BIO
PRESENTATION

T4

May 20, 2004 10:00AM

MEASURING TESTING EFFECTIVENESS USING DEFECT DETECTION PERCENTAGE

Dorothy Graham
Grove Consultants UK

International Conference On
Software Testing Analysis and Review
May 17-21, 2004
Orlando, Florida USA

Dorothy Graham

Dorothy Graham is the founder of Grove Consultants in the UK, which provides advice, training and inspiration in software testing, testing tools and Inspection. Originally from Grand Rapids Michigan, she attended Calvin College and Purdue University, married an Englishman and has lived and worked in the UK for over 25 years.

Dorothy is co-author with Tom Gilb of "Software Inspection", Addison-Wesley, 1993, and co-author with Mark Fewster of "Software Test Automation", Addison-Wesley, 1999. She also initiated and co-authored several editions of "The CAST Report", covering Computer Aided Software Testing tools available in Europe.

Dorothy was Programme Chair for the first EuroSTAR Conference in 1993. She is on the boards of conferences and journals in software testing, and has been an active member of the British Computer Society's Specialist Interest Group in Software Testing since 1989, including a working party to produce a new software component testing standard. She was a founder member of the Software Testing Board of the Information Systems Examination Board (ISEB) of the British Computer Society.

She has been awarded the European Excellence Award in Software Testing.

Grove Consultants

Grove Consultants provides consultancy, training and inspiration in software testing. It comprises Dorothy Graham, Mark Fewster, Lloyd Roden and Clive Bates.

Grove Consultants bring a unique perspective to any organisation’s software testing processes. Because they specialise exclusively in testing, they have a great depth of knowledge in this field. With a broad range of clients, they bring an industry-wide perspective on testing with a focus on a European perspective. Consultancy assignments have included audits of testing practices, reviews of testing methodologies, assessment of testing effectiveness and efficiency, and advice on all aspects of testing, test automation and Inspection.

Grove Consultants also provide training courses in a range of software testing topics and are the UK's leading accredited training provider for the ISEB Software Testing Foundation and Practitioner Certificates, Europe's first formal qualification for testers.

For more information visit our web site: www.grove.co.uk

Measuring test effectiveness using Defect Detection Percentage (DDP)

Prepared by

Grove Consultants

Dorothy Graham
Grove House
40 Ryles Park Road
Macclesfield
SK11 8AH UK
Tel +44 1625 616279
dorothy@grove.co.uk

Mark Fewster
Llwyncynhwyra
Cwmdu
Llandeilo
SA19 7EW UK
Tel +44 1558 685180
mark@grove.co.uk

Lloyd Roden
95 Stonebridge
Orton Malborne
Peterborough
PE2 5NT UK
Tel +44 1733 239183
lloyd@grove.co.uk

Clive Bates
2 Nursery Close
Hilperton Village
Wiltshire
BA14 7RP UK
Tel +44 1225 761670
clive@grove.co.uk

www.grove.co.uk

Contents

Questions you may be asked

How good is the testing anyway?
Can you prove you are doing a good job?

Your testing can still be just as good in less time, can't it?
(That deadline pressure really didn't matter, did it?)

Is the testing any better for this release?
(Have we learned anything?)
(Have we really improved our testing?)

How many bugs have we missed?

Are we better or worse in our testing
compared to other groups/organizations?

