Case Studies in New Project Governance Quantitative Techniques:

lessons-learned give rigour & efficiency to ordnance, aircraft & ship testing

UNSW AUSTRALIA Canberra
Never Stand Still

Capability Systems Centre, School of Engineering and Information Technology

Dr Keith Joiner
k.joiner@adfa.edu.au
+61 499 202 284

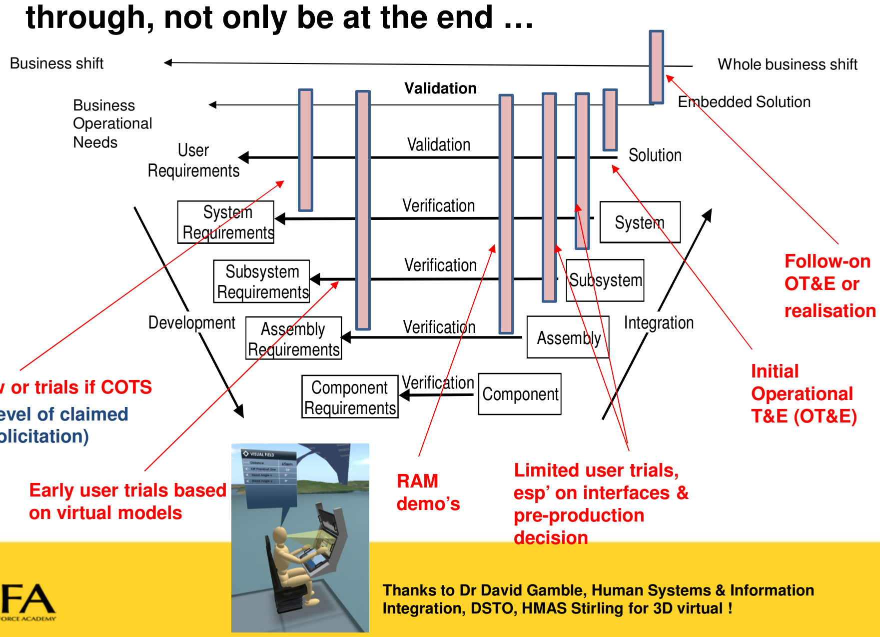
1) **Why** we need better quantitative test & evaluation techniques for project governance (10 min)
2) **What** are six-sigma test techniques used by U.S. DoD & major industries (5 min)
3) **Building** competencies in test design & analysis with the techniques (3 min)
4) Short video of revised subject (7 min)
5) **Case studies** from Defence (10 min)
6) Questions (10 min)

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

1

1) Why

Project Governance needs to Realise Benefits (validate) all the way through, not only be at the end …



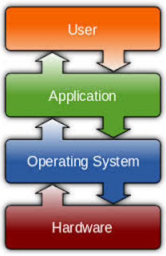
ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

Thanks to Dr David Gamble, Human Systems & Information Integration, DSTO, HMAS Stirling for 3D virtual !

UNSW AUSTRALIA Canberra

2

1) Why Information Technology Difficulties pervades software-intensive systems & functionality



IT projects have high risk of failure due to project scoping not examining before contract:

- technical risks &
- user expectations

Research by Flyvberg & Budzier (2011) into 1,471 IT projects showed that averages were not unremarkable to other projects (i.e., cost overrun 27%) but there was a 'fat tail' of risk

> '*Fully one in six of the projects in the sample was a Black Swan, with a cost overrun of 200%, on average, & a schedule overrun of almost 70%.*'

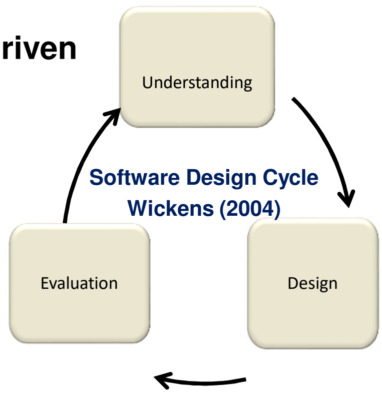
Software development very capable of rapid prototyping & thus Preview T&E

Wickens et. al. (2004) cite early research showing:

- user performance improves about 12 percent with each design iteration &
- average time to perform software-based tasks decreases about 35 percent from the first to the final iteration.

3

1) Why Iterative Usability T&E is key for software-driven functionality



- Use human factors design principles (for computer interaction);
- Use key software usability metrics;
- PT&E before contract
- DT&E regular usability testing against agreed usability metrics, to include rating all software functions according to:
  - frequency of use,
  - whether mandatory or discretionary, &
  - knowledge-level of the user required;
- Ideal number of participants in usability testing of software is **5-6 persons** per evaluation.
- Recommended number of iterations to achieve an optimum cost/benefit ratio is indicatively **five** for a typical software system involving 1000 users
- Critically important to run with 'commanders/managers' for trusted autonomous systems

4

1) Why Software Testing Metrics

- **Effectiveness**
- Percent of tasks completed
- Ratio of success to failure
- Number of features or commands used

- **Efficiency**
- Time to complete a task
- Time to learn a task
- Time spent on errors
- Percent of errors
- Frequency of help features or documentation
- Repetition of failed commands

- **User Satisfaction**
- Rating scale for usefulness of software
- Rating scale for satisfaction with features/functions
- Number of times user registers dissatisfaction
- Rating scale for user versus computer control of task
- User perception that software supports tasks as needed

