HANDREIKING

Verantwoordelijk marktgedrag en de kunst van een gewogen oordeel

Laura Bosman, MSc MSc
Ir. Jonna van der Krift
Em. Prof. Jan Telgen

code verantwoordelijk marktgedrag

Code Verantwoordelijk Marktgedrag, 2021

Voor het realiseren van deze handreiking heeft de Code Verantwoordelijk Marktgedrag een subsidie ontvangen van het Ministerie van BZK.

Auteurs Public Procurement Research Center
Laura Bosman, MSc MSc
Ir. Jonna van der Krift
Em. Prof. Jan Telgen

Informatie info@codevm.nl

Versie juli 2021

code verantwoordelijk marktgedrag

Inhoudsopgave

1. Introductie 1

1.1. Verantwoordelijk marktgedrag reikt tot in de techniek van het beoordelen 1

1.2. Opzet van de handreiking 2

2. Effect van de gekozen beoordelingsmethodiek op de uitkomst 3

2.1. Wat is een beoordelingsmethodiek? 3

2.2. Het belang van een goedgekozen scorerange 3

2.3. Doen de weegfactoren hun werk? 5

3. Valkuilen die leiden tot onbedoeld grote nadruk op prijs 6

3.1. Het onderscheidend effect van de beoordeling op kwaliteit is te klein 6

3.1.1. De gehanteerde scorerange is te breed 6

3.1.2. Er wordt op een onhandige manier gewerkt met een minimumscore voor kwaliteit 6

3.1.3. De gehanteerde prestatiecategorieën zijn te algemeen 7

3.1.4. Van individuele beoordeling naar eindbeoordeling in een beoordelingscomité 8

3.1.5. Last but not least: de beoordelingscriteria zelf zijn niet onderscheidend 8

3.2. Het onderscheidend effect van de beoordeling op prijs is te groot 8

3.2.1. Verkeerd gekozen scorerange leidt tot te grote scoreverschillen 8

3.3. Het effect van een beoordelingsformule kan ook leiden tot onbedoelde nadruk op prijs 9

4. De valkuilen van relatief beoordelen 11

4.1. Relatief beoordelen met kans op rank reversal en bid rigging 11

4.2. Veelgebruikte relatieve beoordelingsmethodieken die niet (meer) kunnen 12

4.3. Veilig relatief beoordelen: verschil in scores afhankelijk van een constante 13

5. Een aantal absolute beoordelingsmethodieken 14

5.1. Beoordelingsmethodiek met een scoregrafiek met horizontale lijnen 14

5.2. Beoordelingsmethodiek met een scoregrafiek die als parabool lopen 14

5.3. Beoordelingsmethodiek met een scoregrafiek met negatieve waarden 15

5.4. Beoordelingsmethodiek met een scoregrafiek met duidelijke knik 15

5.5. Beoordelingsmethodiek met een scoregrafiek van een omgekeerde hyperbool 15

5.6. Voorbeeldformules bij scoregrafieken met rechte lijn(en) 16

6. Alternatieve gunningsmethodieken 18

6.1. Gunnen op waarde 18

6.2. Value for money 19

6.3. Beoordeel eerst zelf met dummy's en deel de methodieken met aanbieders 19

1. Introductie

Het uitbesteden van diensten moet voor alle betrokkenen een positieve uitkomst hebben. De opdrachtgever krijgt waar voor zijn geld, de opdrachtnemer realiseert een gezond rendement en de medewerkers van opdrachtnemer worden gewaardeerd voor hun vakmanschap. Een dergelijke succesvolle uitbesteding is het resultaat van verantwoordelijk marktgedrag.

De Code Verantwoordelijk Marktgedrag stelt zich tot doel om dit gedrag te stimuleren en zo een continue impuls te geven aan verbetering van het uitbestedings- en inleenbeleid in Nederland. De Code doet daartoe een moreel appèl op opdrachtgevers, opdrachtnemers, inleners, vakbonden en intermediairs. Elk van deze partijen heeft een rol in het sociaal verantwoord en met respect voor kwaliteit uitzetten, aannemen en uitvoeren van opdrachten.

1.1 Verantwoordelijk marktgedrag reikt tot in de techniek van het beoordelen

In ons land worden dagelijks vele opdrachten gegund. Vaak gebeurt dit op basis van de uitslag van een offertetraject of aanbesteding. Daarin worden aanbiedingen beoordeeld om te bepalen welke het beste past. Soms is het uitgangspunt dat de 'beste aanbieding' die met de laagste prijs is. Meestal is het uitgangspunt dat de 'beste aanbieding' die met de beste prijs-kwaliteitverhouding is, ook wel de economisch meest voordelige inschrijving genoemd. Kwaliteit bestaat dan uit een aantal criteria, zoals duurzaamheid, communicatie met opdrachtgever en innovatief vermogen. Kwaliteit en prijs worden elk beoordeeld met een score en om tot het eindoordeel te komen, krijgen deze beide criteria een weging mee. Kwaliteit weegt bijvoorbeeld voor 70% mee en prijs voor 30%. Toch blijft vaak het gevoel achter dat prijs zwaarder heeft meegewogen in het eindoordeel dan beoogd. Onderzoek bevestigt dit gevoel.

De Code richt zich nu voornamelijk op aspecten die te maken hebben met houding en gedrag. Minder ver uitgewerkt zijn specifiek technische aspecten van het marktgedrag, zoals de manier waarop opdrachtgevers aanbiedingen beoordelen. Nu kunnen juist in deze aspecten effecten optreden die de marktwerking verstoren zoals het (al dan niet bedoeld) extra gewicht toekennen aan de prijsverschillen. Voorbeelden van technisch falen zijn het hanteren niet-onderscheidende kwalitatieve criteria, waardoor de verschillen in prijs bepalend zijn, en het gebruik van relatieve scores. Deze technische aspecten komen zowel naar voren in de keuze van te hanteren criteria als in de manier waarop wordt beoordeeld.

Onze doelstelling is het stimuleren van eerlijke concurrentie met goede voorwaarden voor de mensen op de werkvloer en goede resultaten voor de opdrachtgever. Hierbij past het gunnen van opdrachten aan de dienstverlener met de beste prijs-kwaliteitverhouding, zonder onbedoelde nadruk op prijs. Met 'kwaliteit' wordt dan zowel de kwaliteit van de dienstverlening bedoeld, als de kwaliteit van het werkgeverschap voor de werknemers.

In deze handreiking lichten wij toe hoe dat kan en geven we handvatten om te voorkomen dat prijs zwaarder meeweegt dan beoogd. Deze handreiking is geschikt voor gebruik in de publieke en de private sector. Voor lezers uit de publieke sector gaan we er vanuit dat kennis van de Aanbestedingswet en de Gids Proportionaliteit aanwezig is. De primaire doelgroep wordt gevormd door de inkopers van de opdrachtgevers, omdat zij de aanbestedingen vormgeven. Maar ook vertegenwoordigers van opdrachtnemers nodigen wij van harte uit deze handreiking te gebruiken om over dit belangrijke onderwerp op een positief kritische manier in gesprek te gaan met hun (potentiële) opdrachtgevers.

Er zijn ook handreikingen beschikbaar over het formuleren en hanteren van onderscheidende criteria voor kwaliteit. Voor elk van onze sectoren is er een eigen handreiking.

1.2 Verantwoordelijk marktgedrag reikt tot in de techniek van het beoordelen

Om een aanbieding te beoordelen op een individueel criterium, is een beoordelingsmethodiek nodig. Dit is de basis voor elke beoordeling. Het is de taak van inkopers om met de stakeholders vast te stellen op welke gunningscriteria aanbiedingen moeten worden beoordeeld. Daarbij hoort het vaststellen van de verdeling van punten voor elk van die criteria. Wanneer krijgt een aanbieding een hoge beoordeling, wanneer een lage en wanneer iets daar tussenin? Dit kan gevisualiseerd worden in grafieken: hoe verloopt de puntenscore van 'uitstekend' naar 'slecht'? Het is vervolgens aan de inkoper om de methodiek uit te werken: met welke formule bereken je de feitelijke score, gegeven de gewenste scoregrafiek?

