

Avaliação objetiva, mensurável e predizível de qualidade de Software através das métricas DRE (*Defect Removal Efficiency*) e DP (*Defect Potentials*) aplicado ao 'Projeto Banco'

Por
Prof. MSc. **Martin Tornquist**
T&M Teste de Software

Novembro de 2017
v1.3

Resumo Executivo

Este estudo apresenta e comenta a 'Qualidade' do 'Projeto Banco' – atualmente em fase de testes de Sistema e Homologação – à luz de valores de mercado.

É apresentada uma definição objetiva, quantificável e predizível de 'Qualidade de Software' e seus indicadores associados de qualidade, DRE (*Defect Removal Efficiency*), DDE (*Defect Detection Efficiency*) e DP (*Defect Potential*) são apresentados e exemplificados.

Finalmente a 'Qualidade' do 'Projeto Banco' é comparada com *benchmarks* para projetos de software de organizações do tipo 'Banking' e 'Financeira' e simula e comenta sua aplicação e implicações, com valores altos, médios e baixos, no 'Projeto Banco'.

Referências

Além da nossa própria base histórica de indicadores de qualidade de +300 projetos de Software/Teste que mantemos como empresa especializada em Testes há mais de 36 anos, utilizamos e citamos neste estudo explicitamente dados externos, internacionais e bastante atualizados (2014, 2016 e 2017) de duas das principais fontes de referência, estudos e métricas da indústria de software mundiais: Capers Jones (Namcook Analytics LLC) e ISBSG (*International Software Benchmark Standard Group*).

Referência 1:

"Exceeding 99% in defect. Removal Efficiency (DRE) for Software",
September 2016 by Capers Jones (Namcook Analytics LLC)

Referência 2:

"Special Analysis Report on Software Defect Density",
2014, by ISBSG (*International Software Benchmark Standard Group*)

Referência 3:

"Software Economics and Function Point Metrics – Thirty years of IFPUG Progress",
April 2017 by Capers Jones (Namcook Analytics LLC)

Histórico de mensurações objetivas, quantificáveis e predizíveis de qualidade de software

Todas as definições propostas para avaliar qualidade de software ('conformidade aos requisitos', 'adequação à finalidade' e 'confiabilidade / manutenibilidade / adequabilidade / segurança / conformidade / portabilidade / desempenho e compatibilidade' (ISO 25.010)) sofrem de ambiguidade, subjetividade e de dificuldades objetivas para sua mensuração e aplicação.

Um bom indicador de qualidade de software deve ser objetivo (para que exista uma maneira precisa de calculá-lo), quantificável (para que possa ser mensurado APÓS a entrega do software em produção) e predizível (para que possa ser utilizado para estimar a qualidade do software ANTES da entrega em produção).

Definição de Qualidade de Software

Uma definição objetiva, mensurável e predizível de qualidade de software seria:

“Qualidade de Software é a ausência de defeitos que ou provocam o não funcionamento do sistema/aplicação ou provocam a produção de resultados incorretos”.

Por esta definição, qualidade de software foca na prevenção e na eliminação de defeitos o que permite estabelecer as seguintes métricas objetivas, quantificáveis e predizíveis: o DP (*Defect Potential*) e o DRE (*Defect Removal Efficiency*). [ref 3, pág 15-16].

Definição da 1ª métrica: DP (*Defect Potential*)

DP (*Defect Potential*) vem a ser a soma dos defeitos provavelmente introduzidos nas fases de requisitos, arquitetura, desenho, codificação, documentação, testes, erros de correção de bugs nos testes e produção.

Isto se mede objetivamente coletando os defeitos produzidos em ou atribuíveis a cada fase para todos os projetos. Com esta métrica é possível prever o número potencial de defeitos em uma aplicação.

DP (*Defect Potential*) costuma ser medido utilizando Pontos de Função (PF) como medida do 'tamanho' do projeto pois assim é possível prever o número de defeitos logo na 1ª fase do projeto (onde os pontos de função já podem ser estabelecidos).

