



de Rechtspraak

Raad voor de
rechtspraak

TOETSINGSCOMMISSIE CIVIELE VONNISSEN

datum 9 december 2010
auteur P. Neleman en P.J. Neijt

Inhoudsopgave

1	Samenvatting	3
2	Aanleiding voor het onderzoek	4
3	Instelling en samenstelling Commissie	5
4	Doelstelling(en) van het onderzoek	5
4.1	Opdracht	5
4.2	Geen representativiteit van de testresultaten	6
5	Werkwijze	7
5.1	Inleiding	7
5.2	Ontwikkeling van een toetsingskader	8
5.3	Toetsing in koppels	9
5.4	Aangeleverde dossiers	11
6	Het toetsingskader getoetst	12
6.1	Inleiding	12
6.2	Gebruikte gegevens	12
6.3	Aanpak van de bespreking van de testresultaten	13
6.4	Overwegingen bij de bespreking van de testresultaten	15
6.5	Bespreking van de testresultaten	16
7	Het toetsingskader getoetst door drie hoven	20
8	Bruikbaarheid van het toetsingskader	21
8.1	Mogelijke toepassingen	21
8.2	Organisatorische aspecten	24
9	Conclusies en aanbevelingen	26
11	Bijlagen	28

1 Samenvatting

In dit rapport wordt verslag gedaan van een door de Raad voor de rechtspraak opgedragen experimenteel onderzoek ("pilot") inzake de vraag of het mogelijk is een kader te ontwikkelen met behulp waarvan de ambachtelijke kwaliteit van civiele vonnissen van rechtbanken kan worden getoetst.

De Presidentenvergadering heeft tot de uitvoering van de pilot besloten op basis van een door de Projectgroep Visitatieprotocol ontworpen opzet, die voorziet in de instelling van een toetsingscommissie en een aantal aspecten formuleert die voor de beoordeling van de kwaliteit van een vonnis van bijzonder belang zijn.

Ter uitvoering van de opdracht is een toetsingscommissie - hierna: de Commissie - ingesteld, bestaande uit acht (deels: oud-)leden van de rechterlijke macht en twee (deels: oud-)advocaten. Daarnaast is een Stuurgroep ingesteld ter begeleiding van het onderzoek. De opdracht aan de Commissie luidt als volgt:

1. Ontwikkel een systematiek voor het toetsen van de kwaliteit van vonnissen.
2. Geef een terugkoppeling van (opvallende) bevindingen en een conclusie over wat op basis van de bevindingen uit de gehouden steekproef naar het oordeel van de commissie gezegd kan worden over de kwaliteit van vonnissen van de hele populatie waaruit de steekproef is getrokken. De commissie zal, maar uitsluitend voor zover de commissie dat gezien de omvang van de steekproef verantwoord acht, deze terugkoppeling toespitsen op elke van de betrokken sectoren afzonderlijk.

Het is mogelijk gebleken een werkbaar toetsingskader te ontwikkelen waarvan de inhoud wordt gedragen door de Commissieleden. Met dit kader kan een beeld van de kwaliteit van een civiel vonnis in een handelszaak met verweer worden verkregen. Het betreft zowel een totaalbeeld als een beeld op onderdelen die de kwaliteit van een civiel vonnis in belangrijke mate bepalen.

De Commissie benadrukt in haar rapport dat haar onderzoek primair is gericht op het ontwikkelen van een toetsingskader en dat, reeds op grond van het beperkte aantal vonnissen dat in het onderzoek zijn betrokken, uit de resultaten van de experimentele toetsing van de vonnissen geen enkele conclusie mag worden getrokken ten aanzien van de kwaliteit van de rechtspraak als geheel, van de betrokken rechtbanken of van de individuele rechters. Daarop was het onderzoek ook niet gericht.

Op basis van door de voorzitter en secretaris opgestelde concepten heeft de Commissie een toetsingsformulier vastgesteld, met behulp waarvan een zo veel mogelijk objectieve toetsing van het vonnis mogelijk zou zijn. Daartoe is gekozen voor een opzet waarin de verschillende in het formulier omschreven aspecten van een vonnis en het eindoordeel daarover kunnen worden voorzien van een cijfer, variërend van 1 (slecht) tot 5 (uitstekend), met de mogelijkheid in een toelichting de waardering nader te motiveren. Bij het testen van het formulier is gebruik gemaakt van dossiers die door drie rechtbanken van verschillende grootte – Alkmaar, Arnhem en 's-Gravenhage – ter beschikking zijn gesteld. De met dit formulier opgedane ervaringen hebben, nadat ongeveer de helft van het aantal vonnissen was onderzocht, geleid tot aanpassing van het formulier. Deze tweede versie van het formulier heeft gediend als basis voor de toetsing van de resterende vonnissen. Nadien heeft de Commissie het formulier nogmaals aan een kritische beschouwing onderworpen, hetgeen heeft geresulteerd in een derde versie van het formulier en de toelichting, die de Commissie als haar definitieve voorstel presenteert.

Met het oog op de bevordering van de objectiviteit van de toetsing is voorts uitgangspunt geweest dat deze zou worden verricht door uit twee leden van de Commissie bestaande koppels in telkens

wisselende samenstelling. Dit is aldus uitgewerkt dat de leden - in eerste instantie onafhankelijk van elkaar - het vonnis hebben gezien en het formulier hebben ingevuld, en dat zij eerst daarna overleg hebben gevoerd over mogelijke heroverweging van hun oordeel. Onderkend is dat deze werkwijze arbeidsintensief is. Uit schattingen van de leden kan worden afgeleid dat de toetsing van een vonnis een gezamenlijke werktijd van ongeveer 13 uur vergde.

De Commissie heeft in haar rapport de bevindingen bij het testen van het toetsingskader opgenomen, niet omwille van de testresultaten zelf, maar om te demonstreren dat een werkwijze als door haar gevolgd in beginsel een geschikte methode is om de kwaliteit van vonnissen te toetsen. Met het oog op de bruikbaarheid van deze werkwijze in de praktijk heeft de Commissie vervolgens verschillende mogelijke varianten op haar werkwijze onder ogen gezien, die evenwel geen van alle aanbeveling verdienen.

Tegelijkertijd met het onderzoek van de Commissie hebben de hoven Amsterdam, Arnhem en 's-Gravenhage een aantal vonnissen van de betrokken rechtbanken in hun ressort getoetst. Dit betrof vonnissen in andere zaken dan door de Commissie zijn getoetst. Hierbij is gebruik gemaakt van de eerste en tweede versie van het toetsingsformulier. De indruk bestaat dat de hoven de formulieren op adequate wijze hebben kunnen gebruiken. Bij de toetsing van de kwaliteit van het vonnis lijkt de beoordeling van het ingestelde appel voor de hoven een rol te spelen. Verder zijn er geen opvallende verschillen in het gebruik van de formulieren door de hoven en de Commissie.

De Commissie heeft met het oog op het gebruik van de door haar aanbevolen werkwijze in een visitatie nog een aantal mogelijke nadere onderzoeksvragen geformuleerd, met name ten aanzien van het formulier, het aantal in het onderzoek te betrekken vonnissen en de mogelijkheid beter zicht te krijgen op de comparitie.

Tot slot beveelt de Commissie aan het toetsingskader niet alleen te beschouwen als een kwaliteitsinstrument maar ook als een instrument voor kennisaccumulatie. Het toetsingskader heeft dan in potentie een scala aan toepassingsmogelijkheden.

