

# GQM Case Study:

The use of GQM to set performance objectives for software delivery

Hans Gruber  
Christoph Haas  
Thomas Kiessling  
Ricki Koinig  
Clifford Shelley

UKSMA, October 2010

**bwin**<sup>com</sup>

# Content

- Introduction
- Background
- Story
- Results
- Summary

## Introduction: 1

- What is GQM ?
- The Goal / Question / Metric Paradigm
  - Was developed by Victor Basili and colleagues at the Software Engineering Laboratory \*
  - Is probably the most useful measurement technique for software developers given:
    - A lack of standard, 'off the shelf' measurement definitions and models of use
    - A scarcity of software measurement data for comparison purposes\*\*.
  - So software professionals have to design their own measures and analyses.
  - It provides a framework for deriving measures and analyses for specific purposes.
  - Caveat 1: It can be difficult (but can be easy too)
  - Caveat 2: GQM was developed in a 'high maturity' organization to develop 'systemic' measures – this is, at best, difficult in the wider software development community.
- \* See The TAME Project: Towards Improvement-Oriented Software Environments, by V.R. Basili and H.D. Rombach, IEEE Trans. On S/W Eng., Vol 14, No. 6, June 1988
- \*\*See Making Sense of Metrics Research Debate: Metrics data is only meaningful within its environment, by B. Kitchenham, Software Reliability and Metrics Club, 26 November 1993

## Introduction: 2

- Goal / Question / Metric has three elements:
  - Guidelines for goal definition (*measurement goals* ?)
  - Guidelines for product and process related quantifiable questions
  - Guidelines for metrics, data collection and interpretation

...that provide for 'tractability of the (top down) definitional quantification process, they also provide for the interpretation context (bottom up).'
- (Also required is:
  - Implementation of data collection and analyses, and data verification
  - Use of data!! - for decision making and actions
  - Validation of metrics definitions )



## Background

- *bwin* is a s/w company providing online services
  - multi-user (250,000 concurrent users)
  - multi- transactional international (20 languages) real-time financial transactions (over 1 million transactions per day)
  - require secure implementation
  - require rapid response (to changing commercial opportunities and legal environment)
- Two sites Vienna and Stockholm
  - Different cultures (Vienna acquired Stockholm organization)
  - Dev. approach is super agile (Scrummish, with interrupts to sprints allowed)

## Measurement Context: 1

- Organization wanted a novel agile capability:
  - to co-ordinate delivery of system components produced by delivery teams...
  - ...to deliver major system changes rapidly...
  - ...teams to be dependable 'delivery units' working together to a common 'rhythm'.
- Delivery teams were (and are) dependent on each other:
  - there were delays
  - it was unclear what was holding up delivery 'upstream'
  - team activity was opaque to other teams and business – inter team and business/team 'friction'
  - cascade further delays

## Measurement Context: 2

- Performance characteristics and objectives of delivery teams, set (in priority order)

**1. higher delivery predictability**

**2. increased responsiveness**

**3. high frequency and regularity of delivery (enabled by 1)**

**4. (while sustaining systems quality – implicit)**

## Measurement Need (Goals)

1. To monitor day to day status of development/test activity, to identify and manage issues as early as possible, and track against plan and inter delivery team dependencies (tactical/enumerative)
2. To monitor operational delivery predictability from sprint to sprint (operational/enumerative)
3. To analyse delivery team performance (in meeting business need) in the longer term (strategic/analytic)

## Questions:

1. What is 'status'?
2. What is:
  - Predictability,
  - Responsiveness,
  - Frequency and Regularity,
  - Quality?
3. What is 'improvement'?

## **M**etrics 1:

- What is 'status'?

**Effort Burn-down in Storypoints**

**Defects during sprint**

**Stories committed**

**Scope variance during sprint**

**Traffic light reporting on Sprint Status, and**

**Traffic light reporting for ventures committed in sprint**

(basic 'scrummy' metrics and some others)

## Metrics 2:

What is:

- **Predictability,**
  - teams to deliver within 7 days of planned delivery
  - teams to deliver 90% of storypoints\* committed to (after first year)
- **Responsiveness,**
  - Increased ability to respond rapidly by committing (mid sprint) to deliver  
(storypoints consumed & functionality delivered that was requested mid sprint)
- **Frequency and Regularity**
  - Enabled by predictability
  - Establishing a dependable organization wide delivery 'rhythm'
- **Quality**
  - As judged empirically from very rapid response from very many users  
(what is quality? – they recognize it when they see it)

## Metrics 3:

- What is 'improvement'?
- Targets:
  - When: 90% of all teams' deliveries within 7 days of planned delivery
  - What: 90% of teams' deliveries to deliver 90% of storypoints \*

## Tactical: Daily Reporting

- Most important reported metrics :
  - Effort Burn-down in Storypoints (*predicted delivery & deviation – most useful*)
  - Defects during sprint – (*by priority, very useful*)
  - Stories committed
  - Scope variance during sprint – (*actually found not to be very useful*)
  - Traffic light reporting on Sprint Status, and
  - Traffic light reporting for ventures committed in sprint

### Sprint Status:

Green	Amber	Red
	Amber	

### X-Backlog Venture Status:

Status	Venture I.D. / Description	Details (VersionOne) / Comments
Green	GAM-WGA-07-03_DiceArena (incl. Backgammon)	File-roll synchronised with GameE. Planned for 25.03.09 circa 10:00.
Green	MAR-CRM-07-48_VIP - VIP Offer Page	Detailed Investigation.
Red	PRT-CON-08-14_Final Rebranding Phase 2	Mobile Pages - Detailed Investigation.
Green	PRT-CON-08-14_Final Rebranding Phase 2	Detailed Investigation.
Green	PRT-CON-08-14_Final Rebranding Phase 2	Detailed Investigation.
Green	PRT-SER-08-06_KYC - Age Verification	Detailed Investigation.
Red	PRT-SER-07-16_Sitecore (SC) 4 Elimination	The team are unable to commit to this venture as a large part of the committed sprint requires heavy development resources.

