Cyreath.co.uk

Empirical Pragmatic Testing

Measuring Quality.

A model for assigning measurable values to elements of software quality.

Discussion Document

By Mark Crowther, Empirical Pragmatic Tester.

1.0 INTRODUCTION

How do we measure Quality? How do we measure the value of one bug over another? If we can measure this how do we relate its impact to Quality? To measure something we have to have a certain level of control over it. If we can measure its impact then what can we do with this new, measurable and so controlled insight?

1.1 Measuring Quality

The general approach is to assess the requirements; the expectations for the software or web property, then perform requirements driven testing to ensure the expectations are met. This sounds simple but is often not the case for a number of reasons:

- The corporate or production environment doesn't insist on defined requirements
- They are partially created, with others derived from standing documents
- The speed or volume of work for test does not allow time for such formal documents

It may be that the organisation is sat firmly in Level 0 CMMi and turning to the test team to provide insight into quality and how to remediate issues seen.

Whatever the situation, managing the situation is vital. As a professional test team can't just wing-it every time, sometimes we'll have to, but we can't do it all the time.

Before you can measure the value of a bug or defect you will need to have agreement from the team and internal customer about your definition of quality. Expect no more than reasonably subjective definitions, like those shown for severity in the next section. With this agreement you can start measuring those bugs.

This paper should be read in conjunction with the Excel workbook Bug Report. Instructions for using this are provided within the workbook. A copy is available from:
http://www.cyreath.co.uk/templates/Cyreath_Bug_Report_Template.xls

In writing this paper I acknowledge many things:

- That the approach is essentially backwards to the more standard requirements driven test approach. However with the lack of these requirements should not be the driver for a lack of a considered test approach.
- That 'everything else' such as Requirements Definition, Test Plans, Test Models and Techniques, etc. are not addressed.

That's because the paper takes a view that the newly arrived Test Manager has been dropped in to save the day. This discussion document offers a possible way to adjust the test strategy and engage the customer in thinking about quality and ways to improve it. The end objective of course is a sensible and mature test function and quality system. The implementation of which is a luxury rarely afforded to us from day one, especially when being driven to provide value as early as possible.

Mark Crowther.

2.0 THE QUALITY VALUE (QV)

2.1 Measuring and Defining Quality, numerically.

Having no definitive list of requirements and basing our impending interpretation of the product's quality on subjective assessment, each product is assumed perfect on arrival to QA. It may be a leap of faith but this first step sets the scene for our test approach. Consider how this aligns to the test function's primary aim, not to prove conformance to stated requirements, but to find errors. Your chosen test techniques, not covered in this document, should be focused on finding errors.

In reality at this point who can say if the product is good or bad? A brief review may reveal issues but exactly where is the product quality wise? We've no definitive list of requirements and for web properties expect no previous testing, such as unit testing, to have been performed.

This assumedly perfect product is assigned a **100 Quality Value (QV)** score.
Now we begin testing.

Each bug found is assigned a severity, defined as follows:[1]

1) System Crash, data loss, data corruption
2) Operational error, wrong result, loss of functionality
3) Minor problem, misspelling, UI layout, rare occurrence

Each severity of 1, 2 or 3, are assigned a **-QV (Negative Quality Value)** and these are:

- Severity 1 : -3QV
- Severity 2 : -2QV
- Severity 3 : -1QV

The worksheet contains a formula for calculating the -QV each time an 'open' bug is assigned a severity. Totals are calculated in the summary table.

The totals of the **-QVs** are subtracted from the opening **100 QV** score. These are chipping away at the overall quality of the item under test and erode the original assumption of 100 QV.

2.2 Test Cycle Calculations.

At the end of the test cycle we have a good indication of a more realistic QV for the product as it was sent to us for test. Charted, these are a key metric that can be used to base quality improvements on, in addition to the **Cost of Quality (CQ)**.

As bugs are closed the QV is recalculated on the bug report and adds back the -QVs, raising the QV of the product now being re-tested. So the Bug Report will show a true QV for the product as it exists now after fixes and completion of this test cycle. It may be that the QV does not rise greatly due to new bugs being found, as is often the case when fixes are delivered. If during testing the QV falls below the agreed **Lower Control Limit (LCL)** the test team can act, this is discussed later in the document.

A customer may sign off bugs so 'closing' them on the retest report, but the –QVs are still reflected on the Bug Report, the bug is still there so the -QVs will remain. Again, if charted over time, we will see the **Measure of Quality (MQ)** a customer can achieve over multiple releases of product and this can be used to qualify rework or launch decisions.