2 Aanleiding voor het onderzoek

De Raad voor de rechtspraak toetst kwaliteit van de rechtspraak op verschillende onderdelen. De visitatiecommissie van de Raad voor de rechtspraak bekijkt met name de invulling van de procesmatige en organisatorische randvoorwaarden voor kwalitatief goede rechtspraak en formuleert aandachts- en verbeterpunten ten behoeve van de kwaliteitszorg van de rechterlijke organisatie. De Raad voor de rechtspraak heeft daarnaast de wens te onderzoeken of een toetsing van de ambachtelijke kwaliteit van civiele vonnissen mogelijk is. Op dit moment wordt de kwaliteit van een civielrechtelijk vonnis nog uitsluitend getoetst doordat collega rechters binnen dezelfde rechtbank een (concept)vonnis meelesen. De toetsing die de Raad voor de rechtspraak in de toekomst voor ogen heeft, zou onderdeel kunnen worden van de visitatie.

In de Presidentenvergadering van 28 september 2009 is besloten tot uitvoering van een pilot kwaliteitstoetsing rechterlijke uitspraken. Dit besluit is genomen op basis van een door de Projectgroep Visitatieprotocol ontworpen "Opzet pilot kwaliteitstoetsing van rechterlijke uitspraken". Deze opzet voorziet in de instelling van de onderhavige Commissie, die zou moeten onderzoeken of het toetsen van de kwaliteit van civiele uitspraken mogelijk is. De Projectgroep heeft in samenspraak met de Presidentenvergadering de volgende hoofdlijnen geformuleerd:

- De toetsing van de kwaliteit van civiele vonnissen zou moeten worden onderzocht in de vorm van een ‘peer review’. De Commissie zou moeten bestaan uit acht leden, van wie twee uit de advocatuur, één uit de hoven, één uit de Hoge Raad en vier uit de rechtbanken. Met de leden uit de advocatuur worden ook externen betrokken en wordt de objectiviteit van de toetsing verhoogd.
- De Commissie moet haar aandacht richten op de ontwikkeling van een beoordelingskader.
- Het beoordelingskader zou de ambachtelijke kwaliteit van een vonnis centraal moeten stellen, waarbij valt te denken aan: vaktechnisch goed in elkaar gezet, leesbaar en helder, niet langer of diepgravender dan nodig, consistent, procedureel en materieelrechtelijk foutloos, uitmondend in een overtuigende en aanvaardbare beslissing.
- De bestudering van dossiers hoort bij de toetsing van de kwaliteit van een vonnis. Dit is tijdrovend, maar noodzakelijk.
- Het rapport van de Commissie zal openbaar zijn. Het ligt in de rede dat leden van de Commissie daarnaast de resultaten vertrouwelijk en meer gedetailleerd aan de betrokken rechters terugkoppelen.

De opzet voorziet in een gelijktijdige toetsing van (andere) vonnissen door een drietal hoven. Deze toetsing heeft plaatsgevonden buiten de verantwoordelijkheid van de Commissie. Op de toetsing door de hoven wordt ingegaan in hoofdstuk 7.

3 Instelling en samenstelling Commissie

De Raad voor de rechtspraak heeft de Commissie ingesteld. De Commissie bestaat uit tien leden die werkzaam zijn (of zijn geweest) in de rechterlijke macht of de advocatuur en ruime ervaring hebben met de civiele rechtspraak. De samenstelling van de Commissie is vermeld in **bijlage 1**.

De Projectgroep heeft daarnaast een Stuurgroep ingesteld, die tot taak had de Commissie te adviseren en waar nodig vragen te beantwoorden over de wijze waarop invulling aan de opdracht moet worden gegeven. De samenstelling van de Stuurgroep is eveneens vermeld in **bijlage 1**.

4 Doelstelling(en) van het onderzoek

4.1 Opdracht

De definitieve opdracht aan de Commissie werd gegeven op 31 maart 2010 (**bijlage 2**). De opdracht houdt het volgende in:

1. Ontwikkel een systematiek voor het toetsen van de kwaliteit van vonnissen.
2. Geef een terugkoppeling van (opvallende) bevindingen en een conclusie over wat op basis van de bevindingen uit de gehouden steekproef naar het oordeel van de commissie gezegd kan worden over de kwaliteit van vonnissen van de hele populatie waaruit de steekproef is getrokken. De commissie zal, maar uitsluitend voor zover de commissie dat gezien de omvang van de steekproef verantwoord acht, deze terugkoppeling toespitsen op elk van de betrokken sectoren afzonderlijk.

In deze opdracht worden de navolgende resultaten omschreven:

1. Een beoordelingskader voor het toetsen van vonnissen.
2. Een openbaar eindrapport met daarin de uitkomsten van de kwaliteitstoetsing, een beschrijving van de gehanteerde methoden van toetsing, zoals toegepast door de Commissie en de

- deelnemende hoven, een vergelijking tussen beide methoden en een advies over de inzet van het ontwikkelde instrumentarium.
3. Een terugkoppeling naar de deelnemende rechtbanken.
 4. Een evaluatie van de gehanteerde werkwijze: wat is de opbrengst en meerwaarde van een dergelijke toetsing van vonnissen?

De opdracht is tot stand gekomen op basis van een door de Projectgroep geformuleerde eerdere versie van de opdracht. Een belangrijk verschil tussen de oorspronkelijke en definitieve opdracht is dat niet langer een rapport wordt gevraagd waarin de uitkomsten van de toetsing per rechtbank worden weergegeven. Daarmee is de nadruk komen te liggen op het ontwikkelen van een beoordelingskader, door de Commissie aangeduid als toetsingskader, en de vraag op welke wijze een dergelijk toetsingskader zou kunnen worden toegepast.

Een belangrijke aanleiding voor de wijziging van de opdracht is een overleg met TNO geweest op 5 maart 2010. Tijdens dit overleg zijn de doelstellingen uit de oorspronkelijke opdracht besproken aan de hand van een plan van aanpak dat de Raad voor de rechtspraak had opgesteld. Dit plan van aanpak wordt nader toegelicht in hoofdstuk 5. TNO heeft een advies uitgebracht over het onderzoeksmodel, de representativiteit van de resultaten, de verwerkbaarheid van de resultaten en de uitvoerbaarheid (tijdsplanning) van het project. TNO heeft geadviseerd de ontwikkeling van het beoordelingskader (toetsingskader) te scheiden van de toepassing daarvan. De reden hiervoor is dat het aantal vonnissen dat in de pilot wordt betrokken te klein is om een uitspraak over de kwaliteit van de vonnissen in de onderzochte rechtbanken te doen, die enige representativiteit zou hebben. Het advies van TNO luidt:

“De Raad voor de rechtspraak wil dat de kwaliteit van vonnissen wordt getoetst. Daarvoor is het nodig om een beoordelingskader te ontwikkelen. Er is voor dit werk nagedacht over een onderzoeksopzet waarmee het beoordelingskader kan worden getoetst. De gedachte in de notities is dat het beoordelen van 56 zaken kan aantonen dat het beoordelingskader hanteerbaar is en goed kan werken. Ons advies is om de toepassing van het beoordelingskader te scheiden van de ontwikkeling van het kader. De ontwikkeling dient te gebeuren in een interbeoordelaarsbetrouwbaarheids-onderzoek. Daarvoor is nodig dat het beoordelingskader wordt neergezet, dat de beoordelingsregels precies worden geformuleerd en dat met een goede opzet de toepasbaarheid van de regels wordt onderzocht. Zulk onderzoek is lastig op zich. De resultaten van de beoordelingen die tijdens een dergelijk onderzoek [worden verkregen,] mogen niet zomaar als een indicatie worden gezien van de kwaliteit van de vonnissen in de onderzochte rechtbanken. Daar is een andere onderzoeksopzet voor nodig.”