# The DM Process Sample Daily Report:

## Delivery Management Reporting - LBAI Sprint 10-01

**OVERALL SPRINT STATUS: Green**

### Schedule

- Sprint Name: LBAI Sprint 10-01
- Sprint Start: 13.01.2010
- Sprint End: 10.02.2010
- Rollout Date: 24.02.2010

### News of the Day

- rollout date for LBAI RB27 (majority of items of this sprint) set for 24.2.2010
- due to last-minute meeting overflow of our product/backlog owners, the planning meetings have been moved by one day, the sprint was extended accordingly to keep the overall schedule
- due to some illnesses, there is a slight chance that not all of the investigation type stories will be finished on time, functional stories are all in good shape

### Venture status

The following List is an Overview of the Ventures committed to the Sprint. Status of the delivery of those Ventures is shown by the traffic light report (green: everything fine, orange: risk medium for not delivery, or other unsolved problems, red: risk high, not delivery, or other big impediments)

Venture	Status	Comment
BMK-07-06-1_BCM-LBAI	Green	
BMK-08-07-1_MultiPrice	Green	

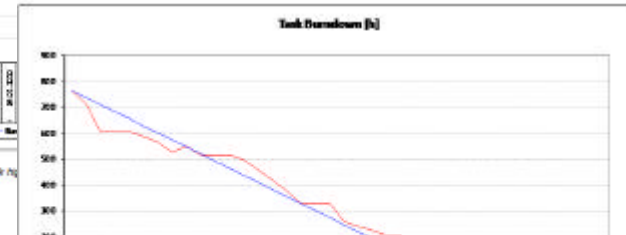
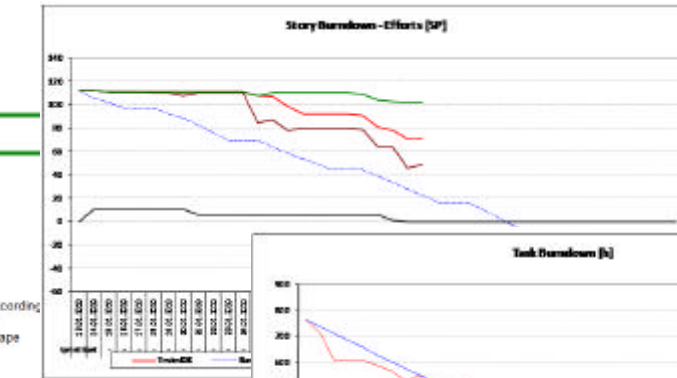
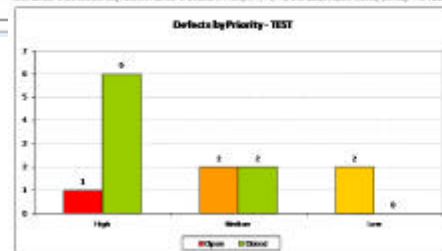
### Detailed Stories Status

The following List is an Overview of the Sprint Scope. All Stories of the actual Sprint backlog are listed. The Type, Status, Title and other properties of the Stories are listed. The Columns may vary as follows:

Number	Backlog	Title	Type	Status
B-22119	LBAI (LBAI)	SEARCH enhancements	Story	6 - Done
B-22153	LBAI (LBAI)	Connu: additional score information	Story	6 - Tested OK
B-16140	LBAI (LBAI)	MULTILABEL: Link Scoreboards for linked events	Story	6 - Ready for testing
Q-45948	Share: LBAI - Livebook	Exception when event goes idle	Defect	On hold
B-19135	LBAI (LBAI)	MULTILABEL: Change icons that indicates child/parent	Story	6 - Tested OK
B-10134	Share: LBAI - Scoreboards	support front-end display of splitted events	Story	6 - Tested OK
Q-46081	Share: LBAI - Scoreboards	fatal scoreboard: additional message lead in case of player names	Defect	6 - Done
B-22159	LBAI (LBAI)	FRANCE: BetCorrection test case creation prio 1	Story	6 - Done
B-22121	LBAI (LBAI)	FRANCE: BetCorrection test case creation prio 1	Story	6 - Done
B-22174	LBAI (LBAI)	OddsWizard: add preview on main market odds	Story	4 - In Progress
B-22144	LBAI (LBAI)	OddsWizard: ignore odds for unsupported templates	Story	6 - Tested OK
B-22114	LBAI (LBAI)	OddsWizard: enhancements	Story	6 - Tested OK
B-20630	LBAI (LBAI)	BCM: inside "Manage Combo Preventions" wizard	Story	6 - Tested OK
B-21177	Share: LBAI - Scoreboards	extend snooker scoreboard message for RING	Story	6 - Tested OK
Q-46127	Share: LBAI - Scoreboards	table tennis: tournament games countdown message	Defect	6 - Done
B-16108	Share: LBAI - IT driven	Resolve "service-double-start" issue	Story	6 - Tested OK
B-22182	Share: LBAI - IT driven	solve exceptions	Story	6 - Ready for testing