Om te beginnen leggen we in hoofdstuk 2 uit wat een beoordelingsmethodiek is en wat belangrijke aandachtspunten zijn om zo'n methodiek goed te laten werken. In hoofdstuk 3 gaan we in op veelvoorkomende valkuilen bij het opzetten en uitvoeren van beoordelingen en de averechtse effecten daarvan. Natuurlijk leggen we ook uit hoe de valkuilen zijn te voorkomen.

Hoofdstuk 4 gaat over relatief beoordelen. Deze manier van beoordelen, die erg veel wordt gebruikt, kent een aantal risico's en is onder bepaalde omstandigheden zelfs onrechtmatig. Die belichten we en we geven een veilig alternatief. Wij adviseren sterk om altijd een absolute beoordelingsmethodiek te gebruiken. In hoofdstuk 5 leggen we er een aantal uit.

Hoofdstuk 6 gaat over gunningsmethodieken. De gunningsmethodiek is de manier waarop de scores die een aanbieding heeft gekregen op de individuele beoordelingscriteria, worden vertaald naar de totaalscore van die aanbieding. In deze handreiking gaan we steeds uit van de zogenaamde gewogen-factormethode. Daarbij wordt per beoordelingscriterium een weegfactor vastgesteld. Voor ieder criterium wordt de score bepaald door de gekozen beoordelingsmethodiek toe te passen en het resultaat te vermenigvuldigen met de betreffende weegfactor, dit resulteert in de gewogen score. De gewogen scores worden bij elkaar opgeteld tot de totaalscore van de aanbieding. Naast de gewogen-factormethode zijn er nog tal van gunningsmethodieken. In hoofdstuk 6 lichten we twee alternatieven toe die interessant kunnen zijn.

2. Het effect van de gekozen beoordelingsmethodiek op de uitkomst

Aanbiedingen worden op elk van de gehanteerde gunningscriteria beoordeeld met een score. De manier waarop de score voor een specifiek criterium wordt bepaald, noemen we de beoordelingsmethodiek.

2.1 Een goede opdrachtomschrijving is de basis om passende aanbieders te selecteren

Aanbiedingen worden op elk van de gehanteerde gunningscriteria beoordeeld met een score. De manier waarop de score voor een specifiek criterium wordt bepaald, noemen we de beoordelingsmethodiek. Een beoordelingsmethodiek bestaat uit een scorerange, een scoregrafiek en een formule. De scorerange is het bereik van de laagst haalbare score (vaak 0) tot de hoogst haalbare score (vaak 10 of 100). De formule geeft aan hoe de aangeboden prestatie moet worden omgezet naar een score. De scoregrafiek laat zien welke score bij welke prestatie hoort. Door de weegfactor wordt bepaald hoe zwaar de behaalde score meetelt in de totaalscore. Door de weegfactor te vermenigvuldigen met de behaalde puntenscore, ontstaat de gewogen score. De weegfactor kan ook in de formule zijn verwerkt.

Bijvoorbeeld: het beoordelingscriterium is responstijd. Voor een responstijd van 0 minuten wordt de maximale score gegeven van 100 punten. Voor een responstijd van 100 minuten of langer wordt 0 punten gegeven.

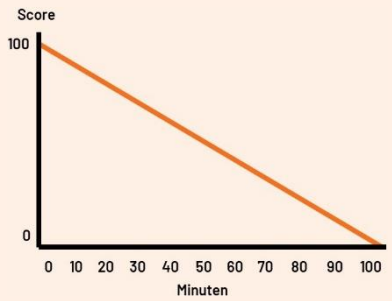
> De scorerange: 0 – 100 verdeeld over 0 tot 100 minuten.

> De bijbehorende formule:
Score = 100 – (1*aanbieding in minuten).

> De scoregrafiek is weergegeven in figuur 2.1. De weegfactor voor dit criterium is 70%, dus de hoogst haalbare gewogen score is 70% * 100 = 70 punten.

Figuur 2.1

Scoregrafiek bij een scorerange van 100 punten over een bereik van 100 minuten.



De toegepaste beoordelingsmethodieken bepalen de uitkomst van de beoordeling. In dit hoofdstuk leggen we eerst het effect van de gekozen scorerange uit en daarna het effect van weegfactoren.

2.2 Het belang van een goedgekozen scorerange

We leggen het effect van een scorerange uit aan de hand van voorbeeld 1. Een valkuil die we vaak zien, is het gebruik van een te brede scorerange..

Voorbeeld 1:

Er is een aanbesteding gedaan voor alarmopvolging. Er zijn twee criteria: prijs en kwaliteit (responstijd) met als weegfactoren respectievelijk 30% en 70%. Voor zowel prijs als kwaliteit kunnen maximaal 100 punten worden gescoord. De maximale gewogen scores voor prijs en kwaliteit zijn dus respectievelijk 30 en 70 punten.

In tabel 2.1 is te zien welke aanbiedingen zijn ontvangen.

Tabel 2.1

Ontvangen aanbiedingen voorbeeld 1.

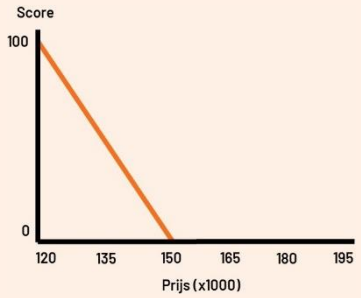
	Prijs	Responstijd
Beveiligings-bedrijf A	€ 145.000	15 minuten
Beveiligings-bedrijf B	€ 130.000	24 minuten

De beoordelingsmethodiek voor kwaliteit gaat uit van een scorerange van 100 punten bij een responstijd van 0 minuten en 0 punten bij een responstijd van 100 minuten of langer. De bijbehorende scoregrafiek is weergegeven in figuur 2.1.

Bij de beoordelingsmethodiek voor prijs wordt er een scorerange genomen van 100 punten bij een prijs van € 120.000 en 0 punten bij een prijs van € 150.000 of hoger, zie de grafiek in figuur 2.2.

Figuur 2.2

Scoregrafiek bij een scorerange van 100 punten over een bereik van € 30.000.

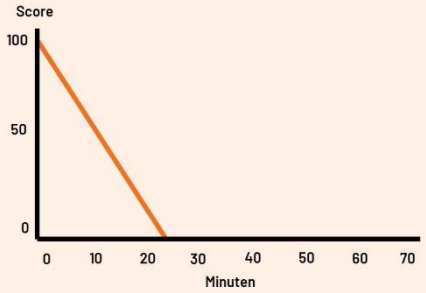


Deze beoordelingsmethodieken leiden tot onderstaande scores (zie tabel 2.2) en dienstverlener B krijgt de opdracht.

Stel nu dat voor kwaliteit nog steeds maximaal 100 punten kunnen worden gehaald, maar dat de scorerange voor responstijd wordt veranderd. Deze wordt 100 punten bij 0 minuten en 0 punten bij 25 minuten of langer. De formule wordt dan 100 – (4 * aantal minuten), zie de grafiek in figuur 2.3.

Figuur 2.3

Scoregrafiek bij een scorerange van 100 punten over een bereik van 25 minuten.



Dat leidt tot de scores in tabel 2.3, dienstverlener A krijgt nu de opdracht.

Dit voorbeeld laat zien dat de gekozen scorerange met het bijbehorende bereik, weergegeven in de scoregrafieken, veel effect heeft. Bij de vlakkere scoregrafiek in figuur 2.1 is het puntenverschil op basis van kwaliteit tussen dezelfde aanbiedingen kleiner dan bij de steilere scoregrafiek in figuur 2.3. Voor een onderscheidende beoordeling is het dus belangrijk te kiezen voor een voldoende steile scoregrafiek.

Tabel 2.2

Aanbiedingen en beoordelingsresultaat voorbeeld 1.