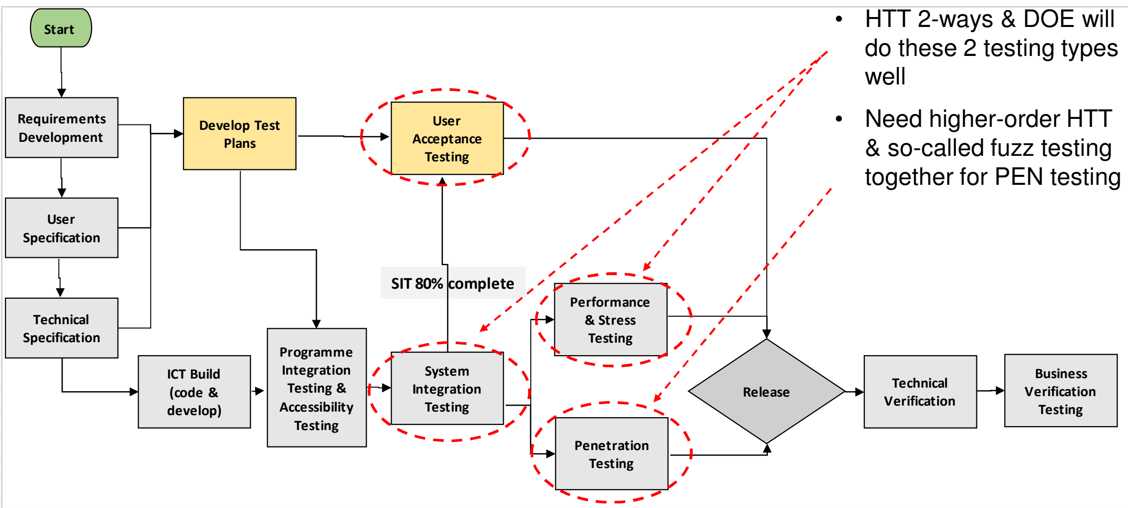
ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

Adapted from Wickens et. al. 2004

UNSW AUSTRALIA Canberra

5

1) Why Example ICT or Software-Intensive Project Lifecycle



- HTT 2-ways & DOE will do these 2 testing types well
- Need higher-order HTT & so-called fuzz testing together for PEN testing

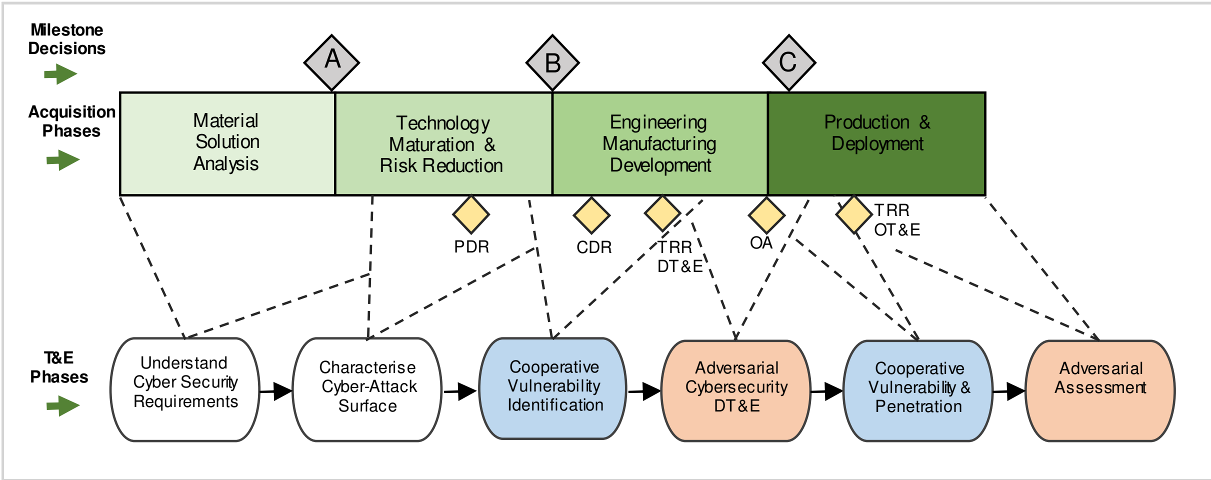
ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