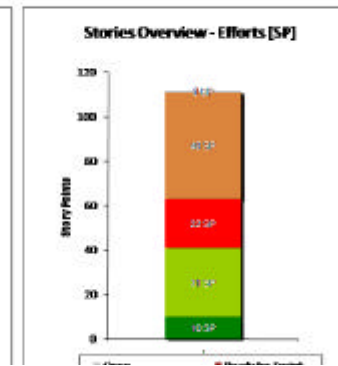
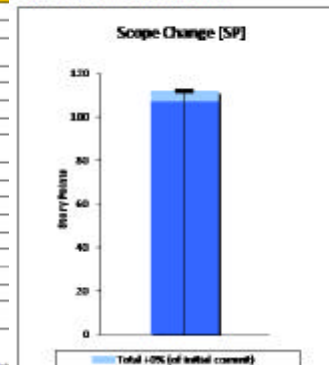
### Defect Overview Test (Number of Defects) on TEST Environment

The Defects by Priority Chart gives a Overview of the actual Defect situation in the Sprint. It shows test may Defects are Open/Closed and/or Ready for test Defects. The Defects are New/Close the Sprint. The chart shows the total number of Defects per Day, within Number of Defects in Priority P1, P2, P3 and P4 and Defects Count, Closed per Day. The Values shown are used for the TEST Environment.



### Scope Variance [Story Points] & Sprint by Status Chart [Story Points]

The Scope variance Chart shows the initial Sprint commit in Efforts (Story Points) (black line). The 3 different blue colours show the change in Scope. The Story Overview Chart shows the Sprint Status measured in Story Points) is Open, Ready for Sprint, Ready for Testing, Tested OK, Ready for Rollout.



### Cross team dependencies

The Cross team dependencies give an overview of dependencies within the committed sprint scope. Stories with a bold title are dependent upon the stories in the rows below marked with a "x" sign. In the "backlog column" you see the backlog dependent story is located.

Number	Backlog	Title	Status	Sprint	X Team
B-10134	Share: LBAI - Scoreboards	support front-end display of splitted events	6 - Tested OK	LBAI Sprint 10-01	TRUE
B-09503	Betting Livebook (BLB)	Splitted events - main, specials, play by play	1 - Done	BLB sprint 10-01 (Current Livebook)	TRUE
B-19567	Share: LBAI - Scoreboards	table tennis scoreboard: best of three matchtype	6 - Tested OK	LBAI Sprint 10-01	TRUE
B-19573	Betting Livebook (BLB)	Table Tennis scoreboard: best of three matchtype	4 - In Progress	BLB sprint 10-01 (Current Livebook)	TRUE

# Operational: Sprint by Sprint: 1

Overall Releases - KPI Achievement								
Release Version/Name	Planned Rollout Date	Effective rollout Date	Delay in Days	Successful Releases <sup>4</sup> , KPI 90%	Time KPI	Target	Comment	Storypoints committed <sup>1</sup>
SB-3.9	15/01/2008	22/01/2007	7	1	100%	90%		320
G12/07	19/12/2007	19/12/2007	0	1	100%	90%	On time	
NDC Wireless Sprint 7	14/01/2008	14/01/2008	0	1	100%	90%	On time. Excellent progress by off-site team in addition to good delivery from on-site, lead to a high amount of deliverables. Christmas-New Year Sprint was 1 week longer than standard Sprint to take into account reduced days of full team interaction. Thr	189
CSM 4.2.5	16/01/2008	21/01/2008	6	1	100%	90%	CSM 4.2.5 delivered with limited scope, due to technical problems with multiple system support (MSS) integration.	
FL-SB-3.9	14/01/2008	21/01/2008	7	1	100%	90%	Delay caused due to Sportsbook delay (roll within the sportsbook roll)	
LM-1.3	14/01/2008	21/01/2008	7	1	100%	90%	Release delay by external dependencies - sportsbook release late	129
LS-1.1	14/01/2008	21/01/2008	7	1	100%	90%	Release delay until SB release since downtime is required, and an additional downtime period is unwanted.	
G1/08	23/01/2008	23/01/2008	0	1	100%	90%	On time	
C01/08	11/02/2008	11/02/2008	0	1	100%	90%		65
SB-3.10	12/02/2008	12/02/2008	0	1	100%	90%		255
NDC Wireless Sprint 8	11/02/2008	12/02/2008	1	1	100%	90%	Positive Sprint with excellent commitment. Development outpacing testing capacity to some degree however. Additional "testers" from development assisted. New dedicated 4th tester to bring testing arm up to full-strength is anticipated to start in March.	214
LM-1.4	12/02/2008	12/02/2008	0	1	100%	90%	Difference in committed vs. delivered due to external delivery too late - stories moved to next sprint	175
LS-1.3	12/02/2008	12/02/2008	0	1	100%	90%		
G2/08	13/02/2008	14/02/2008	6	1	100%	90%	Fantasy Sports project was delayed. BGI part of the project and web parts had to be rescheduled	
CSM 4.2.6	20/02/2008	21/02/2008	1	1	100%	90%	One week pre-planned delay due to delay of 4.2.5 and to finalize MSS. Included scope of 4.2.5. Major problems after release with the MSS, which caused production system to be unusable for more than a day and contributed greatly to the failure of CSM DB de	
C02/08	26/02/2008	27/02/2008	1	1	100%	90%		70
FL-SB-3.11	12/02/2008	04/03/2008	19	0	94%	90%	Delay caused due to 1)backend issues regarding session deletion on beta 2) longer bogfixing phase on beta	140
NDC Wireless Sprint 9	11/03/2008	11/03/2008	0	1	94%	90%	Good delivery from dedicated Scrum team. More can be delivered into the future especially when we have a much needed 4th tester on board.	306
C03/08	11/03/2008	11/03/2008	0	1	95%	90%		87
SB-3.11	11/03/2008	12/03/2008	1	1	95%	90%		458
CSM 4.2.7	12/03/2008	12/03/2008	0	1	95%	90%	Effective short sprint to come back on track with other release cycles. Focus on functional enhancements. Problems with missing database rights of some components at the live rollout.	285
LM3-08	11/03/2008	12/03/2008	1	1	95%	90%		232
LS3-08	11/03/2008	12/03/2008	1	1	96%	90%		
G3/08	12/03/2008	14/03/2008	2	1	96%	90%		
C04/08	26/03/2008	26/03/2008	0	1	96%	90%		70
C05/08	09/04/2008	09/04/2008	0	1	96%	90%		84
SB-3.12	08/04/2008	10/04/2008	2	1	96%	90%		475