	Prijs	Responstijd	Gewogen score Prijs	Gewogen score Kwaliteit	Totaal-score	Ranking
Beveiligingsbedrijf A	€ 145.000	15 minuten	5	60	65	2
Beveiligingsbedrijf B	€ 130.000	24 minuten	20	53	**73**	**1**

Tabel 2.3

Aanbiedingen en alternatief beoordelingsresultaat voorbeeld 1.

	Prijs	Responstijd	Gewogen score Prijs	Gewogen score Kwaliteit	Totaal-score	Ranking
Beveiligingsbedrijf A	€ 145.000	15 minuten	5	28	32	1
Beveiligingsbedrijf B	€ 130.000	24 minuten	20	3	23	2

2.3 Doen de weegfactoren hun werk?

Weegfactoren moeten ervoor zorgen dat prijs en kwaliteit ieder in een bepaalde mate bepalend zijn in de totaalscore van een aanbieding. Stel dat voor zowel prijs als kwaliteit 100 punten kunnen worden gescoord, dan geef je met weegfactoren het belang aan dat je aan prijs en kwaliteit hecht. Bijvoorbeeld een verhouding van 30 punten : 70 punten bereik je dan met weegfactoren 30% en 70%. Als echter voor kwaliteit een beoordelingsmethodiek wordt gebruikt waarbij slechts een beperkt deel van de scorerange wordt benut, dan zal geen enkele aanbieding de hoogste mogelijke score van 100 punten krijgen. Na vermenigvuldiging met de weegfactor zal het belang van kwaliteit in de totaalscore dan ook geen 70% zijn. Stel dat voor prijs wél met een beoordelingsmethodiek wordt gewerkt waarbij de hoogst mogelijke puntenscore kan worden gehaald, dan zal het aandeel van het criterium prijs in de totaalscore wel 30 punten kunnen zijn. Dit betekent dat de feitelijke weegfactor van prijs veel hoger wordt dan de beoogde 30% en de score op kwaliteit dus veel minder gewicht in de schaal legt.

In de praktijk zien we veel relatieve beoordelingsmethodieken waarbij de laagste prijsaanbieding het maximumaantal punten krijgt (zie hoofdstuk 4) én komen we vaak te vlakke scoregrafieken voor kwaliteit tegen. Het komt dus vaak voor dat de werking van de gekozen weegfactoren teniet wordt gedaan.

3. Veelvoorkomende valkuilen die leiden tot onbedoeld grote nadruk op prijs

In hoofdstuk 2 is uitgelegd wat het effect kan zijn van de gekozen beoordelingsmethodiek en de scoregrafiek als onderdeel daarvan.

In dit hoofdstuk leggen we uit hoe dit kan leiden tot een te grote onderscheidende werking van de beoordeling op prijs respectievelijk een onvoldoende onderscheidende werking van de beoordeling op kwaliteit, en hoe dit te voorkomen.

3.1 Het onderscheidend effect van de beoordeling op kwaliteit is te klein

Een van de veelvoorkomende valkuilen is dat het effect waarmee een aanbieder zich kan onderscheiden in de beoordeling op kwaliteit te klein is.

3.1.1 De gehanteerde scorerange is te breed

Om onderscheidende scores te krijgen, is het belangrijk om een voldoende steile scoregrafiek te kiezen, zoals uitgelegd in 2.1. Voor kwaliteit wordt in de praktijk vaak een vlakkere score grafiek gebruikt dan voor prijs, waardoor de beoordeling op kwaliteit een minder onderscheidende werking heeft dan bedoeld. Dit leidt tot onbedoelde nadruk op prijs.

Om een passende scoregrafiek te bepalen, is het noodzakelijk marktkennis te hebben. Welke aanbiedingen mogen worden verwacht en welke aanbieding geldt dan voor de inkopende organisatie als onacceptabel, goed of zelfs uitstekend? Het kan dus nodig zijn om marktonderzoek te doen voordat de beoordelingsmethodiek kan worden vastgesteld. Dan kun je een reële scoregrafiek bepalen die past bij je behoefte.

3.1.2 Er wordt op een onhandige manier gewerkt met een minimumscore voor kwaliteit

Opdrachtgevers willen graag zeker weten dat aanbieders in ieder geval een bepaald niveau van kwaliteit garanderen.

Ze beoordelen dan alle aanbiedingen op kwaliteit en laten vervolgens die aanbiedingen afvallen, waarvan de score op kwaliteit onder een vastgesteld minimum ligt. Het komt geregeld voor dat dit op een onhandige manier wordt gedaan. We illustreren dit met voorbeeld 2.

Voorbeeld 2

Er is een schoonmaakaanbesteding gehouden. In de beoordeling van de aanbiedingen is de prijs-/kwaliteitverhouding 33% prijs en 67% kwaliteit. Kwaliteit wordt beoordeeld op een aantal gunningscriteria die worden beoordeeld met een cijfer tussen 0 en 10. Dit leidt tot een totaalcijfer voor kwaliteit, ook tussen 0 en 10 punten. Het behaalde totaalcijfer wordt vervolgens omgerekend naar een score tussen 0 en 100 punten. Na toepassing van de weegfactor geeft dit een score tussen 0 en 66 punten, zie de oranje lijn in de grafiek in figuur 3.1. Er wordt een minimumbeoordeling van de kwaliteit van de aanbieding geëist: tenminste een 7. De inkopende organisatie wil er zo voor zorgen dat alleen een aanbieding die minimaal aan dit kwaliteitsniveau voldoet, kan winnen. Dit betekent feitelijk dat het onderscheid tussen de niet-afgevallen aanbiedingen maximaal 20 punten kan zijn (van 46 punten tot 66 punten, zie grafiek 3.1). Voor het criterium prijs wordt wel de gehele scorerange van 33 punten gebruikt. Dit betekent dat een onbedoelde nadruk op prijs is ontstaan. De weegfactoren zijn niet 33% en 66%, maar (omgerekend) 62% en 38% voor prijs respectievelijk kwaliteit.

Hoe had dit beter gekund? Door wel de volledige gewogen scorerange van 66 punten te benutten, zoals weergegeven met de blauwe lijn in grafiek 3.1. Deze lijn begint op 0 punten bij de minimumkwaliteit van 7, in plaats van op 46 punten. Op deze manier wordt voor kwaliteit wel de beoogde scorerange van 66 punten gebruikt en wordt de beoogde prijs/kwaliteitverhouding wel gebruikt.

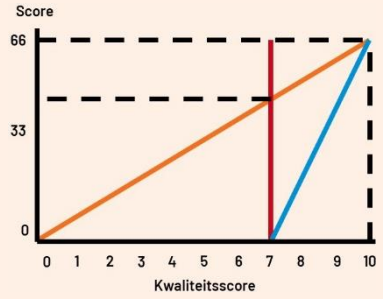
3.1.3 De gehanteerde prestatiecategorieën zijn te algemeen

De voorbeelden die we tot nu toe hebben besproken, gaan over criteria die kwantitatief kunnen worden gemaakt. De aangeboden prestaties worden uitgedrukt in een hoeveelheid, doorlooptijd, percentage et cetera.

Dit is bij kwaliteitscriteria niet altijd mogelijk. Een plan van aanpak, bijvoorbeeld, laat zich lastiger beoordelen op een puntenschaal. Opdrachtgevers gebruiken dan vaak waarderingen in de vorm van een prestatiecategorie. Doorgaans zijn dat waarderingen die lopen van 'onvoldoende' tot 'uitstekend'. Er wordt een beoordelingscomité ingesteld en de individuele beoordelaars bepalen voor elk van de betreffende criteria in welke prestatiecategorie een aanbieding vallen. Om tot een score te komen, worden aan deze categorieën scores gekoppeld, zoals rapportcijfers. Deze categorieën zijn vaak relatief breed, waardoor de kans groot is dat veel van de aanbiedingen in dezelfde categorie vallen, terwijl er wel degelijk verschil in kwaliteit is.

Figuur 3.1

Vlakke en steile scoregrafiek bij een vereiste minimumkwaliteit.