Adapted from Terrell (2016), GM of PPO at DHS, Aust' Gov' from PGCS 2016

UNSW AUSTRALIA Canberra

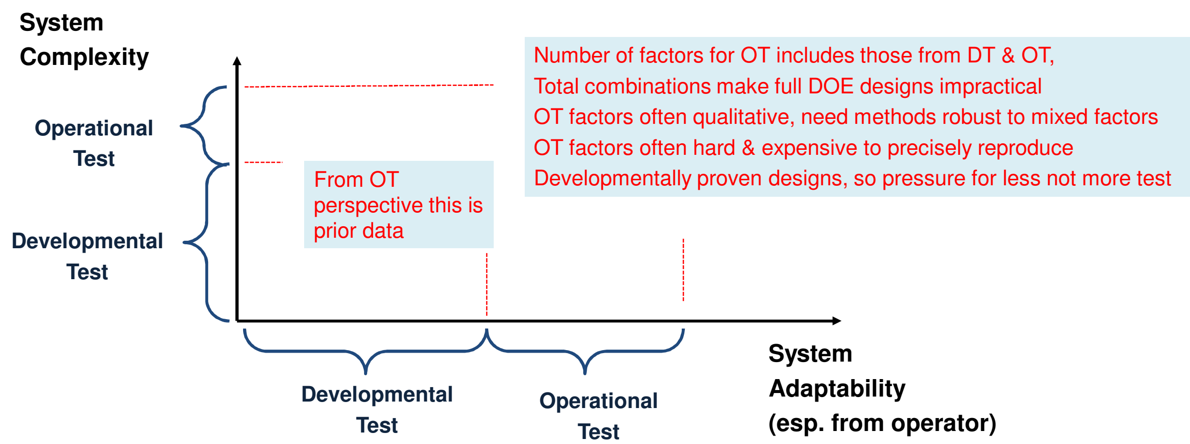
6

1) Why Cybersecurity Governance activities in U.S. DoD Lifecycles



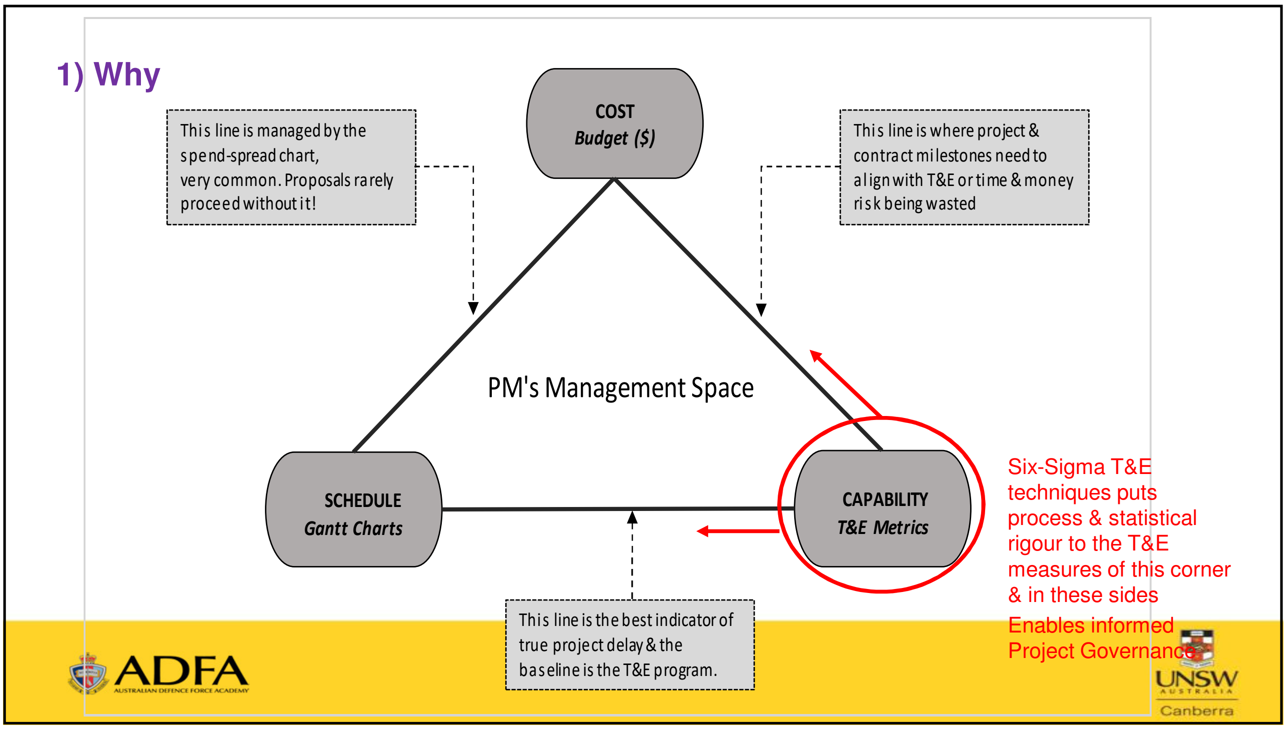
7

1) Why Operational complexity in systems-of-systems interconnectedness

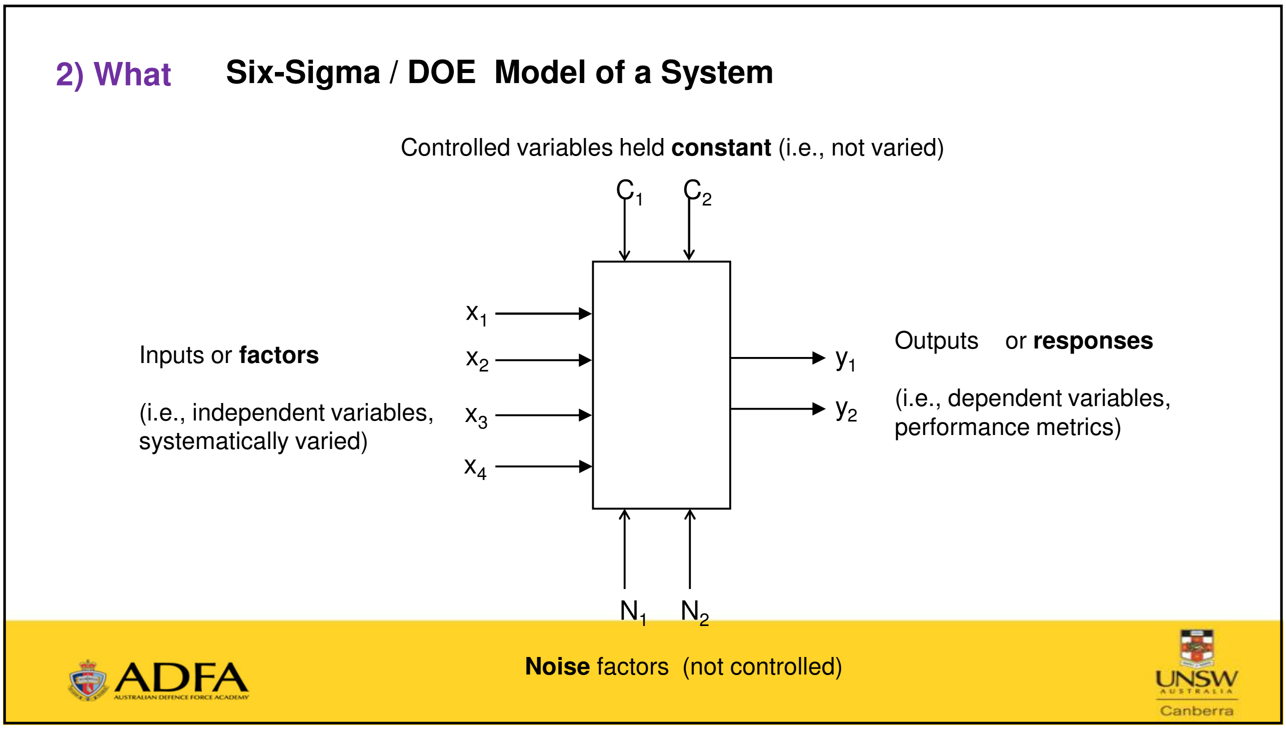


See also Kass (2015) in ITEA Journal

8


1) Why
COST
Budget ($)
This line is managed by the spend-spread chart, very common. Proposals rarely proceed without it!
This line is where project & contract milestones need to align with T&E or time & money risk being wasted
PM's Management Space
SCHEDULE
Gantt Charts
CAPABILITY
T&E Metrics
This line is the best indicator of true project delay & the baseline is the T&E program.
Six-Sigma T&E techniques puts process & statistical rigour to the T&E measures of this corner & in these sides
Enables informed Project Governance
ADFA
UNSW
Canberra

9


2) What
Six-Sigma / DOE Model of a System
Controlled variables held **constant** (i.e., not varied)
C_1 C_2
Inputs or **factors**
(i.e., independent variables, systematically varied)
x_1
x_2
x_3
x_4
y_1
y_2
Outputs or **responses**
(i.e., dependent variables, performance metrics)
N_1 N_2
Noise factors (not controlled)
ADFA
UNSW
Canberra

10

2) What

Screening is to:

- Determine the significant factors that effect responses & focus future testing & effort on those (**efficient, rigourous**)
- Determine if the factors effect absolute value (mean) or variation (spread-shifters), or both (**smart**)
- Factors that only effect variation are insidious, usually not intuitive & only DOE/DFSS approach find these (**operational advantage**)