4.2 Geen representativiteit van de testresultaten

Het onderzoek van de Commissie was gericht op de vraag of het mogelijk is een formulier met toelichting te ontwikkelen waarmee in de toekomst kwalitatief onderzoek naar civiele vonnissen kan worden gedaan. De Commissie heeft hierbij onder meer gebruik gemaakt van vonnissen en procesdossiers in afgedane handelszaken die zijn aangeleverd door drie rechtbanken. Zoals in paragraaf 4.1 is weergegeven, staan de opzet van het onderzoek van de Commissie en het relatief geringe aantal getoetste zaken eraan in de weg om aan de hand van de onderzochte verzameling vonnissen toereikend onderbouwde algemene uitspraken te doen over de kwaliteit van civiele vonnissen. In totaal zijn 54 zaken onderzocht (zie hoofdstuk 6).

De Commissie benadrukt dan ook dat de uitkomsten, opgenomen in dit rapport, niet representatief zijn voor de kwaliteit van vonnissen in handelszaken en louter dienen ter illustratie van de bruikbaarheid van het ontwikkelde toetsingskader. De bespreking van de testresultaten in hoofdstuk 6 staat geheel in het teken van de ontwikkeling en de bruikbaarheid van het toetsingskader. De bespreking bevat geen algemeen oordeel over de kwaliteit van civiele vonnissen en dit kan er gezien de onderzoeksopzet ook niet aan worden ontleend.

Dit strookt met het plan van aanpak van de Raad voor de rechtspraak, waarin al werd gesignaleerd dat de uitkomsten van de pilot niet mogen worden gegeneraliseerd tot een oordeel over de kwaliteit van de rechtspraak als geheel, de betrokken rechtbanken of rechters. Het behoort niet tot de opdracht van de Commissie een dergelijk oordeel te geven en dat was, gelet op de beperkte hoeveelheid zaken die werd gebruikt voor het ontwikkelen van een toetsingskader, ook niet mogelijk. Zoals in het overleg met TNO naar voren is gekomen, is het aantal vonnissen te beperkt om enige relevante uitspraak te kunnen doen over de kwaliteit van de betrokken rechtbanken, afzonderlijk of gezamenlijk. Bovendien is bij de selectie van zaken, zoals in hoofdstuk 5 wordt toegelicht, zo veel mogelijk een spreiding over verschillende rechters gemaakt en zijn niet meer dan één of enkele vonnissen van een individuele rechter in het onderzoek van de Commissie betrokken. Uit de toetsing van vonnissen in het kader van de ontwikkeling van het toetsingskader kan dus ook geen enkele conclusie worden getrokken over de kwaliteit van de betrokken rechters.

De omvang van het onderzoek is al met al beperkt geweest en heeft zich gericht op de ontwikkeling van een kader waarbinnen de kwaliteit van vonnissen kan worden getoetst. De kwaliteitsaspecten die de Commissie in de toetsing heeft betrokken, zijn nader uiteen gezet in hoofdstuk 6. Het betreft kort gezegd aspecten waarmee kan worden beoordeeld of een civiel vonnis tot stand is gekomen volgens de standaard van *best practices*. Andere kwaliteitsaspecten die een rol spelen in de kwaliteit van de behandeling van een zaak, zoals doorlooptijd, bejegening, het verloop van de comparitie e.d., zijn gezien de opdracht en de opzet van het onderzoek niet betrokken bij de ontwikkeling van het toetsingskader.

5 Werkwijze

5.1 Inleiding

De Raad voor de rechtspraak heeft een plan van aanpak opgesteld, waarin in hoofdlijnen de navolgende onderwerpen werden vastgelegd over de wijze waarop de Commissie te werk zou moeten gaan:

- De vonnissen die de Commissie gebruikt voor het ontwikkelen van het toetsingskader, zijn afkomstig van de rechtbanken te Alkmaar, 's-Gravenhage en Arnhem.
- De Commissie is zelf verantwoordelijk voor de opzet van het onderzoek. Daarbij heeft de Commissie de mogelijkheid een onderzoeker in te schakelen voor advies en specifieke uitvoeringstaken. Het is van belang dat de onderzoeksopzet geen aanleiding geeft tot twijfels over de uitkomsten. Een aantal aspecten is echter gegeven, waaronder het aantal vonnissen dat kan worden gebruikt, de spreiding over verschillende rechters en de anonimiteit van de dossiers. De leden van de Commissie werken in koppels, hetgeen inhoudt dat elk vonnis wordt beoordeeld door twee leden van de commissie.
- Er zullen gelijktijdig (andere) vonnissen van voornoemde rechtbanken worden getoetst door de hoven te Amsterdam, 's-Gravenhage en Arnhem. Deze toetsing vindt gelijktijdig plaats, maar valt buiten de verantwoordelijkheid van de Commissie.

De voorzitter en secretaris van de Commissie hebben op basis van de opdracht en het plan van aanpak voorstellen geformuleerd voor de praktische uitwerking daarvan en de wijze waarop het toetsingskader zou kunnen worden ontwikkeld en beproefd. Bij de beschrijving van de werkwijze is tot uitgangspunt genomen dat de opdracht aan de Commissie in hoofdzaak is het opstellen en testen van een toetsingskader. De werkwijze is hierop gericht geweest. De beschrijving van de werkwijze van de Commissie is opgenomen in **bijlage 3**.

5.2 Ontwikkeling van een toetsingskader

De hiervoor in hoofdstuk 2 genoemde 'Opzet kwaliteitstoetsing van rechterlijke uitspraken' houdt in de eerste plaats in dat de Commissie haar aandacht moest richten op de ontwikkeling van een toetsingskader, dat de ambachtelijke kwaliteit van het beoordeelde vonnis centraal zou moeten stellen.

Als een van de doeleinden van de pilot zou men kunnen aanmerken het ontwikkelen van een methode waarmee de toetsing van vonnissen (zo veel mogelijk) zo wordt geobjectiveerd dat iedere beoordelaar tot eenzelfde oordeel over de kwaliteit van een vonnis komt, althans dat verschillen in beoordelingen binnen enge grenzen liggen. Een zodanige objectivering kan door een aantal maatregelen worden bevorderd, maar niet volledig worden verwezenlijkt, omdat de beoordeling van de kwaliteit van een vonnis onvermijdelijk mede een subjectief karakter draagt.

In de eerste plaats zijn ter bevordering van de objectiviteit van de toetsing de in de opzet geformuleerde vereisten uitgesplitst in afzonderlijke onderdelen, waarin concrete aandachtspunten zijn omschreven. Deze konden van een score van 1 (slecht) tot 5 (uitstekend) worden voorzien. Een en ander is nader toegelicht in hoofdstuk 6. Het toetsingskader is opgenomen in de bijlagen (toetsingsformulier versie 1 (**bijlage 4**), versie 2 (**bijlage 5**) en versie 3 (**bijlage 6**), telkens met toelichting). Langs deze weg wordt echter niet bereikt dat de aan afzonderlijke onderdelen en aandachtspunten toegekende waardering met een eenvoudige rekenkundige bewerking kan worden herleid tot een eindoordeel over het vonnis. Hieraan staat in de weg dat afhankelijk van de zaak het ene aspect zwaarder weegt dan het andere en het niet goed doenlijk is tevoren en in abstracto een soortelijk gewicht aan elk aspect toe te kennen. De eindoordelen zijn dus wel gebaseerd op, maar niet rechteaf te leiden uit de oordelen over de afzonderlijke aspecten. Ook dit zal nader worden toegelicht aan de hand van de bespreking van de testresultaten in hoofdstuk 6.