Dit kan op twee manieren worden voorkomen. De eerste manier is om een ruim aantal prestatiecategorieën te definiëren. De beoordelaars hebben dan meer mogelijkheden om hun beoordeling tot uitdrukking te brengen dan wanneer er bijvoorbeeld vier of vijf categorieën zijn. De tweede manier is om de beoordelaars de ruimte te geven voor het aanbrengen van nuances. Stel er wordt gebruik gemaakt van zes prestatiecategorieën met als bijbehorende scores 0-2-4-6-8-10. Als een beoordelaar vindt dat een aanbieding te goed is om in de categorie met score 8 te vallen, maar niet goed genoeg voor de categorie met score 10, dan kan hij kiezen voor een 9.

Tabel 3.1

Individuele beoordelingen en eindbeoordeling op basis van consensus.

	Cateringbedrijf A	Cateringbedrijf B	Cateringbedrijf C
Lid 1	6	8	8
Lid 2	8	8	10
Lid 3	8	8	8
Eindbeoordeling	8	8	8

Tabel 3.2

Individuele beoordelingen en eindbeoordeling op basis van middelen.

	Cateringbedrijf A	Cateringbedrijf B	Cateringbedrijf C
Lid 1	6	8	8
Lid 2	8	8	10
Lid 3	8	8	8
Eindbeoordeling	7,4	8	8,7

3.1.4 Van individuele beoordeling naar eindbeoordeling in een beoordelingscomité

Kwalitatieve kwaliteitscriteria worden doorgaans beoordeeld door een beoordelingscomité. Een dergelijk comité bestaat uit meerdere leden. De leden beoordelen individueel de ontvangen aanbiedingen en noteren hun beoordeling. Vervolgens gaan de leden in gesprek om van de individuele beoordelingen tot een totaalbeoordeling te komen. Hiervoor zijn twee veelgebruikte manieren. De eerste is via consensus: de leden komen samen en overleggen tot ze het eens zijn over de beoordeling. De tweede is door het rekenkundig gemiddelde van de individuele scores te bepalen (middelen). We laten met voorbeeld 3 zien waarom middelen de voorkeur heeft.

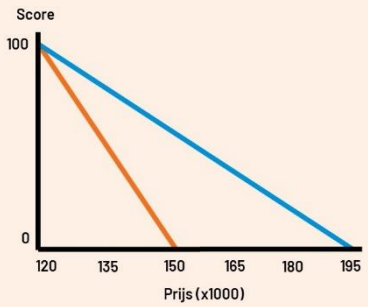
Voorbeeld 3

Er zijn drie aanbiedingen ontvangen in een aanbesteding voor cateringdiensten. Een van de gunningscriteria is de beoordeling van een plan van aanpak. Het beoordelingscomité bestaat uit drie leden die elk de ontvangen plannen van aanpak beoordelen. De scoremogelijkheden zijn zes prestatiecategorieën van 'onvoldoende' tot en met 'uitstekend', die worden vertaald naar puntenscores van 0, 2, 4, 6, 8 respectievelijk 10 punten. De individuele beoordelingen en het eindoordeel zijn weergegeven in tabel 3.1.

Lid 1 plaatst de kwaliteit van cateringbedrijf A in de categorie waar puntenscore 6 bij hoort, leden 2 en 3 komen beiden uit op de categorie met puntenscore 8. In het consensusoverleg komen zij uit op puntenscore 8. Hetzelfde gebeurt voor cateringbedrijf C. Dit leidt ertoe dat alle drie de aanbieders dezelfde eindbeoordeling krijgen, terwijl er wel verschil zit in de kwaliteit.

Figuur 3.2

Kies een passende scoregrafiek voor prijs.



Dit voorkom je door de puntenscores te middelen. Na de individuele beoordeling overlegt het comité, waarna leden de kans hebben om hun individuele beoordeling eventueel nog aan te passen. Vervolgens wordt het gemiddelde berekend van de behaalde puntenscores. Deze manier van werken leidt in ons voorbeeld tot andere eindbeoordelingen, zie tabel 3.2. Nu komt het verschil in kwaliteit wél naar voren.

3.1.5 Last but not least: de beoordelingscriteria zelf zijn niet onderscheidend

Om dienstverleners de kans te geven een aanbieding te maken waarin zij zich goed van hun concurrenten kunnen onderscheiden, is het belangrijk om beoordelingscriteria te hanteren die dat mogelijk maken. Dit klinkt logisch, in de praktijk blijkt het echter vaak lastig. Bij het uitbesteden van een verhuizing wil je niet dat inventaris beschadigd raakt, bij schoonmaakwerkzaamheden mag de bedrijfsvoering niet worden verstoord, in het bedrijfsrestaurant moet voldoende gezond assortiment worden aangeboden en als er een alarm afgaat, wil je dat dit snel wordt opgevolgd. Dit zijn echter niet per se criteria waarop dienstverleners zich onderscheiden, dit kunnen zij allemaal heel goed. Als je op dergelijke criteria gaat scoren, krijgt elke aanbieding ongeveer dezelfde waardering.

Het is dus zaak om tot criteria te formuleren die je als opdrachtgever belangrijk vindt en waarop de aanbiedingen wél zullen gaan verschillen.

Hiervoor hebben wij vier aparte handreikingen beschikbaar: voor verhuis-, schoonmaak-, catering- en beveiligingsdiensten.

3.2 Het onderscheidend effect van de beoordeling op prijs is te groot

Een andere veelvoorkomende valkuil is dat het effect waarmee een aanbieder zich kan onderscheiden in de beoordeling op prijs te groot is.

3.2.1 Verkeerd gekozen scorerange leidt tot te grote scoreverschillen

In 3.1.2 hebben we uitgelegd dat voor kwaliteit vaak te vlakke scoregrafieken worden gebruikt. In tegenstelling hiermee zien we voor prijs vaak (te) steile scoregrafieken. In figuur 3.2 geven we twee voorbeelden van scoregrafieken voor prijsaanbiedingen.

In het eerste voorbeeld, weergegeven met de blauwe lijn, wordt een scorerange van 33 punten gekoppeld aan prijsaanbiedingen van € 120.000 tot € 195.000. In het tweede voorbeeld, weergegeven met de oranje lijn, wordt dezelfde scorerange van 33 punten gekoppeld aan prijsaanbiedingen van € 120.000 tot € 150.000. Bij de scoregrafiek uit het tweede voorbeeld zal een aanbieding die dichterbij € 120.000 ligt, minder scorepunten krijgen dan in het eerste voorbeeld. Andersom betekent het ook dat een prijsaanbieding die richting € 150.000 gaat, in het tweede voorbeeld een veel lagere puntenscore krijgt dan in het eerste voorbeeld.

Om een goede beoordelingsmethodiek op te stellen voor de prijsaanbieding, is het ten eerste van belang om dit effect goed te begrijpen. Vervolgens is het essentieel om te bepalen wat reële verwachtingen zijn voor wat betreft de te ontvangen prijsaanbiedingen. Liggen die, in geval van ons voorbeeld, tussen € 120.000 en € 150.000 of tussen € 120.000 en € 195.000? En wat kan beschouwd worden als een goede prijs bij de gevraagde kwaliteit? Is dat een prijs die in het midden van de verwachte range aan aanbiedingen ligt, of is die eerder lager of hoger dan dat midden? Met deze marktkennis kan een passende, reële scoregrafiek worden vastgesteld. Mogelijk is het nodig om marktonderzoek te doen om deze kennis te verzamelen en verwachtingen te toetsen.

3.3 Het effect van een beoordelingsformule kan ook leiden tot onbedoelde nadruk op prijs

In 2.2 hebben we uitgelegd dat het belangrijk is een passende scorerange te kiezen in combinatie met het verwachte bereik van de aanbiedingen. Er is nog een ander aspect van belang. Bij dezelfde scorerange, die hetzelfde bereik van de aanbiedingen afdekt, kunnen er verschillende uitkomsten in scores zijn. De uitkomst van de beoordeling hangt af van de beoordelingsformule die wordt gebruikt. In voorbeeld 4 laten we zien wat het effect is van een andere beoordelingsformule bij dezelfde scorerange.