## Operational: Sprint by Sprint: 2

1. Release name
2. Planned roll out date
3. Effective roll out date
4. Delay (in days)
5. Schedule KPI flag (1/0)
6. Derived metric
7. Target metric
8. Comment
9. Storypoints committed
10. Storypoints delivered
11. Storypoints added
12. Percentage delivered/committed
13. Scope KPI flag (1/0)
14. Derived metric
15. Target metric

## Strategic Improvement – 2007 to 2010

How?

- Require a single process shared across all teams
  - ...a 'standard' and 'defined' process (no tailoring)...
  - ...that is understood by all...
  - ...making a sprint's status, including emerging issues, at any time, visible and understandable to all
- Using 'rule 11'
  - "Document processes, practices and activities 'as is' and use this to establish a credible process baseline which can then be developed. (Do not - ever - document processes as you would wish them to be and attempt to introduce them, or roll them out.)"*
- Third time lucky (this wasn't easy)
- In effect introducing 'process discipline' to an agile environment
  - (cf Systematic Software Engineering - introducing agile to a CMMI ML5 organization reported in *'Scrum and CMMI Level 5: The Magic Potion for Code Warriors'*)

## Strategic Analysis - Results: 1

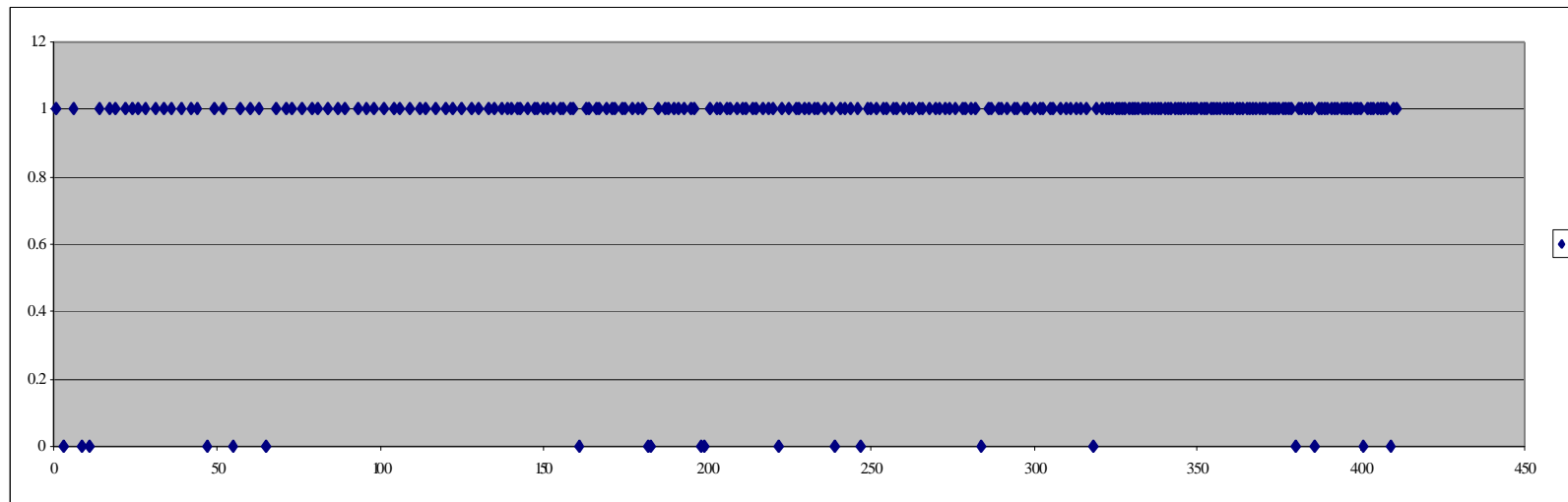
- Data collected from mid 2007 to end of 2009
- 6 teams using defined process in 2007, 10 in 2008, 16 (all teams) in 2009
- Analysis performed using day-to-day operational data (as it should be)
- Some additional data wanted by analyst, but not available  
(too bad!)
- No filtering or stratification of data performed, outliers retained, and analyses kept as simple as possible (with one, exploratory exception)

## Results: 2

- Schedule predictability (1) KPI :

