



TESTEN DUUR? NIET TESTEN IS DUURDER!

Bepaal het testkostenoptimum met eenvoudige kengetallen

*auteurs: Leo van der Aalst en Corné de Koning
gebaseerd op de originele publicatie in: Informatie*



© 2010, Sogeti Nederland B.V. te Vianen.

Niets uit deze uitgave mag verveelvoudigd en/of openbaar worden gemaakt (voor willekeurig welke doeleinden) door middel van druk, fotokopie, microfilm, geluidsband, elektronisch of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van Sogeti Nederland B.V.

TMap® is een geregistreerd handelsmerk van Sogeti Nederland B.V.

TESTEN DUUR? NIET TESTEN IS DUURDER!

Bepaal het testkostenoptimum met eenvoudige kengetallen

*auteurs: Leo van der Aalst en Corné de Koning
gebaseerd op de originele publicatie in: Informatie*

Op de vraag "wat vindt u van het testproces?", volgt steevast het antwoord "duur!". Dit antwoord kan zelden op waarde worden geschat, omdat het vaak gevoelsmatig wordt gegeven: cijfermatige onderbouwing ontbreekt. "Testen is duur." Ja, dit is waar als hierbij alleen de testkosten, en niet de testbaten in ogenschouw worden genomen. Testkosten zijn onder andere gebaseerd op de kosten van testinfrastructuur, de bestede uren van de testers en hun tarieven. De kwaliteit van het systeemontwikkelingsproces speelt ook een rol. Is deze minder, dan zijn er meer fouten aanwezig. Dit betekent dat er grondiger, met een hogere dekkinggraad, getest moet worden om ze tijdens het testproces te detecteren. Fouten die in productie optreden kunnen leiden tot gevolgschade en hogere herstelkosten. De voorkomen gevolgschade en herstelkosten vormen de testbaten. Het hieraan gekoppelde dilemma is: moet ik testen tot alle fouten eruit zijn, voordat ik met de applicatie in productie kan gaan? Los van het feit dat nooit alle fouten voor die tijd worden gevonden, is het heel goed mogelijk dat het vooraf willen vinden van een fout duurder is dan hem in productie te laten optreden. Er moet hierover dus een afweging gemaakt worden.

"Testen duur? Niet testen is duurder!" De titel van dit artikel suggereert dat er een optimum in de kosten is te vinden. [Juran, 1988] laat zien dat er een optimum van de totale kwaliteitskosten bestaat. In dit artikel willen we laten zien dat het optimum berekend kan worden met gebruik van eenvoudige kengetallen, en zonder dat historische gegevens nodig zijn. In tegenstelling tot Juran zetten we de kosten niet af tegen de kwaliteitsgraad, maar tegen de tijd. Dit gebeurt om het tijdstip te kunnen bepalen waarop het niet meer (kosten)rendabel is om door te gaan met testen. Op dit bepaalde moment, het *kostenoptimum*, zijn niet alle fouten van het systeem gevonden, maar toch is het vanuit kosten oogpunt hét moment om te stoppen met testen en in productie te gaan.¹

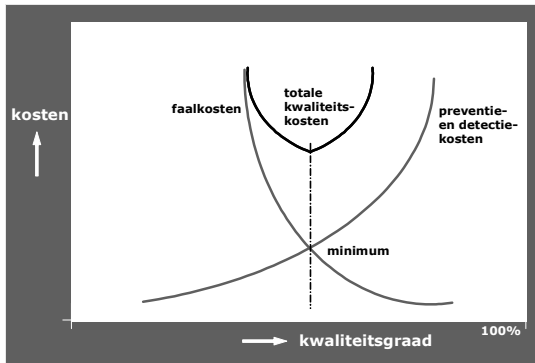
UITGEDIEPT

Kosten kwaliteitszorg

De maatregelen die in het kader van kwaliteitszorg moeten worden genomen, kosten geld. Het totaal van deze verschillende soorten kosten worden kwaliteitskosten genoemd. Hieronder worden de volgende kosten verstaan:

- Preventiekosten
de kosten voor het nemen van preventieve maatregelen
- Detectiekosten
de kosten voor het nemen van detectieve maatregelen
- Faalkosten
de kosten voor het nemen van correctieve maatregelen of de kosten als gevolg van onvoldoende kwaliteit (gederfde omzet, extra servicekosten, garantieclaims, schadeclaims).

