

# Testmetrieken

## In 10 vragen en antwoorden

Uitkomsten van de Valori Meeting of Minds, 13 december 2011  
Versie 1, opgesteld door Derk-Jan de Groot en Egbert Bouman



***Not everything that can be counted counts, and not everything that counts can be counted***  
(Einstein)

***Je hebt leugens, grove leugens en statistieken". "Als het dan toch moet, dan liever mijn eigen leugens"***  
(Churchill)

***You can't control what you can't measure***  
(Tom DeMarco vroeger)

***The more you focus on control, the more likely you deliver something of relatively minor value.***  
(Tom DeMarco nu)



## AAN DE LEZER

Beste bezoeker van de Valori Meeting of Minds,

13 december stonden we in Lage Vuursche stil bij het thema “Slimme Metrieken”. Het stuk wat voor u ligt biedt de beloofde inhoudelijke terugkoppeling. We hebben uw en onze visie gestructureerd aan de hand van de volgende vragen:

- 1 Waarom gelooft Tom DeMarco niet meer in metrieken?
- 2 Zijn metrieken inderdaad zinloos?
- 3 Het gaat toch om het verhaal, niet om de cijfers?
- 4 Wat zeggen de bekende aanpakken en modellen over metrieken?
- 5 KPI's en metrieken, is dat hetzelfde?
- 6 Kan ik met metrieken mijn testafdeling krachtiger positioneren?
- 7 Is het invoeren van een testmetriekenprogramma moeilijk?
- 8 Welke testmetrieken adviseert Valori?
- 9 Metrieken en Scrum, hoe gaat dat samen?
- 10 Hoe zit het met mens en psychologie?

Hartelijk dank voor uw bijdrage, wij zijn geïnspireerd, u hopelijk ook. Reacties en verdere vragen zijn van harte welkom en als u verder met ons wil sparren kan dat natuurlijk ook.

We horen het graag!

Egbert Bouman en Derk-Jan de Groot

*(Product Managers Valori Testing & Acceptance)*

### Valori Meeting of Minds, wat is dat?

*Een Valori Meeting of Minds is niet zomaar een interessante vakinhoudelijke bijeenkomst, maar een sociale/culinaire ervaring waarbij u na een inspirerende inleiding in klein comité met vakgenoten van gedachten wisselt in een persoonlijke, informele en prettige setting. In kort bestek krijgt u de stand van zaken rondom een thema mee en gaat u met inspirerende ideeën en concrete handvatten naar huis.*



## 1 Waarom gelooft Tom DeMarco niet meer in metriecken?

Tom DeMarco is de aartsvader van het meten en besturen van ICT-projecten. Zijn boek “Controlling Software Projects: Management, Measurement and Estimation” met de klassieke openingszin “You can’t control what you can’t measure” is het standaardwerk over softwaremetriecken.

Echter: DeMarco zelf gelooft niet meer in zijn eigen verhaal.

In een artikel in IEEE Magazine stelt hij zichzelf drie vragen:

1. Was mijn toenmalige advies correct?
2. Is het nog relevant?
3. Geloof ik nog steeds in de noodzaak van metriecken?

DeMarco: “My answers are: no, no and no”. Hij is tot het inzicht gekomen dat de toegevoegde waarde van op metriecken gebaseerde besturing marginaal is en totaal geen succesfactor. En als we dan toch gaan meten dan gaat het mis.



Hij is in goed gezelschap, want Einstein zei het al: “Not everything that can be counted counts, and not everything that counts can be counted”.

### *Continue transformatie*

En daar zit veel waars in. DeMarco stelt dat we bijna altijd de simpele dingen meten en niet dat waar het om gaat: een proces van continue transformatie van de business. Ondersteund door ‘agile’ software teams, die niet aan harde kwantitatieve targets werken maar creatief, met visie en experimenterend de reis meemaken en faciliteren.

## 2 Zijn metriecken inderdaad zinloos?

Wij van Valori denken dat we heel goed naar DeMarco moeten luisteren. Want deze man heeft geleerd wat ‘lean’ en ‘agile’ daadwerkelijk betekenen voor ‘Business IT Optimization’.

