Innovation in teaching software development skills

William Pugh
Univ. of Maryland

Dec 2nd, 2011
SUNY Stony Brook

Big idea topics

- Whither lecture?
- Programming exercises
- Programming projects
- Inscrutable Errors
- Software quality: more than passing tests

Problems

- Academic dishonesty and the internet
- Training/skills for faculty and TAs
- Huge range of student ability
  - position, velocity, acceleration
- Lack of respect for computer science education research and innovation
- Scaling it out to all the places world wide that will need to teach programming and software development

Don't have time to talk about

- Programming at the high school or lower level, or similar content at the college level for non-majors
- Skills other than software development (e.g., discrete math, NP-completeness, algorithm design)
- What topics to cover, or how they get grouped into courses/threads/concentrations/majors

Common state of the art

- Standard lecture format
- 2 week projects; detailed specs, some test cases
- projects submitted to electronic drop box
- after project deadline, TA runs all submissions against provided test cases and secrets tests
- grades given back a week after submission

Audience responses included

- After my talk at Suny StonyBrook, audience members gathered around and put up comments and questions on several huge post-it notes (3x2 foot)
  - This provided more opportunity for group discussion than I could do in a typical question period
  - I photographed each comment/question sheet and wrote responses as best I could. Those sheets and my responses are included in this slide deck

Big picture topics

Whither lecture?

- In the age of iTunesU, Univ. of Phoenix, 100,000 students on-line classes and TED talks, what is the future of lectures?
- If a lecture is good, can it be put online and reused?
  - once you get to 40, 50, 100+ students, how much value do students get out of attending a live lecture?
- What else can we do with classroom time?

Whither Lecture?

Record them and give students clickers to mark spots so they can later generate questions and make it more interactive.

Class participation in TA recitations sometimes drops after 15 people in class. How effective is it in larger ones?

How do you keep students off Facebook?

Perhaps do masterclasses [time best spent.

Give sample code, let students code. Only lecture students' mistakes. Lecture

* Khan academy is doing nice experiments how to record software programming lecture with interact code editor that execute the code.

- Clickers in the classroom are interesting, but I've always been hesitant about requiring students to go out a buy a gadget
- When watching a video lecture, having students click to add emoticons and questions is a good idea
- Keeping students off Facebook: good question. Ideas:
  - set expectations
  - walk around during class, including to back of room
  - tell students when to close laptops

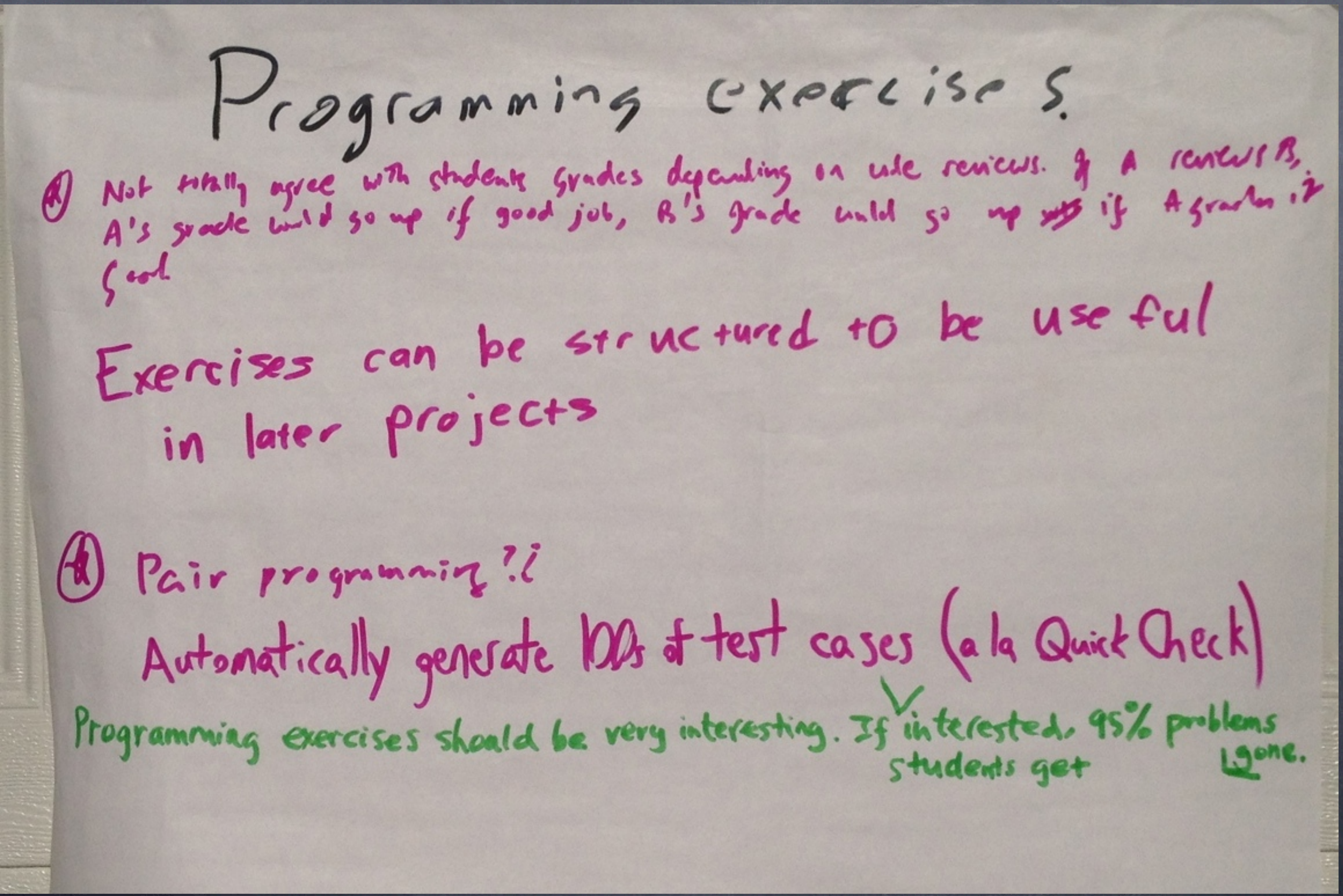
- Kahn academy: They are worth following; haven't yet looked at what they are doing with programming
- What to do with lecture time: some kind of active learning, perhaps some time students working on coding where they can be in a supportive environment and get help from others. Perhaps pair programming

