

---

# Why Tests Don't Pass (or Fail)

PNSQC October 28, 2009

Douglas Hoffman  
Software Quality Methods, LLC.  
Doug.Hoffman@acm.org  
[www.SoftwareQualityMethods.com](http://www.SoftwareQualityMethods.com)

---

## Definitions

- Test
  - An exercise designed to surface bugs
  - Includes a validation mechanism
- Pass
  - A test that behaves as expected per validation
- Fail
  - A test where validation fails
  - A test with unexpected behavior
- Oracle
  - Principle or mechanism by which we recognize a problem
  - Answers “why” when unexpected things happen

---

## Pass?

- What results are we checking?
- How do we know the expected outcome?
- Are we checking all of the results?
- What outcomes are we not checking?
- How would we know misbehaviors for the outcomes we are not checking?

---

## A Programming Challenge

Given a program and a set of tests for it, modify the code [to do anything] in a way that the tests don't detect.

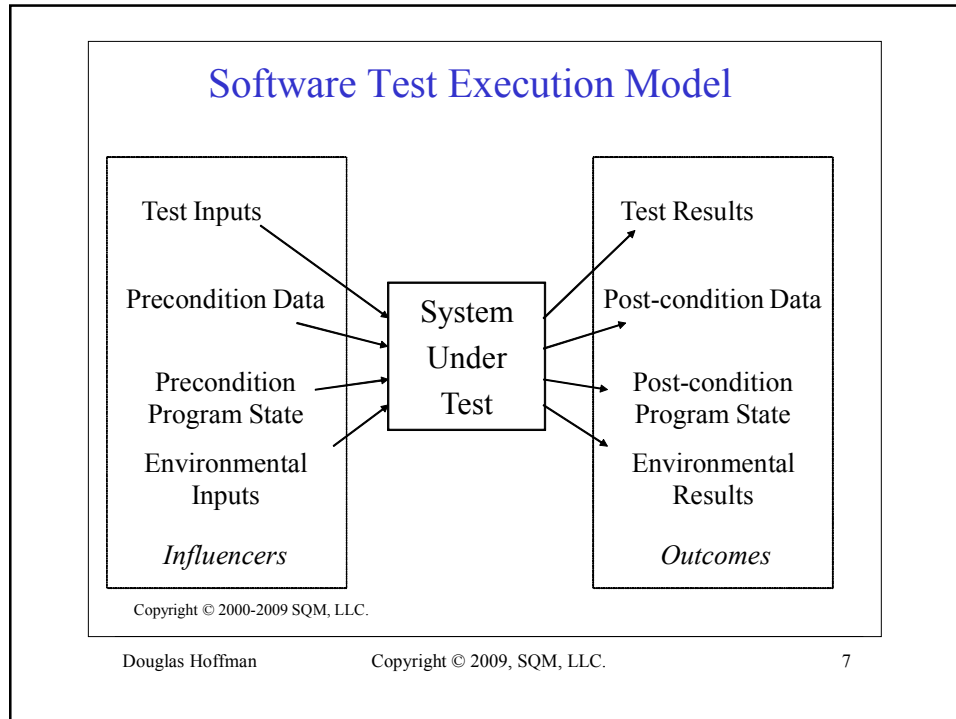
This change is an undetected bug.

## Test Result Possibilities

<b>Situation</b>	<b>No Error in SUT</b>	<b>Error in SUT</b>
<b>Test Result</b>		
<b>As Expected (Pass)</b>	<b>Correct Pass</b>	<b>Silent Miss</b>
<b>Unexpected (Fail)</b>	<b>False Alarm</b>	<b>Caught it!</b>

## Fail?

- The test finds the bug it's designed for!
- Something unexpected happens while the test is running.
- Possible false alarms.
- Possible unnoticed failures.



## Questions to Ask Ourselves

- What values/conditions should influence the SUT?  
How do we set/cause them?
- Are we controlling or monitoring the best influencers?  
(What influencers are we not controlling or monitoring?)
- Are we checking the most important outcomes? (What outcomes do we know we are we not checking?)
- How do we know the expected outcomes?
- What gives us confidence the test isn't missing bugs?

---

## So What

- Be open minded about (skeptical of) “Pass”
- Understand that a “Fail” could mask errors
- We aren’t checking all of the results
- We don’t know the outcomes from arbitrary errors
- Pass/Fail metrics don’t really give us interesting information

---

## What Can We Do

- Influencers
  - Set values
  - Monitor
- Outcomes
  - Check what’s practical
  - Compare pre- with post-
  - Real time monitoring
  - Sanity checks

---

## Summary

- The SUT doesn't really pass or fail
  - “Pass” means we noticed only expected behaviors
  - “Fail” means we noticed something not according to plan
- There are lots of other possible reasons for “passing” or “failing” other than a bug in the SUT