In de tweede plaats is met het oog op de objectiviteit ervoor gekozen dat voor elk vonnis het toetsingsformulier door telkens twee leden van de Commissie in koppels van wisselende samenstelling - in eerste instantie onafhankelijk van elkaar - wordt ingevuld. Zie hieromtrent ook paragraaf 5.3.

Hierbij past nog de aantekening dat een onderlinge vergelijking van de vonnissen aan de hand van de eindoordelen verder wordt bemoeilijkt door tenminste twee factoren die voor de interpretatie van het eindoordeel over het vonnis van belang zijn, maar niet systematisch konden worden onderzocht en dus in het algemeen niet in het eindoordeel konden worden verdisconteerd, te weten de kwaliteit van de processtukken alsmede het uitzonderlijke karakter en/of de complexiteit van een zaak. Kort gezegd: ook indien twee vonnissen eenzelfde eindoordeel hebben, is het vonnis in de ene zaak kwalitatief niet zonder meer gelijk te stellen aan het vonnis in de andere zaak.

Versie 1 van het toetsingskader (**bijlage 4**), bestaande uit het toetsingsformulier en de toelichting, is op 4 juni 2010 voorgelegd aan de Stuurgroep en op 8 juni 2010 toegezonden aan de deelnemende hoven. De Commissie en de hoven hebben deze versie gebruikt voor het onderzoek van een eerste gedeelte van de vonnissen.

Versie 1 van het toetsingsformulier is zowel op papier als digitaal aan de leden van de Commissie gezonden. Door het gebruik van het digitale formulier bleek een vrij algemeen door de leden naar voren gebracht bezwaar: door de fijnmazigheid bleek het vaak lastig te overzien op welke plaats een bepaalde waardering en/of opmerking zou moeten worden opgeschreven. De inhoudelijke dimensie van dit bezwaar is betrokken in de ontwikkeling van versie 2 van het toetsingsformulier. Afgezien daarvan hebben de leden naar aanleiding van de toetsingswerkzaamheden ook andere, veelal inhoudelijke, opmerkingen gemaakt en daarover vragen gesteld aan de voorzitter en de secretaris. Ook deze vragen en opmerkingen betroffen meestal de vraag hoe en op welke plaats een kwestie in het formulier moest worden verwerkt.

In de vergadering van de Commissie, die is gehouden op 29 juni 2010 nadat ongeveer de helft van de vonnissen was beoordeeld, zijn de ervaringen met het formulier besproken, mede aan de hand van de opmerkingen die op de ingezonden formulieren of in de begeleidende e-mails waren gemaakt. Ook hier was de belangrijkste conclusie dat het formulier eenvoudiger zou moeten zijn, terwijl voorts enkele onduidelijkheden en lacunes in het formulier werden gesignaleerd.

De voorzitter en de secretaris hebben op 2 juli 2010 aan de Stuurgroep verslag gedaan van de eerste bevindingen. Hierbij is onder meer gesproken over de opmerkingen van de leden van de Commissie over versie 1 van het toetsingsformulier.

Dit heeft geleid tot versie 2 van het toetsingsformulier en de bijbehorende toelichting (**bijlage 5**). De resterende vonnissen zijn met gebruikmaking van dit formulier en deze toelichting onderzocht. Ook de hoven hebben deze versie voor een tweede tranche van de vonnissen gebruikt.

Reeds tijdens de toetsing van deze tranche bleek dat dit tweede formulier naar het oordeel van vele leden een verbetering ten opzichte van de eerste versie was. Dit neemt niet weg dat in de vergadering van de Commissie van 31 augustus 2010 nog op een aantal punten verbeteringen zijn voorgesteld. Zo is voorgesteld de instructie (verrichtingen van de rechter die de zaak gereed maken voor het wijzen van vonnis), die van wezenlijk belang is voor de verdere behandeling van de zaak en de kwaliteit van het vonnis, een prominentere plaats in het formulier te geven.

Naar aanleiding van de verschillende suggesties is het formulier wederom op een aantal punten gewijzigd. Het gewijzigde formulier met de bijbehorende toelichting is als versie 3 bij dit rapport gevoegd (**bijlage 6**). Versie 3 is ontwikkeld aan de hand van de ervaringen die zijn opgedaan met de eerdere versies, maar heeft - gegeven de opzet van het onderzoek - geen onderwerp van nadere toetsing bij de beoordeling van vonnissen kunnen zijn. Gelet op de opgedane ervaringen acht de commissie het echter verantwoord deze versie als haar uiteindelijke voorstel te presenteren.

5.3 Toetsing in koppels

Bij de opzet van het onderzoek is steeds uitgangspunt geweest dat voor elk vonnis het toetsingsformulier zou worden toegepast door telkens twee leden van de Commissie. Dit is aldus uitgewerkt dat uit de leden van de Commissie - met uitzondering van de voorzitter en de secretaris - 28 combinaties van verschillende samenstelling ('koppels') zijn gevormd, die elk twee vonnissen zouden kunnen toetsen.

Aan deze opzet liggen verschillende, ten dele samenhangende overwegingen ten grondslag:

1. De kwaliteit en volledigheid bij het gebruiken van het formulier worden hiermee bevorderd: twee zien meer dan één. Zoals consciëntieus meelezen van een vonnis de kwaliteit daarvan kan verhogen, zal ook toetsing van een vonnis door twee beoordelaars de kwaliteit van die toetsing naar verwachting verhogen.
2. Toetsing door twee beoordelaars beperkt naar verwachting de kans dat persoonlijke voorkeuren de beoordeling onevenredig kleuren. Dit is van belang met het oog op de objectiviteit van de toetsing (zie paragraaf 5.2), al kan vanzelfsprekend niet worden gezegd dat met toetsing in koppels de objectiviteit van de toetsing per definitie voldoende is gewaarborgd.
3. De wisselende samenstelling brengt mee dat een lid van de Commissie telkens met een ander van de overige leden van de Commissie een toetsing heeft uitgevoerd. Aldus wordt ook voorkomen dat, zoals zich bij vaste samenstelling van de koppels laat denken, gemeenschappelijke, al dan niet uitdrukkelijke opvattingen van de twee koppelgenoten een overheersende rol gaan spelen.

Deze opzet is praktisch aldus uitgewerkt dat een exemplaar van het te onderzoeken vonnis met het bijbehorende dossier is verstrekt aan ieder van beide leden die een koppel vormen. Ieder van beide leden heeft vervolgens onafhankelijk van de andere beoordelaar een toetsing van het vonnis verricht, waarna de resultaten zijn uitgewisseld met het oog op onderling overleg tussen beide beoordelaars.