Programming exercises

- You can't learn how to program by reading about it
- Projects are too big: too infrequent, too easy to get lost
- Want students to test whether they can put to use the skills/techniques they just learned
- Even more important if we move to on-line lectures and/or self-paced instruction

Programming exercises

- Codecademy http://www.codecademy.com/
- CodingBat http://codingbat.com/
- TurningCraft http://turingscraft.com/
- Gradiance http://www.newgradiance.com/



Programming exercises.

Ⓐ Not totally agree with students grades depending on code reviews. If A reviews B, A's grade would go up if good job, B's grade would go up ~~if~~ if A grades it good

Exercises can be structured to be useful in later projects

Ⓐ Pair programming?¿

Automatically generate 100s of test cases (a la QuickCheck)

Programming exercises should be very interesting. If students get interested, 95% problems gone.

- Re: code reviews (don’t fully understand q)
    - Can’t have students grade other students, but can give feedback and learn from looking at other solutions
    - Code reviews from TA assign part of grade for things other than functional correctness
- exercises can be useful for projects and/or very interesting, but I think it is important that they are optimized for learning and mastering small bits of material covered

- Pair programming is a great idea. We had trouble implementing it when we expected students to do pair programming on their own. Works much better when done in lab sesssions.
- Test generation is an interesting idea. See Pex For Fun

Two Kinds of Programming Projects

- Everyone implements the same project spec
  - common tests can be used to assess functional correctness
- Independent projects
  - students working on different problems
  - students given lots of liberty in their implementation, must be evaluated independently

Programming projects

- Learn what students are doing
- Give better and prompt feedback to students and instructors
- Encourage students to start early, work in bursts rather than in a concentrated slog
- Automate much of the tedium for instructors

Marmoset

- Developed by Jaime Spacco for his Ph.D. work at UMD
- Used in almost all programming courses at UMD, and at a few other universities
- open source, apache license

Using Marmoset

- public test cases distributed with project
- Submit project from IDE, command line, or through web interface
- all tests cases run on server when project is submitted
- If submission passing all public tests, students may request release testing

Marmoset
Release Testing

- Students are told # of release tests passed and failed
  - and names of first two release tests that failed
- For example, on Poker project, might be told that they "fail FourOfAKind, FullHouse, and 2 other tests"
- Release testing consumes a token
  - students receive 2 tokens
  - tokens regenerate 24 hours after used

Advantages of release testing

- Encourages students to think, develop their own tests
- Gives students an indication of where they are, whether that are having trouble
- Gives students an incentive to start working early
- Instructors get live feedback about student progress before project deadline

Marmoset Research Study

- Students agree to participate in research study
  - 72 students from Fall 2004
- Get snapshots of every save as they develop their projects
  - 41,000 compilable snapshots from Fall 2004
- On each snapshot, run all test cases, all bug finders, etc.
- Is there a correlation between changes that remove bug warnings and changes that cause more test cases to pass? yes!

More questions

- We can actually look at why student projects fail
  - e.g., we can look at all places where a NPE exception occurred
  - Can we find new detectors that are helpful to students?