'1' indicates delivery within 7 days of planned, 0, greater than 7 days

Note increasing density of '1' data points with consistent density of '0' points



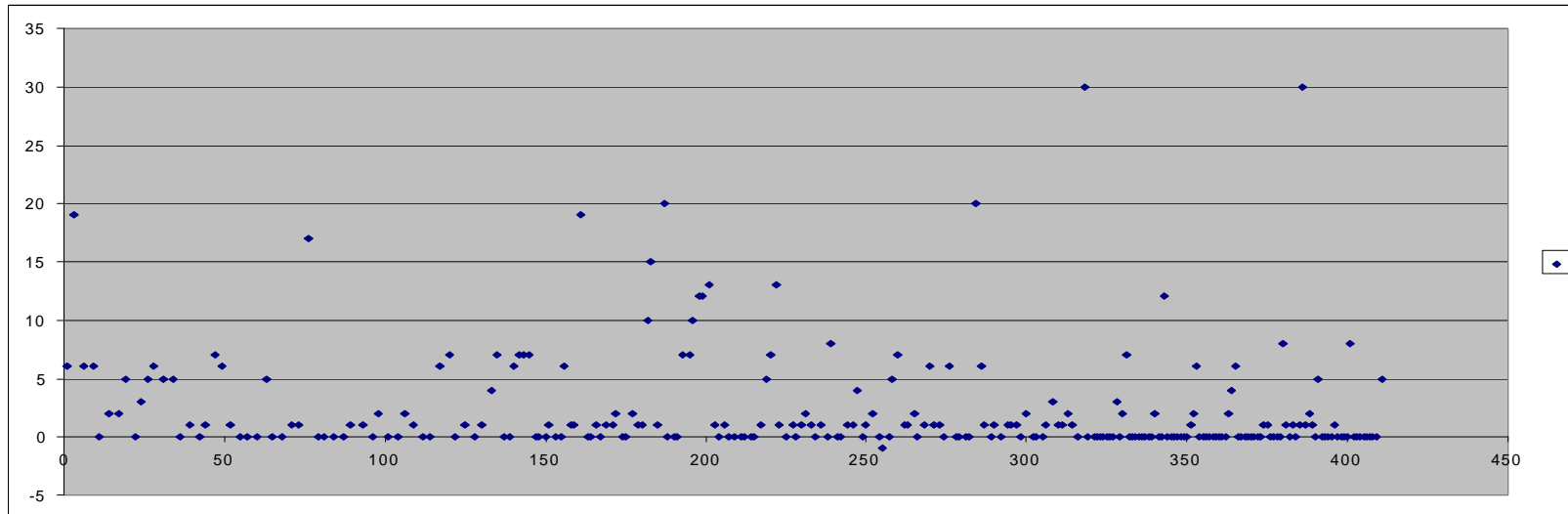
## Results: 3

- Schedule predictability (2) Days deviation:

Plotted with actual days deviation

Note number of deliveries with zero deviation towards and of reporting period

(performance better than indicated by KPI)



## Results: 4

If the number of days delay for releases in the first four months are considered then the average delay per release is 4.25 (85 days delay/20 releases).

For the last four months the average is 1.45 days per release (61/42), and if the exceptional delay of 30 days, for SP 09-06 FRB 3.25, is removed this reduced to 0.74 days per release.

The ratio between the first four and last four months is a 2.93 times improvement in schedule predictability, or, excluding a SP 09-06 FRB 3.25, **a 5.74 fold improvement in schedule predictability**

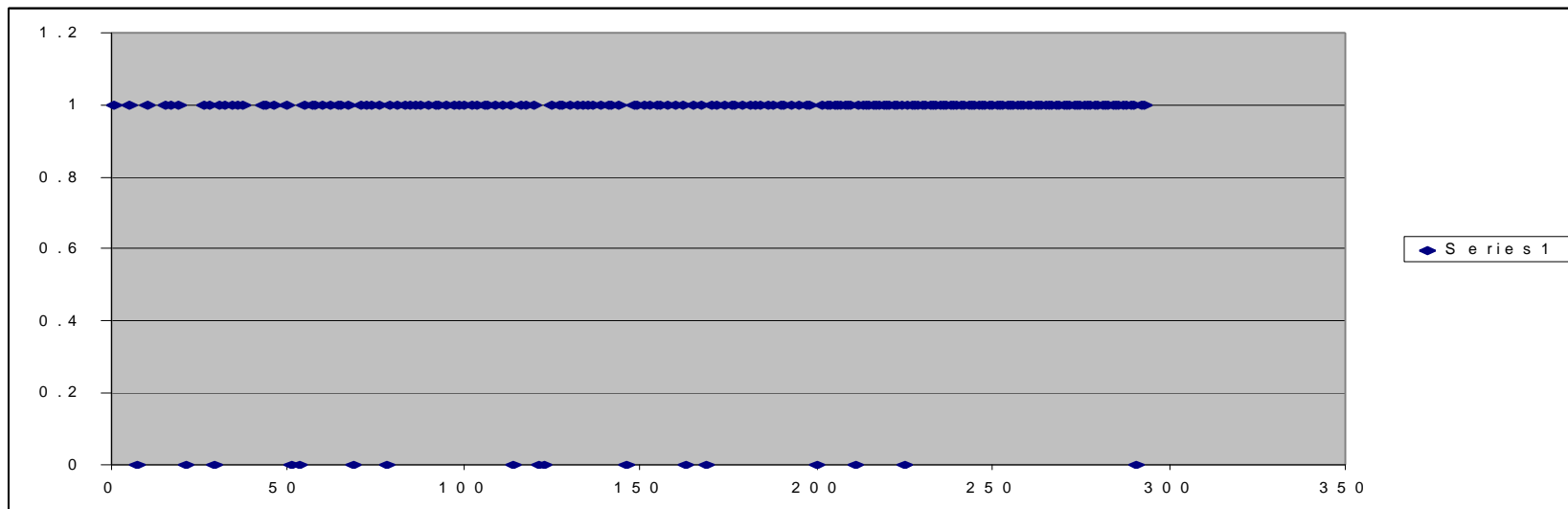
## Results: 5

- Scope predictability (1): KPI

'1' indicates 90% or more of committed storypoints delivered, '0', less

Note the increasing density of '1' over time, and decreasing number of '0's

(NB – period from beginning 2008 to end 2009)