3

Some questions for you

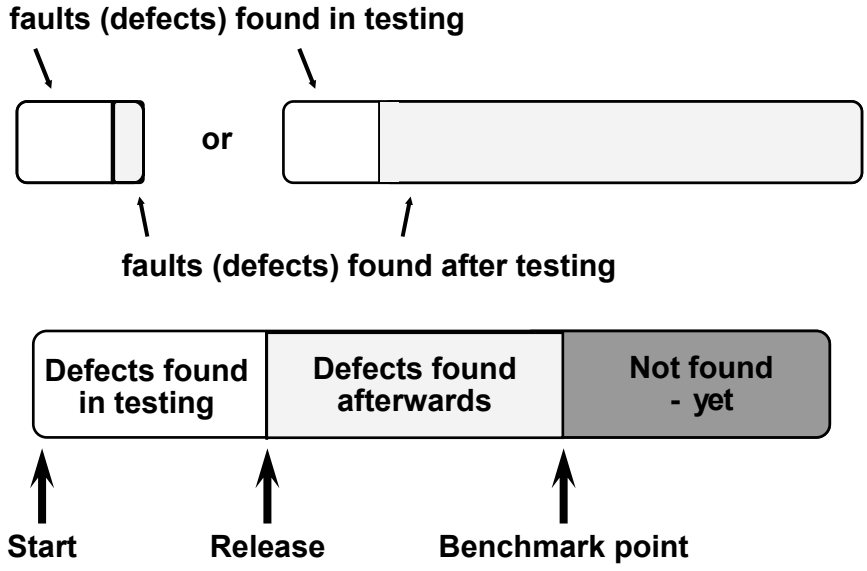
- **Do you keep track of defects?**
  - defects found in testing
    - different test stages,
      - e.g. system test, user acceptance test
    - different releases
      - e.g. testing for an incremental release in RAD
  - defects found in live running
    - reported by users / customers
- **Can you find these numbers from a previous project and your current project?**
- **Do you have a reasonable number of defects found?**

if so, you can use DDP to measure your test effectiveness

4

Contents

How effective are we at finding faults?



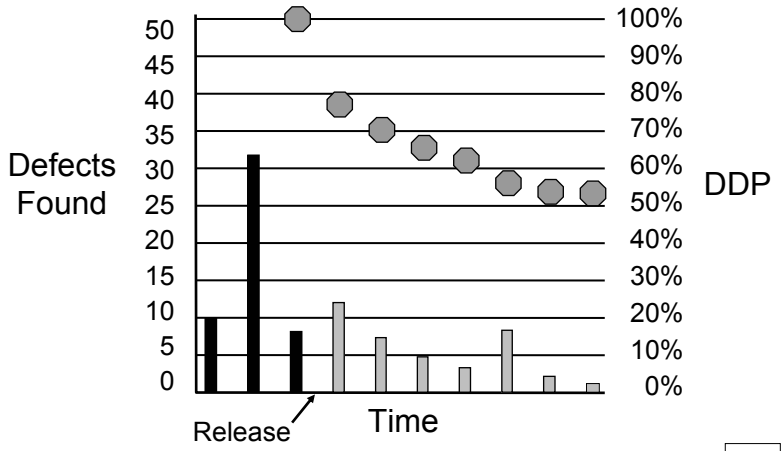
Defect Detection Percentage (DDP)

$$\frac{\text{Defects found by this testing}}{\text{Total defects including those found afterwards}}$$

- **"this" testing could be**
  - a test stage, e.g. component, integration, acceptance, regression, etc.
  - all testing for a function or subsystem
  - all testing for a system

7

Effectiveness at finding defects



Defects found in testing: 50 Defects found after testing: 38 Total defects found: 88

$$\text{DDP} = \frac{50}{88} = 57\%$$

8

www.grove.co.uk
011 44 8702 406172

Contents

- Introduction: some questions for you
- What is DDP
- Case studies
- Examples of calculating DDP
- Advice about DDP & Conclusion

Case studies from clients

	1 mo	10 mo
Year 1	70%	50% est
Year 2	92%	

Finance (insurance)

Operating system

System Test Group DDP = 38%
(before performance testing)

