PRH Consulting, Inc.

leveraging know-how for performance!

Test Strategies

Verifying Capability to Perform

ISPI Conference
April 7, 2008

PRH Consulting, Inc.

Peter R. Hybert
20 Danada Square West, #102
Wheaton, IL 60187
ph: 630.682.1649
web: www.prhconsulting.com

PRH Consulting, Inc.

leveraging know-how for performance!

Test Strategies

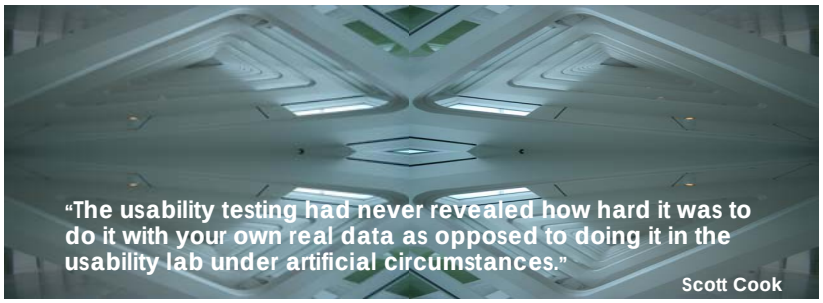
Verifying Capability to Perform

ISPI International Conference
April 7, 2007

PRH Consulting, Inc.
Peter R. Hybert
20 Danada Square West, #102
Wheaton, IL 60187
phone: 630.682.1649
web: www.prhconsulting.com

TABLE OF CONTENTS

Overview: We Can Do a Better Job with Employee Testing



The Opportunity: Effective Verification of Capability

Premises

1. People generally dislike testing as a "gate" (but love it as a game show)
2. There is an increased interest in testing (pre and post) in business lately...
3. ...and, there are a lot of misconceptions about testing
4. Testing itself doesn't add value (to the performer/performance) but it does add cost
5. All testing involves decisions and trade-offs—what to sample, how much to sample, method

Success Factors—How We Can Realize the Opportunity

1. Test strategically—select targeted areas
2. Choose the easiest path only when it is also the right path
3. Focus on performance where possible, knowledge when required

The key to the above is a holistic test strategy

Session Objectives

By the end…you will be able to

- Describe the benefits of performance testing over knowledge testing
- Describe considerations for "chunking" and specifying performance tests
- Create a (partial) performance test for a simple "duty"

Testing Misconceptions, Principles, and Ideas

Common Practices

- Rely on SME judgment for thresholds and "cut-scores"
- Rely on "objective tests" because they are
  - Easier to create
  - Easier to grade
  - Easier to deliver
  - More familiar
- Testing to accumulate records for compliance purposes

Principle: There is No Such Thing as an Objective Test

All tests involve (at least some) judgment—no test is completely objective. Decisions affecting accuracy include

- Individual items
- Test conditions
- Number of items
- Number of attempts allowed/required
- Scoring
- Politics

Principle: Tests Test Whatever You Ask the Performer to Do

Testing Like This…	Tests
Multiple choice tests	Test guessing and "process of elimination" skills
Recall tests	Test ability to remember
"Going Through the Compliance Motions"	Level of cynicism
Observing performance	Ability to perform

Concept: Norm-Referenced vs. Criterion-Referenced

Criterion-referenced is not the same as "performance-based"

- May have a knowledge test that is labeled as "criterion-referenced"
- Norm-referenced tests may still be standardized

The real issue is what is being tested

- Performance
- Knowledge

And what are you measuring

- Accuracy?
- Fluency?
- Recall?
- Problem-solving?

Principle: Any Test Used as a Gate May be Challenged

Key legal issues

- Validity
  - External: Is your test appropriate for the performance for which it is assessing? Can you show the link between requirements and test items?
  - Internal: Does the test really measure what it is trying to measure?
- Bias—Will people from different population segments (e.g., gender, race, age) score differently on the same test?