Maar we moeten metriecken niet afschaffen, in elk geval niet voor testers die een onafhankelijke vinger aan de pols moeten houden. Ook hier geldt dat onze bijdrage aan het “transformatieproces” natuurlijk dat intuïtie en gevoel belangrijker zijn dan we vaak denken en de “blauwe ogen” van de test manager en de *key users* annex de acceptanten kunnen de doorslag geven. Maar een objectieve en liefst kwalitatieve onderbouwing van ons oordeel over kwaliteit en voortgang horen ons inziens bij een professioneel testproces.

### *Voldoende is voldoende*

De algemene conclusie van de Meeting of Minds was dat het belangrijk is om metriecken te hebben die technisch voldoende betrouwbaar zijn. Dat “voldoende” is cruciaal want als organisatie moet je metriecken gericht inzetten ter ondersteuning van je doel en je informatiebehoefte. Vaak is een gemakkelijke en goedkopere metriek voldoende. Testen is domweg niet volwassen genoeg om alle relevante resultaten te kunnen kwantificeren. Dit geldt overigens voor systeemontwikkeling in zijn algemeenheid, en de vraag is of dat kan en moet veranderen. Zie DeMarco!



#### *Over de sloot springen*

Betrouwbare metrieken zijn primair belangrijk als je in de kritische zone zit. Stel je kunt ongeveer twee meter ver springen. Als je over een sloot moet springen van bijvoorbeeld 60 cm, of 4 meter hoef je niet precies te weten hoe breed de sloot is. Als je springt ben je zeker van een nat pak. Alleen als de sloot ongeveer 2 meter is, loont het om je meetlint te pakken en je kansen nauwkeurig te berekenen.

#### *Garantie voor mislukking*

Tot slot: Metrieken zijn nooit een middel om kansloze projecten alsnog succesvol te maken. Sommige projecten zijn bij voorbaat kansloos. Meindert Munnik (UVIT): een project dat langer dan negen maanden duurt met een team van meer dan 10 mensen is bijna een garantie voor mislukking. Dat ga je met metrieken en een strakke besturing echt niet opvangen.

### **3 Het gaat toch om het verhaal, niet om de cijfers?**

Inderdaad, uiteindelijk gaat het om het verhaal: geloven we erin of geloven we er niet in? Durven we de vrijgave aan of durven we het niet aan? Maar dat sluit metrieken niet uit, want die versterken het verhaal.

#### *Trust én Proof*

In alle gevallen gaat het om "Trust én Proof", zoals Bert Bouwmeester van Equens het tijdens de Meeting of Minds verwoordde. Ook al omdat interne en externe auditors en wetgeving die 'proof' simpelweg eisen. Een professioneel kwaliteitsoordeel en vrijgaveadvies kunnen dus niet zonder een zo goed mogelijke kwantitatieve onderbouwing.

#### *Innovatieve projecten*

Bij nieuwe projecten is er weinig ervaring. De metrieken die je hier presenteert zijn anders dan bij bijvoorbeeld een maandelijkse release. Bij de laatste is de trend belangrijk en kun je sturen op cijfers. Echter, dit is omdat men vanuit ervaring het verhaal er achter de cijfers kent. Bij een innovatief project is de informatiebehoefte meer contextgedreven. Metrieken moeten daar antwoord geven op vragen als "Gaat het goed?" en "Weten waar we mee bezig zijn?" Het gaat hier dus primair om het verhaal en de metrieken dienen als ondersteuning.

### **4 Wat zeggen de bekende aanpakken en modellen over metrieken?**

Over metrieken is al veel gezegd en geschreven. Zowel Prince2 en CMMi als testmethodieken en volwassenheidsmodellen als SmarTEST, TMap, TestGoal, TMMi en TPI zeggen iets over testmetrieken. Op de TMMi volwassenheidsschaal hoort het procesgebied "Test Measurement" goed ingevuld te zijn om op het vierde niveau "Measured" (de naam zegt het al) in aanmerking te komen.

#### *RUP*

Rational Unified Process (RUP) adviseert voor elk systeemontwikkeltraject minimaal 3 "key metrieken": Earned Value, Test Progress en Defect Trend. Wat daaraan opvalt is dat twee van deze drie metrieken *test*metrieken zijn. Het testproces is dus de belangrijkste leverancier van metrieken voor projectbesturing.