Project Snapshots

students	73
projects	8
student projects	569
snapshots	51,502
compilable	41,333
unique	33,015
Test outcomes	505,423
not implemented	67,650
exception thrown	86,947
assertion failed	115,378
passed	235,448

Size of incremental deltas

size	instances
1	12,873
2	5,484
3-4	4,726
5-8	3,608
9-16	2,503
17-32	1,229
33-64	612
65+	352

Other information

- We have a timestamp for each snapshot
- No idea what the students were doing between snapshots
- No idea why the students saved their file
- We know when the student performed Submit, when they performed release testing

Exceptions thrown

Exception	Student projects	Test cases
NullPointer	323	7,289
ClassCastException	158	2,524
IndexOutOfBounds	124	1,859
StackOverflow	123	1,728
StringIndexOutOfBounds	120	2,592
IllegalState	104	1,075
OutOfMemory	81	2,608
IllegalArgument	78	2,119
RuntimeException	73	951
NoSuchElement	65	1,024
ArrayIndexOutOfBounds	63	1,111

Can Find False Positives and False Negatives

- Strong correlation between certain errors and certain warnings
  - the warnings are designed to catch the errors
- Can run all student test cases, see when we generate exceptions and when we generate warnings

Infinite Recursive Loops

- Surprisingly common

-
  ```
  public WebSpider() {
    WebSpider w = new WebSpider();
    }
  ```
- Wrote a detector
  - caught a number of cases in student code
    - caught 3 cases in Sun's Java JDK

Correlation

		Stack Overflow Exception	
		yes	no
Recursive Loop Warning	yes	279	2
	no	1,140	31,594

Manually examined false negatives

- We manually examined some of the false negatives
- Identified some cases we could easily catch with simple static analysis
- Augmented bug detector

Enhanced recursive loop warning

		Stack Overflow Exception	
		yes	no
Recursive Loop Warning+	yes	935	?
	no	793	?

Class Class Exception

		Class Cast Exception	
		yes	no
Bad Cast Warning	yes	362	1,306
	no	1,775	30,878

Enhanced detector

		Class Cast Exception	
		yes	no
Bad Cast Warning+	yes	1,047	?
	no	1,090	?

Null Pointer detector

		Null Pointer Exception	
		yes	no
NP Warning	yes	267	382
	no	5,863	26,503

New work on Marmoset

New work

- working to make it easier for other schools to get running
- Code reviews
- Tournaments

Code reviews

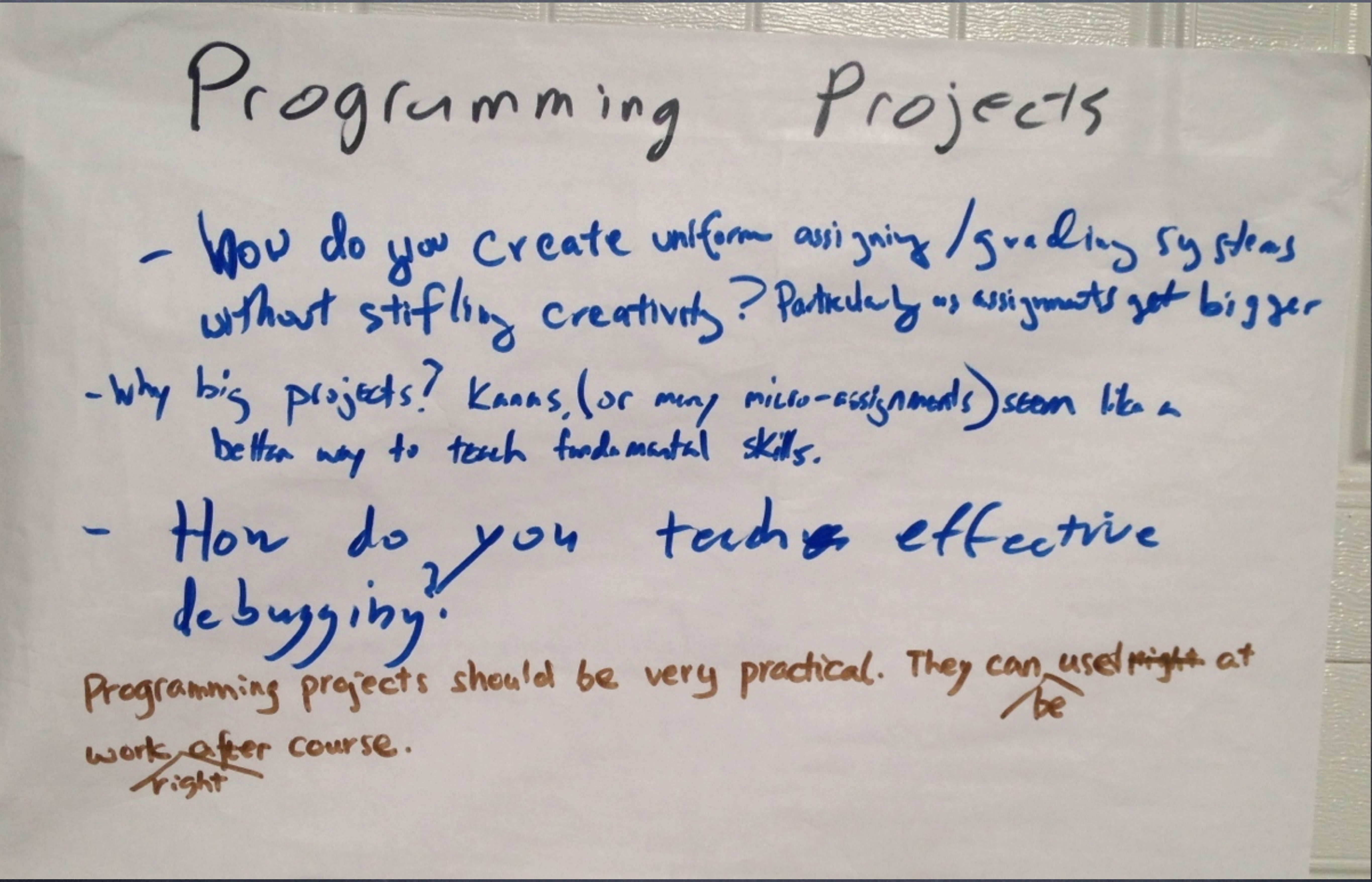
- Web based system for doing code reviews
- Look at code, double click on a line of code to add a comment there
- reviews can include rubrics
    - does the code use XYZ?
- authors can respond to comments