In onderstaand figuur van Juran wordt de samenhang tussen die kosten weergegeven.



Een belangrijke boodschap van Juran is dat het nastreven van perfecte kwaliteit economisch niet rendabel is, omdat steeds meer moet worden geïnvesteerd om een klein beetje extra kwaliteit te krijgen.

BEPALEN TESTKOSTENOPTIMUM

Voor het bepalen van het optimum zijn diverse gegevens nodig. Deze gegevens worden in figuren gepresenteerd in plaats van in cijferreeksen, aangezien figuren in de praktijk verhelderder werken.²

De figuren worden opgebouwd uit drie stappen:

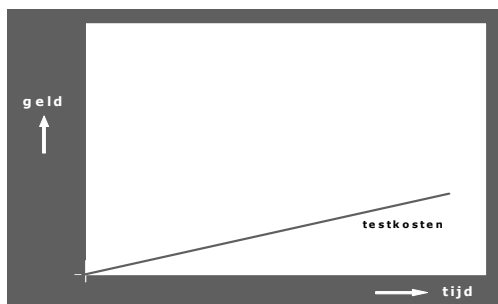
1. berekenen testkosten
2. berekenen voorkomen schade
3. bepalen kostenoptimum.

Berekenen testkosten

Testkosten zijn opgebouwd uit diverse kosten. Naast personeelskosten zijn er onder andere ook kosten voor infrastructuur, testhulpmiddelen, beheer en huur van kantoorruimte. Het is niet eenvoudig om op deze manier testkosten te berekenen. In de praktijk blijkt het product van de bestede testuren en het uurtarief een goed alternatief te zijn. Een bijkomend voordeel is dat hiermee de leesbaarheid van de figuren toeneemt. In formulevorm ziet de berekening van de testkosten er als volgt uit:
 $testkosten = bestede\ testuren \times uurtarief$

Met testuren worden naast de testuitvoeringsuren ook de uren bedoeld die zijn besteed tijdens de overige fasen in het testproces: planning, beheer, inrichting en beheer infrastructuur, voorbereiding, specificatie en afronding [Aalst, 2006].

In figuur 1 "Testkosten" zijn de testkosten geïllustreerd. Er is gebruikgemaakt van genoemde formule, waarbij aangenomen is dat het testteam een constante bezetting kent. Dit betekent dat de lijn van de testkosten door het snijpunt van de kosten- en tijdas loopt en als een rechte lijn is getekend.



Figuur 1: Testkosten

Berekenen voorkomen schade

De voorkomen schade is de som van de potentiële gevolgschade van alle gedane bevindingen. Het vaststellen van de bevinding is geen probleem: deze is rechtstreeks afkomstig uit het testproces. Het bepalen van de gevolgschade per bevinding, indien deze in productie zou zijn opgetreden, is lastiger. In de praktijk blijkt hier toch een redelijk betrouwbaar antwoord op gegeven te kunnen worden: in een select gezelschap, bestaande uit een ontwikkelaar (analist), een gebruikersvertegenwoordiger en een vertegenwoordiger van het rekencentrum, worden de bevindingen één voor één op gevolgschade beoordeeld. Elke vertegenwoordiger gaat te werk vanuit zijn eigen verantwoordelijkheid. De ontwikkelaar geeft een schatting van de extra herstelkosten die nodig zijn als de bevinding niet tijdens het testproces, maar in productie zou zijn gevonden. De gebruiker geeft aan wat de verwachte schade is ten gevolge van gederfde inkomsten, schadeclaims, wachttijd gebruikers, enzovoort. Het rekencentrum geeft onder andere de verwachte schade aan ten gevolge van performanceverlies en restore-activiteiten.