Modelling is to:

- Optimise – either the design (**iterate**) or its later use (operational advantage)
- Manage system through life especially adaptive threats & with other systems (hopefully also modelled)
- Take deterministic models (theory) into real-World (probabilistic)

Validating Performance

- Confirms system & model work in representative environment, users & missions, meaning with representative sampling (statistical)

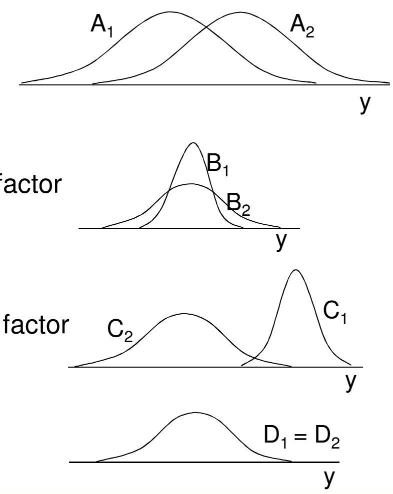
ADFA AUSTRALIAN DEFENCE FORCE ACADEMY UNSW AUSTRALIA Canberra

11

2) What

Screening

1) Average-shifting factor
2) Variance or spread-shifting factor
3) Average & variance shifting factor
4) Factor with no effect

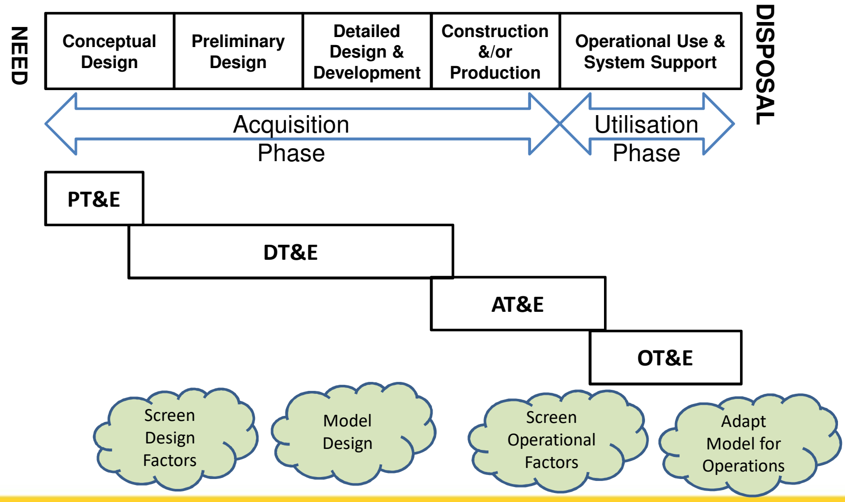


Figures from Reagan & Kiemele (2007)

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY UNSW AUSTRALIA Canberra

12

2) What System Life Cycle

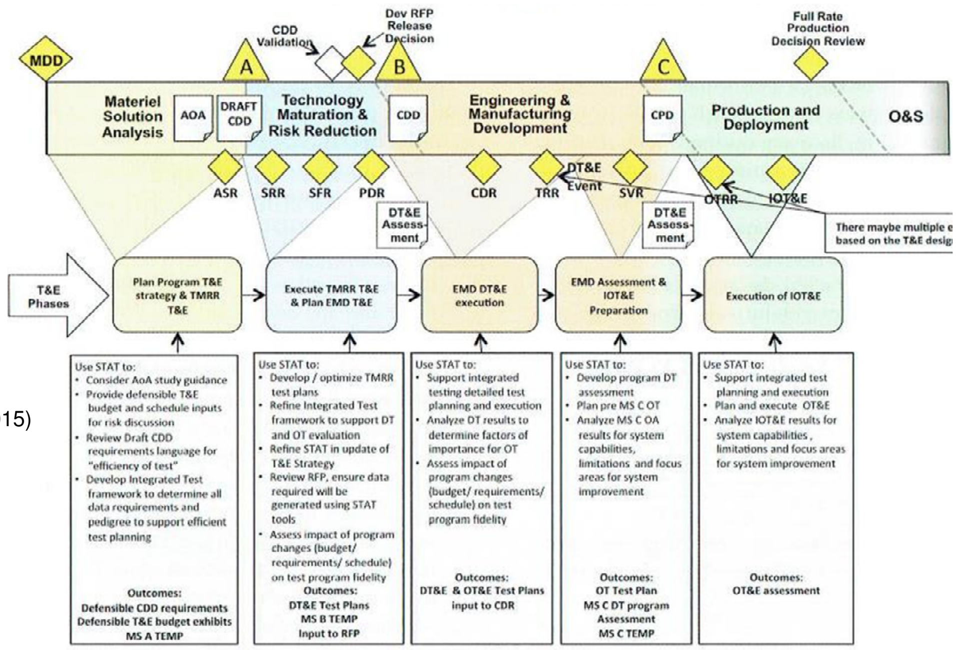


ADFA

UNSW Canberra

13

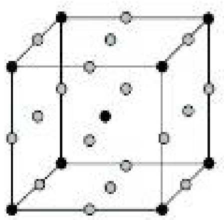
2) What

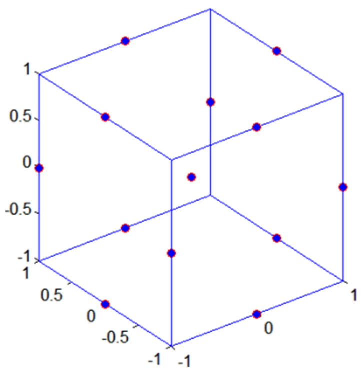


Murphy et. al. (2015)
ITEA Journal

14

2) What Multi-factor test design & analysis





Box-Benhken Design

Compared to one-factor-at-a-time methods used in older '*standards-based*' protocols:

- Significantly less test points = highly **efficient**
- Improved **rigour**
- Confidence **limits** as well as means
- Orthogonality gives **independence** in test factor effects & interactions = **certainty** for optimising
- In industry gives **competitiveness**
- In military gives **overmatch** for near-peers
- Builds better operational **models** for through-life
- Applies equally to service industries

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

Right picture courtesy Mathworks: http://au.mathworks.com/help/stats/bbdesign.html?s_tid=gn_loc_drop