Tabel 3.4

Ontvangen aanbiedingen voorbeeld 4.

	Prijs	Score kwaliteit
Verhuisbedrijf A	€ 4.000	60
Verhuisbedrijf B	€ 4.500	70
Verhuisbedrijf C	€ 4.600	71,5

Voorbeeld 4

Er zijn drie aanbiedingen ontvangen in een aanbesteding voor verhuisdiensten. De weegfactor voor zowel prijs als kwaliteit is 50%, waarbij er 100 punten voor prijs en 100 punten voor kwaliteit te behalen zijn. Na toepassing van de weegfactoren is de maximum score voor zowel prijs als kwaliteit dus 50 punten. In dit voorbeeld werken we twee beoordelingen uit. De beoordelingsmethodiek voor kwaliteit blijft steeds gelijk.

Voor prijs worden twee verschillende formules gebruikt om de score op prijs te bepalen, elke met dezelfde scorerange:

- De maximale score van 100 punten wordt gegeven bij een prijsaanbieding van € 3.800.
- De minimale score van 0 punten wordt gegeven bij een prijsaanbieding van € 11.400. Bij een prijsaanbieding van € 7.600 hoort een score van 50 punten.
- Beoordelingsformule methodiek 1:
  Score = 150 - 50 * (prijs / 3.800)
- Beoordelingsformule methodiek 2:
  Score = 100 * (3.800 / prijs)

Tabel 3.4 geeft de aanbiedingen weer. Tabellen 3.5 en 3.6 geven de scores voor beoordelingsmethodiek 1 respectievelijk 2. Ze leiden elk tot een andere winnaar.

Tabel 3.5

Aanbiedingen en beoordelingsresultaat methodiek 1.

	Prijs	Gewogen score Prijs	Gewogen score Kwaliteit	Totaal-score	Ranking
Verhuisbedrijf A	€ 4.000	48,7	30,0	78,7	3
Verhuisbedrijf B	€ 4.500	45,4	35,0	80,4	2
Verhuisbedrijf C	€ 4.600	44,8	35,8	**80,6**	**1**

Tabel 3.6

Aanbiedingen en beoordelingsresultaat methodiek 2.

	Prijs	Gewogen score Prijs	Gewogen score Kwaliteit	Totaal-score	Ranking
Verhuisbedrijf A	€ 4.000	47,5	30,0	**77,5**	1
Verhuisbedrijf B	€ 4.500	42,2	35,0	77,2	2
Verhuisbedrijf C	€ 4.600	41,3	35,8	77,1	3

Dit voorbeeld laat zien dat het delen van alleen de weegfactoren en scoreranges met het bijbehorende bereik de aanbieders nog onvoldoende informeert. Wij raden nadrukkelijk aan om de hele beoordelingsmethodiek, inclusief de formule, expliciet in de offerteaanvraag op te nemen. Dit leidt tot een meer passende aanbieding, doordat aanbieders niet meer hoeven te gissen naar de voorkeur van de inkopende partij. De beoordelingsmethodiek kan het beste worden weergegeven met de scorerange, de beoordelingsformule, de scoregrafiek en de weegfactor. De aanbieder kan met deze informatie bepalen wat de beste aanbieding is die hij kan doen.

4. De valkuilen van relatief beoordelen

Als je geen goed beeld hebt van de te verwachten prijsaanbiedingen, is het lastig om een goede scorerange te kiezen. Je weet immers niet tussen welke waarden de aanbiedingen waarschijnlijk liggen en hoe je de scores daarover zou willen verdelen. Relatief beoordelen biedt dan uitkomst.

Bij een relatieve beoordelingsmethodiek is de score van een aanbieding afhankelijk van de andere aanbiedingen die zijn ontvangen, bijvoorbeeld voor prijs of responstijd. De maximale puntenscore wordt dan toegekend aan de beste ontvangen aanbieding, zoals de laagste prijs of de snelste responstijd. De puntenscores van de andere aanbiedingen worden gerelateerd aan die beste aanbieding.

Er zijn verschillende manieren om een relatieve beoordeling vorm te geven. Niet elke manier is bruikbaar bij het beoordelen van inschrijvingen. We onderscheiden binnen de relatieve beoordelingsmethoden twee groepen methoden: de methoden die niet tot rank reversal kunnen leiden en de methoden die wel tot rank reversal kunnen leiden. De laatste groep is niet bruikbaar voor beoordelingen van inschrijvingen. De Commissie van Aanbestedingsexperts heeft geconcludeerd dat relatief beoordelen waarbij rank reversal kan optreden in strijd is met de Aanbestedingswet[1] (zie 4.1).

In dit hoofdstuk leggen wij uit wat rank reversal is en geven een aantal voorbeelden van manieren van relatief beoordelen die ongeschikt zijn en op welke manier relatief beoordelen wel kan. Als er relatief wordt beoordeeld, is dat vrijwel altijd voor het gunningscriterium prijs. Daarom nemen we het in het vervolg van dit hoofdstuk steeds de prijsaanbieding als voorbeeld.

4.1 Relatief beoordelen met kans op rank reversal en bid rigging

Kans op rank reversal

Bij een relatieve beoordelingsmethodiek is de score van de ene aanbieding afhankelijk van de andere aanbiedingen. Het maximaal aantal punten wordt toegekend aan de aanbieding met de laagste prijs.

Als de aanbieding met de laagste prijs niet meer meedoet, bijvoorbeeld omdat deze ongeldig wordt bevonden, dan kan dit leiden tot een andere winnaar van het offertetraject: de ranking wijzigt. De ranking kan wijzigen doordat er dan gebruikt wordt gemaakt van een ander ijkpunt en hierdoor de scores ander uitvallen. Dit wordt rank reversal genoemd en kan optreden bij bepaalde relatieve beoordelingsmethodieken. We leggen het principe van rank reversal uit met voorbeeld 5 (zie tabel 4.1).

Relatief beoordelen met kans op rank reversal zoals geschetst in voorbeeld 5, is volgens de Commissie van Aanbestedingsexperts in strijd met de beginselen van transparantie en gelijke behandeling en leidt niet tot de keuze van de economisch meest voordelige inschrijving. Oftewel, relatief beoordelen met kans op rank reversal is in strijd met de Aanbestedingswet.

Kans op bid rigging

Een relatieve beoordelingsmethodiek met de kans op rank reversal kan ook leiden tot bid rigging, waarbij aanbieders die kans benutten voor hun eigen voordeel. In voorbeeld 5 (zie tabel 4.1) is het voor dienstverlener A wel erg gunstig als C een heel lage prijs aanbiedt. Dat heeft tot gevolg dat het verschil in scores voor de prijsaanbiedingen van concurrenten en die van dienstverlener A, kleiner wordt. Dit vergroot de kans dat dienstverlener A de opdracht wint.

De kans op rank reversal doet zich voor bij een aantal veelgebruikte relatieve beoordelingsmethodieken die we toelichten in de volgende paragraaf.

[1] Advies 504 d.d. 18-12-2020, zie commissievanaanbestedingsexperts.nl.

Tabel 4.1

Voorbeeld 5: relatief beoordelen met 3 aanbieders.

	Prijs	Score prijs	Score kwaliteit	Totaal-score	Ranking
Dienstverlener A	€ 500	40	95	**135**	**1**
Dienstverlener B	€ 400	50	80	130	2
Dienstverlener C	€ 200	100	20	120	3

Voorbeeld 5: Het effect als de aanbieding met de laagste prijs wegvalt.