Hierbij verdient aantekening dat bij de voorbereiding van het onderzoek is onderkend dat het overleg afbreuk zou kunnen doen aan de onafhankelijke toetsing van de beide koppelgenoten en de vraag onder ogen is gezien of met het oog hierop dit overleg wel wenselijk was. Ofschoon dit aspect niet zonder gewicht is, is uiteindelijk toch gekozen voor overleg tussen beide beoordelaars, in het bijzonder omdat daarmee de kwaliteit van de toetsing zou worden verhoogd, doordat daarmee bijvoorbeeld aspecten die een van de koppelgenoten niet heeft gezien of niet belangrijk heeft gevonden, alsnog aan de orde kunnen komen, of doordat het overleg anderszins kan leiden tot heroverweging. Overigens is uit opmerkingen van de leden gebleken dat de mogelijkheid van overleg in het algemeen als positief is ervaren en dat in vele gevallen de resultaten van de toetsing zo dicht bij elkaar lagen dat geen aanpassing nodig was. In een enkel geval heeft wel aanpassing plaatsgevonden.

In de Commissie is voor aanvang van de toetsing afgesproken dat het overleg tot doel had dat de koppelgenoten van gedachten wisselden over overeenkomsten en verschillen in hun beoordeling en nagingen of de afwegingen van de ander aanleiding waren om de eigen toetsingsresultaten te wijzigen. Het doel was niet dat de koppelgenoten het eens werden wanneer zij van mening verschilden.

De definitieve toetsingsresultaten van de koppelgenoten zijn toegezonden aan de voorzitter en de secretaris van de Commissie. Daarbij werd vermeld of het overleg tussen de koppelgenoten tot bijstelling van de toetsingsresultaten had geleid. De voorzitter en de secretaris hebben de uitkomsten gezien in verband met het desbetreffende vonnis en het dossier, maar zij hebben het niet tot hun taak gerekend de toetsingsresultaten, als het ware als een hogere instantie, aan een inhoudelijke herbeoordeling te onderwerpen.

De leden hebben niet exact bijgehouden - dat was hun ook niet gevraagd - hoeveel tijd elke beoordeling vergde. Schattingen kwamen neer op gemiddeld ongeveer zes uur per vonnis voor ieder van beide beoordelaars, te vermeerderen met een uur voor het overleg en de uitwerking, derhalve in totaal 13 uur per te toetsen vonnis. Indien de thans gehanteerde methode ook in het kader van bijvoorbeeld visitaties zou worden gebruikt, moet derhalve bij de planning worden uitgegaan van een aanzienlijk tijdsbeslag.

5.4 Aangeleverde dossiers

Bij het testen van het toetsingskader tijdens de ontwikkeling daarvan is gebruik gemaakt van een beperkt aantal dossiers. De Commissie heeft de dossiers aangeleverd gekregen van de sectorvoorzitters civiel van de rechtbanken te Alkmaar, Arnhem en 's-Gravenhage.

De Projectgroep heeft bepaald welke rechtbanken aan de pilot zouden deelnemen. De Projectgroep heeft bij de keuze van de rechtbanken de veronderstelling laten meewegen dat rechtbanken een verschillende interne werkwijze hanteren. Zij heeft daarbij gekeken naar de omvang van de civiele sectoren van de rechtbanken. Er is gekozen voor een grote, middelgrote en kleine civiele sector. De volgende gegevens geven enig inzicht in de omvang van de verschillende sectoren. In de sectoren civiel recht van de rechtbanken Alkmaar, Arnhem en 's-Gravenhage werden in 2009 respectievelijk 402, 680 en 1049 eindvonnissen in handelszaken met verweer gewezen (totaal: 2131). De Commissie heeft zoals gezegd in totaal 54 zaken in de toetsing betrokken. In 51 daarvan is het eindvonnis gewezen in 2009, in de overige 3 in 2010. Dat geeft een indruk van de beperkte omvang van het aantal vonnissen dat in het onderzoek is betrokken: 50 van 2131 is 2,3%. Bij het toetsen van de vonnissen heeft de Commissie overigens geen verschillen opgemerkt die mogelijk samenhangen met de grootte van de sector.

Het is aan de sectorvoorzitters overgelaten te bepalen welke dossiers zouden kunnen worden onderworpen aan het onderzoek door de Commissie. Zij hebben aan de Commissie bericht hoe zij hun keuze hebben gemaakt. Dit is weergegeven in de werkwijze (**bijlage 3**). Uit de toelichtingen van de sectorvoorzitters blijkt dat de wijze waarop zij vonnissen hebben geselecteerd voor het doel van de pilot - de ontwikkeling van een toetsingskader - voldoende willekeurig is geweest. Er heeft niettemin wel enige selectie plaatsgevonden, met name om ervoor te zorgen dat de vonnissen zoveel mogelijk van verschillende rechters afkomstig waren.

De rechtbanken hebben elk 24 dossiers aangeleverd, met dien verstande dat:

- alle dossiers betrekking hebben op een zaak waarin een eindvonnis is gewezen;
- (eind)vonnissen van zoveel mogelijk verschillende rechters voor het onderzoek worden aangewezen. Hiervoor is gekozen om "oververtegenwoordiging" van een of meer rechters te voorkomen;
- twee dossiers worden aangeleverd waarin door een meervoudige kamer vonnis is gewezen;
- geen zaken worden aangeleverd op het gebied van het familierecht, of behorend tot de huidige competentie van de sector kanton.

5.4.1 Aantal dossiers

Uit de door de rechtbanken ingezonden dossiers heeft de Commissie zelf de 54 meest bruikbare dossiers geselecteerd. De zogenaamde 'bagatel'-zaken werden uit de aangeleverde selectie verwijderd. Het betrof hier bijvoorbeeld zaken waarin de advocaat van de gedaagde zich aan de zaak onttrok nog voordat een conclusie van antwoord was genomen. Een eventuele oververtegenwoordiging van een bepaalde rechter werd geredresseerd. Er zijn voorts enkele reservedossiers geselecteerd. Uiteindelijk zijn 54 zaken in de toetsing betrokken, zoals is uiteengezet in hoofdstuk 6.

5.4.2 Anonimiteit

De rechtbanken is verzocht ervoor te zorgen dat de naam van de rechter die een vonnis heeft gewezen, niet uit de dossiers zou blijken. De leden van de Commissie hebben in enkele gevallen toch kunnen waarnemen wie de betrokken rechters zijn geweest. Hoewel de (tussen)vonnissen vrijwel altijd waren geanonimiseerd, bleek de naam van die rechter en van andere bij de zaak betrokken rechters uit correspondentie, het proces-verbaal van de comparitie of aantekeningen die in het dossier zaten. De

toetsing heeft dus niet volledig anoniem plaatsgevonden. Voor zover de naam van een rechter uit het dossier bleek, is deze informatie verder niet gebruikt.

5.4.3 Inhoud van de dossiers

De Commissie heeft de vonnissen beoordeeld aan de hand van de volledige dossiers, dat wil zeggen niet alleen de vonnissen, het proces-verbaal van de comparitie en processtukken, maar ook de producties die partijen hebben overgelegd. Het is de Commissie opgevallen dat bij het archiveren van dossiers door de rechtbanken niet steeds de producties worden bewaard. Het gevolg hiervan voor de pilot is geweest dat in een aantal zaken alsnog producties van de gebruikte dossiers moesten worden opgevraagd bij de (advocaten van) de procespartijen.

De bevindingen bij de toetsing door de koppels zijn gebaseerd op het eindvonnis en eventuele tussenvonnissen, in samenhang gezien. Het is immers zeer wel mogelijk dat in een tussenvonnis werd beslist op de belangrijkste geschilpunten tussen partijen. Voor de toetsing maakt dit niet uit, omdat daarbij is uitgegaan van het geheel van eventuele tussenvonnissen en het eindvonnis gezamenlijk.