Left picture from: http://www.gmpua.com/World/Manu/07/i.htm

UNSW AUSTRALIA Canberra

15

2) What Example L_{12} screening – testers should know

Quantum XL

Taguchi Design

L12 (11 two-level factors array)

11 Factors in 12 Runs

Number of replicates: 4

	USL	
enter spec limits (optional) -->	LSL	

	A	B	C	D	E	F	G	H	I	J	K
Run	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	-1	-1	1	1	1	1	1	1
3	-1	-1	1	1	1	-1	-1	-1	1	1	1
4	-1	1	-1	1	1	-1	1	1	-1	-1	1
5	-1	1	1	-1	1	1	-1	1	-1	1	-1
6	-1	1	1	1	-1	1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	1	-1	1	-1
8	1	-1	1	-1	1	1	1	-1	-1	-1	1
9	1	-1	-1	1	1	1	-1	1	1	-1	-1
10	1	1	1	-1	-1	-1	-1	1	1	-1	1
11	1	1	-1	1	-1	1	-1	-1	-1	1	1
12	1	1	-1	-1	1	-1	1	-1	1	1	-1

optional) --> *<-- enter fact* *data here -->*

Output 1					
Y1	**Y2**	**Y3**	**Y4**	**Y-bar**	**S**
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####
				####	####

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

OFAT would require 2^{11} = 2048 runs

UNSW AUSTRALIA Canberra

16

3) Building SS Test competencies in U.S. DoD mandatory by OSD from 2009

- Applies to 5.7% of SE workforce in projects & test units
- Administered by DAU
- Necessary for 8 of 25 (32%) of the T&E competencies involving data rigour

Field	T&E Competence	Number & Type of Competency Elements
T&E Planning	2) Capabilities assessment	4 elements: translating capability requirements into evaluation criteria, adequacy of capability requirement definition for testing, & determining data & T&E infrastructure requirements.
Test Execution	8) Test control management	4 elements: confirm integration of data collection tools, instrumentation, M&S & system under test, monitor security & safety compliance, organise test rehearsals & executions, & control test schedule to complete scenarios with regard to priorities & objectives.
	9) Data management	3 elements: verify data is collected, documented & archived securely, ensure validity of data against objectives & distribute for appropriate analysis
Analysis	10) Data verification & validation	2 elements: translate outputs from test instrumentation & identify gaps & variances in raw test data to determine voids or outliers.
	11) Data reduction & assimilation	3 elements: reduce, translate & analyse raw test data, conduct data scoring, & align data to test objectives.
Evaluation	12) Determining test adequacy	2 elements: confirm M&S & tests credible support test objectives, especially accredited M&S supplementing live data.
	13) Validation of test results	2 elements: determine M&S & test data credibly supports T&E metrics.
	14) Evaluative conclusions	3 elements: confirm test data can support the evaluation framework in the TEMP, relate T&E results to performance results & operational significance, & examine integration of systems & consequence to larger systems of systems.

17

3) Building Australian Test Design competency efforts (2012-2015)

- 2009 US DoD (DOT&E) make DOE competencies & methods mandatory
- 2011 ITEA conference: AFOTEC, MCOTEA, COMOPTEVFOR, ATEC & JITC abuzz with the reform
- 2012-2014 > 60 ADF T&E students trained in US DoD methods
- No policy back-up, no assessment or mentoring
- 2015 - one forced trial (Joiner, McAuliffe & Kiemele) indicating:

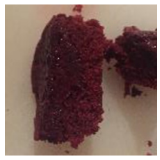
- Great rigour & discrimination
- Very good efficiency
- Collaborative basis for shared testing
- Need for early reinforcing use with training

18

3) Building

Revised Curriculum

- 2016 University of New South Wales adopts course (Joiner & Brewster) with:
  - Collaborative test assessment in intensive week
  - Follow-up assessed research assignment – work or hobby
  - Targeted attendance to new inductees to test units
  - Selective attendance by senior testers



ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

Full video at https://youtu.be/RVBYWHOarRE

UNSW AUSTRALIA Canberra

19

UNSW CANBERRA

Advanced Test and Evaluation Techniques Course (ZEIT 8034)

Postgraduate Coursework (Master)

School of Engineering and IT
and Capability Systems Centre
University of New South Wales, Canberra

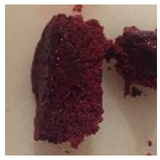
V180109

20

3) Building

Revised Curriculum

- 2018 Initiatives
  - Autonomous robot replaces catapults
  - Defence applications
  - Quantum XL (more capable)
  - Cybersecurity application explained
  - Logistics regression necessity in Defence explained



ADFA Full video at https://youtu.be/RVBYWHOarRE UNSW Canberra

21

2) More What

Even more efficiency in screening & validation testing …

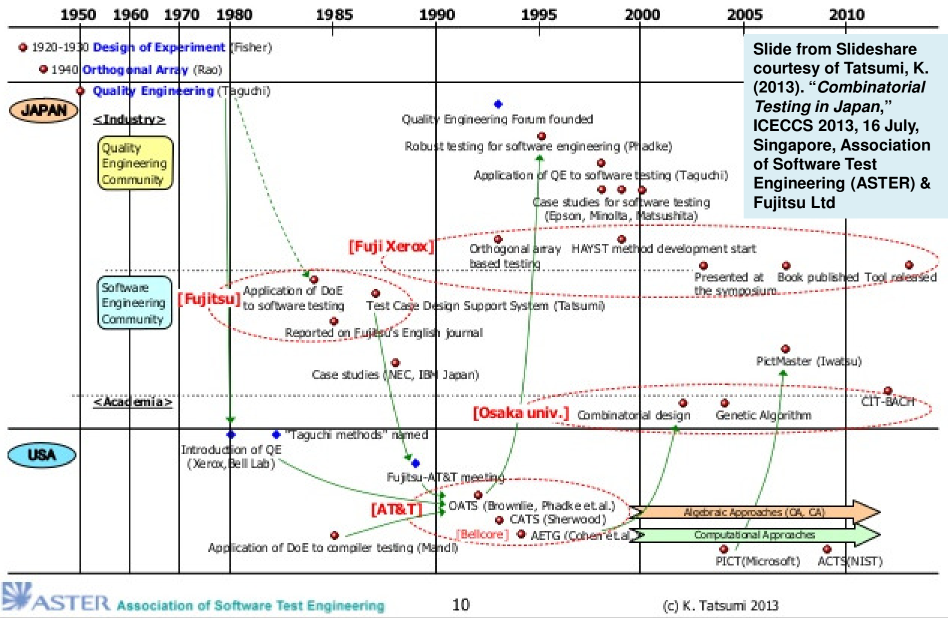
Known as:

