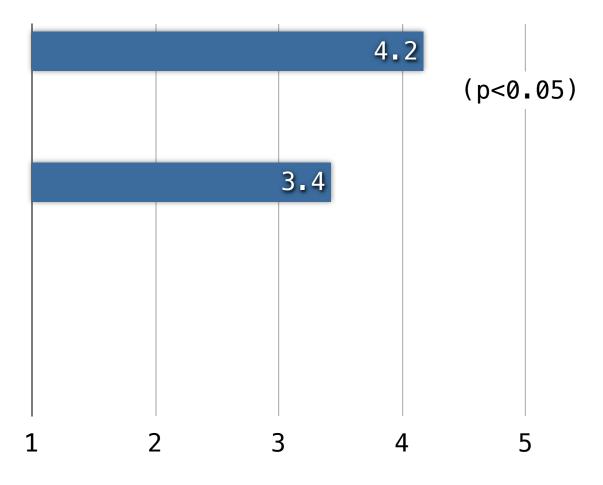
#### Fall 2012 Average Responses (Scale from 1 to 5)

How much did you read of the composition feedback that you received on your project submissions?



How much did you read of the composition feedback that you received on your project submissions?

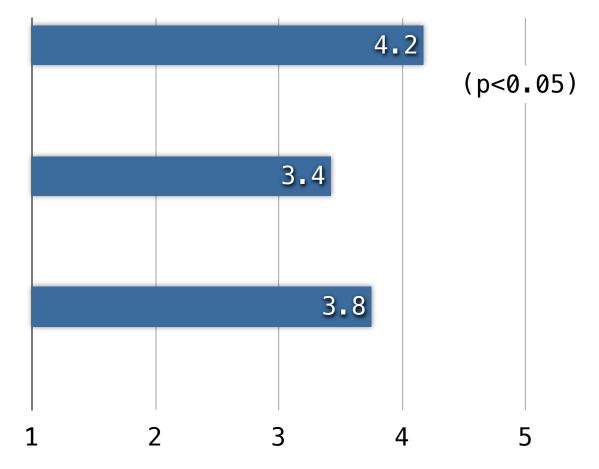
How helpful was the composition feedback you received?



How much did you read of the composition feedback that you received on your project submissions?

How helpful was the composition feedback you received?

How well did the composition feedback software system work for you?

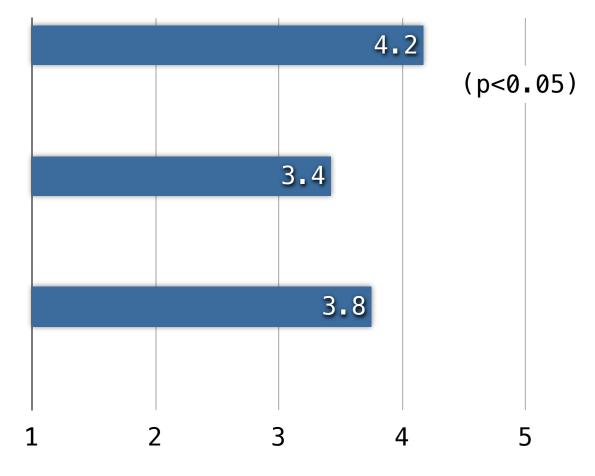


How much did you read of the composition feedback that you received on your project submissions?

How helpful was the composition feedback you received?

How well did the composition feedback software system work for you?

Total (626 responses) Female (134 responses) Male (485 responses)



How much did you read of the composition feedback that you received on your project submissions?

How helpful was the composition feedback you received?

How well did the composition feedback software system work for you?

Total (626 responses)
Female (134 responses)
Male (485 responses)