VOORBEELD

Bij een bank wordt een 'real-time' fraudedetectiesysteem (FDS) getest. Dit systeem bepaalt op basis van historische gegevens of een uitgevoerde transactie een mogelijk fraudegeval is. Tijdens de testuitvoering wordt geconstateerd dat de fraudescore (hoe hoger de score, hoe groter de kans dat het een fraudegeval betreft) die bij een bepaalde transactie hoort, niet wordt verwijderd bij het bepalen van de fraudescore van een volgende transactie, maar erbij wordt opgeteld. De ontwikkelaar schat de extra herstelkosten, onder andere het maken van een conversieprogramma, in op 1000 euro. De gebruiker constateert dat er meer transacties als fraudegevoelig worden bestempeld dan er in werkelijkheid zijn. Dit betekent extra uitzoekwerk voor de gebruikers van het FDS. Dit wordt geschat op 5000 euro. Het rekencentrum schat het installeren van het nieuwe programma en het draaien van een conversieprogramma op 500 euro. De totale geschatte gevolgschade bedraagt hiermee 6500 euro.

Het schatten van de voorkomen schade voor elke gevonden bevinding zoals bij het FDS is gedaan, is een arbeidsintensieve werkwijze. Wij hebben ervoor gekozen om een schatting te maken van het *gemiddeld* voorkomen schadebedrag per bevinding. De gevolgschade per bevinding kan nogal verschillen: de ene bevinding leidt tot een eenvoudige cosmetische schermaanpassing, terwijl bij een andere geen enkele gebruiker toegang heeft tot het systeem. Om tot een gemiddeld voorkomen schadebedrag te komen, is gebruikgemaakt van een weging per bevinding. De gehanteerde weegfactoren zijn:

- oorzaak (testbasis: 3, omgeving: 2, testobject: 1)³
- ernst (zwaar: 8, middel: 6, licht: 2, wens: 1)
- kwaliteitsattribuut (continuïteit: 10, functionaliteit: 8, beveiliging: 7, performance: 5, gebruikersvriendelijkheid: 1, controleerbaarheid: 1).

Bij toekenning van weegfactoren geeft de ontwikkelaar aan wat de oorzaak van de bevinding is, de gebruiker noemt de ernst van de bevinding en ontwikkelaar, gebruiker en rekencentrum kennen samen het kwaliteitsattribuut (de eigenschap van het informatiesysteem waarop de bevinding betrekking heeft) toe.

VOORBEELD

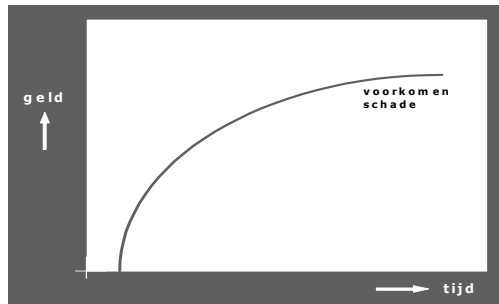
De ontwikkelaar stelt vast dat de bevinding uit het vorige voorbeeld is veroorzaakt door een fout in het testobject. De gebruiker beoordeelt de ernst ervan als middelmatig en ontwikkelaar, gebruiker en rekencentrum samen bepalen dat de bevinding betrekking heeft op het kwaliteitsattribuut functionaliteit. Het totaal aantal weegpunten komt hiermee op: $1 \times 6 \times 8 = 48$.

Van een aantal bevindingen wordt zo nauwkeurig mogelijk bepaald wat de gevolgschade in geld zou zijn. Vervolgens wordt dit bedrag gedeeld door het aantal toegekende weegpunten, zodat er een schadebedrag per weegpunt wordt vastgesteld. Het berekenen van de gevolgschade van de bevindingen houdt vanaf dit moment in dat het totaal aantal weegpunten per bevinding wordt vastgesteld, welke vervolgens wordt vermenigvuldigd met het schadebedrag per weegpunt.