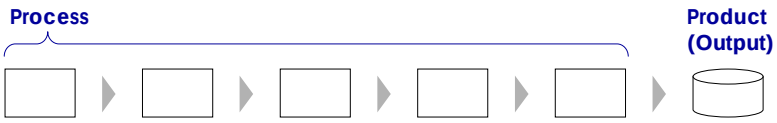
Principle: There are Always Trade-Off Decisions

Practical Considerations

- Time
  - Trainee
  - Evaluator
  - Developer
- Control of test content and answers
- Fidelity vs. convenience
- Organizational tolerance for testing

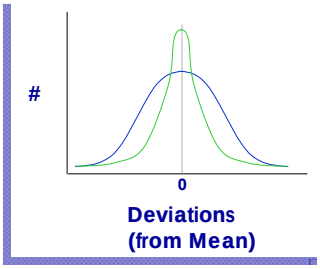
How Engineers Test Software and Products

Testing Strategies in Manufacturing



Manufacturing testing tends to focus on predicting the likelihood of producing defective products.

- 100% inspection vs. statistical sampling
- Finished vs. in-process inspection vs. critical sensor/process safeties checks
- Accept/reject of entire batch based on sampling results
- Stringent inspection of incoming materials



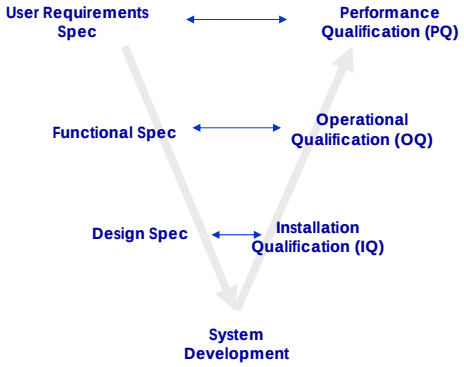
Performance versus Functional or Operational Testing (i.e., how the system performs it's mission vs. how well it operates)

Tests begin in the design process.

Performance: Does the system do what it was intended to?

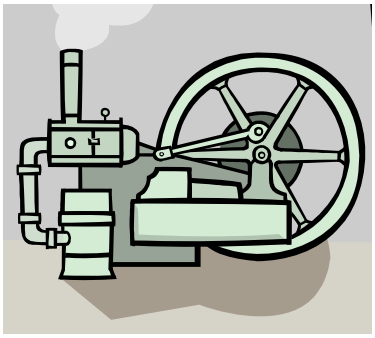
Functional: Does the part/system work the way we intended?

Design: Is the part built as designed (e.g., dimensions)?



"Bottom-up" versus "Top-down" Testing—NASA Example

One of the practices alleged to have contributed to the Challenger shuttle* explosion was the discontinuation of "bottom-up" testing.



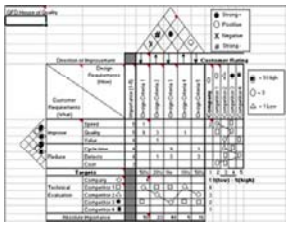
- Why would this practice have been discontinued?
- What is the parallel in training or HPT solution developmental testing?

* Minority/dissenting opinion by Rogers Commission member and Nobel Prize-winning physicist Richard Feynman—included in the Appendix of the commission's report.

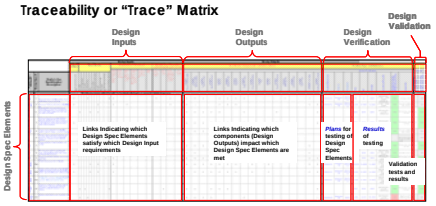
Requirements "Traceability"

How product design engineers identify key characteristics and develop test plans for critical functions and attributes

Quality Function Deployment (QFD)