## 5 KPI's en metrieke, is dat hetzelfde?

In de ruime betekenis van het woord wel, maar concrete metrieke voor het ontwikkel- en testproces zijn niet altijd even makkelijk aan de KPI's op organisatieniveau te koppelen. Op organisatieniveau is vaak sprake van Kritische Succesfactoren en Key Performance Indicators, al dan niet gecombineerd en gepresenteerd in een overall "Business Balanced Scorecard" (BBS) met de kwadranten klant, organisatie, financiën en innovatie. Het software testproces is indirect leverancier voor een deel van deze cijfers. Om de relevantie van het testproces en de uitkomsten beter uit te kunnen dragen is het van belang om een metriekeprogramma te relateren aan de KPI's op organisatieniveau. En dat is heel goed mogelijk, vooral als we in staat zijn betekenisvolle cijfers over rendement en efficiëntie van het testproces te verschaffen.

### *Metrieke helpen je te leren en te verbeteren*

Het structureel bijdragen aan KPI informatie vraagt het verzamelen van metrieke over de projecten heen. Na elk project gaat echter veel informatie verloren. Mensen vinden het blijkbaar leuker om het wiel op nieuw uit te vinden, dan zich te verdiepen in evaluaties. Benchmarken over projecten heen en het uitvoeren van trend analyses is iets wat theoretisch mooi klinkt, maar vaak in het gedrang komt. Zolang er geen pijn wordt gevoeld gaat dit niet veranderen: als er niemand last van heeft wordt er ook niet gemeten. Dus kies je ambities bewust en begin hier alleen aan als er breed gevoelde pijn is!

## 6 Kan ik met metrieke mijn testafdeling krachtiger positioneren?

Ja, zie ook het antwoord op bovenstaande vraag. Een testafdeling die niet in staat is om zijn relevantie voor organisatiebrede doelstellingen en KPI's aan te tonen heeft op den duur geen bestaansrecht. De maatschappij vraagt om transparantie en feitelijke controleerbaarheid en het testproces is bij uitstek in staat daarin te voorzien.

### *Balans*

De uitdaging hierbij is de balans vinden tussen twee dingen:

- verschaffen van kwantitatieve data over kwaliteit en voortgang
- voluit meedraaien in het creatieve ('agile') transformatieproces waarin je niet alles kunt en wilt meten.

De testteams en –afdelingen die hierin goed positie kiezen zijn verzekerd van support door zowel de directie als de werkvloer van de onderneming.

## 7 Is het invoeren van een testmetriekeprogramma moeilijk?

Nee, niet moeilijker dan "gewoon goed testen". En bij gewoon goed testen hoort onlosmakelijk dat je objectief rapporteert over voortgang en kwaliteit, met cijfers en percentages. Wat wel moeilijk kan zijn is het met cijfers onderbouwen dat testen rendabel is en een hoge ROI heeft.

### *Slim en simpel*

Je kunt sterven in schoonheid, en dat is precies wat er met te ingewikkelde metriekeprogramma's gebeurt. Dus doe het slim en hou het simpel. Met een set



basismetrieën die er niet “ook nog eens bijkomen”, maar die sowieso van direct belang zijn voor het testproces. Zie de volgende vraag.

## 8 Welke testmetrieën adviseert Valori?

Tijdens de MoM is een set met een vijftigtal vragen gepresenteerd. Stuk voor stuk vragen die in de praktijk ook gesteld worden en waar je een eenduidig antwoord op wilt kunnen geven, niet alleen kwalitatief maar ook kwantitatief, met metrieën dus.

### *Vier hamvragen*

Analyse van de vragenlijst leert dat het uiteindelijk om vier grote “hamvragen” gaat:

4. Hoe ver ben je?
5. Wat heb je gevonden?
6. Wat heeft het gekost?
7. Wat heb ik eraan?

Volgens de GQM methode kunnen we hier een set metrieën van afleiden.