Valores típicos de DP (*Defect Potential*)

O DP (*Defect Potential*) por Ponto de Função, em 2016, por fase do desenvolvimento, nos projetos estudados pelo Capers Jones é o seguinte: [ref 1 pág. 3]

Table 4: U.S Average Ranges of Defect Potentials Circa 2016
(Defects per IFPUG 4.3 function point)

Defect Origins	Best	Average	Worst
Requirements	0.34	0.70	1.35
Architecture	0.04	0.10	0.20
Design	0.63	0.95	1.58
Code	0.44	1.15	2.63
Security flaws	0.18	0.25	0.40
Documents	0.20	0.45	0.54
Bad fixes	0.39	0.65	1.26
TOTAL	2.21	4.25	7.95

Um projeto de 1.000 PF produziria, portanto, algo entre $1.000 \times 2,21 = 2.210$ defeitos ('*best case*') e $1.000 \times 7,97 = 7.970$ defeitos ('*worst case*') e, para a maioria dos projetos, $1.000 \times 4,25 = 4.250$ defeitos ('*average case*')!

As variações entre '*best*', '*average*' e '*worst*' deve-se a variações do tipo e da maturidade dos processos de desenvolvimento bem como da *expertise* de seus participantes.

Na ausência de Pontos de Função (desconhecemos o tamanho do 'Projeto Banco' em Pontos de Função) podemos utilizar o número de Casos de Testes como uma aproximação para prever o potencial de defeitos.

Quando se utilizam Casos de Teste em vez de PF a conversão é direta: Cada Caso de Testes tem o potencial de detectar 1 defeito, portanto, para 1.000 Casos de Testes o DP (*Defect Potential*) é de 1.000 defeitos.

Definição da 2ª métrica: DRE (*Defect Removal Efficiency*)

DRE (*Defect Removal Efficiency*) é o percentual de defeitos localizados e corrigidos ANTES do *release* do software aos clientes.

Caso por exemplo, antes do *release* tenham sido localizados e corrigidos 95 defeitos e, nos primeiros 90 dias em produção tenham sido localizados e corrigidos 5 defeitos adicionais, nossa DRE (*Defect Removal Efficiency*) será de $95 / (95 + 5) = 95/100 = 95\%$

Segundo o Capers Jones [ref 1 pág 7] valores típicos para o DRE (*Defect Removal Efficiency*) seriam:

Table 6: U.S. Software Average DRE Ranges by Application Size

Function Points	Best	Average	Worst
1	99.90%	97.00%	94.00%
10	99.00%	96.50%	92.50%
100	98.50%	95.00%	90.00%
1000	96.50%	94.50%	87.00%
10000	94.00%	89.50%	83.50%
100000	91.00%	86.00%	78.00%
Average	95.80%	92.20%	86.20%

É evidente que todo projeto importante e crítico deveria apresentar um DRE (*Defect Removal Efficiency*) superior a 99% (pois só isso garante pouquíssimos defeitos em produção e o menor custo global dos projetos), mas poucos projetos o alcançam!

O que significa uma DRE de 92,20%? (*average* da coluna *average* da tabela acima)

Supondo a existência de 1.000 defeitos em um software; para um DRE (*Defect Removal Efficiency*) de 92,2%, um total de 922 defeitos foi localizado ANTES da entrega do software (durante os requisitos, desenho, arquitetura, programação e testes) e um total de 78 defeitos foi detectados APÓS a entrega (*nos primeiros 90 dias de operação do software*)!

Setenta e oito (78) defeitos em produção é um número absoluto bastante elevado e inadequado na maioria dos projetos e uma DRE de 92,2% é uma DRE inadequada (muito baixa) para qualquer projeto crítico (segundo a tabela acima os melhores projetos conseguem uma DRE média de 95,8%).

Severidade das falhas em Produção

Felizmente os mesmos estudos indicam que a severidade das falhas em produção, na média [ref 1, pág 9] é a seguinte:

Here too it is painfully obvious that defect volumes go up with application size. However table 8 shows all severity levels of delivered defects. **Only about 1% of delivered defects will be in the high-severity class of 1 and only about 14% in severity class 2. Severity class 3 usually has about 55% while severity 4 has about 30%.**

Os potenciais 78 defeitos em produção do exemplo acima se reduzem a $78 \times 1\% = 1$ defeito de severidade 1 (software não funciona), $78 \times 14\% = 11$ defeitos de severidade 2 (grandes funcionalidades inoperantes) e o restante, $78 - 1 - 11 = 66$ defeitos de severidade 3 ou 4 (falhas menores que não impedem a utilização ou falhas cosméticas).