- Combinatorial Testing or
- High Throughput Testing

Powerful large factor test analysis technique

Used especially in U.S. for cybersecurity testing vice fuzz testing

Easy to teach six-sigma testers, similar skills



22

2) More What

Includes Representative Sampling for efficient but rigorous V&V

i.e., **Nearly Orthogonal Latin Hypercube Design (NOLHD)**

- especially useful for checks of contractor models, simulations or limited checking real systems
- Example shown cover a four factor aircraft survivability envelope in 28 test points

NOLHS Design

Name	Aspect	Airspeed	Altitude	Threat Range
Lower	0	150	1000	5
Upper	180	300	10000	20
1	0	150	10000	12.5
2	30	150	4000	7.5
3	60	200	8500	12.5
4	90	300	1000	20
5	120	150	5500	20
6	150	250	2500	10
7	180	175	1000	5
8	0	275	4000	15
9	30	200	8500	17.5
10	60	275	10000	15
11	90	175	7000	12.5
12	120	200	1000	12.5
13	150	175	4000	17.5
14	180	275	8500	7.5
15	0	175	1000	17.5
16	30	300	5500	7.5
17	60	300	7000	15
18	90	225	2500	20
19	120	225	2500	5
20	150	225	10000	17.5
21	180	300	7000	15
22	0	225	5500	5
23	30	250	2500	10
24	60	275	4000	7.5
25	90	200	10000	10
26	120	250	8500	5
27	150	250	5500	20
28	180	150	7000	10

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

UNSW AUSTRALIA Canberra

23

3) Building

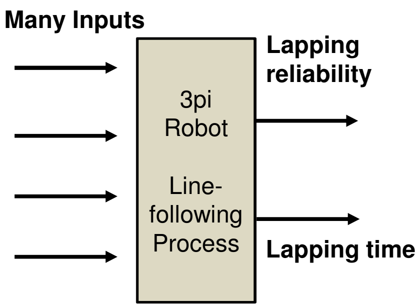
Course Resources

Course Texts:

- Antony (2014), *DOE for Engineers and Scientists*
- Reagan & Kiemele (2008), *Design for Six Sigma* (Blue text)

Software:

- Quantum XL (test works)
- rdExpert (Lite) (test works)
- Example Datafiles
- Each syndicate uses:
  - Atmel 7.0 [*]
  - Pololu Programmer Library Package



ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

UNSW AUSTRALIA Canberra

24

4) Case Studies

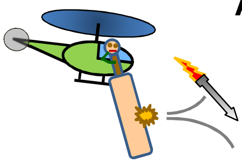
1. Brooks, M. & Joiner, K. F., 2017. Small Arms Terminal Effects: Implementation of a DOE methodology in order to design small arms terminal effects tests, PARARI 2017.
2. Grafton, R.; Brown, G.; Novakovic, Z. & Joiner, K.F., 2018. Developments in test design and analysis techniques for aircraft survivability assessments, SETE 2018
3. Hanly, G. & Joiner, K. F., 2018. Test and Evaluation of Factors affecting Facial Recognition Performance of Systems, SETE.
4. Henry, R. & Joiner, K. F., 2017, Improving Small Arms Ammunition Qualification with Statistical Test Techniques from U.S. Defense, PARARI 2017
5. Sisson, G. & Joiner, K. F., 2018. Applying new analysis techniques to calibrate use of a projectile locating system in ammunition testing, SETE 2018.
6. Wernas, M. & Joiner, K. F., 2018 (submitted). Screening the important factors in Supportability Test and Evaluation activities for Ships‘ SETE 2018

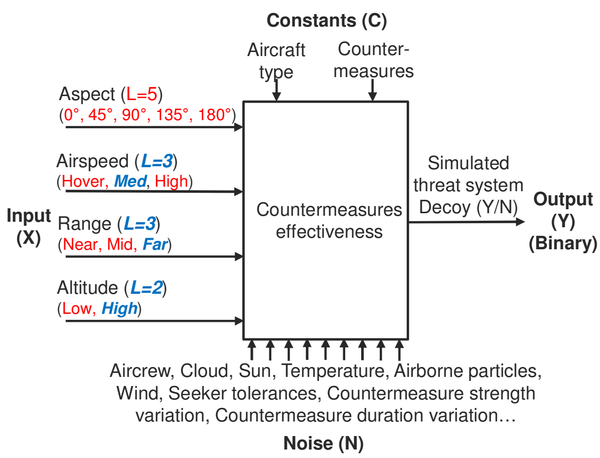
ADFA
AUSTRALIAN DEFENCE FORCE ACADEMY

25

4) Case Studies

Aircraft Survivability





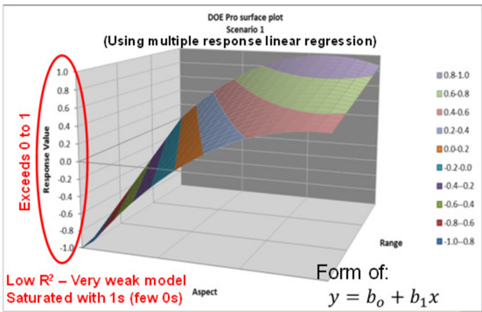
Indicative test matrix

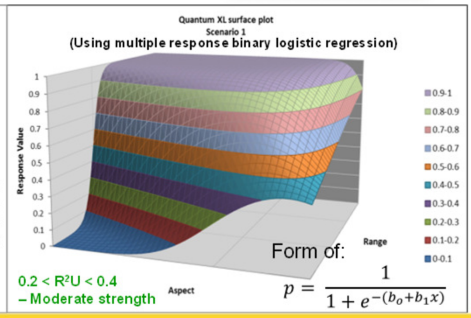
Source	Test No.	A : Aspect	B : Airspeed	C : Range	D : Altitude
DOE	1	180	Hover	Near	Low
	2	180	Med	Mid	High
	3	180	High	Far	Low
	4	135	Hover	Near	High
	5	90	Hover	Mid	Low
	6	90	Med	Far	High
	7	45	Hover	Far	High
	8	45	Med	Near	Low
	9	45	High	Mid	High
	10	0	High	Near	High
	11	135	High	Far	Low
	12	90	High	Near	Low
	13	0	Hover	Mid	Low
	14	0	Med	Far	Low
	15	135	Med	Near	High
	16	135	Med	Mid	High
Legacy (must do)	17	180	Hover	Mid	Low
	18	180	High	Mid	Low
	19	135	Hover	Mid	Low
	20	135	High	Mid	Low
	5	90	Hover	Mid	Low
	21	90	High	Mid	Low
	22	45	Hover	Mid	Low
	23	45	High	Mid	Low
	13	0	Hover	Mid	Low
	24	0	High	Mid	Low