VOORBEELD

De gevolgschade van bevinding 1 wordt geschat op 6500 euro. Oorzaak is testobject, ernst is middelmatig en kwaliteitsattribuut is functionaliteit.
Aantal weegpunten bedraagt: $1 \times 6 \times 8 = 48$.
Gevolgschade bevinding 2 wordt geschat op 1000 euro. Oorzaak is testbasis, ernst is licht en kwaliteitsattribuut is gebruikersvriendelijkheid.
Aantal weegpunten bedraagt: $3 \times 2 \times 1 = 6$.
Gevolgschade bevinding 3 wordt geschat op 12.000 euro. Oorzaak is testbasis, ernst is middelmatig en kwaliteitsattribuut is performance.
Aantal weegpunten bedraagt: $3 \times 6 \times 5 = 90$.
Het totaalbedrag aan geschatte gevolgschade bedraagt: 19.500 euro. Totaal aantal weegpunten bedraagt: 144. Het schadebedrag per weegpunt komt hiermee op 135 euro. 'Scoort' een nieuwe bevinding 60 weegpunten, dan wordt het schadebedrag geschat op 8100 euro (60×135 euro).

In figuur 2 "Voorkomen schade" is de som van de geschatte gevolgschade van alle gevonden bevindingen uitgezet tegen het tijdstip waarop de bevindingen zijn gevonden. Doordat de bevindingen zijn gevonden vóór het systeem in productie is genomen, spreken we van voorkomen schade.



Figuur 2: Voorkomen schade

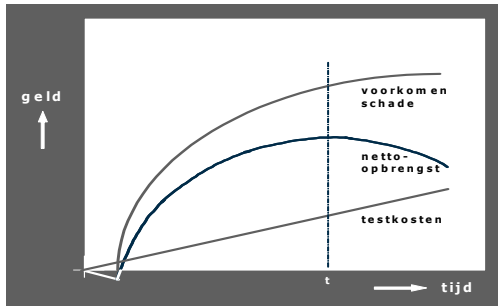
De voorkomen schadecurve in figuur 2 begint niet bij tijdstip 0, omdat het testproces niet direct tot het vinden van bevindingen leidt. In de fase Planning worden geen bevindingen gevonden. Tijdens de fasen Voorbereiding en Specificatie gebeurt dit wel, maar tijdens de testuitvoering worden relatief de meeste bevindingen gevonden. In figuur 2 is aangenomen dat alle bevindingen tijdens de testuitvoering worden gevonden. Door een goed uitgevoerde risicoanalyse en teststrategiebepaling loopt de curve aan het begin steil op. Het doel van de teststrategie is zo vroeg mogelijk de belangrijkste fouten vinden tegen de minste kosten. (De belangrijkste fouten veroorzaken immers de grootste gevolgschade en bij ontdekking tijdens het testproces wordt daardoor de grootste schade voorkomen.) De vorm van de curve wordt ook bepaald doordat er in het begin van een testproces meer bevindingen worden gedaan dan naarmate het proces vordert.

Bepalen kostenoptimum

Het bepalen van de netto-opbrengst van het testproces (zie figuur 3) is nu een relatief eenvoudige exercitie. Door de testkosten van de voorkomen schade af te trekken,

ontstaat de netto-opbrengstcurve. Het kostenoptimum is in de figuur aangegeven met een verticale stippellijn.

Het tijdstip (t) waarop het optimum wordt bereikt, is tevens het tijdstip waarop overwogen kan worden om met het testproces te stoppen. Immers, *testen moet doorgaan zolang de kosten van het vinden en herstellen van een fout in de test lager zijn dan de kosten verbonden aan het optreden van de fout in productie.*



Figuur 3: Netto-opbrengst testproces

Als er vóór het optimum wordt gestopt met testen, kan men concluderen dat dit te vroeg is: de maximale netto-opbrengst is nog niet bereikt. Hier tegenover staat dat als men na het bereiken van het optimum nog steeds aan het testen is, de netto-opbrengst afneemt. Het is vrijwel onmogelijk om óp het optimum te stoppen met testen. Ook te vroeg stoppen kan grote consequenties hebben. Het is daarom aan te bevelen om met testen te stoppen, zodra het duidelijk is dat de netto-opbrengstcurve daadwerkelijk aan het dalen is. Het hierbij genomen verlies ten opzichte van het stoppen op het optimum, is te prefereren boven de risico's die genomen worden door te vroeg met testen te stoppen.