Traceability or "Trace" Matrix

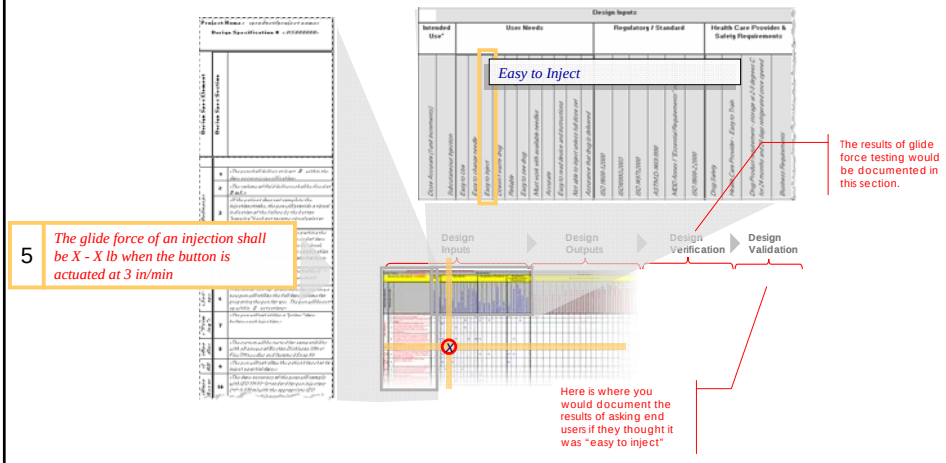


Failure Modes and Effects Analysis (FMEA)

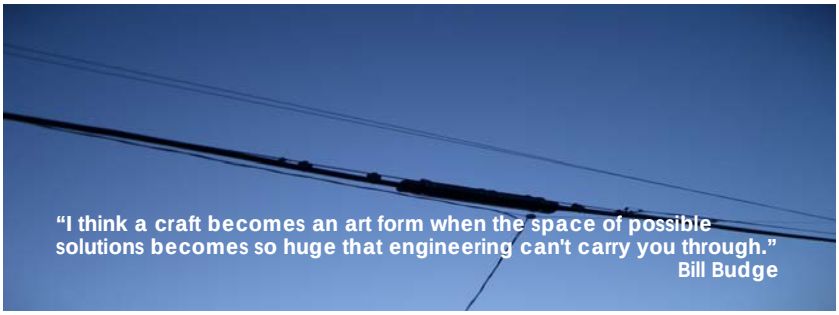
Item #	Step/Station Number	Step/Station Name	Step Function	Potential Failure Mode	Hazard	Potential Causes	O C C	Effect of Failure: Effect on Product	Effect of Failure: Effect on User	S E V	Controls: Prevent	Controls: Detect	Current Control Verification	D E T	R P N	Risk Acceptability	Recommended Actions/Status/ Owner	Resulting: SEV	Resulting: OCC	Resulting: DET	Resulting: RPN	FMEA Rev
A	Station 30 - Injecting & Deep	Step V.1 Press Injection Button and count slowly to 5	Complete injection stroke and wait for full delivery	Body part C moves on Body parts A and B (label intact)	N/A	Patient damages glue joint during handling	2	Body C moves side to side for Label holds Body C in linear direction enough to maintain dose accuracy	Patient Dissatisfaction	5	None	None		6	60		Update Manual	5	1	4	20	
B				Body C moves on Body A and B (label not intact)	N/A	Patient damages glue joint during handling	1	Body C capture point moves relative to lead screw head	Underdose	10	None	None		8	80		Update Manual	10	1	4	40	

Example of Using Requirements to Drive Testing

This example zooms in on a single user need "easy to inject" and shows how it is converted to "engineering language."