26

4) Case Studies Aircraft Survivability

- Most outcomes decoy/no decoy - poor linear regression results.
- DOE PRO XL™ marginal means can still show general trends.
- Binary logistic regression using Quantum XL™.





ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

UNSW AUSTRALIA Canberra

27

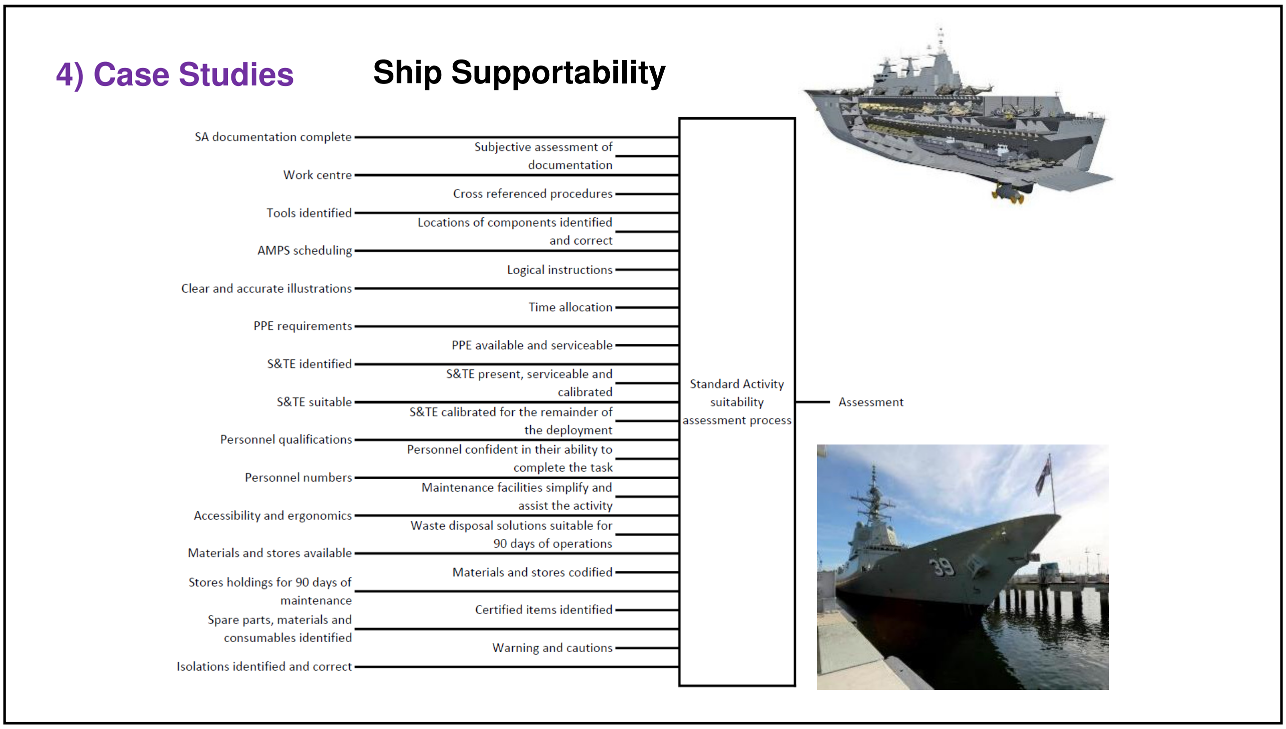
4) Case Studies Gaining Traction in JEWOSU

Year	2016	2017	2018 and beyond
Phase	Test case	Confidence building	Rollout → Full implementation
Application	• 1 trial	• 4 trials (mixed legacy & DOE in critical cases) • 2 historical analyses	• All trials must consider DOE • In-depth historical analyses • Countermeasure optimisation
Regression type	• Linear	• Linear • Binary logistic	• Linear • Binary logistic • Ordinal logistic
Training	• Advanced T&E techniques, UNSW Canberra	• Advanced T&E techniques, UNSW Canberra • Ad-hoc binary logistic regression	• Advanced T&E techniques, UNSW Canberra • Design for Six Sigma, Air Academy Associates