### *Set basismetrieën*

Valori adviseert een set basismetrieën die het antwoord geven op deze hamvragen. In de “Goal Question Metrieën” aanpak leiden deze vragen tot de volgende set metrieën:

Goal	Question	Metric
Voortgang en dekking meten	<b>Hoe ver ben je?</b>	<u>Percentage uitgevoerde testen</u> (testgevallen/scenario's/...) Voorwaarde: controleerbare dekking van risico's, acceptatiecriteria en requirements (RAR) door benoemde testen.
Resultaten! Inzicht in risico's en kwaliteit	<b>Wat heb je gevonden?</b>	<u>Aantallen bevindingen</u> En vooral hun trend in de tijd. Met relevante impactclassificatie, actuele status en andere attributen, conform de best practices van bevindingenbeheer.
Weten wat het kost	<b>Wat heeft het gekost?</b>	<u>Bestede uren</u> Plus eventuele kosten van testomgevingen, toollicenties, et cetera.
Het rendement bepalen	<b>Wat heb ik eraan?</b>	Verhouding <u>gevonden in test / in productie</u> DDP: Defect Detection Percentage

In een normaal volwassen testaanpak is het consequent bijhouden van deze metrieën goed te doen. In de Meeting of Minds zijn twee wat lastiger uitdagingen benoemd, namelijk hoe meet je voortgang en hoe vergelijk je test met productie.

### *1. Hoe meet je voortgang?*

De voortgangsvraag hangt nauw samen met de dekkingsgraadvraag. Hier geldt bij uitstek dat een kaal getal helemaal niets zegt. “We zijn 70% klaar”. Klaar met wat? Wat heb je geraakt, wat niet? Zowel voortgang als dekkingsgraad kun je op verschillende manieren meten:

- Uitgevoerde testen: test cases, test scenario's, test scripts, ...



- Afgedekte risico's, acceptatiecriteria en requirements (RAR)
- Code coverage: lines, branches, conditions, statements, classes, objecten, ...
- Gevalideerde (business) processen
- Enzovoort

Valori adviseert om voortgang te meten aan de hand van uitgevoerde testen, *mits*:

- die duidelijk gerelateerd zijn aan relevante (business) doelen, risico's en requirements
- er een weging wordt gemaakt: het ene testgeval weegt zwaarder dan het andere
- de presentatie aan de business weer in business termen plaatsvindt, conform de vrijgavekaart

Dit alles vraagt wel een minimum aan documentatie, nodig om voortgang te kunnen meten. Minder focus op documentatie, bijvoorbeeld bij een exploratief testen aanpak, hoeft op zichzelf niet slecht te zijn maar als zelfs de charters of missies die bij een dergelijke aanpak horen niet beschikbaar zijn dan wordt het wel moeilijk.

## 2. Hoe vergelijk je test met productie?

Het bepalen van de DDP stelt eisen aan de vergelijkbaarheid van de bevindingenadministratie in test en de incident/probleem registratie in productie. Hier wordt in de literatuur weinig aandacht aan besteed maar het überhaupt kunnen bepalen van het DDP staat of valt hiermee. De volgende tips kunnen van dienst zijn:

- Maak afspraken met beheer:
  - Testen eindigt niet met technische go-live
  - Testen eindigt pas na een periode van "Begeleide productie" (bv 3 maanden)
  - Hanteer het aloude ITIL IPW model: Incident, Probleem, Wijziging
- Zorg dat de Problemen goed worden geadmineistreerd
  - Op dezelfde manier als bevindingen uit test
  - En liefst in dezelfde tool...

### *Projecten verschillen sterk*

Wouter Smith (TNT Express) merkt op dat het moeilijk blijft omdat de projecten zo verschillend zijn en niet onderling vergelijkbaar. Jeroen van Seeters (Menzis) sluit daarbij aan en voegt eraan toe dat IT projecten nauwelijks meer Software Development zijn, maar pakketinrichting.

### *Tooltip hierbij: QSM*

Een van de deelnemers kwam met een tip: de de tool "Quantitative Software Metrics", die metrieken bevat van tienduizenden softwareprojecten wereldwijd, ook pakketimplementaties. Je kunt benchmarkcijfers samenstellen op basis van projecten die vergelijkbaar zijn met de jouwe. Anand Bisoon (KPN, nu zelfstandig consultant) bevestigt dat KPN deze tool al meer dan 10 jaar met succes gebruikt. Niet voor elk project, maar wel regelmatig.

### *Hanteer bewezen vuistregels en verhoudingscijfers*

Een voorbeeld is het SmarTEST UTBM model, dat op basis van overweldigende "proof" uit projecten en onderzoeken stelt dat systeem- en acceptatietest samen meestal circa 30% van het budget beslaan. Dit wordt unaniem als relevant en zinvol gezien, mits in de juiste context toegepast. Dit voorbeeld geldt voor een "gemiddeld" administratief IT project. Meindert Munnik (UVIT): Philips hanteert weer andere cijfers voor medical systems en de NASA voor ruimtevaarttechnologie waar mensenlevens mee gemoeid zijn ook. Daar loopt het aandeel testen op tot 80 a 90%.