3.0 MEASURE OF QUALITY (MQ) AND LEVEL OF QUALITY (LQ)

3.1 An ongoing measure.

What do we get from the MQ? What is this actually telling us?

As with all good measures it can be applied and acted on both locally and globally. Locally an averaged sum of the QVs is calculated to arrive at a single figure for the MQ. Production and development teams can address ways to bring improvements to their individual product QVs and so improve the departmental average figure for the MQ, over time.

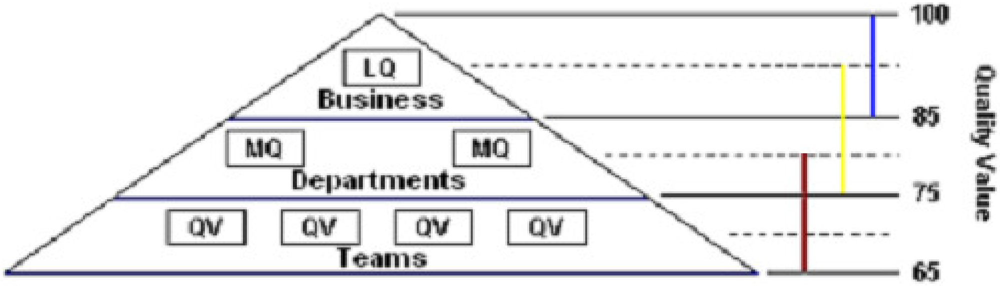
3.2 Level of Quality (LQ)

Globally, the business can set a **Level of Quality** for the service or product based on achievable MQ figures and set targets for improvement. By assessing exactly what can be achieved by production teams and investigating what is a cost effective value to achieve.

It would be reasonable to assign higher MQ expectations to high value service areas or web properties, lower expectations for lower value areas. This will bring the time and resources from test to focus on where it is most valuable to the business. Bringing the different MQs together will allow measurement of the global **LQ** of the business.

Each team is given the flexibility to deliver its QV within a range. Teams will need to balance their acceptance of low and high QV products as constantly low QVs mean the averaged value won't hit the departmental MQ. The key is there's flexibility to respond to internal and external environmental influences and not work with the assumption production process is fully defined.

Diagram 1 – Quality Value bands for each Measurement



In diagram 1 we can see there are three tiers, three coloured Quality Bands and Upper and Lower Control Limits (UCL and LCL)*:

- Level of Quality – Blue band – UCL100 – LCL 85. Median 90.
- Measure of Quality – Yellow band – UCL85 – LCL75. Median 80.
- Quality Value – Red band – UCL75 – LCL65. Median 70.

For a collection of products from the teams to achieve the MQ of the department their must naturally be products with QVs at the top of the red band. If not the QV average wont achieve the 85 – 75 value required. However, we have the flexibility to let some products only achieve the LCL of 65 and others meet or exceed the UCL of 75, just so long as when averaged we're achieving the MQ.

This same logic applies to MQ values helping the business achieve its desired LQ. Each measurement feeds into the next and, as expected, the cumulative affect of earlier quality activities mean the difference between meeting or failing the departmental MQ and then the global business LQ.

This again demonstrates the flexibility available through the system but encourages future process improvements to safeguard the ability to meet expected standards. Failure to achieve these can be quickly recognised and the causes identified for corrective action.

* Values are illustrative and should be modified for the production environment

4.0 USING THE APPROACH

4.1 Chart the Data

Now that we can measure the Quality Value of a product how will we use this information?
Each QV can be plotted on a **Control Chart**, as the chart develops we will be able to see the contribution to the departmental MQ and variation of the teams averaged QV.

With data on this chart we can easily set realistic and achievable levels of expectation for quality and agree what happens if these expectations are not met. A great step forward from the lack of understanding and ability to observe and improve the work we do.

4.2 The Control Chart

As we are going to collect data we need to record it in a fashion that helps analysis. In quality we have the control chart and this can be applied with our approach here.

Below is an example of a completed Control Chart. This shows the data gathered over more than thirty product tests. But to get here we need to have done some groundwork.

4.3 The UCL

The **UCL** is the reference point for the Internal Customer. This is the QV level they have declared they can and will achieve, before sending individual products to test or releasing to live. It's a way of saying 'it's good enough'.

This is a powerful measure in supporting launch decisions, if a product is business critical, albeit not perfect, perhaps we should release it now? Clearly we would want to look at bugs, business impact, etc. but now we have something solid to base our 'down to the wire' decisions.