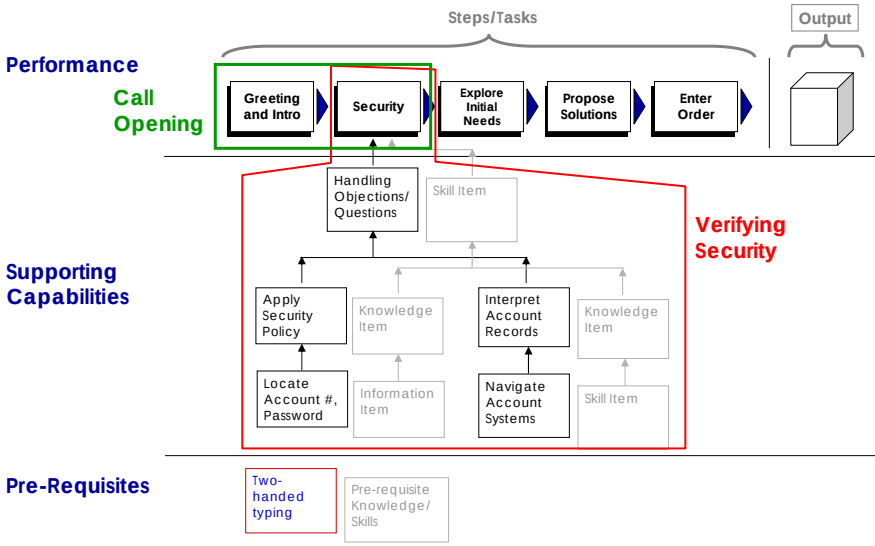


Guidelines for Testing—A Practitioner's View—Which Test Strategy to Use?



 Page 18

The Elements of Performance—Remember the Learning Hierarchy?



 Page 19

Testing the Limits

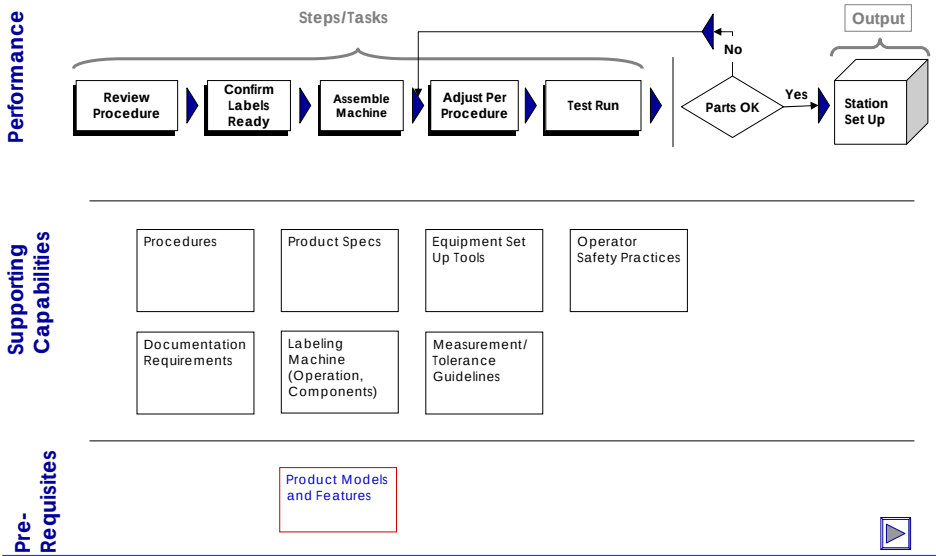
From FDA Guidelines on Process Validation:

"The test conditions for these runs should encompass upper and lower processing limits and circumstances, including those within standard operating procedures, which pose the greatest chance of process or product failure compared to ideal conditions; such conditions have become widely known as ***'worst case'*** conditions." (emphasis added)