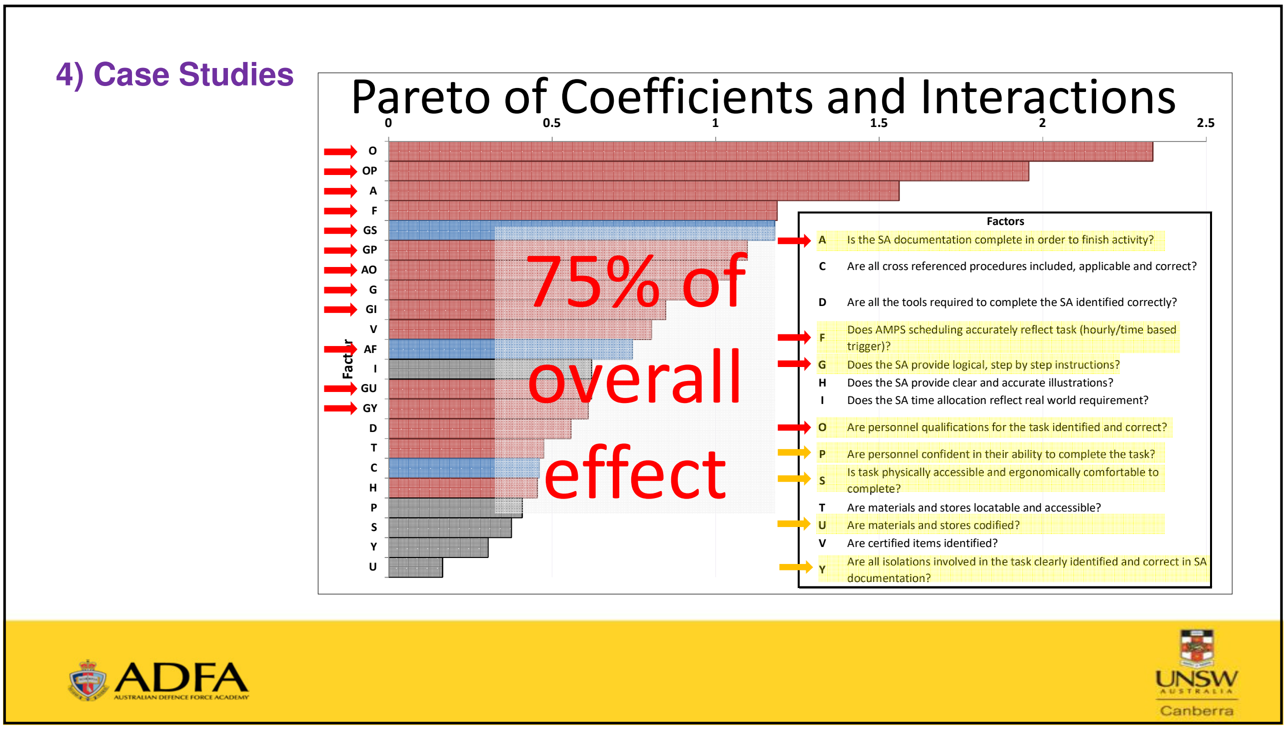
ADFA AUSTRALIAN DEFENCE FORCE ACADEMY

UNSW AUSTRALIA Canberra

28


4) Case Studies
Ship Supportability
SA documentation complete
Work centre
Tools identified
AMPS scheduling
Clear and accurate illustrations
PPE requirements
S&TE identified
S&TE suitable
Personnel qualifications
Personnel numbers
Accessibility and ergonomics
Materials and stores available
Stores holdings for 90 days of maintenance
Spare parts, materials and consumables identified
Isolations identified and correct
Subjective assessment of documentation
Cross referenced procedures
Locations of components identified and correct
Logical instructions
Time allocation
PPE available and serviceable
S&TE present, serviceable and calibrated
S&TE calibrated for the remainder of the deployment
Personnel confident in their ability to complete the task
Maintenance facilities simplify and assist the activity
Waste disposal solutions suitable for 90 days of operations
Materials and stores codified
Certified items identified
Warning and cautions
Standard Activity suitability assessment process
Assessment

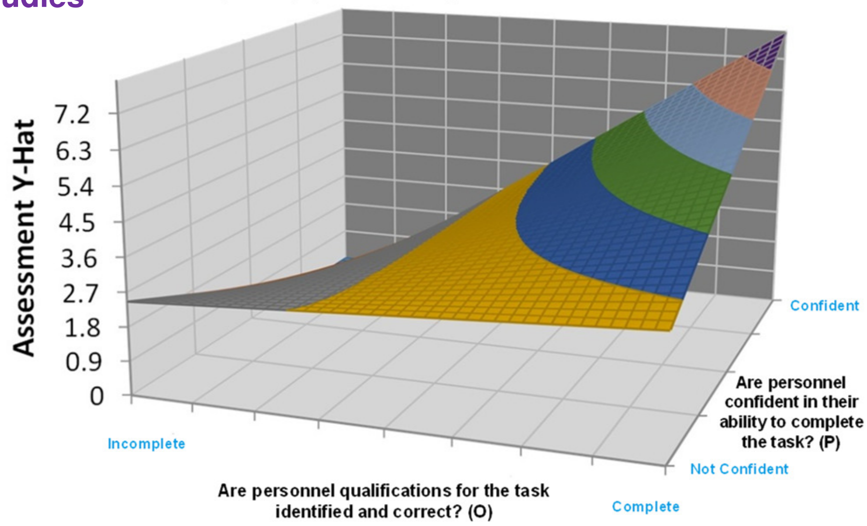
29


4) Case Studies
Pareto of Coefficients and Interactions
0
0.5
1
1.5
2
2.5
O
OP
A
F
GS
GP
AO
G
GI
V
AF
I
GU
GY
D
T
C
H
P
S
Y
U
Factor
75% of overall effect
Factors
A Is the SA documentation complete in order to finish activity?
C Are all cross referenced procedures included, applicable and correct?
D Are all the tools required to complete the SA identified correctly?
F Does AMPS scheduling accurately reflect task (hourly/time based trigger)?
G Does the SA provide logical, step by step instructions?
H Does the SA provide clear and accurate illustrations?
I Does the SA time allocation reflect real world requirement?
O Are personnel qualifications for the task identified and correct?
P Are personnel confident in their ability to complete the task?
S Is task physically accessible and ergonomically comfortable to complete?
T Are materials and stores locatable and accessible?
U Are materials and stores codified?
V Are certified items identified?
Y Are all isolations involved in the task clearly identified and correct in SA documentation?
ADFA
AUSTRALIAN DEFENCE FORCE ACADEMY
UNSW
AUSTRALIA
Canberra

30

4) Case Studies

Ship Supportability



ADFA AUSTRALIAN DEFENCE FORCE ACADEMY UNSW AUSTRALIA Canberra

31

4) Case Studies

So What

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY UNSW AUSTRALIA Canberra

32

Conclusions

- No real efficiency or rigour without new methods
- Inherently empowering in characterising, understanding & operational modelling
- Must involve real application in class & in workplace
- Improves T&E units & projects
- Inherently aligned to U.S. DoD
- Part of DNA of new cybersecurity T&E (at least US DoD & not the DSTG versions)
- Need the language at the very least

What next?

- The elusive Flight test units
- The elusive & critical project engineering managers (come on down CASG)
- Defence Industry (goes where Defence goes)
- The forever next year cybersecurity workforce
- T&E Policy backup

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY UNSW AUSTRALIA Canberra

33

Questions

Action & collaborative learning with line-following robots on ZEIT 8034 from CMU video:

https://youtu.be/RVBYWHOarRE

ADFA AUSTRALIAN DEFENCE FORCE ACADEMY UNSW AUSTRALIA Canberra

34