Priority 1 & 2 only: DDP = 31%

23% to 87% by application

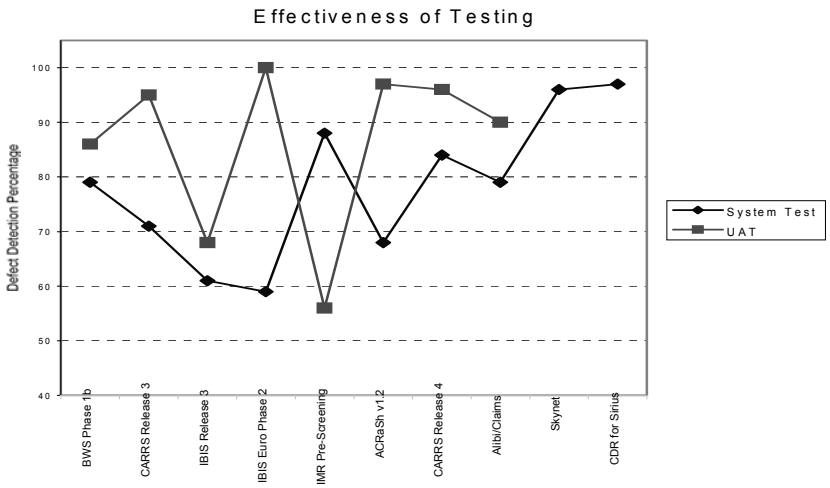
Defects: 1 / 4 160 / 40

Not useful for low numbers of defects

Scientific software
(chemical analysis)

www.grove.co.uk
011 44 8702 406172

RESULTS SO FAR



Source: Dave Norman, EuroSTAR02 DDP Advanced Workshop, with permission

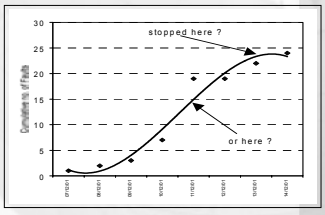
---

AON Information Technology

MESSAGES

Conclusions

- UAT more variable than ST – mainly personnel
- Target zone for ST : 75 - 90%
- Factors behind the figures
  - size, complexity, tester experience, time, documentation
  - whether UAT started before ST was finished
  - where on the S-curve when stopped
- Figures don't tell you
  - cost, severity of those you missed
  - cost of finding



DDP Summary for AP Europe

Project or App.	Months	DDP	DDP Status	Comments
Before New Testing Process				
S4		**50%**	**ESTIMATED**	
After New Testing Process				
R1	**3**	**81%**	**FINAL**	*Major re-engineering*
LBS	**4**	**91%**	**FINAL**	
CP	**7**	**100%**	**FINAL**	*Reporting System*
DS	**3**	**95%**	**FINAL**	
APC	**4**	**93%**	**FINAL**	
ELCS	**4**	**95%**	**FINAL**	*Eur impl. of US system*
SMS	**3**	**96%**	**FINAL**	*Enhancement Release*
C	**4**	**96%**	**FINAL**	
E7 (US)	**5**	**83%**	**FINAL**	*Global Enhancements*
E7 (Eur)	**1**	**97%**	**NEW**	*Global Enhancements*

Source: Stuart Compton, Air Products plc

Rolling DDP

Software Testing Defect Detection Percentage Measure

(rolling quarterly produced values looking back four quarters)

Period under review	# Projects Analysed	Target	Defects in Testing	Total Defects	Prod'n Bugs	DDP
Historical Estimate		**n/a**				**50**
Rolling 1 Qtrs DDP to Q1 Y1	**2**	**n/a**	1111	1400	289	**79**
Rolling 2 Qtrs DDP to Q2 Y1	**1**	**n/a**	1171	1466	295	**80**
Rolling 3 Qtrs DDP to Q3 Y1	**1**	**n/a**	1211	1508	297	**80**
Rolling 4 Qtrs DDP to Q4 Y1	**2**	**n/a**	1492	1807	315	**83**
Rolling 4 Qtrs DDP to Q1 Y2	**3**	**80**	2034	2129	95	**96**
Rolling 4 Qtrs DDP to Q2 Y2	**0**	**80**	1974	2063	89	**96**
Rolling 4 Qtrs DDP to Q3 Y2	**3**	**80**	2086	2204	118	**95**
Rolling 4 Qtrs DDP to Q4Y2	**2**	**80**	1976	2087	111	**95**
Rolling 4 Qtrs DDP to Q1 Y3		**90**				?
Rolling 4 Qtrs DDP to Q2 Y3		**90**				?
Rolling 4 Qtrs DDP to Q3 Y3		**90**				?
Rolling 4 Qtrs DDP to Q4 Y3		**90**				?