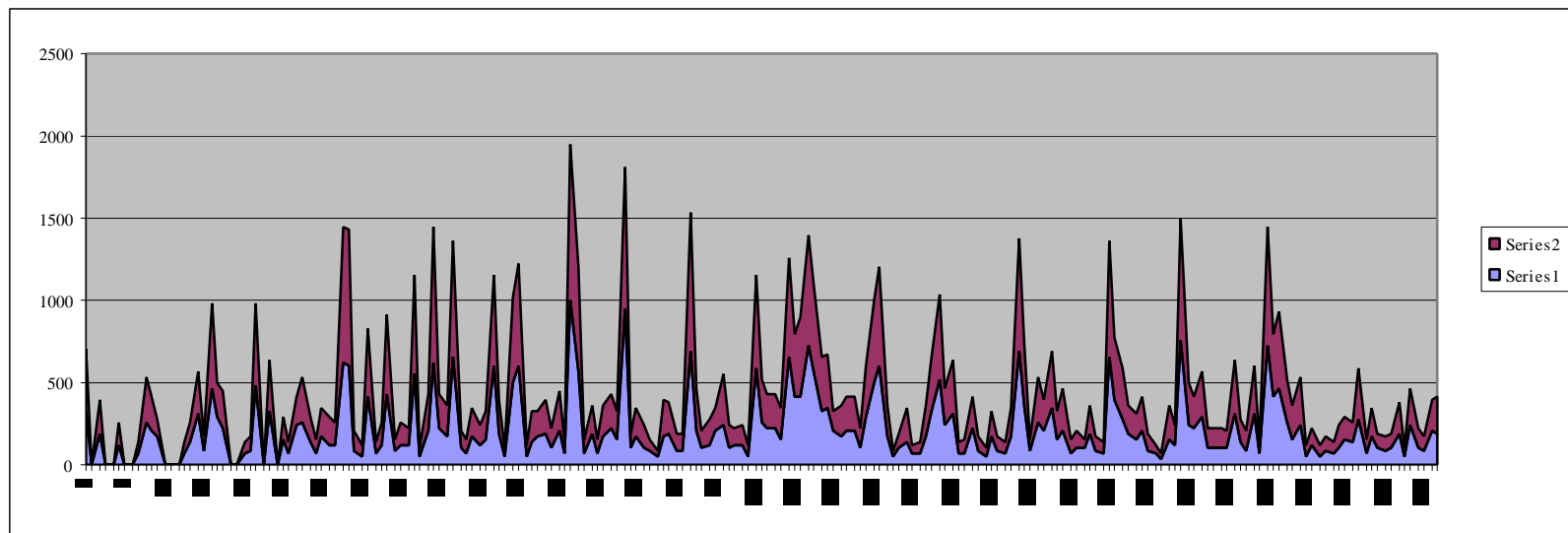
## Results: 6

- Scope predictability (2) Storypoints:

Shows the numbers of committed and delivered storypoints

Storypoints are calibrated locally (in teams)

Difficult to discern any pattern or trend

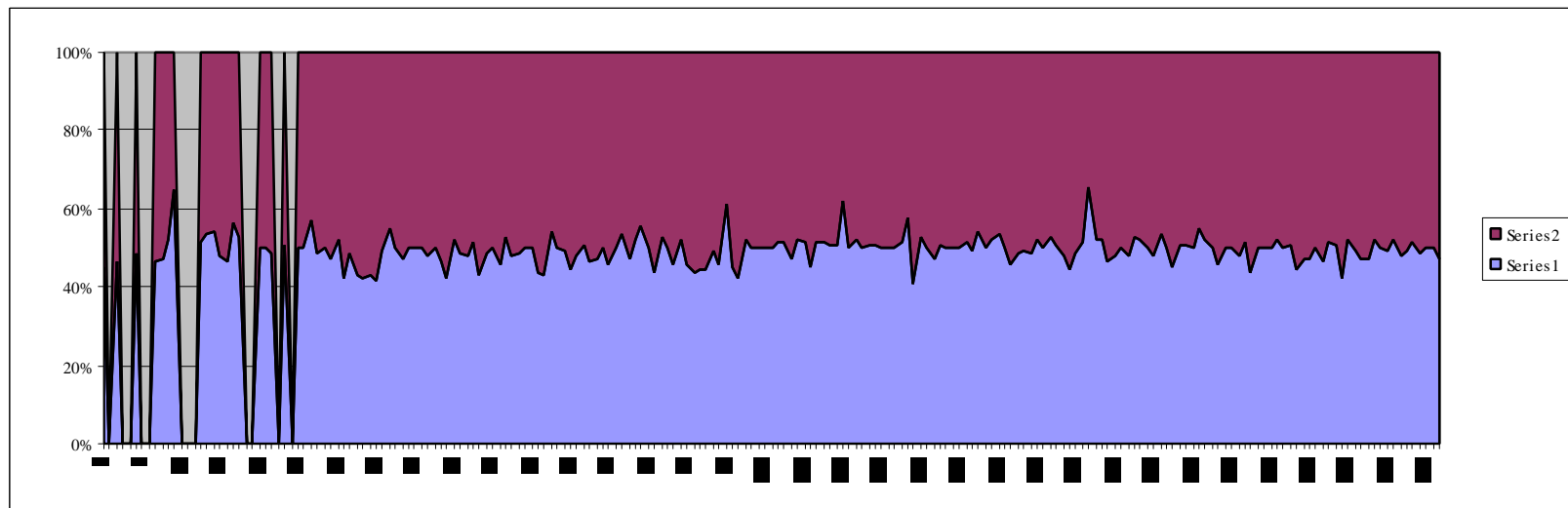


## Results: 7

- Scope predictability (3) Storypoints:

Plot the ratio (or proportion) of committed/delivered storypoints  
(normalized!!)