Assigned code reviews

- Students can be assigned to review other student's submission
  - might typically be assigned to review 2 other students' submission
- Can also be asked to fill out checklist
  - can compare them for consistency

Tournaments

- For projects such as implementing logic for tic-tac-toe or connect-4, can run tournaments of submissions that pass public tests
- In early trials, good motivation for some students
- May also allow test case tournaments
    - can your tests find problems in other student's code?



Programming Projects

- How do you create uniform assigning/grading systems without stifling creativity? Particularly as assignments get bigger

- Why big projects? Kaaas, (or many micro-assignments) seem like a better way to teach fundamental skills.

- How do you teach effective debugging?

Programming projects should be very practical. They can be used ~~might~~ at work right after course.

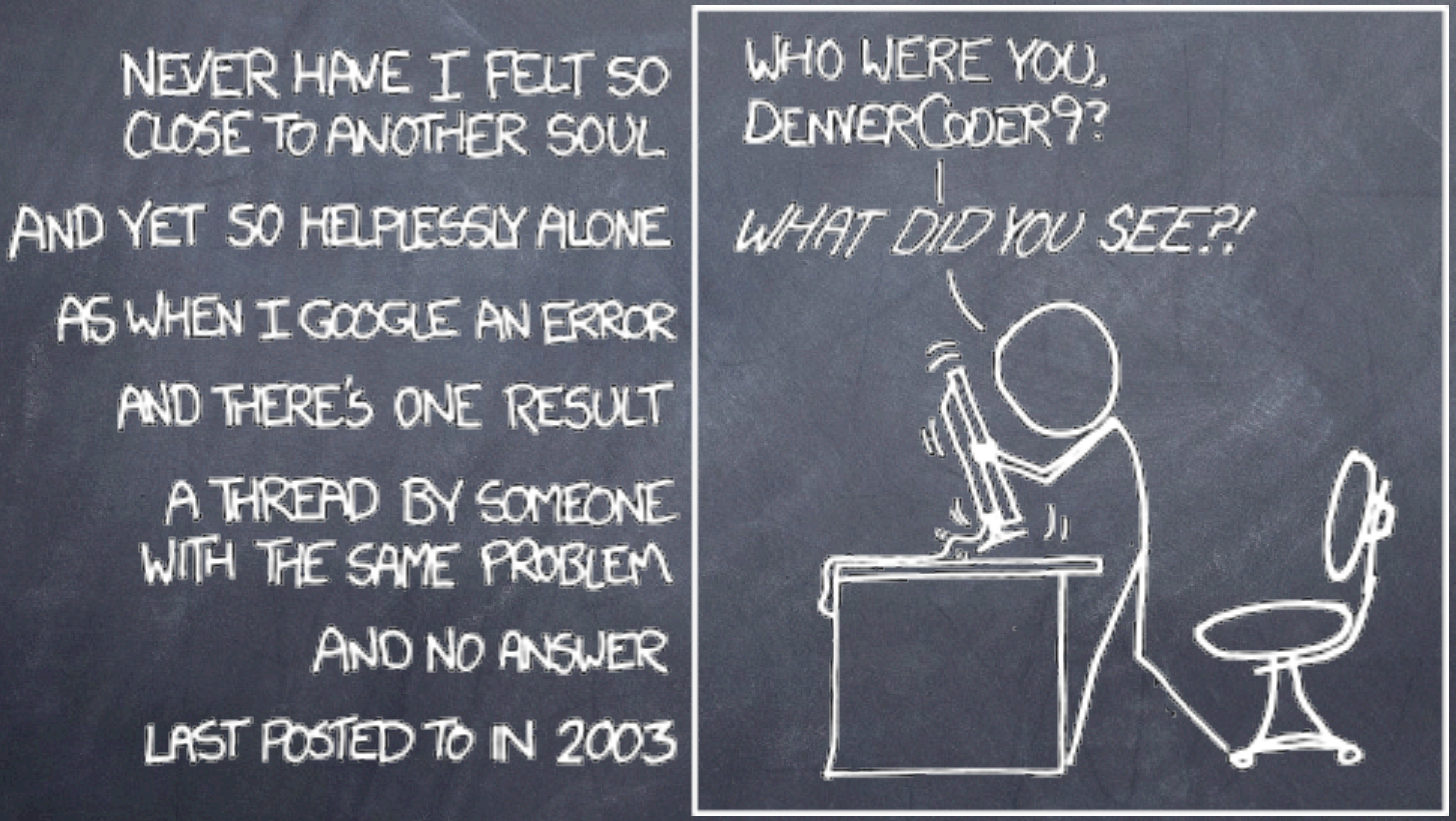
- Automated grading vs. stifling creativity:
    - Students should do some coding where they have to design API's and features; such projects are useful but also difficult to scale to a class with 100 students.
    - Coding to an API and dealing with unit tests is something every software developer is going to need to learn
    - Didn't say Marmoset is the only way to do projects; but better than submitting to a black box, having a TA run test scripts after the deadline, and getting grades back a week after deadline