## 9 Metrieken en Scrum, hoe gaat dat samen?

Een sterk punt van Scrum is dat een beperkt aantal metrieken wordt gehanteerd die een concrete rol spelen in de daily cycle en de sprint. Scrum schrijft het meten van *velocity* voor en het gebruik van de *burndown chart*. Wat niet iedereen zich realiseert is dat dit slechts voortgangsmetrieken zijn. Scrum heeft weinig te bieden op het punt van kwaliteitsmetrieken.

### *Kwaliteit zonder metrieken?*

Scrum kent andere mechanismes om kwaliteit te waarborgen zoals korte feedbackloops en gebruikersparticipatie. Die mechanismes zijn bewezen effectief en het Scrum team hoeft niet altijd metrieken op te leveren, zeker als er sprake is van een acceptatietest met metrieken, buiten het Scrum proces. De deelnemers waren overigens van mening dat dit een goede zaak is en dat de acceptatietest in de meeste situaties niet binnen het Scrum proces thuishoort. Het is een aparte activiteit ná oplevering door het Scrum team, conform het SmarTEST W-model.

### *Complete set Scrum metrieken*

Desondanks: de beperktheid van de standaard Scrum metrieken wordt breed gevoeld. Voor een oriëntatie op meer Scrum metrieken adviseren wij de whitepaper over "Measuring Scrum", hier te vinden: <http://rapidscrum.com/Agile2010-ScrumMetrieken.pdf>

## 10 Hoe zit het met mens en psychologie?

Metrieken zijn bij uitstek geschikt om gedrag te sturen. Dat pakt niet per definitie goed uit. Naarmate metrieken een belangrijkere rol spelen in onze beoordeling, onze reputatie of onze mogelijkheden tot beïnvloeden worden ze belangrijker en zijn we eerder bereid er iets mee te doen of ons steentje bij te dragen.

### *Leugens*

Churchill zei het al: "Je hebt leugens, grove leugens en statistieken". Hij voegde eraan toe: "Als het dan toch leugens zijn dan liever mijn eigen leugens". Metrieken kunnen dus heel opportunistisch en oneigenlijk worden gebruikt, bijvoorbeeld om je in te dekken.

### *Transparantie en openheid van zaken geven*

Maar ook om transparantie te creëren, te laten waar je mee bezig bent, hoe het gaat. Mensen hebben soms een sterk gevoel bij metrieken, en niemand wil graag geassocieerd worden met een slechte score. Ondanks dat er vaak een verhaal bij deze score hoort, onthouden mensen vaak wel de score maar niet het verhaal. Wil je metrieken gebruiken als verbetering, kan het effectief zijn om ze te indexeren. Bij de nulmeting zet je de score op 100, en vandaar monitor je ontwikkeling. Dit is minder gevoelig.





### Sturing beïnvloedt gedrag

In sommige organisatie heerst het geloof dat je complexe processen kunt beheersen aan de hand van een paar KPI's. Het gaat echt om het verhaal, en sturen op KPI's gaat hieraan voorbij. Bijgaande cartoon



behoeft geen toelichting. Het artikel "Rethinking Software Metrics" van Cem Kaner werkt het gevaar en (soms) de schijnzekerheid van metrieken verder uit. Een deelnemer van de Meeting of Minds vertelt van een sales manager die afgerekend werd op de korting die hij kon bedingen. De leveranciers hielp hem door de initiële prijs te verhogen en de salesmanager een hoge korting te geven en daarmee te scoren op zijn KPI. De organisatie was niet goedkoper uit. Zeer herkenbaar. Een andere deelnemer vertelt het verhaal dat er op testkosten werd bespaard door goedkope testers in te huren. Na verloop van tijd begonnen mensen te klagen dat de testers onvoldoende kennis en ervaring hadden. Kortom als het doel niet duidelijk is, en als het verhaal achter de cijfers niet verteld wordt, kunnen vele ongewenste effecten ontstaan.

*Tot zover. Wilt u verder sparren? Neem gerust contact met ons op.*

*Egbert Bouman en Derk-Jan de Grood, Valori*