6 Het toetsingskader getoetst

6.1 Inleiding

Zoals gezegd kunnen de resultaten van het testen van het toetsingskader gezien de onderzoeksopzet en het aantal onderzochte zaken niet worden beschouwd als een indicatie van de kwaliteit van civiele vonnissen in het algemeen. Het doel van dit hoofdstuk is om inzichtelijk te maken of het toetsingskader in beginsel een geschikt instrument is voor de toetsing van de kwaliteit van civiele vonnissen. In verband hiermee komen in dit hoofdstuk ook ervaringen met de werkwijze van de Commissie aan de orde, omdat die van invloed kunnen zijn op de geschiktheid van het toetsingskader.

In dit licht worden de testresultaten hierna besproken. De gegevens uit de toetsingsformulieren zijn daartoe verzameld, onderworpen aan een analyse en besproken door de Commissie. Dit hoofdstuk is als volgt ingericht. Eerst volgt een toelichting op de gebruikte gegevens. Vervolgens wordt de aanpak van de bespreking van de testresultaten beschreven. Daarna volgen enkele overwegingen tegen de achtergrond waarvan de uitkomsten worden besproken. Tot slot worden de uitkomsten besproken en aan een nadere beschouwing onderworpen. Een korte conclusie maakt hiervan deel uit.

6.2 Gebruikte gegevens

Uit de dossiers die zijn aangeleverd door de drie rechtbanken, zijn 54 zaken geselecteerd (zie paragraaf 5.4). Na de verdeling over de leden van de Commissie werden twee zaken alsnog te omvangrijk geacht in verhouding tot de overige zaken. Deze zijn vervangen door andere zaken van dezelfde rechtbank.

Uiteindelijk zijn dus 54 afgedane zaken betrokken in het testen van het toetsingsformulier. In 27 van die 54 zaken is versie 1 van het toetsingsformulier gebruikt en in 27 versie 2. Het betrof steeds handelszaken met verweer: de procespartijen zijn verschenen in de procedure en er is een schriftelijk partijdebat geweest. In alle zaken was er ook een mondelinge behandeling (comparitie en/of pleidooi).

6.3 Aanpak van de bespreking van de testresultaten

In de bespreking van de bevindingen bij het gebruik van de toetsingsformulieren is tot uitgangspunt genomen dat het kader steeds is de opdracht aan de Commissie om te onderzoeken of een bruikbaar toetsingsinstrument kan worden ontwikkeld.

6.3.1 Overzicht ‘toetsing per zaak per koppel’

Bij de beschrijving van de bevindingen in paragraaf 6.4 wordt gebruik gemaakt van het ‘Overzicht toetsing per zaak per koppel’. In dit overzicht is per zaak het oordeel per koppelgenoot (door middel van twee cijfers naast elkaar) weergegeven over het vonnis in zijn geheel (hierna: totaalscore) en op de afzonderlijke onderdelen. Tabel 1 in dit overzicht betreft de zaken waarin is getoetst met gebruik van versie 1 van het toetsingsformulier; tabel 2 betreft het gebruik van versie 2.

Tabel 1 – Overzicht toetsing per zaak per koppel versie 1 toetsingsformulier

Zaak	1	2	3	4	5	6	7	8
A	5 5	4 4	5 4	4 4	3 4	4 4	4 4	3 5
B	5 5	4 4	5 5	5 4	4 4	4 4	4 4	4 4
C	5 5	5 4	4 4	3 4	4 3	4 4	4 4	4 4
D	5 5	4 4	4 4	4 4	4 4	4 4	3 4	4 4
E	5 5	4 4	4 4	4 4	4 4	4 4	4 4	4 4

Zaak	9	10	11	12	13	14	15	16
A	4 4	2 2	3 4	4 4	4 4	3 4	2 2	3 2
B	4 4	3 3	4 4	3 3	3 3	3 3	2 2	2 2
C	4 4	2 2	4 3	3 3	3 3	3 3	2 2	2 2
D	4 4	3 3	4 3	4 4	4 2	4 4	4 4	3 3
E	4 4	3 3	3 3	3 3	3 3	3 3	2 2	2 2

Zaak	17	18	19	A = vaststelling feiten	5 = uitstekend
A	2 2	2 2	4 3	B = juridische beoordeling	4 = goed
B	3 2	2 2	2 3	C = motivering en beslissing	3 = voldoende
C	3 2	2 2	2 2	D = leesbaarheid	2 = onvoldoende
D	3 3	3 3	4 4	E = totaalscore	1 = slecht
E	2 2	2 2	2 2		

Zaak	20	21	22	23	24	25	26	27
A	5 5	4 5	3 3	4 4	5 4	4 4	4 2	3 3
B	5 4	5 4	4 4	4 4	4 2	4 1	4 4	3 3
C	5 4	4 3	4 3	4 4	4 2	2 2	3 3	3 2
D	5 4	5 4	3 4	4 4	4 4	4 4	4 4	3 3
E	5 4	5 4	4 3	4 3	4 2	3 2	3 2	3 2

6.3.2 Afzonderlijke onderdelen versie 1

Bij versie 1 betreffen de afzonderlijke onderdelen de volgende onderwerpen: vaststelling van de feiten die relevant zijn voor de beslissing, juridische beoordeling, aspecten van motivering en beslissing, en leesbaarheid.

**Tabel 2 – Overzicht toetsing per zaak per koppel
versie 2 toetsingsformulier**

Zaak	28	29	30	31	32	33	34	35
F	4 5	4 4	4 4	4 4	4 4	4 4	4 4	4 4
G	4 5	4 4	4 4	4 4	3 3	4 4	4 4	4 4
H	4 5	4 4	4 4	4 4	4 4	4 4	4 4	4 4
I	4 5	4 4	4 4	4 4	4 4	4 4	4 4	4 4
J	3 2	4 4	4 4	4 4	4 4	4 4	4 4	3 3
K	3 5	3 4	4 4	3 4	4 4	4 4	4 4	4 4
L	4 4	4 4	4 4	4 4	4 4	4 4	4 4	4 4

Zaak	36	37	38	39	40	41	42	43
F	3 5	3 3	4 3	4 4	3 3	3 4	3 3	4 4
G	4 5	3 3	3 3	4 4	3 3	3 4	3 3	3 3
H	3 4	2 2	4 3	3 3	4 4	3 4	3 3	3 2
I	4 4	4 4	4 3	4 3	3 2	3 4	4 4	4 4
J	4 4	4 4	3 4	3 3	3 3	3 3	3 3	3 3
K	3 4	3 3	3 4	3 3	3 3	3 3	3 3	3 3
L	4 4	3 3	3 3	3 3	3 3	3 3	3 3	3 3

Zaak	44	45	46	47	48	49	50
F	2 3	4 4	3 2	3 3	3 4	3 3	2 2
G	4 4	3 3	4 4	3 3	1 2	3 3	3 3
H	3 3	3 4	3 3	2 3	2 3	2 2	2 2
I	4 4	3 3	4 4	3 3	4 4	4 4	3 4
J	3 3	4 4	3 3	3 2	4 3	4 4	2 2
K	2 2	3 3	2 2	2 3	3 2	2 2	2 2
L	3 3	3 3	3 3	3 3	2 2	2 2	2 2

Zaak	51	52	53	54	F = vaststelling feiten	5 = uitstekend
F	5 4	4 3	4 4	4 3	G = juridische	4 = goed
G	5 4	4 4	3 2	4 3	beoordeling	3 = voldoende
H	5 4	4 4	4 4	4 3	H = motivering en	2 = onvoldoende
I	5 5	4 3	4 3	4 4	beslissing	1 = slecht
J	5 3	4 4	4 4	4 3	I = inkleiding vonnis	
K	5 4	4 4	4 4	4 3	J = algemene aspecten	
L	5 4	4 3	4 3	4 3	K = aanvaardbaarheid	
					L = totaalscore	

6.3.3 Afzonderlijke onderdelen versie 2

In versie 2 komen grotendeels dezelfde onderwerpen aan de orde als in versie 1, maar ze zijn – aan de hand van de ervaringen met het testen van versie 1 – hier en daar meer uitgesplitst of specifieker omschreven.