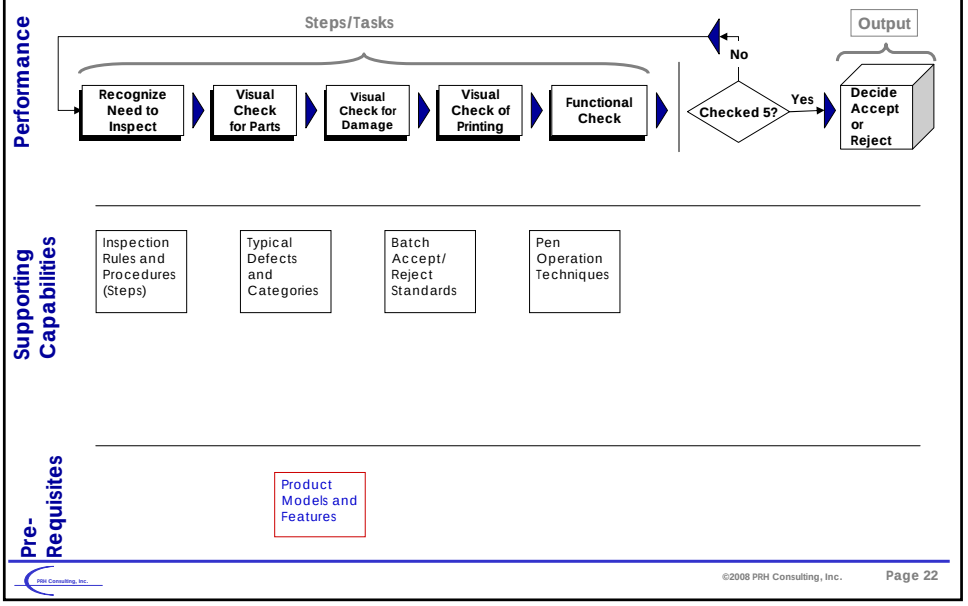
This means we need to

- Identify likely points of failure
- Create test items that verify capability in "worst case" (or boundary) conditions

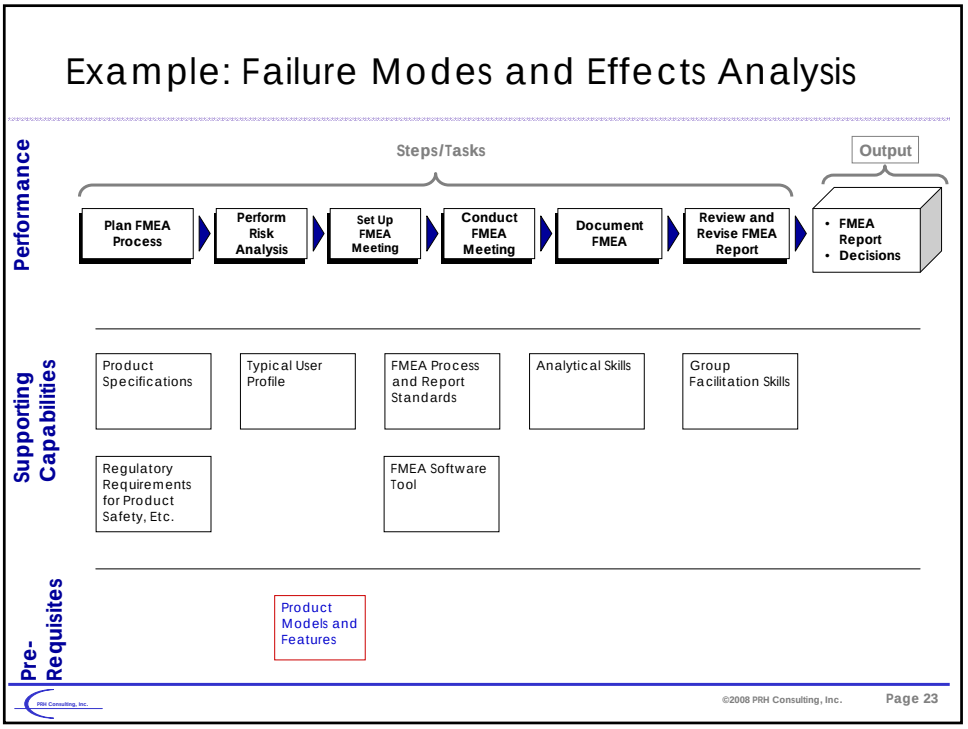
Example: Set Up Labeling Equipment



Example: Pen Inspection



Example: Failure Modes and Effects Analysis



Considerations for Performance Testing

The primary focus should be on performance of steps per defined criteria. Occasionally you may need to

- Include "what-if's?" and "why's?"
- Distinguish between critical and non-critical