Assim podemos prever tanto a quantidade de defeitos, por fase do desenvolvimento, quanto à sua severidade!

Avaliação da 'Qualidade' do 'Projeto Banco' em função do número e severidade de falhas encontrados pelos Testes.

Em um total de 1.200 casos de Testes (*potencialmente 1.200 falhas*) foram encontradas 116 falhas (9,6% dos casos de teste falharam) até 28 novembro de 2017. Destes, 3,4% (4 falhas) são falhas de Severidade 1 (*show-stoppers*) e 31% (36 falhas) são falhas de severidade 2.

Estes percentuais são aproximadamente o triplo e o dobro respectivamente das falhas desta severidade esperados (3,4% em vez de 1% e 30% em vez de 14%)!

As consequências diretas deste nível de qualidade abaixo do esperado é que serão necessários + retrabalhos, + ciclos de (re)Teste, + custos e + prazo para alcançar um produto de alta qualidade em produção.

DRE (Defect Removal Efficiency) PRÉ-TESTES e DRE dos TESTES

Neste estudo vamos dividir o DRE (*Defect Removal Efficiency*) final de um projeto em um **DRE Pré-Testes** (o % de defeitos localizados e corrigidos ANTES da fase de Testes de Sistema e Aceite, ou seja: dos Requisitos aos testes unitários) e no **DRE dos Testes** (o % dos defeitos localizados e corrigidos DURANTE os testes de Sistema e Aceite).

Os dois se combinam (vide tabela a seguir) para determinar a DRE (*Defect Removal Efficiency*) final (o % de defeitos localizados e eliminados ANTES do *release* aos clientes) de um projeto.

Como seria o 'Projeto Banco' com valores de mercado "best in class" de DRE?

A planilha a seguir ilustra os resultados predizíveis para o 'Projeto Banco' utilizando valores altos de DRE's (*Defect Removal Efficiency*) [na tabela utilizamos valores "best in class" medidos e publicados pelo Capers Jones, ref 1 páginas 16-18]:

Utilizamos, para efeito de simulação e simplificação do entendimento um DP (Defect Potential) de 1 defeito por caso de Testes. No 'Projeto Banco' temos em torno de 1.200 Casos de Teste o que daria um potencial máximo de 1.200 defeitos.

Modelo de DRE (Defect Removal Efficiency) Pré Testes e dos Testes		
Total de Defeitos Potenciais:	1200	
DRE Pré-Testes ('best in class')	91,69%	
DRE dos Testes ('best in class')		89,78%
Total de Defeitos Removidos Pré-testes e nos Testes:	1100	90
Total de Falhas remanescentes em produção:	10	
DRE combinada das DRE's pré-testes e dos testes:		99,2%

Na tabela acima, uma DRE Pré-Testes de 91,6%, combinada com uma DRE dos Testes de 89,78% (ambos valores tidos como 'best in class' pelo Capers Jones) gera uma DRE final de 99,2%!

Apenas 10 defeitos potenciais chegariam à produção (= alta qualidade) e para estes podemos prever que haverá menos de um defeito de severidade 1 (1% x 10), menos de 2 defeitos de severidade 2 (14% de 10) e até 8 defeitos de severidades 3 e ou 4!

Entendendo a planilha:

- A DRE pré-testes localizou 91,6% dos potenciais 1.200 defeitos ou seja 1.100;
- Para os testes sobram portanto $1.200 - 1.100 = 100$ defeitos a serem localizados;
- A DRE dos Testes localizou 89,78% destes 100 defeitos ou seja, 90 defeitos;
- Em produção sobram, portanto $1.200 - 1.100 - 90 = 10$ defeitos em potencial;

Os melhores projetos ('best in class') sempre apresentam DRE's altas

Em todos os levantamentos realizados e publicados pelo Capers Jones, os projetos de alta qualidade pressupõem DRE's (tanto pré-testes quanto dos testes) individuais em torno de ou maiores que 90% com uma DRE final de 99% ou maior (maiores informações sobre DRE's dos mercados 'Banking' e 'Financeira' no final deste estudo (página 7)).