POTENTIËLE FAALKOSTEN

De voorkomen schade wordt direct afgeleid van de bevindingen. Het is echter onjuist om te veronderstellen dat als de voorkomen schadecurve horizontaal begint te lopen – dus als geen bevindingen meer gedaan worden – ook alle bevindingen in het systeem zijn gevonden. Wellicht is er tijdens de strategie bepaling ten onrechte vanuit gegaan dat bepaalde functies met een lagere dekking, of zelfs helemaal niet, getest moeten worden. In zo'n geval is het mogelijk dat er tijdens productie nog vele bevindingen worden gedaan.

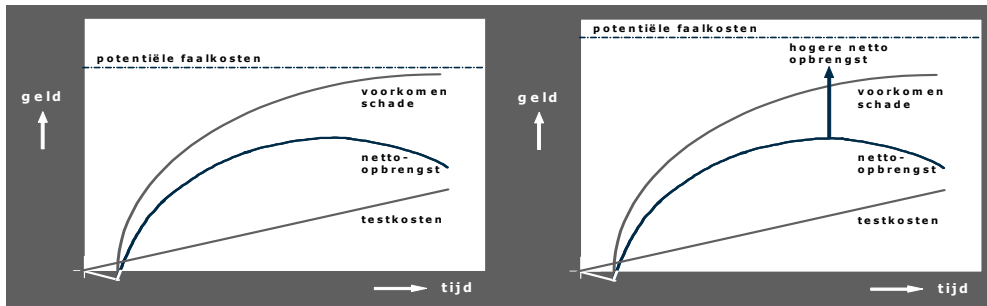
In feite wordt tijdens het testproces slechts een percentage van alle aanwezige bevindingen gevonden. De som van de tijdens test gevonden bevindingen en de in productie optredende bevindingen, is het totaal aanwezige bevindingen in het systeem. Hierbij worden eventuele bevindingen die wel in het systeem aanwezig zijn, maar die niet productie zijn opgetreden of niet kunnen optreden, niet als bevinding meegenomen.

Waarom is het interessant om te weten hoeveel bevindingen in het systeem aanwezig zijn? Als men 'slechts' de voorkomen schade tot zijn beschikking heeft, kan er geen oordeel gegeven worden over de verhouding van de voorkomen schade met de potentiële faalkosten. Dit is in feite een maat voor de kwaliteit van het testproces en de -strategie. De potentiële faalkosten worden berekend door de schade te schatten voor alle tijdens het testproces gevonden bevindingen (= maximaal voorkomen schade). Vervolgens wordt de schade die opgetreden is tijdens de eerste drie maanden productie hierbij opgeteld. De potentiële faalkosten zijn dus pas na *drie maanden productie* bekend. Dit kan te laat zijn om, indien nodig, de teststrategie van het project nog te wijzigen. Wel zijn deze gegevens bruikbaar voor de schatting van potentiële faalkosten bij toekomstige projecten.

In figuren 4a en 4b wordt de relatie tussen voorkomen schade en potentiële faalkosten verduidelijkt.

In figuur 4a is te zien dat de voorkomen schade de potentiële faalkosten nadert. In dit geval is het niet te verwachten dat er nog veel meer te voorkomen schade in het systeem aanwezig is. De nog aanwezige schade bedraagt maximaal het verschil tussen de potentiële faalkosten en de voorkomen schade. Met andere woorden, de netto-opbrengst is een juiste afspiegeling van de werkelijkheid en het is een juiste keuze om op het optimum stoppen.

In figuur 4b is te zien dat de potentiële faalkosten aanzienlijk hoger liggen dan de voorkomen schade. Het verschil tussen de potentiële faalkosten en de voorkomen schade is maatgevend voor de nog in het systeem aanwezige schade. Met een andere teststrategie is het wellicht mogelijk om met de voorkomen schade de potentiële faalkosten te benaderen. Dit betekent dat de netto-opbrengst hoger kan uitvallen. (Deze is immers gedefinieerd als het verschil tussen voorkomen schade en testkosten.)



