

Software Test Metrics.

Key metrics and measures for use within the test function.

Discussion Document
By Mark Crowther, Empirical Pragmatic Tester

1.0 INTRODUCTION

An important deliverable from the test team is information that conveys the progress of test work and the status of test activities within the project. This information can include metrics, statistics and summaries and take the form of graphs, charts and written reports.

Various partners within the business require this information in support of their own activities and decision making. As such any information provided by the test team should be useful to the business partner and be published in a timely and consistent manner.

This discussion document outlines the key Metrics that will be provided by the test team, where the information will be drawn from and how it will be published. The document excludes statistics and summaries except where this is mentioned for context or reference.

2.0 SOFTWARE METRICS / SOFTWARE TEST METRICS

The term used to describe a measurement of a particular attribute of a software project is a Software Metric. The Software Metrics that the QA team produce are concerned with the test activities that are part of the Test Phase and so are formally known as *Software Testing Metrics*. Other teams will produce and publish their own Software Metrics during the project.

2.1 Test Process and Test Product Metrics

QA recognise two subsets of Software Testing Metrics:

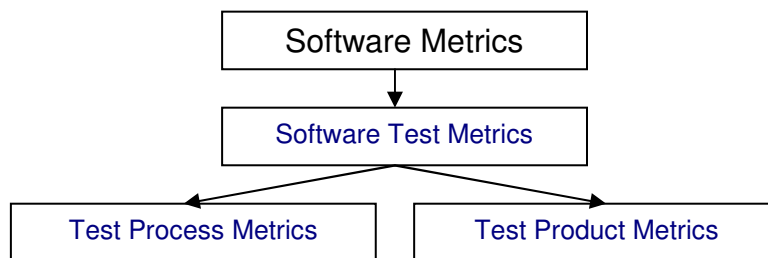
- **Test Process Metrics**

These measures provide information about preparation for testing, test execution and test progress. They don't provide information about the test state of the product and are primarily of use in measuring progress of the Test Phase.

- **Test Product Metrics**

These measures provide information about the test state of the product and are generated by test execution and code fixes or deferralment. Using these metrics we can gauge the products test state and indicative level of quality, useful for product release decisions.

Diagram 1: Software Metrics breakdown



2.2 Metrics within the Test Plan

At the Initiation Phase of a project QA will issue a Test Plan which will define the metrics to be captured and published at each stage of the Test Phase. The Test Phase is broken down into two key stages; Test Preparation and Test Execution. The Test Plan includes both Test Process and Test Product metrics as they provide equally useful information to business partners.

The Test Plan also defines the milestone dates for key deliverables such as the Test Plan and these are metrics captured for ongoing statistical process analysis across successive projects.

3.0 TEST PHASE METRICS

For all projects the following metrics will be captured and published by the QA team during the Test Phase. Metrics that look at Functional Areas/Requirements check for test coverage and consistency of test effort.

3.1 Test Process Metrics

The following are provided during the Test Preparation stage of the Test Phase:

- **Test Preparation**
 - Number of Test Requirements Vs Functional Areas/Requirements (Test coverage)
 - Number of Test Cases Planned Vs Ready for Execution
 - *Total Time Spent on Preparation Vs Estimated Time*

The following are provided during the Test Execution stage of the Test Phase:

- **Test Execution and Progress**
 - Number of Test Cases Executed Vs Test Cases Planned
 - Number of Test Cases Passed, Failed and Blocked
 - Total Number of Test Cases Passed by Functional Areas/Requirements
 - *Total Time Spent on Execution Vs Estimated Time*