- Why programming projects: there are a lot of skills you never pick up just doing exercises: how to debug, how to structure a program, etc. They have to move beyond exercises at some point.
- How do you teach debugging?
  - a lot of students just don't know the capabilities of a debugger or how to use them. Exercises and labs could help with that.
  - After that, perhaps some live demonstrations and strategy discussions

Inscrutable Errors

- One of the big changes over the decades I’ve been coding is that we are working with much more complicated systems
- The error messages are often incrutable and confusing
    - Googling the error message is sometimes helpful

Googling for errors


NEVER HAVE I FELT SO CLOSE TO ANOTHER SOUL
AND YET SO HELPLESSLY ALONE
AS WHEN I GOOGLE AN ERROR
AND THERE'S ONE RESULT
A THREAD BY SOMEONE WITH THE SAME PROBLEM
AND NO ANSWER
LAST POSTED TO IN 2003
WHO WERE YOU, DENVERCODER9?
WHAT DID YOU SEE?!

Error several students got when using Java ForkJoin framework

```
java.lang.ClassCastException: java.lang.Thread cannot be cast to
jsr166y.ForkJoinWorkerThread
    at jsr166y.ForkJoinTask.fork(ForkJoinTask.java:591)
    at cmsc433.p4.explorers.ForkJoinExplorer
$ParallelraphExploration.<init>(ForkJoinExplorer.java:50)
    at cmsc433.p4.explorers.ForkJoinExplorer.explore(ForkJoinExplorer.java:38)
    at cmsc433.p4.Tests.buildExplorer(Tests.java:34)
    at cmsc433.p4.PublicTests.testCircularLinkedList(PublicTests.java:43)
```

The problem?

- The problem was that the code written by the student was calling fork from a main thread to start a fork join task
    - fork should only be called from inside a ForkJoin task. From outside a ForkJoin task, you should call ForkJoinPool.invoke
- The ForkJoin library should catch that error and throw an error that explains the problem

Notes

- Beginning students get this all the time
- Experienced developes get this when they try to use a new language/framework/library
  - Ever tried to use GWT/Hadoop/JSP?

What can we do?

- Pay a lot more attention to giving useful error messages
- Datamine errors
    - a lot could be done here in for student projects