What's more the UCL can also help to define the Level of Quality the business expects. Unlike the averaged MQ a department is aiming for, the LQ is fixed, nothing goes into the market unless it can support the LQ the business requires, allowing for the above of course.

4.4 The LCL

The **LCL** is a reference point for QA, the LCL is the point at which QA will cease testing of a product. In the chart the LCL of 70QV has been set realistically against a baseline MQ of nearly 80. QA and the Customer know that 70QV is below that which can be expected and so the product is returned for the Internal Customer to work on again.

This is a key event that alerts us to two possibilities:

- Firstly, if we were base lining the QV, it could show the production process is unable to meet our expectations or if this was along the chart, it shows the process has somehow begun to fail and we would need to address why.
- Secondly, the product simply isn't ready to release and therefore should not be in testing, so we send it back to the customer.

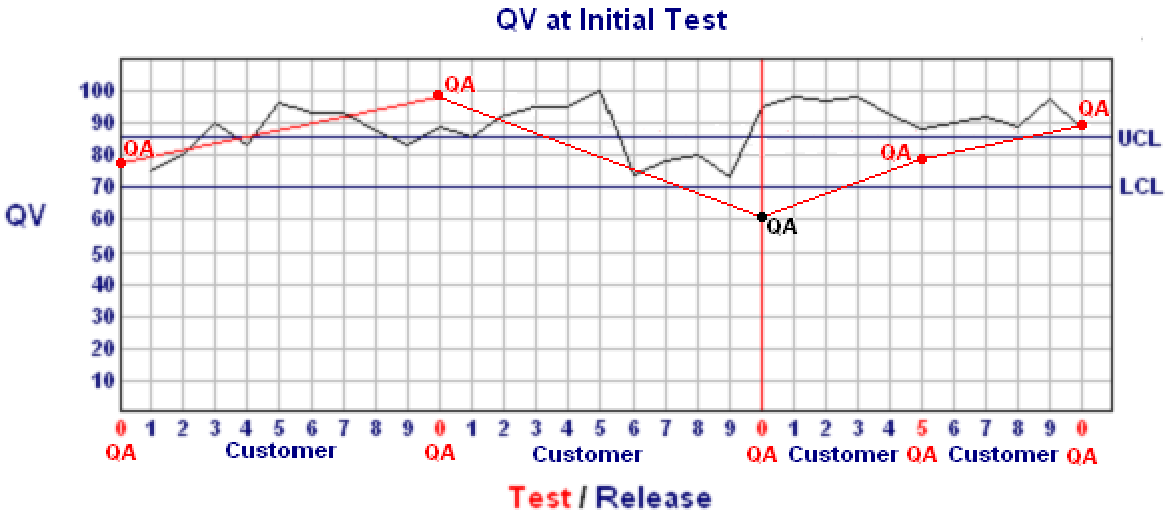
Anything between the LCL and the UCL is the usual test-fix cycle. Alternatively we could say anything between the LCL and 100QV, it may depend on the LQ the business has set.

These two measures can be applied however it helps us best, the flexibility is there to provide just the level of control needed to ensure products meet our expectations.

4.5 Base-Lining the QV

When commencing to use this approach our first activity is to find out where we are quality wise. To do this we need to plot the QVs for an agreed number of product tests. These will then demonstrate what we should expect of the upper and lower QVs of the team. From this we can set our expectation for the departmental MQ. The Upper and Lower control limits tell us when the process is out of control. If the QV is unusually high or low it can act as a trigger to agree stopping test or perhaps releasing the product as is.

Diagram 2 – Control Chart



Looking at the chart we can see the test frequency of 'QA' and 'Customer' is quite different. This shows we have agreed the customer will test and report on their products, QA monitor the process. At test 20 however QA find the QV below the agreed LCL. An incident line is put on the chart and QA double their test frequency. Test frequency is explained later in the paper.

4.6 Data Sources

Data for the chart comes from two sources, Test Team Bug Reports and the Internal Customers feedback. On the chart the testing activities and frequency for QA are the points labelled 'QA' and shown in red, with the unlabeled blue points for the Customer, along the 'Test / Release' line. Each time testing is performed a point is plotted on the chart.

5.0 TEST FREQUENCY

5.1 Setting the Start Frequency
How often QA need to test products from a customer can be drawn from and shown on the Control Chart. Analysing the chart used to declare the opening QV for this period sets the starting Test Frequency.