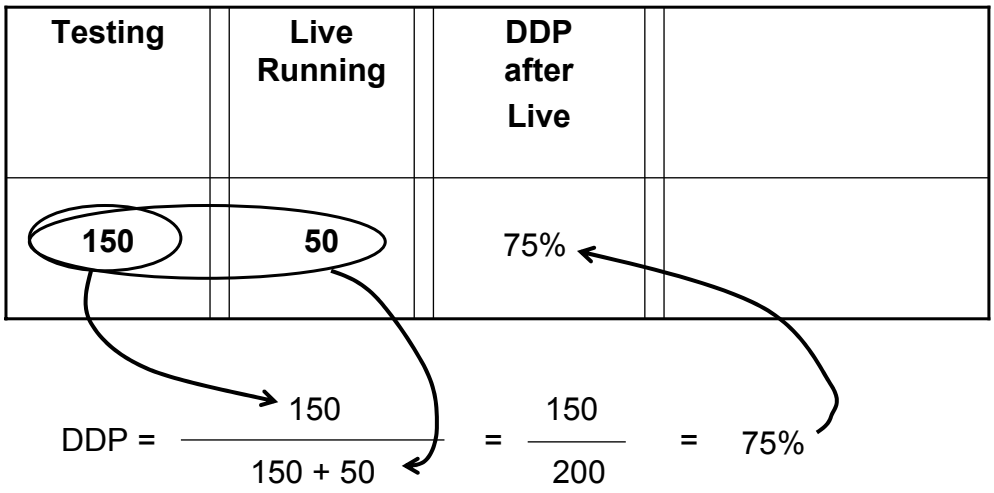
Source: Stuart Compton, Air Products plc

What does it mean?

- **DDP is very high (> 95%)**
  - testing is very good?
  - system not been used much yet?
  - next stage of testing was very poor?
    - e.g. ST looks good but UAT was poor, ST after UAT is high – but live running will find many defects!
- **DDP is low (< 60%)**
  - testing is poor?
  - requirements were very poor, affecting tests?
  - poor quality software (too many to find in the time)?
  - deadline pressure – testing was squeezed?

Contents

DDP example



17

DDP is not percent of total defects

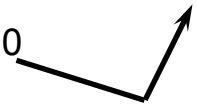
Stage of testing	defects	Percentage
"Official" (Mod & Int)	299	75%
"Tool" & development	40	10%
Release testing	19	5%
User Acceptance test	10	2.5%
Pilot	9	2.5%
Live use (1 mo.)	20	5%
		100%

Source: Client

this is <u>not</u> DDP (it's %) 18

DDP compares testing processes

Stage of testing	defects	Percentage
"Official" (Mod & Int)	299	75%
"Tool" & development	40	40%
Release testing	19	33%
User Acceptance test	10	25%
Pilot	9	31%
Live use (1 mo.)	20	



live use determines DDP

DDP example: ST DDP after UAT

ST	UAT	Live	ST DDP after UAT	ST DDP after Live	UAT DDP after Live	All test DDP after Live
100	**50**		67%			

$$\text{DDP} = \frac{100}{100 + 50} = \frac{100}{150} = 67\%$$

DDP example: ST DDP after Live

ST	UAT	Live	ST DDP after UAT	ST DDP after Live	UAT DDP after Live	All test DDP after Live
100	50	100	67%	40%		

$$DDP = \frac{100}{100 + 50 + 100} = \frac{100}{250} = 40\%$$

21

DDP example: UAT DDP after Live

ST	UAT	Live	ST DDP after UAT	ST DDP after Live	UAT DDP after Live	All test DDP after Live
100	50	100	67%	40%	33%	