List criteria as objectively as possible

Action Item No.	Activity	Expected Response(s) and/or Comments	Met	Not Met	PA Facilitator Initials (If Action Item is Performed Correctly)
PERFORM: SIMULATED FREQUENCY CHECKS AFTER MACHINE D, continued					
1.	OBSERVE check of Body A and B	• Identified any defects that were present	❑	❑	
2.	OBSERVE check of the dial	• Identified any defects that were present	❑	❑	
3.	ASK "Please describe three examples of dial defects for which you are inspecting."	• Answer included the following: o Arrow not centered o Arrow not completely visible o Damage or deformities to dial	❑	❑	

Describe the steps the ***assessor*** should take

Include space for documentation

Considerations for Knowledge Testing

Try to keep as focused on performance as possible

- Scenarios and situations
- Use of photos, hands-on materials
- Open book

Downside is that you often need to

- Create multiple tests or test items
- Control (hide) the questions and answers
- Make the right answer a "clear-cut" decision

How Many Checks Until You Know?

Sampling theory doesn't help us much here…

- Typically, not enough volume
- Typically, not enough opportunity to run the process to determine sampling requirements

Test Every Capability?

- Yes—at least until effectiveness is proven
- Performance tests allow testing of bundled sets of multiple capabilities

Test Only Once?

No—need to test both clear and borderline (or "challenge") conditions.

 Page 26

Knowledge Test Spreadsheet

WORKING

Capability	Type	Test Item	Answer
Identify typical defects	Match	Photo A Photo B Etc.	Obvious defect No defect Borderline defect
Classify defects	Multiple Choice	Photo D Photo E Etc.	Clear category A Clear category B Borderline category
Decide if batch should pass/fail	Binary (Like True-False)	Example Results A Example Results B Example Results C	Clear pass Clear fail Borderline fail
Etc.			

MANAGING

Balance by Capability

- % Items
- # Items

Balance by Question Type

- % Items
- # Items

 Page 27

Summary and Q&A

Summary

- Testing human performance and learning is often an afterthought.
- We should do more testing of performance and less testing of knowledge.
- If we use knowledge tests, still try to target performance.
- We need to be clear about the intent of the test (i.e., gate for performance? For training? Feedback for learner?)
- Test strategically—it adds "cost" but limited value to the learner.

For More Information . . .

Contact

Peter R. Hybert
PRH Consulting, Inc.
20 Danada Square West, #102
Wheaton, IL 60187
phone: 630.682.1649
email: pete@prhconsulting.com
web: www.prhconsulting.com

See Also

Articles available on the website (or by request)

- *"It Only Counts if You Can do the Job!,"* by Peter R. Hybert and Kelly R. Smith, (originally authored in 1999)
- *"Project Profile: Designing a Performance Measurement System"* by Peter R. Hybert, (January 2006)
- *And more…*

Other resources

- Feyman Challenger Report Appendix http://www.ralentz.com/old/space/feynman-report.html
- *"Criterion-Referenced Test Development; Technical and Legal Guidelines for Corporate Training,"* by Sharon Schrock and William Coscarelli
- Practical Assessment Research and Evaluation (on-line, peer-reviewed journal) http://pareonline.net/
- FDA Guidelines on Process Validation http://www.fda.gov/cder/guidance/pv.htm

Session Presenter

Peter R. Hybert

Pete has been in the human performance improvement field since 1984 and has been a consultant since 1989.

His clients include many Fortune 500 firms. He has analyzed, designed, and developed training and development for almost every type of business function and process.

Pete is the author of more than twenty articles and has presented more than twenty times at international conferences and local chapters of ISPI, ASQ, and ASTD. He has also served as the chairperson for ISPI's Awards of Excellence Committee and a President of the Chicago Chapter of ISPI.

- Analyzed over 100 jobs and work processes
- Designed over 40 modular curriculum architectures and developed over 50 training and performance support solutions
- Designed, developed, and implemented qualification systems and instruments for engineers, technicians, validation specialists, project managers, and service engineers
- Designed, developed, and implemented performance-based competency systems, including assessment tools, coaching tools, and links to curriculum for leadership and technical audiences