Simulação 1: Como seria um projeto com DRE medianas (75%)?

Modelo de DRE (Defect Removal Efficiency) Pré Testes e dos Testes		
Total de Defeitos Potenciais:	1200	
DRE Pré-Testes ('median in class')	75,00%	
DRE dos Testes ('median in class')		75,00%
Total de Defeitos Removidos Pré-testes e nos Testes:	900	225
Total de Falhas remanescentes em produção:	75	
DRE combinada das DRE's pré-testes e dos testes:		93,8%

A DRE final, de 93,8% não parece muito menor que os 99,2% do 'best in class' mas:

- Observe a diferença de número de defeitos em produção: de 10 para 75, quase 8 X mais!
- Durante os testes haverá mais de o dobro de falhas a reportar, corrigir e retestar (de 90 para 225)!

DRE's medianas não servem nem para projetos medianos, foram aqui apresentados apenas para contextualizar as demais simulações e para demonstrar, de forma unívoca, que baixos DRE's impactam diretamente os prazos de entrega (ampliados) e os custos (maiores) dos projetos!

Simulação 2: Utilizando as DRE's premissas do 'Projeto Banco'

O 'Projeto Banco' estabeleceu como premissa contratual uma DRE de 90% tanto para as atividades Pré-Testes quanto para as atividades de Testes visando alcançar uma DRE final de 99%, dentro dos parâmetros de projetos considerados 'best in class':

Modelo de DRE (Defect Removal Efficiency) Pré Testes e dos Testes			
Total de Defeitos Potenciais:	1200		
DRE Pré-Testes ('premissa do projeto Bancs ATG')	90,00%		
DRE dos Testes ('premissa do projeto Bancs ATG')		90,00%	
Total de Defeitos Removidos Pré-testes e nos Testes:	1080	108	
Total de Falhas remanescentes em produção:			12
DRE combinada das DRE's pré-testes e dos testes:			99,0%

Isto resulta em um projeto com DRE final de 99%, com elevada qualidade (apenas 12 defeitos previstos em produção, dos quais apenas 1 a 2 de severidade 1 e 2) e permite entregar o projeto dentro do prazo e do orçamento.

Simulação 3: Como seria um projeto com DRE pré-testes mediana (75%) e uma DRE dos Testes elevada (90%)?

Para ilustrar, vamos considerar, finalmente, uma DRE pré-testes de 75% e uma DRE de Testes de 90%. O resultado está na tabela a seguir:

Modelo de DRE (Defect Removal Efficiency) Pré Testes e dos Testes			
Total de Defeitos Potenciais:	1200		
DRE Pré-Testes ('Projeto Banco')	75,00%		
DRE dos Testes ('Projeto Banco')		90,00%	
Total de Defeitos Removidos Pré-testes e nos Testes:	750	225	
Total de Falhas remanescentes em produção:			25
DRE combinada das DRE's pré-testes e dos testes:			97,5%

Novamente a DRE final, de 97,5% parece 'adequada' mas implica, quando comparada com a tabela 'premissa do Projeto Banco' (em azul), no dobro de defeitos em produção (de 12 para 25) e no dobro de defeitos a localizar e eliminar durante os testes (de 108 para 225).

Simulação 4: Situação do 'Projeto Banco' em final de Novembro de 2017

A DRE pré-testes está um pouco abaixo (84%) da premissa, a DRE dos Testes está dentro da Premissa (90%) gerando a seguinte DRE momentânea:

Modelo de DRE (Defect Removal Efficiency) Pré Testes e dos Testes			
Total de Defeitos Potenciais:	1200		
DRE Pré-Testes ('projeto Bancs ATG')	84,00%		
DRE dos Testes ('projeto Bancs ATG')		90,00%	
Total de Defeitos Removidos Pré-testes e nos Testes:	1008	173	
Total de Falhas remanescentes em produção:			19
DRE combinada das DRE's pré-testes e dos testes:			98,4%

Observem que isto está gerando um impacto de aproximadamente 50% a mais de falhas em produção (de 12 para 19) e na quantidade de falhas a localizar e eliminar durante os testes (de 108 a 173) do que o estimado para o projeto!