Still not much of a pattern, although it looks like delivery is more or less  
(literally) what was committed

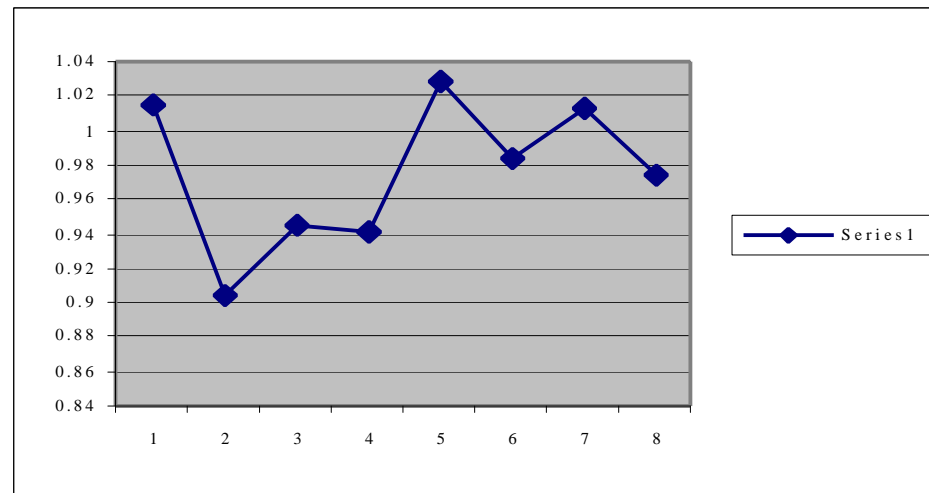


## Results: 8

- Scope predictability (4) Storypoints:

Plot the average of data points 'binned' in three monthly periods

Pattern emerges, but with odd first data point





## **Results: 9**

**In the first four months for which we have data (January – April 2008) 14 of 20 releases achieved the 90% scope target (with 6 failing and no data for 7 others recorded.)**

**For the last four months for which data is available (May to August 2009) 39 of 40 releases succeeded.**

**This gives a ratio of about 12:1 - a twelve times improvement in scope predictability.**

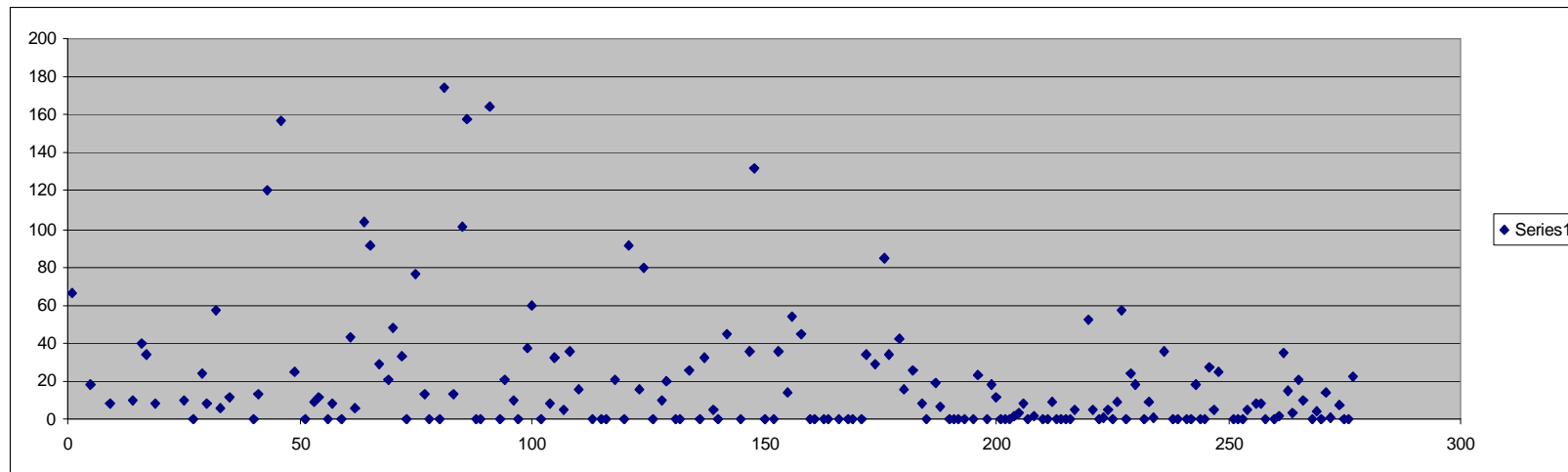
## Results: 10

- Responsiveness

Evaluated by the number of storypoints committed to in mid sprint and delivered

Number *declines* over this period

Interpreted as evidence of increasing use of a well understood, predictable and trusted process by teams and business, and only interrupted by genuine business need