Figuur 4a: Kwaliteit testproces goed

Figuur 4b: Kwaliteit testproces niet goed

AANBEVELINGEN

Vóóraf zijn de potentiële faalkosten niet bekend. Het verschil tussen voorkomen schade en potentiële faalkosten is dus onbekend. De keuze van het moment van stoppen kan daarom niet gebaseerd worden op het verschil tussen deze beide gegevens.

Onze aanbeveling is: *stop niet met testen als de netto-opbrengst nog een stijgende lijn laat zien. Beter is om te stoppen als de netto-opbrengst al enige tijd een daling laat zien.* De kans dat het systeem dan nog veel bevindingen (gevolgschade) bevat, is daarmee kleiner geworden.

In de kaders op de volgende pagina's geven we enkele praktijkvoorbeelden waarbij gebruikgemaakt is van de in dit artikel beschreven werkwijze. Ondanks het feit dat altijd pas achteraf (na drie maanden productie) vastgesteld kan worden of het moment van stoppen met het testproces het juiste is geweest, heeft deze aanpak in de praktijk tot bevredigende resultaten geleid.

PRAKTIJKVOORBEELDEN

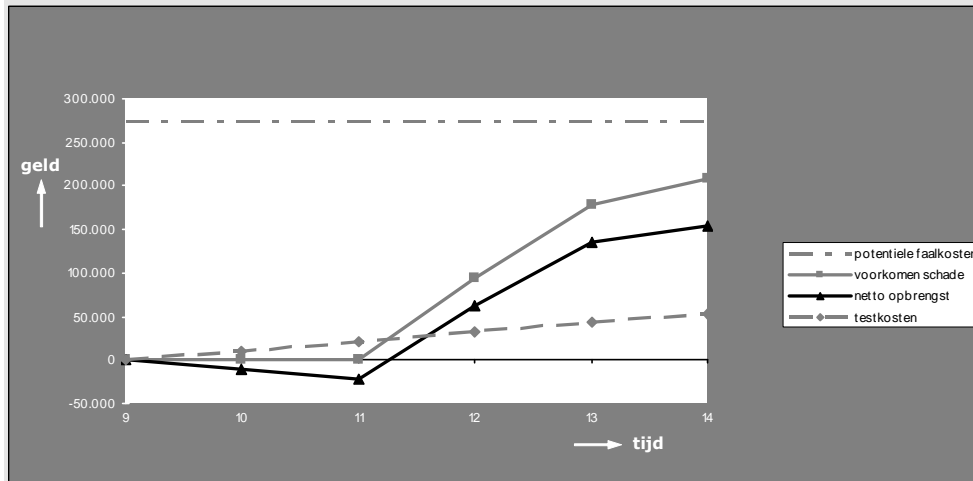
De gegevens uit de praktijkvoorbeelden zijn afkomstig uit een testproces van een administratieve applicatie van een grote financiële instelling.

Voorbeeld 1: te kort getest?

De netto-opbrengst vertoonde nog een stijgende lijn, op het moment (week 14) dat er werd gestopt met testen. Dit betekent dat er te vroeg werd gestopt. Achteraf bleek dat de schade, tijdens drie maanden productie, 65.000 euro bedroeg. Met andere woorden, de netto-opbrengst had maximaal 42 procent (220.000 euro in plaats van 155.000 euro) hoger kunnen zijn. Hierbij moet echter nog rekening worden gehouden met de testkosten die nodig waren geweest om deze productieverstoringen tijdens het testproces te vinden.

Conclusie:

In eerste instantie leek er te vroeg te zijn gestopt. Dit werd na drie maanden productie bevestigd: de netto-opbrengst had hoger kunnen zijn.