Um terceiro e último indicador: 'Bad-fix defects'

Segundo Capers Jones [ref 1, pág 3] o índice de defeitos introduzidos pela correção de outros defeitos é de na média, nos EUA, de 7%.

No caso do 'Projeto Banco' este índice está em 21,9%! Isto também impacta o prazo, custo e qualidade de entrega do Software aos clientes.

Conclusão: 'Quality is free'

Não é sempre intuitivo mas projetos com altos DRE's, tanto pré-testes quanto nos testes, como aqui demonstrado, alcançam uma qualidade maior (*número de defeitos em produção menor*) em prazos e a custos menores.

Quais seriam os DRE's para projetos do Mercado 'Banking' e 'Finacial'

O ISBSG (*International Software Benchmark Standards Group*) publicou um levantamento sobre Densidade de Defeitos [ref 2, página 3] por tipo de organização.

Separámos os DRE's de projetos de organizações tipos 'Banking' e 'Finacial' e os resultados estão na planilha a seguir:

ISBSG 2016	Defects per Function Point in Operation		Defect Potential per Function Point (CMMI 4)	Final DRE		Considering a Test DRE of:	Demands a Pre Test DRE of	
	Mean	Worst		Mean	Worst		Mean	Worst
Banking	0,0100	0,0909	2,5	99,60%	96,36%	90%	96,00%	63,64%
Finacial	0,0050	0,0428		99,80%	98,29%		98,00%	83,09%

Consideramos um nível de maturidade CMMI 4 (*projetos geridos e mensurados*) para determinar o potencial de Defeitos (2,5 defeitos por ponto de função segundo Capers Jones [ref 1 página 12]).

Interpretação da tabela:

Segundo o ISBSG, para projetos em organizações 'Banking' a densidade de defeitos em produção, na média (*mean*), seria de 0,01 defeitos por Ponto de Função. Isto se traduz em uma DRE final de 99,6%¹.

Considerando uma DRE dos Testes de 90%, a DRE pré-testes teria que ser de 96% (em azul)².

No mesmo segmento e 'Banking', para os piores (*worst*) projetos, a densidade de defeitos é 9 x maior (0,0909 (*worst*) na planilha acima) que corresponde a uma DRE de apenas 96,36% e, ainda considerando a mesma DRE dos testes de 90%, gera uma DRE pré-testes de meros 63,64% (em vermelho)!

Este seria um projeto onde a qualidade é assegurada apenas pelos testes de sistema e de aceite não havendo atividades significativas de inspeção, revisão e testes ANTES destes Testes.

O estudo do ISBSG também mostrou que parcela significativa (entre 60% e 70%) dos projetos estudados nas organizações de tipo 'Banking' e 'Finacial' tiveram ZERO falhas em produção (nos primeiros 30 dias) o que significaria uma DRE de praticamente 100% (não é 100% porque podem aparecer falhas em produção após os 30 dias);

A mesma interpretação se aplica aos projetos do tipo 'Finacial'.

Quais seriam as DRE's pré-testes praticadas neste mercado?

No setor 'Banking' varia de 63,64% (*worst case*) a 96,00% (*best case*).

¹ $(2,5 - (2,5 * 99,60\%)) = 0,01$

² $(2,5 - ((2,5 * 96\%) + ((2,5 - (2,5 * 96\%)) * 90\%))) = 0,01$.



No setor 'Finacial', varia de 83,09% (*worst case*) a 98,00% (*best case*).

É evidente que, para assegurar a saúde econômica dos projetos e a sua qualidade, precisamos procurar nos aproximar dos cenários '*best case*'!

Por isso nossa premissa de DRE's de 90%, tanto para os Testes quanto para os Pré-testes, gerando uma DRE final de 99% representam uma ótima relação custo benefício e estão no espectro superior da faixa de DRE's finais apontadas na tabela acima (96,36% - 99,8%) para projetos da área '*Banking*' e '*Finacial*'.

-x-