3.2 Test Product Metrics

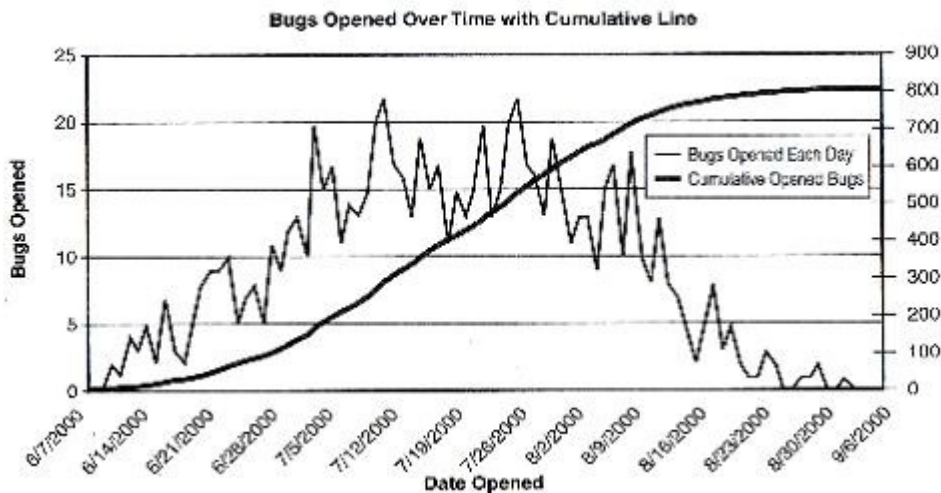
- **Bug Analysis**
 - Total Number of Bugs Raised and Closed per Period
 - Total Number of Bugs Closed Vs Total Number of Bugs Re-Opened (Bounce Rate)
 - Bug Distribution Totals by Severity per Period
 - Bug Distribution Totals by Functional Areas/Requirements by Severity per Period

The metrics in section 3.2 can also be applied to products in their maintenance phase but this is excluded from the scope of this document.

4.0 PUBLISHING TEST METRICS

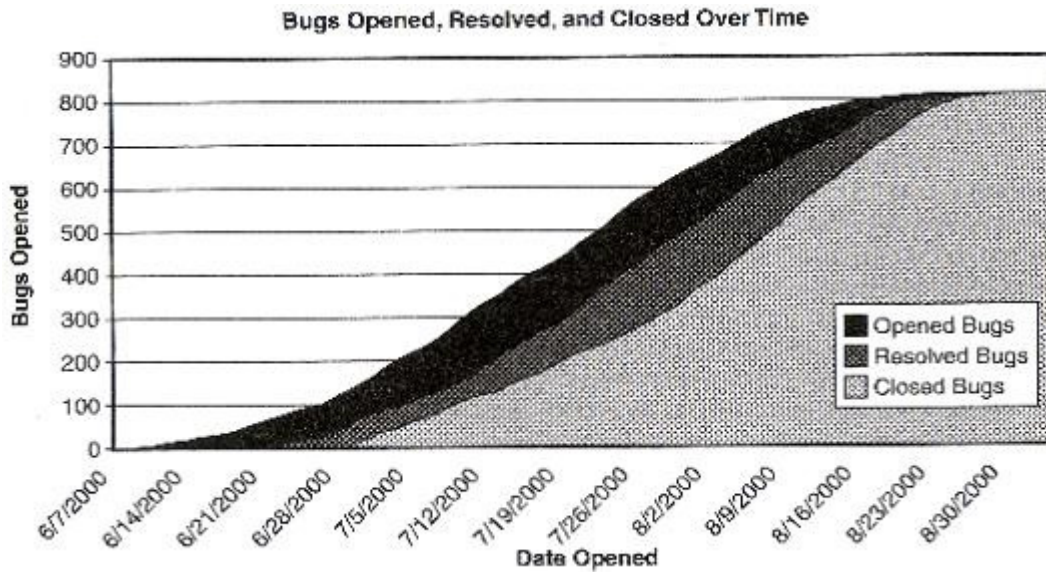
The test team will publish the Test Phase Metrics at the agreed intervals. Primarily this will be a Summary Report which will include a table of Metric values for periodic review. In addition the test team will publish two key charts:

- **Total Number of Bugs Raised Over Time with Cumulative Line**



As the total number of bugs raised each period lowers the products test state should be reaching a higher level of quality. Additionally as the Cumulative Opened Bugs levels off this indicates a decreasing bug-find rate. If this remains consistent it may be further testing is adding little value and the product can be released.