	Prijs	Score Prijs	Score Kwaliteit	Totaal-score	Ranking
Dienstverlener A	€ 500	80	95	175	2
Dienstverlener B	€ 400	100	80	**180**	**1**

4.2 Veelgebruikte relatieve beoordelingsmethodieken die niet (meer) kunnen

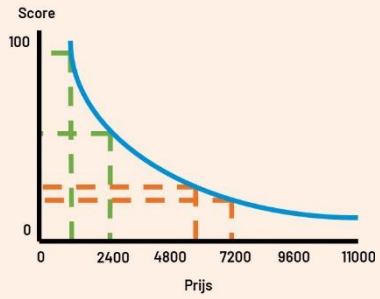
De beoordelingsmethodieken in deze paragraaf creëren kans op rank reversal. We lichten deze wel toe, omdat deze (en varianten erop) veel worden toegepast. Deze relatieve beoordelingsmethodieken kunnen bovendien tot een onbedoelde nadruk op prijs leiden. Dat vergroot de kans op een beoordelingsresultaat waarin prijs zwaarder blijkt te wegen dan bedoeld. Een voorbeeld hiervan is de vaak gebruikte formule:

Score =
maximum score x (beste prijs/aangeboden prijs)

Zie de grafiek in figuur 4.1 voor de bijpassende scoregrafiek (blauwe lijn).

Figuur 4.1

Relatieve beoordeling met kans op rank reversal.



De blauwe lijn loopt in het begin steil naar beneden en vlakt later af. Dit betekent dat een klein prijsverschil in de linkerkant van de grafiek een groot verschil in puntenscore oplevert en dat een even groot prijsverschil in de rechterkant van de grafiek weinig verschil in puntenscore geeft (zie de oranje en groene lijnen). Met deze methodiek kan er onbedoelde nadruk worden gelegd op een lage prijs, doordat de laagste prijs alle punten krijgt en een iets hogere prijs al veel minder punten. Het verschil in puntenscores tussen de laagste prijsaanbieding en de andere aanbiedingen wordt daardoor al snel groot. Dat heeft tot gevolg dat een prijsaanbieding die mogelijk best acceptabel is, toch een (zeer) lage puntenscore krijgt.

Een ander voorbeeld is:

Score =
maximum score x
(2 – (aangeboden prijs/beste prijs))

Bij deze formule wordt een score van 0 punten gegeven bij een prijs die twee keer zo hoog is als de beste aangeboden prijs, zie de grafiek in figuur 4.2. Bij deze beoordelingsmethodiek wordt er al minder onbedoeld nadruk gelegd op de prijs doordat hetzelfde verschil in prijs overal hetzelfde puntenverschil geeft. Er is echter nog steeds een risico op rank reversal (zie 4.1). Omdat je vooraf niet weet welke aanbiedingen worden ontvangen, weet je vooraf niet hoe steil de scoregrafiek blijkt uit te pakken.

4.3 Veilig relatief beoordelen: verschil in scores afhankelijk van een constante

Een relatieve beoordelingsmethodiek kan wel gewenst zijn, bijvoorbeeld omdat de inkopende organisatie zich geen goed beeld heeft kunnen vormen van de prijsaanbiedingen die kunnen worden verwacht. De volgende formule kan dan worden gebruikt:

Score =
maximale score –
(aangeboden prijs – beste prijs) / Q

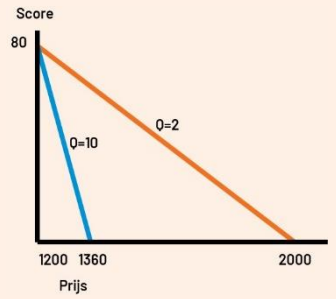
Hierbij krijgt de scoregrafiek de vorm van een rechte lijn. De steilheid van de lijn wordt bepaald door de constante Q. Deze Q is de hoeveelheid geld die de opdrachtgever over heeft voor een score op kwaliteit die 1 punt beter is. Hoe groter de Q, hoe vlakker de scoregrafiek en vice versa. De waarde van Q bepaalt dus hoeveel effect een verschil in prijs heeft op de puntenscore, zie grafiek 4.3. Bij deze manier van relatief beoordelen is de helling van de lijn niet afhankelijk van het laagste bod, maar van de vooraf vastgestelde Q. Bij deze beoordelingsmethodiek is er geen kans op rank reversal.

Figuur 4.2

Relatieve beoordeling met een rechte lijn.

Figuur 4.3

Scoregrafiek bij relatieve beoordeling met constante Q.



5. Een aantal absolute beoordelingsmethodieken

Bij een absolute beoordelingsmethodiek is de uitkomst van de beoordeling niet afhankelijk van andere aanbiedingen, maar zijn de scores voor een bepaalde prijs of kwaliteit van tevoren vastgesteld.

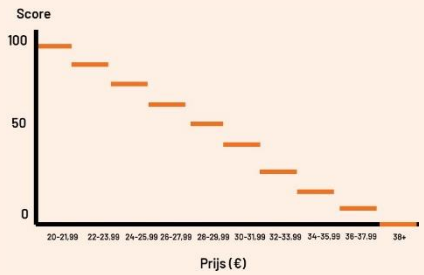
In dit hoofdstuk leggen we een aantal veelvoorkomende absolute beoordelingsmethodieken uit en geven we aan in hoeverre deze geschikt zijn. De beoordelingsmethodieken in dit hoofdstuk zijn die met specifieke scoregrafieken: scoregrafieken met horizontale lijnen, in de vorm van een parabool, met negatieve waarden, met een knik en scoregrafieken in de vorm van een hyperbool.

5.1 Beoordelingsmethodiek met een scoregrafiek met horizontale lijnen

Zoals in 3.1.3 aangegeven worden er voor sommige criteria prestatiecategorieën gebruikt. Bij die categorieën horen dan puntenscores. Aanbiedingen die in dezelfde categorie zijn beoordeeld, verschillen doorgaans van elkaar. Verschillende niveaus in kwaliteit of prijs krijgen dus dezelfde score. Dit geeft een scoregrafiek met horizontale lijnen, zoals in de grafiek in figuur 5.1. Deze scoregrafiek wordt in de praktijk vaak voor kwaliteit gebruikt en in mindere mate voor prijs.

Figuur 5.1

Scoregrafiek horizontale lijnen bij prestatiecategorieën.



Bij deze beoordelingsmethodiek is het belangrijk om in gedachten te houden dat aanbieders altijd in het uiterste gaan zitten van de horizontale lijn die het best past bij wat ze willen aanbieden. In het voorbeeld van de grafiek in figuur 5.1 zijn categorieën van uurtarieven weergegeven.

De uurtarieven lopen van €20 tot €38 en hoger. Je kunt, zoals in dit voorbeeld, voor tien categorieën kiezen, maar bijvoorbeeld ook voor twee of twintig. De categorieën uit het voorbeeld zijn alleen acceptabel als je er daadwerkelijk mee instem dat er (bijvoorbeeld) €31,99 moet worden betaald in plaats van €30 of €31,50. Aanbieders zullen immers met hun prijs helemaal rechts op een horizontale lijn gaan zitten, omdat dit evenveel punten oplevert als een prijs helemaal links.

Voor kwaliteitscriteria komt deze beoordelingsgrafiek vaak voor, in de vorm van prestatiecategorieën zoals 'onvoldoende', 'goed' en 'uitstekend'. Dan speelt het nadeel van bewust aanbieden in het uiterste van een categorie minder, omdat aanbieders vooraf niet zelf kunnen vaststellen in welke prestatiecategorie hun aanbieding zal worden beoordeeld.

5.2 Beoordelingsmethodiek met een scoregrafiek die als parabool lopen

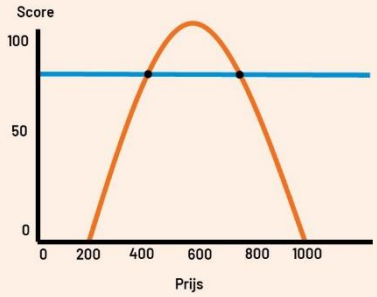
Beoordelingsmethodieken kunnen een scoregrafiek hebben die als een parabool loopt, zoals in de grafiek in figuur 5.2 Hiermee voorkom je zowel te hoge als te lage prijsaanbiedingen.