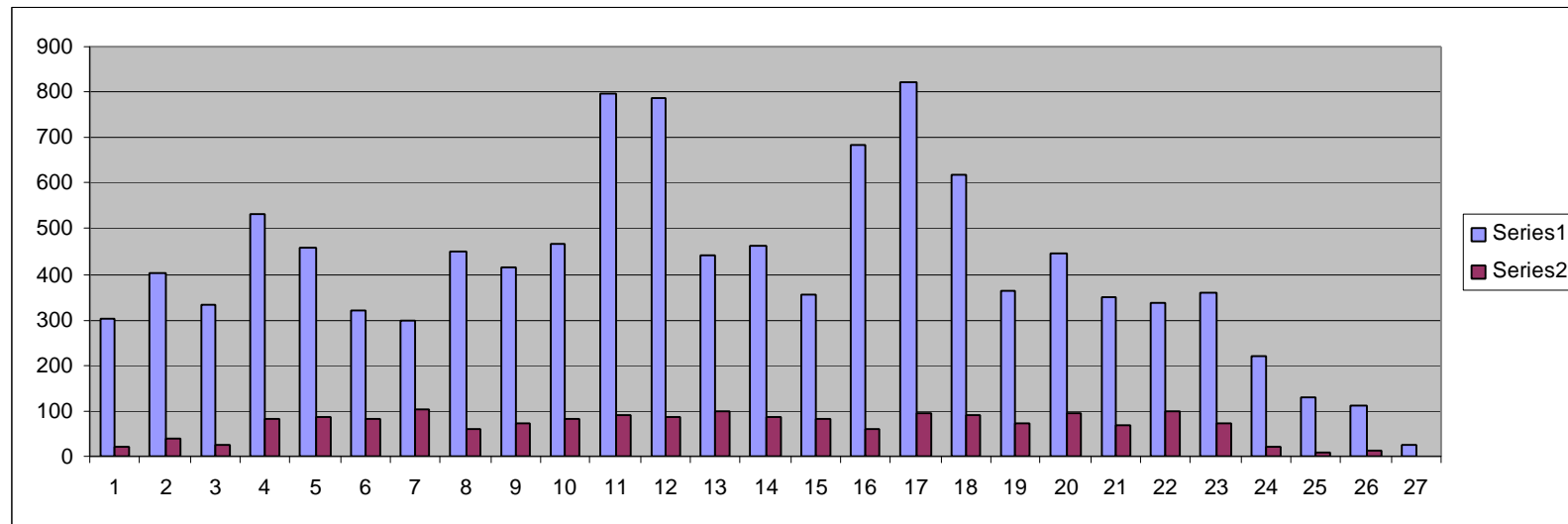
## Results: 11

- Quality (1);

Defects from live operation (red bars) as *proxy* measure of quality

(only data available for all teams over entire period)

Indicates consistent level of quality over 2007 – 2009 period



## Results: 12

- Defect reporting levels suggested high quality
- How to establish level of quality?
- Defect densities have been used:

<i>Released S/W</i>	<i>Defect Density</i>
NASA	0.1 defects / KLOC
Leading Edge	0.2 defects / KLOC
Critical systems	1.4 defects / KLOC
Military	5 – 55 defects / KLOC

## Results: 13

- What is bwin defect density?
- A defect model:

bwin's software is heavily used and most defects will be found quickly.

If bwin did not make changes to these systems, but simply fixed defects as they were found through day to day use the number of defects being found would decrease over a short period of time – say six months.

At present bwin *does* make changes and the number of defects has not decreased over time.

Therefore the defects being found (about 60 – 100 per month) are due (mostly) to changes being made.

Therefore, if the number of changed lines of code can be estimated (extracted from the SCM system) the defect density of these changed lines of code can be estimated.

(Note: If this model is reasonable it also suggests that software quality could be improved quickly and easily by introducing a carefully selected set of quality controls to changed software.)

## Results: 14

- One third of the defects reported from live operations are high priority (about 500 of about 1800). With live defects being reported at about 60 to 100 per month this means about one high priority defect being reported per day.
- *Bwin* software assets were about 400 KLOC\* (thousands of net, executable lines of code) in December 2007. This had increased to about 600 KLOC in December 2008 and to about 1 MLOC in September 2009.
- This increase in LOC, together with consistent defect reporting levels suggests that, in practice, the quality of bwin software is improving.
- In addition, if it is assumed that defects reported are mostly a result of new or changed code then the level of defects reported implies exceptional levels of code quality:
- 200 KLOC of new code (plus changed lines of code) in 2008 gave rise to about 1000 defect reports that year, of which about 400 were high priority. This gives a defect density of about 5 defects/KLOC.

Note: The data indicate that quality has been sustained (and improved) over this period and has not been 'traded off' against improved delivery predictability.

(\* 100 KLOC eqv. 1 KJB (King James Bible))

## Summary:

- A clear business need was identified and characterized...
- ...for novel development capability.
- This required a hybrid mature/process/agile model of development...
- ...supported by carefully selected and tailored tools (incl. Version One)
- With high visibility across teams and the business - and carefully monitored and supported.
- **bwin** met and exceeded performance targets...
- ...and may have a unique new capability



Contact: [shelley@osel.netkonect.co.uk](mailto:shelley@osel.netkonect.co.uk)

**Thank You**

***bwin*** ● **com**

***bwin***<sup>.com</sup>

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

'Anscombe's Quartet'

$$N = 11$$

$$\text{Mean of X's} = 9.0$$

$$\text{Mean of Y's} = 7.5$$

$$\text{Regression line } Y = 3 + 0.5X$$

$$\text{Standard error of estimate of slope} = 0.118$$

$$t = 4.24$$

$$\text{Sum of squares } X - \bar{X} = 110.0$$

$$\text{Regression of sum of squares} = 27.50$$

$$\text{Residual sum of squares of Y} = 13.75$$

$$\text{Correlation coefficient} = 0.82$$

$$R^2 = 0.67$$