$$DDP = \frac{50}{50 + 100} = \frac{50}{150} = 33\%$$

22

www.grove.co.uk
011 44 8702 406172

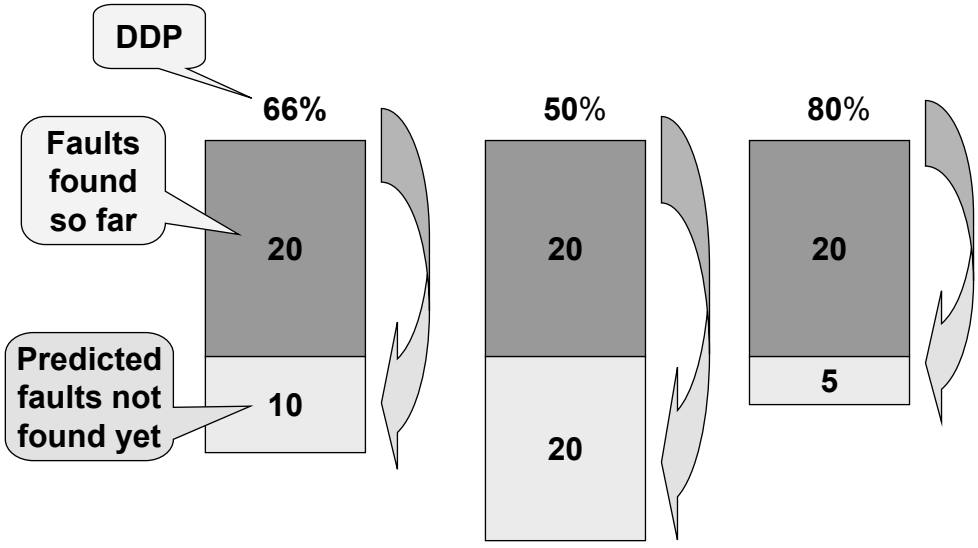
DDP example: all test DDP after Live

ST	UAT	Live	ST DDP after UAT	ST DDP after Live	UAT DDP after Live	All test DDP after Live
100	50	100	67%	40%	33%	60%

$$DDP = \frac{100 + 50}{100 + 50 + 100} = \frac{150}{250} = 60\%$$

23

Prediction of remaining faults



24

www.grove.co.uk
011 44 8702 406172

Contents

- **Introduction: some questions for you**
- **What is DDP**
- **Case studies**
- **Examples of calculating DDP**
- **Advice about DDP & Conclusion**

Advice about DDP

- **What is not important**
  - projects of difference sizes, scope, duration, technology, life cycle approaches
  - how you collect and analyse defect data
  - really good accuracy or detailed classification (e.g. duplicates, where inserted)
  - how well the project actually went
- **What is important**
  - consistency of defect data collection and analysis
    - either use severity or not (same approach)
- **May matter / may not**
  - different people, different test approaches, complexity

DDP limitations

- **not useful unless you have a reasonable number of defects and projects, over time**
- **never use it to measure individuals!**
  - only test efforts
- **don't get sucked into too much detail (too soon)**
  - don't do hard things – do easy things
- **it reflects what is happening**
  - not [just] how good a job you are doing of testing
- **not related to efficiency / cost**

Technical aspects

- **What time frame should I use for defects found in live?**
  - this is arbitrary / whatever makes sense for you
    - many people use 1 month, some use 3 or 6 months
- **Can I measure DDP of different test stages?**
  - you can measure any stage as long as you have defects that came afterwards
    - but don't measure individual people!!
- **Can I use DDP in incremental / RAD development?**
  - you have choices – accumulate, or measure until next release
- **What if different defect tracking systems?**
  - ok to combine for different stages or if consistently recorded

www.grove.co.uk
011 44 8702 406172

How to start using DDP

- **suggested first step**
  - calculate DDP for a release that is now live
- **what DDP to measure first?**
  - most people start with System Test
  - consider looking at highest severity only to start
    - or two DDPs, one for high severity, one for all defects
- **getting data from live running**
  - if you don't normally have live defect data, ask for it
- **data collection & calculation should be easy / automatic**
  - get your test management tool or defect tracking tool to calculate it for you automatically