Inscrutable Errors

Class FORUM

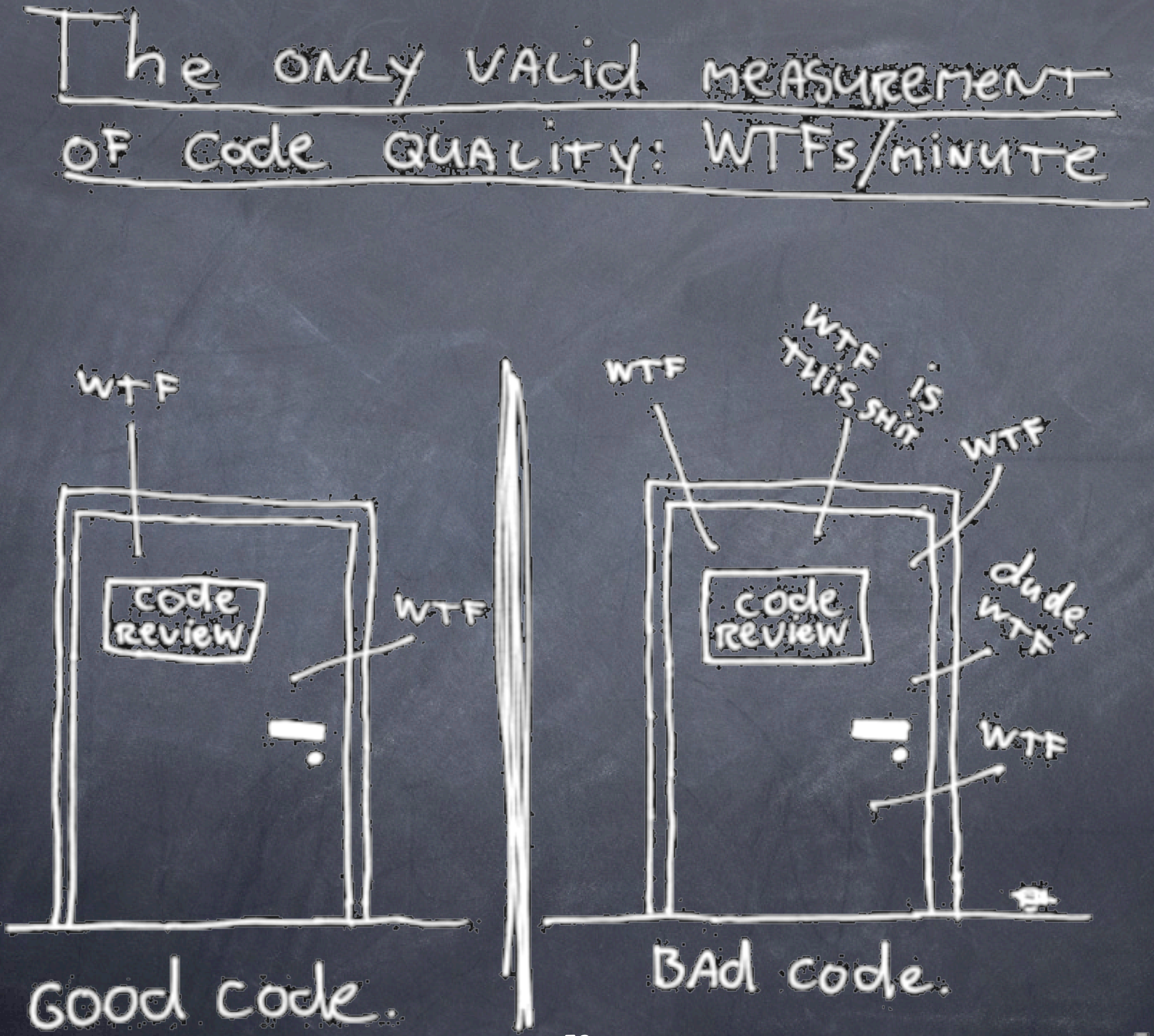
Adopt a better compiler
e.g. LLVM/Clang

Like such errors. Good training.

Which language has better error reporting?

Is not interpreted languages are easier to teach?

- A class forum is works OK. But there is a long tail of errors. And sometimes it isn't obvious that when you get an a error that needs to be researched.
- Better tools that give better error messages help. llvm/clang does give more and better errors than gcc. Perhaps valgrind would also be helpful for C projects.
- Not clear if interpreted languages give better error msgs. Perhaps it is better to have static typing giving a warning. Perhaps something in between (e.g., gradual typing).