De grafiek in figuur 5.2 laat zien dat verschillende prijzen dezelfde score kunnen krijgen, zie de twee zwarte punten op de oranje lijn. Dat is op zichzelf vreemd, omdat je ervan uit zou moeten kunnen gaan dat een lagere prijs tot een hogere score leidt. Het betekent ook dat als de betreffende aanbiedingen gelijk scoren op kwaliteit, de totaalscores voor die aanbiedingen ook gelijk zijn. Dat is onwenselijk.

Ook al zijn de scores voor kwaliteit hetzelfde, er is slechts één aanbieding met een lagere prijs en dat is de economisch meest voordelige inschrijving. Binnen de publieke inkoop is een scoregrafiek zoals deze daarom onrechtmatig.

Figuur 5.2

Scoregrafiek parabool met twee dezelfde scores voor verschillende prijzen.



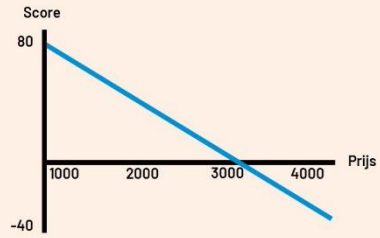
De scoregrafiek in de vorm van een parabool wordt weleens gebruikt om te voorkomen dat er te lage prijzen worden geboden. Het is voor dat doel echter beter om een minimum te definiëren voor de prijsaanbieding. Aanbiedingen die daaronder liggen, worden uitgesloten. Ook een scoregrafiek met negatieve waarden is dan een optie, zie 5.3.

5.3 Beoordelingsmethodiek met een scoregrafiek met negatieve waarden

Ook negatieve scores behoren tot de mogelijkheden als een prijsaanbieding een bepaalde grens overgaat. De grafiek in figuur 5.3 geeft een voorbeeld.

Figuur 5.3

Scoregrafiek met positieve en negatieve scores.



In dit voorbeeld krijgen prijsaanbiedingen hoger dan €3.000 een negatieve score. Als de beoordelingsmethodiek met aanbieders is gedeeld, maakt dit de kans klein op een aanbieding boven deze prijs. Een aanbieding met een negatieve score op prijs maar een (zeer) hoge score op kwaliteit zou nog wel kunnen winnen. Dat is afhankelijk van de gehanteerde weegfactoren.

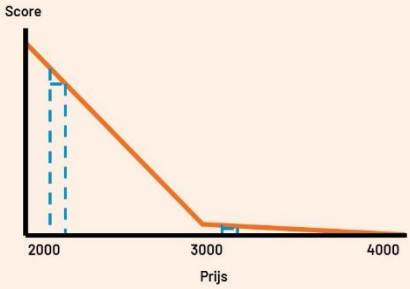
5.4 Beoordelingsmethodiek met een scoregrafiek met duidelijke knik

Een andere scoregrafiek is die met een knik, zie de grafiek in figuur 5.4.

Hierin is te zien dat het effect van een verschil in aangeboden prijzen op de beoordeling niet constant is. Met de scoregrafiek uit het voorbeeld zal een prijsaanbieding van bijna €4.000 slechts een iets lagere score krijgen dan een prijsaanbieding van €3.000. Dit werkt in de hand dat aanbieders een relatief hoge prijsaanbieding doen. Doordat ze hiervoor ook een relatief hoge kwaliteit kunnen leveren, creëert dit het risico dat een (te) dure aanbieding wint.

Figuur 5.4

Scoregrafiek met knik waarbij effect op verschil in punten voor de knik groter is dan na de knik.



Als het doel is om hoge prijsaanbiedingen te weren, kan beter een scoregrafiek zoals in 5.3 worden gebruikt, of een scoregrafiek waarbij de lijn eerst zonder knik doorloopt tot aan de x-as (score 0 punten) en daarna met de x-as samenvalt.

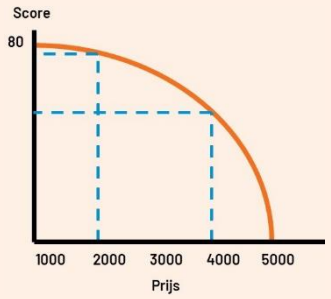
5.5 Beoordelingsmethodiek met een scoregrafiek met duidelijke knik

Als voor een acceptabele prijs een acceptabele kwaliteit wordt gezocht, raden wij de omgekeerde hyperbool aan als methodiek.

Een voorbeeld van een grafiek van de omgekeerde hyperbool is te zien in de grafiek in figuur 5.5.

Figuur 5.5

Scoregrafiek met omgekeerde hyperbool.



Een reden om voor een omgekeerde hyperbool als scoregrafiek te kiezen, is het uitgangspunt dat de aanbiedingen voor prijs en voor kwaliteit verband houden met elkaar. De aanname daarbij is dat een hogere kwaliteit zou moeten leiden tot een hogere prijs en vice versa. Zoals te zien in de grafiek in figuur 5.5, heeft het verhogen van de prijs van een laaggeprijsde aanbieding weinig effect op de prijsscore. Ook al wordt de aanbieding duurder omdat er een betere kwaliteit wordt geboden, de score op prijs wordt niet veel lager. Dit stimuleert dienstverleners om bij laaggeprijsde aanbiedingen zoveel mogelijk kwaliteit te bieden: de score op kwaliteit gaat omhoog, de score op prijs slechts weinig naar beneden.

Aan de andere kant is het zo dat het verlagen van een prijsaanbieding die aan de hoge kant is, juist wel een duidelijk effect heeft op de score van prijs. Dienstverleners die hogere prijzen bieden, worden gestimuleerd om de kwaliteit van hun aanbieding iets te verlagen waardoor ook hun prijs wat lager wordt. De minder goede score op kwaliteit wordt dan gecompenseerd door een betere score op prijs.

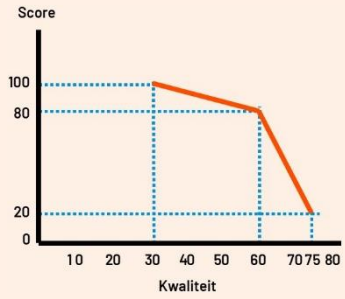
Een omgekeerde hyperbool voor prijs is vergelijkbaar met de zogenaamde 'Superformule' of de 'Argitek-formule'. Deze formules raden wij echter niet aan in verband met de wiskundige complexiteit.

5.6 Voorbeeldformules bij scoregrafieken met rechte lijn(en)

Als duidelijk is welke scoregrafiek gehanteerd gaat worden, kan het nodig zijn de bijbehorende formule vast te stellen. Dat is het geval als de aanbiedingen kwantitatief moeten worden gemaakt en als de score niet eenvoudig van de scoregrafiek zelf valt af te lezen. We geven twee voorbeeldformules voor grafiekdelen die bestaan uit een rechte lijn. Met deze formules kun je een scoregrafiek maken met één rechte dalende of stijgende lijn, of een grafiek die uit meerdere rechte lijnen bestaan, zoals die met een knik uit 5.6.

Figuur 5.6

Scoregrafiek opgebouwd uit stukken rechte lijn.



We gebruiken de grafiek in figuur 5.6 als voorbeeld.

Voor elk stuk rechte, dalende lijn geldt de volgende formule:

Score = max score – (max score – min score) / (minste kwaliteit – beste kwaliteit) x (aanbieding – beste kwaliteit)

Stel we beoordelen kwaliteit, waarbij een hogere waarde een lagere kwaliteit betekent. We berekenen de score voor een aanbieding met een kwaliteit van 40. Dit valt in het eerste stuk van de grafiek. Daarvoor geldt dat de 'max score' 100 punten is en de 'min score' 80 punten. De 'minste kwaliteit' is in dit stuk 60 en de 'beste kwaliteit' is 30. De berekening van de score ziet er dan als volgt uit:

Score = 100 – (100 – 80) / (60 – 30) x (40 – 30)

Score = 100 – 20 / 30 x 10

Score = 100 – 6,7 = 93,3

Op dezelfde manier geldt de berekening van de score voor een aanbieding met een kwaliteit van 70:

Score = 80 – (80 – 20) / (75 – 60) x (70 – 60)

Score = 80 – 60 / 15 x 10

Score = 80 – 4 x 10 = 40

De volgende formule geldt als de lijn van links naar rechts stijgt:

Score = min score + (max score – min score) / (beste kwaliteit – minste kwaliteit) x (aanbieding – minste kwaliteit)

Een combinatie van meerdere rechte lijnen kan worden gebruikt om bijvoorbeeld tot een scoregrafiek te komen die een omgekeerde hyperbool (zie 5.5) benadert. Dit is een optie als het ingewikkeld is om zelf de formule van een dergelijke scoregrafiek vast te stellen.