---

Accuracy of defect data

- **most common "stumbling block"**
  - what about duplicates?
  - what about enhancement requests?
  - what if some aren't really defects?
- **the same answer always applies**
  - it doesn't matter how you do it
    - as long as you do it the same way each time!
- **most useful aspect of DDP**
  - trends, changes over time

www.grove.co.uk
011 44 8702 406172

The technical person's trap

- **we're testers – we can see all the problems!**
  - you will think of lots of "problems" with this metric
  - yes, DDP (as any measure) can be mis-used
    - but that doesn't mean it can't be useful
- **take the high level view**
  - DDP, warts and all, computed simply and consistently, can help you monitor your testing processes
  - and show the effects of both good and bad things

When NOT to use DDP

- **when you don't have many defects**
  - in test or in production (i.e. very high quality software)
- **your defect tracking is immature, purely subjective, untrustworthy, or non-existent**
- **the software products you produce**
  - are never used by anyone (no live running)
  - it doesn't matter how many defects are in them
- **it is impossible to get data on defects found in live running**
  - (difficult is OK!)
- **you're not interested in improving**

DDP benefits

- **DDP can highlight**
  - test process improvements
  - the effect of severe deadline pressure
  - the impact of overlapping test phases
- **can raise the profile of testing**
- **can help predict future defect levels**
- **is applicable over different projects**
  - reflects testing process in general
- **can give on-going monitoring of testing**

Summary: key points

- **DDP requires counts of defects**
  - but does not need great accuracy
- **DDP is a useful measure**
  - easy to calculate
  - based on defect data you probably already have
  - can tell you how effective your testing efforts are
    - and how other things affect it

www.grove.co.uk
011 44 8702 406172

DDP Exercise 3

If you have any project data from your own projects, put the numbers in the relevant columns. Calculate your own DDP using a calculator or using the workshop leader's spreadsheet.

If you do not have real project data, put in your best guess from a previous project, or work with a neighbor on their figures.

Fault Information

Release or project name	System Test or other test	UAT or other test stage	Live running (1 month)		ST DDP after UAT	ST DDP after LR	UAT DDP after LR	All test DDP after LR

How did you get these figures? (show your working out below or in the cells)

DDP Exercise 1

The following data has been recorded for a project.

Calculate the DDP of each testing stage based on all the defect information.

Fault Information

Testing stage	Number of faults	DDP
"Official" testing – module and integration	299	= ---------- = -------- = ___%
"Tool" testing & development	40	= ---------- = -------- = ___%
Release testing	19	= ---------- = -------- = ___%
User Acceptance test	10	= ---------- = -------- = ___%
Pilot	9	= ---------- = -------- = ___%
Live Running (after one month)	20	

Hint: you don't really need a calculator – just round the numbers to the nearest 10 and you will be close enough!

$$DDP = \frac{\text{Defects found in this stage of testing}}{\text{Defects found in this and all subsequent stages of testing}}$$

DDP Exercise 1 Solution: DDP Calculation

Testing stage	No. faults	DDP
1) "Official" testing – module and integration	299	75% 300 / 400
2) "Tool" testing & development	40	40% 40 / 100
3) Release testing	19	33% 20 / 60
4) User Acceptance test	10	25% 10 / 40
5) Pilot	9	33% (or 31%) 10 / 30 (9 / 29)
6) Live Running (after one month)	20	n/a

How did we get these figures?
Remember DDP = Defects found in testing / all subsequent defects

Stage 1 "Official" testing
Test stage 1 found approximately 300 defects – this is the numerator (top)
The sum of all the subsequent stages is = 40 + 20 (rounded up) + 10 + 10 (rounded up) + 20 = 100
So the denominator (bottom of the equation) is 300 + 100 = 400
DDP for Stage 1 is therefore 300/400 or 75%