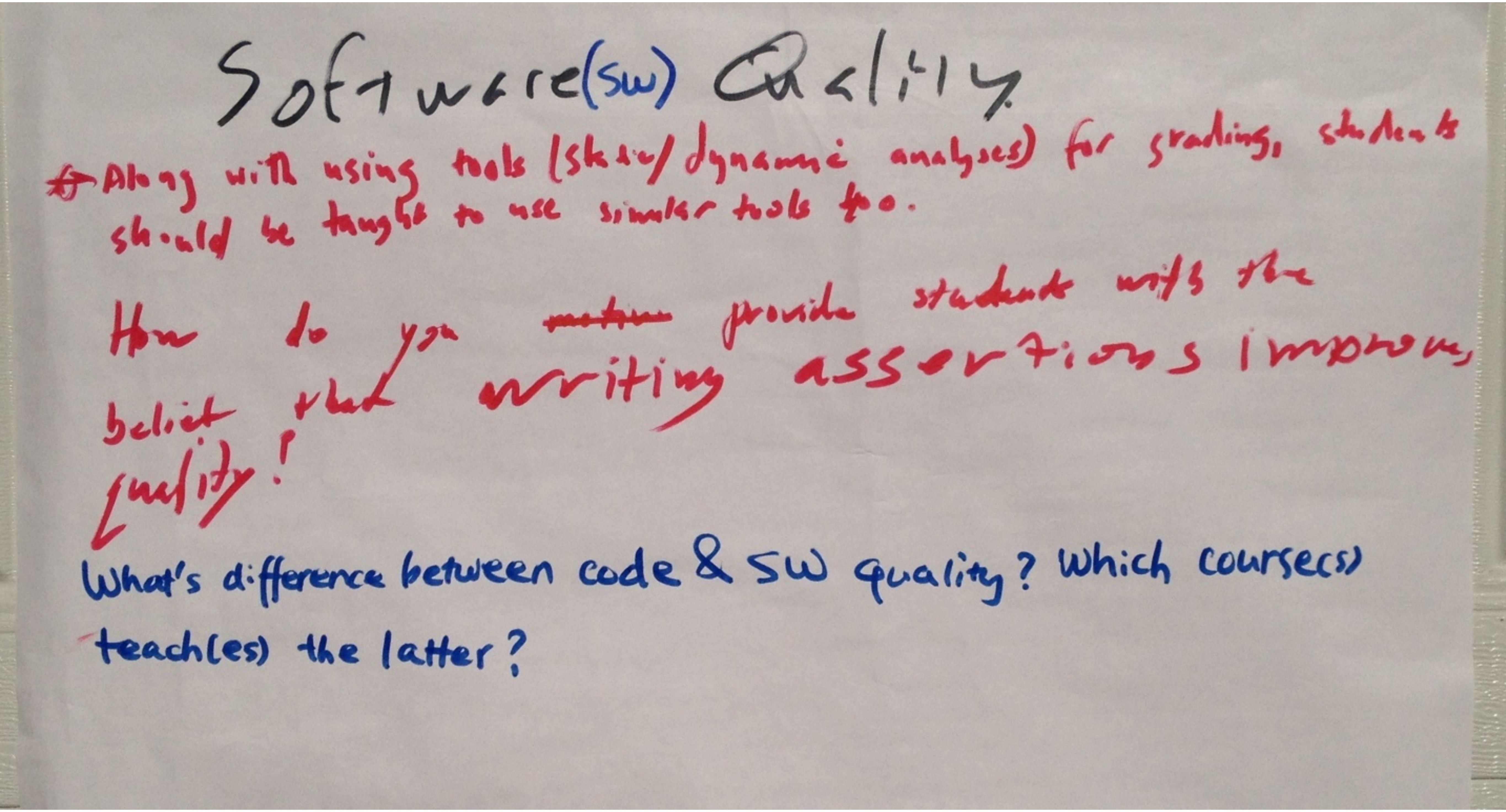

The only valid measurement of code quality: WTFs/minute
WTF
code review
WTF
Good code.
WTF
WTF is this shit
WTF
code review
dude, WTF
WTF
BAd code.

Software quality: more than passing tests

- If you've ever spent much time looking at student code, you've probably been horrified
- Even the students who pass test cases often have horrible code
  - looking at their code sometimes inspires new test cases

Ideas

- use static and dynamic analysis
  - code coverage, FindBugs, lint, etc.
- do code reviews
- get them to write test cases



Software (SW) Quality

* Along with using tools (static/dynamic analyses) for grading, students should be taught to use similar tools too.

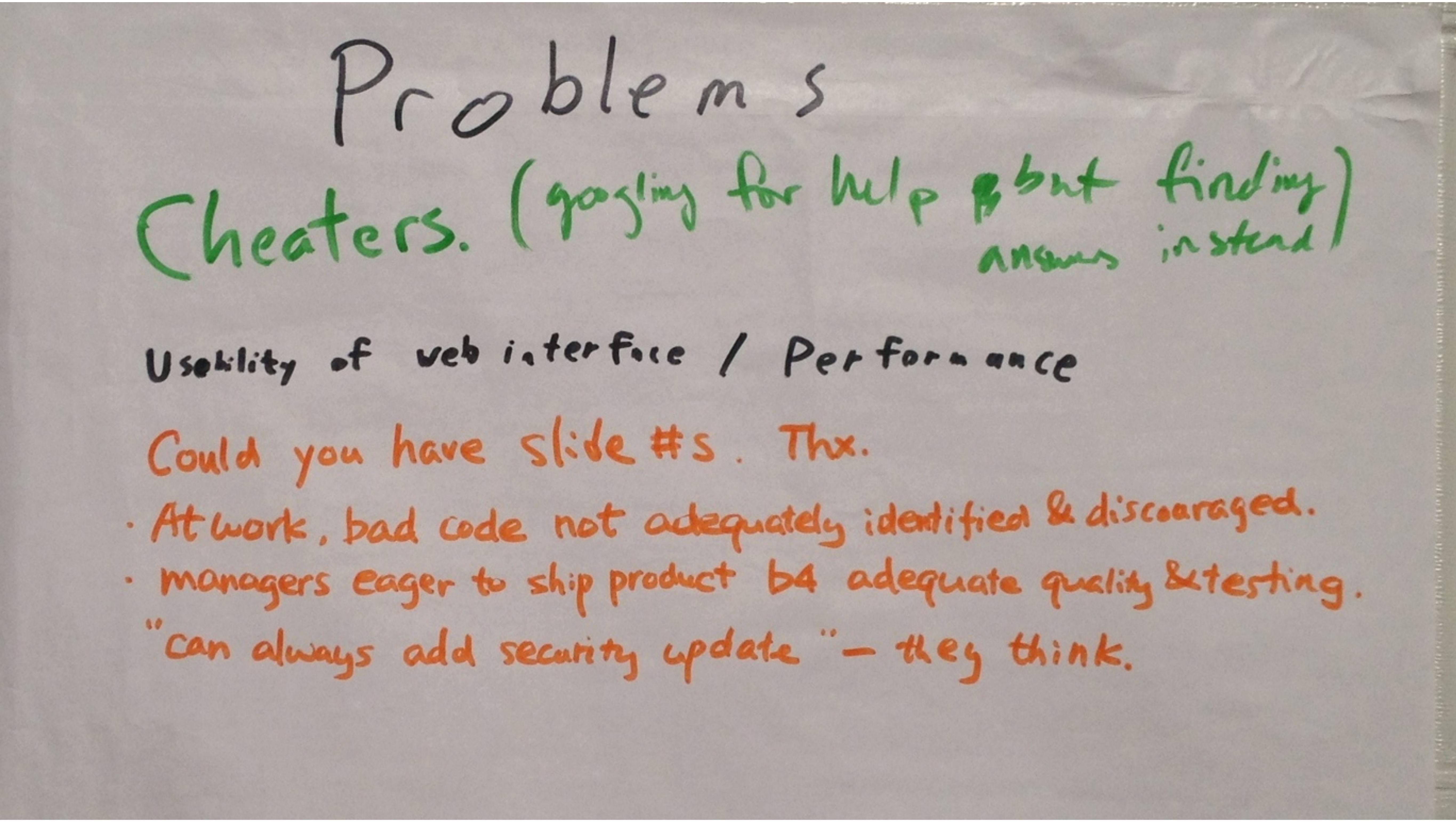
How do you ~~motive~~ provide students with the belief that writing assertions improves quality!