If the chart shown was the one used to baseline we simply check how often the Internal Customer produces work under the LCL. Now we have an indication of how often we need to make sure they are above it.

5.2 Increasing the Test Frequency
On our example chart it is set to every ten test runs, shown as ‘0’ on the bottom scale. The second time test occurs the LCL is exceeded and the Test Frequency remains every ten. However the third time QA test the QV is below the LCL and testing not only stops at that point but also the Test Frequency is increased to every five.

We can select any figure that gives us the control we feel is required. Should the testing reveal the LCL is still not being met we can increase Test Frequency repeatedly until QA are back to testing every product again if that’s the level of control required.

5.3 Reducing the Test Frequency
If the product exceeds expectations consistently during the Test Cycle we can safely assume it should continue to do so. There for we can reduce the Test Frequency as there is no measurable benefit or obvious need to perform testing. How much we reduce this is completely flexible. A reduction brings cost, time and resource savings no matter how small.

Again, referring to the Control Chart and values gives us a more defined way to reduce or increase testing.

6.0 THE COST OF QUALITY

6.1Reduced test brings reduced cost
This approach has provided a way to assign measurable value to product quality. With this insight QA can assess sound reasons to modify their test strategy for a customer.

One common approach is to recognise that there is little benefit in QA testing when it can be seen quality of products will be to the expected standards. Reducing the test frequency, in a measured and considered way, directly reducing the overall cost of quality for the business, improves the ROI to quality.

This also gives strong reason for customers to work with the process. Customers can see they directly affect the overall costs of the business and can be encouraged to achieve high QVs and so keep their associated costs lower. If departments are required to allocate a portion of their budget or are cross-charged then this further reinforces the tie between business costs and the quality of products they produce.

The cost of quality should be placed at the centre of a progressive business. Any activity that contributes to final product quality is a cost to quality, it should be measured and improved.

7.0 DEFINITIONS OF KEY ELEMENTS

7.1 100 Quality Value
When a product is sent to QA for test we have no idea how good it could be, so we assume it's perfect and assign it the 100QV. If we tried the reverse, increasing it's QV as we go when would we know it was 100? We only know what we find in testing and that's what we have available to measure.

7.2 -QValue
Each bug is assigned a -QValue and the value raises with severity. Severity 1, 2 and 3 are assigned a value of 3, 2 and 1 respectively. Get a lot of Sev1 bugs and your QV is dropping fast.

7.3 Lower Control Limit
This is the lowest level of QV that either QA will stop testing or the customer has asked testing to stop if reached. The LCL sets an expected minimum standard for the quality of a product sent for test. The LCL can be set against the QV of a customers products or it can be aligned to the CQ, just how much does it cost to test and retest a product that arrived for test with say 70QV? Perhaps testing should stop until the product is 'quality' enough?

7.4 Upper Control Limit
This is a measure of acceptable quality, you could say 'it's good enough, let it go'. If we consider the CQ again, there may be times setting a UCL will be beneficial.

7.5 x QV
The hidden QV of a product when it arrives for test. Until testing has been completed the true QV can't be known. However, expectations can be set and used to define the LCL. Expectations are based on previous testing.

7.6 Measure of Quality (MQ)
Similar to the LQ but applied to local quality values. What reduces the MQ can be investigated with Root Cause Analysis. Get to the root of the causes for a reduced MQ and watch the quality level rise.

7.7 Level of Quality (LQ)
The LQ is the MQ for the business, for the service we provide the customers. Achieving the LQ can be broken down into MQ expectations, hit these and we hit the LQ.

Notes and references:

Notes
LCL and UCL – QA industry terms for measuring process tolerance, the measurements being data for statistical process control.

Level of Quality and Measure of Quality – Loosely defined and applied terms.

100QV, QV, -QValue and xQV – These are terms and concepts originating from this paper.

Bibliography
① *The Web Testing Handbook,* Splaine, Jaskiel and Savoia, STQE Publishing, 2001**, ISBN 0970436300**

Comments
1. Problem
Complexity of development work across teams isn't reflected in this approach.

1. Response.
The approach is currently geared to multiple teams producing product of a similar complexity. Given complexity will vary across products results over time should normalise. A business would have to weight products given their complexity, values and importance to the business.

Specific corrective action approaches tied into the approach.
Reduction in complexity of the approach.
Training and mentoring of customers in the use of the approach.
-- PowerPoint presentation explaining the steps as phases completed