Stage 2 "Tool" testing
Test stage 2 found 40 defects – this is the numerator (top)
The sum of all the subsequent stages is = 20 (rounded up) + 10 + 10 (rounded up) + 20 = 60
So the denominator (bottom of the equation) is 40 + 60 = 100
DDP for Stage 2 is therefore 40/100 = 40%

Stage 3 Release testing
Test stage 3 found 19 defects (round up to 20) – this is the numerator (top)
The sum of all the subsequent stages is = 10 + 10 (rounded up) + 20 = 40
So the denominator (bottom of the equation) is 20 + 40 = 60
DDP for Stage 3 is therefore 20/60 = 33%

Stage 4 User Acceptance test
Test stage 4 found 10 defects – this is the numerator (top)
The sum of all the subsequent stages is = 10 (rounded up) + 20 = 30 (29 to be exact)
So the denominator (bottom of the equation) is 10 + 30 = 40
DDP for Stage 4 is therefore 10/40 = 25%

Stage 5 Pilot
Test stage 5 found 9 defects (round up to 10) – this is the numerator (top)
The sum of all the subsequent stages is = 20 (the only remaining stage is live running)
So the denominator (bottom of the equation) is 10 + 20 = 30
DDP for Stage 4 is therefore 10/30 = 33% (31% if you calculate 9/29)

There is no DDP for live running, since the live running total goes into the calculation of all the previous DDP's.

DDP Exercise 2

The following data has been recorded for a project.

Calculate the DDP's in the columns on the right. The first one has been done as an example.
(A calculator may be useful for some of these – your mobile phone has one!)

Fault Information

Release	System Test	User Acceptance Test	Live running (1 month)		ST DDP after UAT	ST DDP after LR	UAT DDP after LR	All test DDP after LR
Release 1	100	50	100		*67%*	*40%*	*33%*	*60%*
Release 2	150	50	10					
Release 3	200	50	50					
Release 4	50	25	125					

How did we get these figures?

Release 1

ST DDP after UAT: 100 / (100 + 50) = 100 / 150 = 67%
ST DDP after LR: 100 / (100 + 50 + 100) = 100 / 250 = 40%
UAT DDP after LR: 50 / (50 + 100) = 50 / 150 = 33% (Remember not to include the ST defects here)
All test DDP after LR: (100 + 50) / (100 + 50 + 100) = 150 / 250 = 60%

DDP Exercise 2 Solution: DDP Calculations

Release	System Test	User Accep-tance Test	Live running (1 month)		ST DDP after UAT	ST DDP after LR	UAT DDP after LR	All test DDP after LR
Release 1	100	50	100		***67%***	***40%***	***33%***	***60%***
Release 2	150	50	10		***75%***	***71%***	***83%***	***95%***
Release 3	200	50	50		***80%***	***67%***	***50%***	***83%***
Release 4	50	25	125		***67%***	***25%***	***17%***	***38%***

How did we get these figures?

Release 2

ST DDP after UAT: 150 / (150 + 50) = 150 / 200 = 75%
ST DDP after LR: 150 / (150 + 50 + 10) = 150 / 210 = 71%
UAT DDP after LR: 50 / (50 + 10) = 50 / 60 = 83%
All test DDP after LR: (150 + 50) / (150 + 50 + 10) = 200 / 210 = 95%

Release 3

ST DDP after UAT: 200 / 200 + 50) = 200 / 250 = 80%
ST DDP after LR: 200 / (200 + 50 + 50) = 200 / 300 = 67%
UAT DDP after LR: 50 / (50 + 50) = 50 / 100 = 50%
All test DDP after LR: (200 + 50) / (200 + 50 + 50) = 250 / 300 = 83%

Release 4

ST DDP after UAT: 50 / (50 + 25) = 50 / 75 = 67%
ST DDP after LR: 50 / (50 + 25 + 125) = 50 / 200 = 25%
UAT DDP after LR: 25 / (25 + 125) = 25 / 150 = 17%
All test DDP after LR: (50 + 25) / (50 + 25 + 125) = 75 / 200 = 38%