Bij versie 2 betreffen de afzonderlijke onderdelen de volgende onderwerpen: vaststelling van de feiten die relevant zijn voor de beslissing, juridische beoordeling, aspecten van motivering en beslissing,

inkleding van het vonnis, algemene aspecten en aanvaardbaarheid van het vonnis voor partijen en de maatschappij.

6.3.4 Aandachtspunten bij de afzonderlijke onderdelen

De aandachtspunten die zijn meegewogen in de afzonderlijke onderdelen zijn te lezen in de versies 1 en 2 van het toetsingsformulier (**bijlagen 4 en 5**).

6.3.5 Toelichting bij het toetsingsformulier

Bij elke versie van het toetsingsformulier behoort een toelichting, waarin de toetsing per onderdeel is beschreven. De toelichting was bedoeld om de afspraken van de Commissie over de aard, inhoud en reikwijdte van de toetsing voorafgaand aan de testfase vast te leggen, maar diende ook om in de testfase de uniformiteit te bevorderen bij de toetsing op de afzonderlijke onderdelen (zie paragraaf 5.2).

6.3.6 Scores

De waardering vond plaats met cijfers op een schaal van 5 (uitstekend) tot 1 (slecht) per aandachtspunt, per afzonderlijk onderdeel en voor het vonnis in zijn geheel (de totaalscore). Om vergelijkbaarheid van de toetsingsresultaten mogelijk te maken moesten de waarderingen in cijfers worden uitgedrukt. In paragraaf 5.2 is uiteengezet hoe de totaalscore tot stand kwam.

6.3.7 Opmerkingen op het toetsingsformulier

Op de toetsingsformulieren is bij alle onderdelen ruimte gelaten voor het maken van opmerkingen. De bedoeling daarvan was dat de leden hun in een cijfer uitgedrukt oordeel desgewenst zouden kunnen toelichten of motiveren, ook bijvoorbeeld door verwijzing naar onderdelen van het getoetste vonnis die in het bijzonder grond gaven aan hun oordeel. De bedoeling was ook om te onderzoeken in hoeverre er behoefte was aan het geven van een toelichting. De opmerkingen bleken ook te worden gebruikt om aanvullingen, wijzigingen en verbeteringen van het toetsingsformulier te suggereren (zie paragraaf 6.5).

6.3.8 Rangschikking van zaken in het ‘Overzicht toetsing per zaak per koppel’

De zaken 1 t/m 19 in tabel 1 zijn de zaken waarin de koppelgenoten tot dezelfde uitkomst kwamen bij gebruik van versie 1.

De zaken 20 t/m 27 zijn de zaken waarin dat niet het geval was. De zaken 28 t/m 50 in tabel 2 zijn de zaken waarin de koppelgenoten tot dezelfde uitkomst kwamen bij gebruik van versie 2. In de zaken 51 t/m 54 was dat niet het geval.

De zaken zijn in de beide tabellen van het ‘Overzicht toetsing per zaak’ voor het leesgemak aflopend gerangschikt, dat wil zeggen van een totaalscore van 5 (uitstekend) naar 1 (slecht). De zaaksgegevens zijn geanonimiseerd, onder meer omdat artikel 28 van het Wetboek van Burgerlijke Rechtsvordering aan openbaarmaking in de weg staat.

6.4 Overwegingen bij de bespreking van de testresultaten

6.4.1 Geen toetsing van de juistheid van de uitkomst in een zaak

Met behulp van het toetsingsformulier is getoetst of de beslissingen in een vonnis tot stand zijn gekomen op een wijze die naar de indruk van de Commissie voldoet aan *best practices* (zie hierover ook hoofdstuk 4). Het toetsingsformulier bevat niet de vraag of de uitkomst van de zaak (de beslissing in het eindvonnis) juist was. Dat is een vraag die een procespartij in hoger beroep of cassatie aan de orde heeft kunnen stellen en die in het kader van intercollegiale toetsing weinig zou toevoegen. Over de juistheid van een beslissing kunnen professionals nu eenmaal gemotiveerd van mening verschillen zonder dat de kwaliteit van de zaaksbehandeling in het geding hoeft te komen. Het toetsingsformulier wordt daarentegen juist ontwikkeld om te dienen als een instrument ter bevordering van de kwaliteit van

de zaaksbehandeling. De focus van het toetsingsformulier ligt dan ook op de weg naar de beslissing en niet op de beslissing zelf. Dit betekent tegelijk dat, wanneer de totaalscore of een onderdeel met een onvoldoende is gewaardeerd, daarmee niet is gezegd dat een betere uitkomst van de toetsing tot een andere beslissing in de zaak zou hebben geleid.

6.4.2 Objectief en subjectief

Aan degene die een vonnis in een zaak toetst met behulp van het toetsingsformulier is dus niet gevraagd of hij na kennisneming van de processtukken tot een andere beslissing zou zijn gekomen wanneer hij die zaak zelf zou hebben behandeld en een vonnis zou hebben moeten wijzen. Getoetst is of degene die het vonnis heeft gewezen, de regels en normen heeft toegepast die binnen de beroepsgroep als geldend worden beschouwd voor de totstandkoming van de inhoud van het vonnis.

De toetsing is geschied aan de hand van zo objectief mogelijke maatstaven, die voorafgaand aan de toetsing zijn vastgesteld. De maatstaven houden onder meer in: zijn de voor de beslissing relevante feiten juist en volledig vastgesteld, zijn de toepasselijke regels gebruikt en zijn de algemeen geldende eisen voor de motivering van een beslissing in acht genomen. Daarnaast is in de toetsing de vraag betrokken of het vonnis naar het oordeel van het toetsende Commissielid aanvaardbaar is voor partijen en voor de maatschappij.

De bevindingen van de Commissieleden tijdens de testfase met betrekking tot de uitwerking van de maatstaven in de aandachtspunten per onderdeel in het toetsingsformulier zijn besproken door de Commissie. Aan de hand van de ervaringen met versie 1 is versie 2 ontwikkeld, die is doorontwikkeld tot versie 3.

De Commissie onderkent dat een toetsing zoals deze in enigerlei mate altijd ook subjectieve elementen zal bevatten. Oordeelsvorming is menselijkerwijs niet los te koppelen van eigen opvattingen, ook niet als het gaat over de toepassing van rechtsregels op feiten. De Commissie heeft ernaar gestreefd de toetsing zo veel mogelijk te objectiveren, bijvoorbeeld door te vragen om geanonimiseerde vonnissen, te toetsen in koppels en gebruik te maken van een eenvormige beschrijving van aard, inhoud en reikwijdte van de toetsing (de toelichting) (zie hierover hoofdstuk 5).