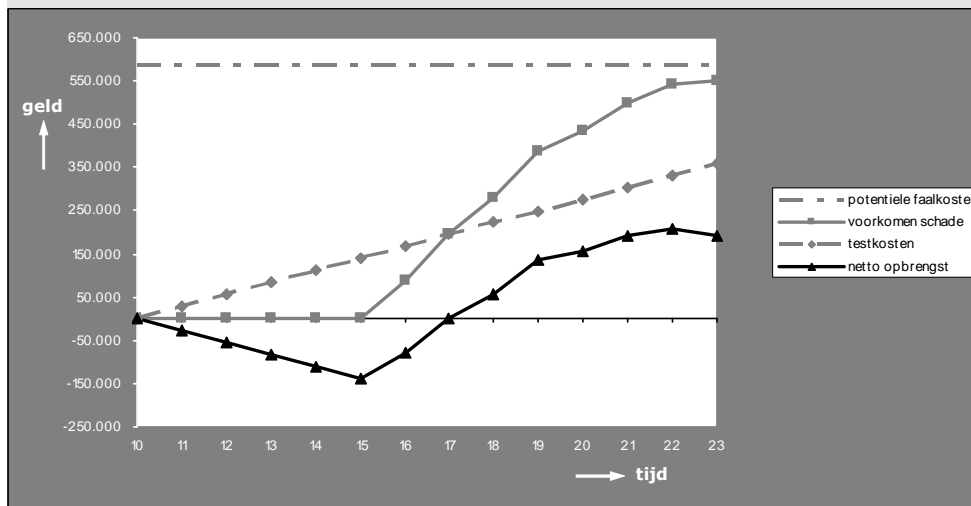
bij voorbeeld 1

Voorbeeld 2: op het juiste moment gestopt?

De netto-opbrengst vertoonde een licht dalende lijn, op het moment (week 23) dat er werd gestopt met testen. Dit zou betekenen dat er op het juiste moment werd gestopt. Achteraf bleek dat de schade, tijdens drie maanden productie, 30.000 euro bedroeg. Met andere woorden, de netto-opbrengst had maximaal 15 procent (220.000 euro in plaats van 190.000 euro) hoger kunnen zijn, los van de testkosten die nodig waren geweest om deze productieverstoringen tijdens het testproces te vinden.

Conclusie:

In eerste instantie leek er al op het juiste moment te zijn gestopt, en na drie maanden productie werd dit bevestigd.



bij voorbeeld 2

LITERATUUR

- [Juran, 1988]
Juran, J.M. (1988), *Juran's Quality Control Handbook*, McGraw-Hill, ISBN 0-070-33176-6
- [Aalst, 2006]
Aalst, L. van der, Broekman, B., Koomen, T., Vroon, M. (2006), *TMap Next, voor resultaatgericht testen*, 's-Hertogenbosch: Uitgeverij Tutein Nolthenius, ISBN 90-72194-79-9

-
- ¹ Het zou natuurlijk nog mooier zijn als het moment van stoppen voor aanvang van het testproject bepaald zou kunnen worden. Hiervoor zijn echter historische gegevens nodig en moeten diverse aannames gemaakt worden. Dit betekent dat het maken van een dergelijke 'voorspelling' een ingewikkelde en omgevingsafhankelijke activiteit is, waardoor de methode maar beperkt toepasbaar is. In dit artikel wordt de registratieve aanpak behandeld, het 'voorspellende' gedeelte is buiten beschouwing gelaten.
 - ² Voor het krijgen van bruikbare resultaten is een goede uren- en bevindingenregistratie onontbeerlijk gebleken. Als dit in uw organisatie of project niet adequaat is geregeld, raden wij u af om uw besluitvorming, omtrent het moment van stoppen met het testproces, te baseren op de beschreven figuren.
 - ³ Fouten in testbasis zijn fouten die (als ze niet ontdekt waren) leiden tot fouten in de programmatuur. Bij het corrigeren van dit soort fouten zijn hoge herstelkosten gemoeid. Fouten op testobject worden als 'normaal' beschouwd. Fouten in de omgeving zijn vaak lastiger op te lossen, deze krijgen daarom de waarde 2.