- **Bugs Open, Resolved (Deferred) and Closed Over Time**



Taking the Cumulative Opened Bugs line from the first chart and comparing this against Resolved (Deferred) Bugs and Closed Bugs it's easy to see what work is left to do. If the black band opens up then the test team are raising bugs far quicker than the developers can fix them. If the dark grey band opens up bugs are being resolved (deferred) faster than the test team can close them. If the bands are converging then a release is in sight.

Both of these charts will allow Product Managers and other key partners to assess the test state of the product and the likelihood that release date will be achieved. Key project milestones may be added to the charts and when analysed against other metrics the overall status of the Test Phase can be assessed.

5.0 ACCEPTABLE QUALITY LEVELS

In the charts and discussion above we mention Resolved (deferred) bugs as one way to help us get to a release point. However it mustn't be forgotten that these bugs remain in the product. This raises the question of how many of what type can be deferred and still provide for a product of acceptable quality for release, in addition to the potentiality that not everything may get tested in the first place.

5.1 What Good Looks Like

The need to define What Good Looks Like (WGLL) has been discussed and it's anticipated that these will be defined in terms of Silver, Gold and Platinum standards. Recognising when the product has reached these standards can be achieved through setting Acceptable Quality Levels (AQLs) for product test attributes.

In order to define these standards with consistency and allow a mechanism for adjustment two sets of Measurable Success Criteria are provided within the Test Plan. Combined into a set of three matrices it will be possible for the test team to declare that a product has reached an AQL that achieves WGLL Standards

5.2 Measurable Success Criteria

Measurable Success Criteria are also known as Hygiene Standards and these help us to understand the test state of the product. They can be used without application to WGLL Standards and provide a realistic measure of when we have agreed a product is good enough to release. They should be based on what has been historically proven as achievable, an agreed commercial or operational level of acceptability or accepted good practice. The examples here are based on good practice within the test community.

- **Percent Pass / Fail by Test Class Criteria**

Each Test Case will be assigned a Test Class Criteria and this can include both the number of Test Cases run and passed as well as those not yet run.

Example Test Class Criteria Pass Rate AQL

Test Class Criteria	Definition	Standard % Pass Rate AQL
Critical	Essential to the Product	100 %
Important	Necessary to the Product	90%
Desirable	Preferred, but not necessary to the Product	70%

There for as a standard we say that all Test Cases for Critical functionality must be run and pass. Similarly at least 90% of Test Cases for Important functionality must be run and pass. More than 90% can be run of course but this is the minimum that must pass of those run. The same logic applies to Desirable functionality.

- **Deferred / Open Defects by Level of Severity**

Each bug is assigned a severity and as the test-fix-retest cycle progresses the quality of the product rises towards achieving the WGLL Standards. It's recommende that all Severity A bugs are fixed as a standard, with Severity B and C able to be closed either by fix or deferrment. The AQL levels for B and C bugs being set within the WGLL Standards Matrices.

Example Test Class Criteria Pass Rate

Severity Level	Definition
A	High – system crash, data loss, no workaround (100%)
B	Medium – operational error, wrong result, work around available
C	Low - minor problems

Together these can be use to provide an overall AQL for the product in test. How Measurable Success Criteria will be defined and if WGLL Standards are to be used will be defined in the Test Plan at the project Initiation Phase.

5.2 WGLL Standard Matrices

The three proposed WGLL Standards of Silver, Gold and Platinum should ideally be used to declare the test state of products during development and not as standards for release candidates into maintenance. It would be my recommendation that the WGLL Platinum Standard be the standard for release Candidates as the overall AQL will be at the highest we belive practical to achieve.

The following three matrices are provided as examples of how AQL levels might be set for each WGLL Standard. It should be possible to interogate DevTrack using a tool such as Access to produce the figures required and go on to display these in a form of a dashboard.

		Silver			Gold			Platinum		
		Critical	Important	Desirable	Critical	Important	Desirable	Critical	Important	Desirable
Bug Severity	A	100	100	100	100	100	100	100	100	100
	B	90	80	70	100	90	80	100	90	90
	C	80	70	60	90	80	70	90	90	80