What's difference between code & SW quality? Which course(s) teach(es) the latter?

- Great to teach students to use sw quality tools on their own, but they can be a little tricky to set up and interpret. Don't want this to be a blocking factor to the students learning what the tools can do
- Assertions can help; so can lots of things, not sure if assertions are the most important thing.
    - can tell students that if an assertion fails on a release test, we'll give them the report of the assertion, helping them debug their code

- Difference between code and sw quality?
  - The two are pretty well meshed together. But I’d say sw quality is about more than just passing test cases. Instead, it covers things like how easy the code is to understand, debug and change, particularly when those tasks might be done by someone else months or years after it is written
  - We try to cover a little of SW quality in all of our courses, but increasing more in latter courses. The SW Engineering course spends a lot of time on SW quality

More responses



Problems

Cheaters. (googling for help but finding answers instead)

Usability of web interface / Performance

Could you have slide #s. Thx.

- At work, bad code not adequately identified & discouraged.
- managers eager to ship product b4 adequate quality & testing.

"can always add security update" – they think.

- But the fact that real developers google for help is a reality that we shouldn't ignore
- If students can find the answer on-line, maybe you shouldn't be assigning it as a project
  - don't give red-black tree projects
- Getting the whole Marmoset infrastructure robust, high performance, secure and with a good UI is a big challenge. We've been working on it for years.
- If you find yourself at a company that is happy with bad code, leave
  - but think about good/bad code from a cost/benefit point of view. There is a fixed amount of engineering time/talent available, and it must be spent carefully.

Things I forgot to talk about.

→ How do we evaluate the effectiveness of these new teaching methods? How much better?

→ Do you use some versioning system? Y N → Why not?

→ Can you provide more detailed feedback for certain test cases using code coverage?

→ Can we automate every aspect of teachers except OH's?

→ Can we have better Debuggers reducing time Running thru cases & finding errors.

→ Does your tools work for other than Java?

→ Measuring quality / measuring simplicity

simpler code ⇝ better code?

→ How do you teach teamwork effectively?

- Measuring the effectiveness of teaching techniques is really hard. Harder than measuring the effectiveness of different software development techniques.
  - variations in teacher and student ability introduce a lot of noise
- Can we automate everything other than office hours?
  - No, but we should automate what we can automate well, and find ways to put classroom and human contact time to better use

- Can we have better debuggers?
    - I’ve been waiting for reversible debuggers for over a decade. Yes, it would be nice.
- Does Marmoset work for languages other than Java?
    - yes; it has a mode where each test case is just a unix executable that either succeeds or fails.
    - but code coverage and some other features only work for Java

- Is simpler code better code?
    - Sometimes, a solution might be more complex because it is trying to be a more general solution
    - But in general, when I see a student solution that is much more complicated than the typical solution, it is a bad solution
- Teaching team work is hard, and evaluating and grading team work is hard. And this isn't just a software development skill. If you want to teach teamwork, you'll need to have team projects.