6. Alternatieve gunningsmethodieken

Naast de gewogen-factorscore, waar de voorbeelden in deze handreiking op zijn gebaseerd, zijn er ook andere gunningsmethodieken. We leggen twee veelgebruikte methodieken toe: 'gunnen op waarde' en 'value for money'.

6.1 Gunnen op waarde

Bij gunnen op waarde wordt er gegund aan de aanbieder die heeft ingeschreven met de laagste fictieve prijs. Deze fictieve prijs wordt vastgesteld door de totaalscore op de kwaliteitscriteria om te rekenen naar een bedrag in euro's. Dit bedrag wordt in mindering gebracht op de aangeboden prijs, het resultaat is de fictieve prijs. Hiervoor wordt de volgende formule gebruikt:

Fictieve prijs = aangeboden prijs –
(maximale waarde kwaliteitscriteria x score)

De maximale waarde van de kwaliteitscriteria is de delta value.

In tabel 6.1 laten we de werking zien met een voorbeeld. De delta value is € 100.000. De maximale score die kan worden behaald op de kwaliteitscriteria is 100 punten. Deze score wordt omgerekend naar een kwaliteitswaarde (in euro's) door de score te vermenigvuldigen met 1.000. Dienstverlener A heeft bijvoorbeeld 50 punten gescoord op kwaliteit, wat hem een kwaliteitswaarde oplevert van 50 x 1.000 = € 50.000. Deze waarde wordt afgetrokken van zijn inschrijfprijs van € 150.000. Daarmee wordt zijn fictieve prijs € 150.000 - € 50.000 = € 100.000.

Bij gunnen op waarde worden de prijsaanbiedingen dus niet beoordeeld: er hoeft geen scoregrafiek voor prijs bepaald te worden. Het is ook niet nodig om weegfactoren op te nemen voor prijs en kwaliteit. Er is wel sprake van een weging, maar die is impliciet meegenomen in de maximale financiële waarde die de kwaliteitscriteria vertegenwoordigen. Hoe hoger de delta value, des te meer gewicht de kwaliteitscriteria krijgen.

Het is bij gunnen op waarde wél nodig om beoordelingsmethodieken vast te stellen voor de kwaliteitscriteria. Hierbij gelden de valkuilen en aanwijzingen om die te voorkomen uit hoofdstuk 3. Om deze gunningsmethodiek goed toe te passen, is ook kennis nodig van prijsstelling. Je moet immers bepalen hoeveel kwaliteitswaarde je per scorepunt wilt toekennen. Met deze methodiek maak je expliciet hoeveel extra je bereid bent te betalen voor een hogere kwaliteit. Als de maximaal te scoren kwaliteitswaarde relatief klein is ten opzichte van de aangeboden prijzen, blijft het effect beperkt met het risico dat je feitelijk op prijs gunt. Andersom, als de maximale kwaliteitswaarde relatief hoog ligt, ontstaat het risico dat je meer betaalt dan je (bij nader inzien) had gewild.

Tabel 6.1

Voorbeeld gunningsmethodiek gunnen op waarde.

	Prijs	Score kwaliteit	Kwaliteits-waarde	Fictieve prijs	Ranking
Dienstverlener A	€ 150.000	50	€50.000	**€100.000**	1
Dienstverlener B	€ 200.000	90	€90.000	€110.000	2
Dienstverlener C	€ 120.000	5	€5.000	€115.000	3

Tabel 6.2

Voorbeeld gunningsmethodiek value for money.

	Prijs	Totaalscore kwaliteit	Value for money	Ranking
Dienstverlener A	€ 150.000	50	3000	2
Dienstverlener B	€ 200.000	90	2222	1
Dienstverlener C	€ 120.000	5	24000	3

6.2 Value for money

De gunningsmethodiek value for money maakt inzichtelijk hoeveel kwaliteitswaarde er per euro wordt aangeboden. Hiertoe worden de aanbiedingen beoordeeld op kwaliteit, net zoals bij de gunningsmethodieken gewogen-factormethode en gunnen op waarde. De scores op de individuele kwaliteitscriteria worden opgeteld tot een totaalscore op kwaliteit. Vervolgens wordt de aangeboden prijs gedeeld door deze totaalscore. Zie tabel 6.2 voor een voorbeeld.

Bij de gunningsmethodiek value for money is dus het niet nodig de prijsaanbieding te beoordelen. Bij deze methodiek is het wel zo dat een aanbieding die twee keer zo hoog scoort op kwaliteit als een andere aanbieding en ook bijna twee keer zo duur is als die andere aanbieding, toch kan winnen. Dit effect kan worden beperkt, door ervoor te zorgen dat de verschillen in kwaliteitsscores niet te ver uiteen kunnen gaan lopen. Bijvoorbeeld geen scorerange van 0 tot 100, maar van 60 tot 100.

6.3 Beoordeel eerst zelf met dummy's en deel de methodieken met aanbieders

In deze handreiking hebben we een aantal beoordelings- en gunningsmethodieken uitgelegd. Om het effect van de methodieken en weegfactoren goed duidelijk te krijgen, raden we sterk aan om zelf een aantal dummy-aanbiedingen te beoordelen. Dit geeft inzicht in het effect van de beoogde methodieken. Kloppen de scores met wat u voor ogen had? Brengen ze onderscheid aan tussen de aanbiedingen? En is de aanbieding die als beste uit de beoordeling komt inderdaad een aanbieding waar u gelukkig mee zou zijn? Voor deze dummy-beoordeling uit vóórdat de offerteaanvraag wordt verstuurd. De te beoordelen aanbiedingen kunt u zelf verzinnen, of misschien liggen er nog oude aanbiedingen uit een eerder offertetraject.

Neem voor elk beoordelingscriterium de beoordelingsmethodiek (formule, scorerange en scoregrafiek) en de weegfactor op in de offerteaanvraag. Dienstverleners zien dan wat uw voorkeuren zijn. Dit hebben ze nodig om hun best passende aanbieding te doen.

code verantwoordelijk marktgedrag

Het uitbesteden van diensten moet voor alle betrokkenen een positieve uitkomst hebben. De opdrachtgever krijgt waar voor zijn geld, de opdrachtnemer realiseert een gezond rendement en de medewerkers van opdrachtnemer worden gewaardeerd voor hun vakmanschap. Een dergelijke succesvolle uitbesteding is het resultaat van verantwoordelijk marktgedrag.

De Code Verantwoordelijk Marktgedrag stelt zich tot doel om dit gedrag te stimuleren en zo een continue impuls te geven aan verbetering van het uitbestedings- en inleenbeleid in Nederland. De Code doet daartoe een moreel appèl op opdrachtgevers, opdrachtnemers, inleners, vakbonden en intermediairs. Elk van deze partijen heeft een rol in het sociaal verantwoord en met respect voor kwaliteit uitzetten, aannemen en uitvoeren van opdrachten.

Onze doelstelling is het stimuleren van eerlijke concurrentie met goede voorwaarden voor de mensen op de werkvloer en goede resultaten voor de opdrachtgever. Hierbij past het gunnen van opdrachten aan de dienstverlener met de beste prijs-kwaliteitverhouding, zonder onbedoelde nadruk op prijs. Met 'kwaliteit' wordt dan zowel de kwaliteit van de dienstverlening bedoeld, als de kwaliteit van het werkgeverschap voor de werknemers.