6.5 Bespreking van de testresultaten

In deze paragraaf worden aan de hand van het ‘Overzicht toetsing per zaak per koppel’ (tabel 1 en 2) de bevindingen beschreven bij het testen van het toetsingskader met behulp van de 54 zaken die daarvoor zijn gebruikt. De beschrijving vindt plaats in het licht van de vraag of het toetsingskader in beginsel een geschikt instrument is voor de toetsing van de kwaliteit van civiele vonnissen.

6.5.1 Voldoende of niet voldoende

In 42 van de 54 zaken (78%) kwam ieder van de koppelgenoten tot een totaalscore van een voldoende of meer (3, 4 of 5). In 4 van de 54 zaken (7%) waren de koppelgenoten verdeeld: de een gaf een totaalscore van een voldoende of meer, de ander kwam niet tot een voldoende (tabel 1, de zaken 24 tot en met 27). In deze 4 zaken werd gemiddeld (deels afgerond) steeds een voldoende gescoord. Zo bezien hebben de koppels in 46 van 54 zaken (85%) een gemiddelde totaalscore van een voldoende of meer gegeven.

In 8 van de 54 zaken (15%) kwam ieder van de koppelgenoten tot een totaalscore van een onvoldoende (2) (tabel 1, de zaken 15 tot en met 19; tabel 2, de zaken 48, 49 en 50). Het heeft zich niet voorgedaan

dat het koppel verdeeld was en de gemiddelde totaalscore onvoldoende was. In geen van de 54 zaken is de totaalscore 'slecht' (1) gegeven.

6.5.2 Vaker eens dan oneens

Zoals gezegd kwam ieder van de koppelgenoten in 42 van de 54 zaken tot een totaalscore van een voldoende of meer. In 34 van die 42 zaken (81%) kwamen zij tot dezelfde totaalscore. In de overige 8 van de 42 zaken (19%) kwamen zij tot een verschillende nuance in de totaalscore: voldoende, goed of uitstekend (tabel 1 de zaken 20 tot en met 23; tabel 2, de zaken 51 tot en met 54).

Bij de bespreking van de ervaringen in de voltallige Commissie is geconstateerd dat het toetsingsformulier uitnodigt om de eigen gedachten over de inhoud van een vonnis te ordenen. Het overleg werd nuttig gevonden om de eigen bevindingen aan te scherpen, waar te nemen of iets over het hoofd was gezien en de toegekende waardering te spiegelen. Het overleg had niet tot doel dat de koppelgenoten het eens werden (zie paragraaf 5.3). Dat er gevallen waren waarin de koppelgenoten ook na overleg niet tot dezelfde totaalscore kwamen, is mede aanleiding geweest voor de Commissie om aan te bevelen de totaalscore in het vervolg steeds toe te lichten in woorden naast de waardering met een cijfer. Ook is denkbaar dat van de koppelgenoten wordt gevraagd inzichtelijk te maken waarom zij het op bepaalde punten niet eens zijn geworden. Een andere mogelijkheid is dat een derde (bijvoorbeeld de voorzitter) met de koppelgenoten eventuele verschillen in waardering bespreekt. De Commissie verwacht dat het overleg van de koppelgenoten met aanpassingen van deze strekking nog meer zal bijdragen aan geobjectiveerde toetsing.

6.5.3 Versie 1 en versie 2

Omdat de pilot is gericht op de ontwikkeling van het toetsingsformulier, worden enkele uitkomsten van versie 1 en versie 2 van het toetsingsformulier ook afzonderlijk belicht. Daarbij verdient vermelding dat de Commissieleden bij gebruik van versie 1 meer opmerkingen hebben gemaakt over de inrichting van het formulier dan bij gebruik van versie 2.

In 27 zaken is versie 1 gebruikt. In 19 van die 27 zaken (70%) kwamen de koppelgenoten tot dezelfde totaalscore. In 14 van die 19 zaken (74%) ging het om een voldoende of meer, in 5 van de 19 zaken (26%) om een onvoldoende.

In 8 van de 27 zaken (30%) kwamen de koppelgenoten tot een verschillende totaalscore. Dat betrof in 4 van de 8 gevallen (50%) een voldoende of meer. In de andere 4 gevallen (50%) kwam telkens één van de koppelgenoten tot een onvoldoende, maar was het gemiddelde per koppel (deels afgerond) steeds voldoende.

In 27 zaken is versie 2 gebruikt. In 23 van de 27 zaken (85%) kwamen de koppelgenoten tot dezelfde totaalscore. In 20 van die 23 zaken (87%) ging het om een voldoende of meer, in 3 van de 23 zaken (13%) om een onvoldoende.

In 4 van de 27 zaken (15%) kwamen de koppelgenoten tot een verschillende totaalscore. Dat betrof in alle gevallen een voldoende of meer. Anders dan bij versie 1, is het bij versie 2 niet meer voorgekomen dat de een op een voldoende of meer uitkwam en de ander niet tot een voldoende kwam.

Voor beide versies samen kwamen de koppelgenoten tot dezelfde totaalscore in 93% van de zaken (de som van 78% voldoende en 15% onvoldoende, zoals hiervoor in 6.5.1 vermeld). In de zaken waarin versie 1 is gebruikt, is dit percentage slechts 70%. Het is bovendien opvallend dat het bij versie 2,

anders dan bij versie 1, niet meer is voorgekomen dat de ene koppelgenoot tot een voldoende totaalscore kwam en de andere niet. De bevindingen zijn samengevat weergegeven in tabel 3.

Tabel 3 – Samenvatting totaalscores per koppel

	<i>zelfde totaalscore</i> voldoende of meer	<i>zelfde totaalscore</i> onvoldoende	<i>totaalscore verschilt</i> voldoende of meer	<i>totaalscore verschilt</i> 1 onvoldoende 1 voldoende	<i>totaalscore verschilt</i> onvoldoende
versie 1	14 zaken	5 zaken	4 zaken	4 zaken	n.v.t.
versie 2	20 zaken	3 zaken	4 zaken	n.v.t.	n.v.t.

De vergelijking van de bevindingen van de koppelgenoten met gebruik van versie 1 en versie 2 laat zien dat de koppels (a) het bij gebruik van versie 2 veel vaker eens waren over de totaalscore dan bij gebruik van versie 1 en (b) bij gebruik van versie 2 geen verschil van inzicht meer hebben gehad op de grens van wel en niet voldoende. De Commissie heeft, gezien deze uitkomsten en de bespreking van de ervaringen in de voltallige Commissie, de indruk dat versie 2 betere ondersteuning biedt bij consistente, geobjectiveerde oordeelsvorming dan versie 1, zeker wanneer bij het gebruik de toelichting ernaast wordt gehouden. Versie 2 specificeert meer en in een nog logischer volgorde waarmee zoal rekening kan worden gehouden bij het bestuderen van het vonnis en de processtukken in een zaak. Versie 2 biedt mogelijk meer kans op dezelfde uitkomsten in de oordeelsvorming van de koppelgenoten. De uitwisseling van dergelijke ervaringen in de Commissie is mede aanleiding geweest versie 2 tot slot hier en daar nog bij te stellen en tot versie 3 te komen.

6.5.4 Totaalscore in relatie tot afzonderlijke onderdelen

Zoals gezegd was in de Commissie vooraf afgesproken dat ieder de totaalscore zou geven na een zaaksgebonden eigen afweging van het soortelijk gewicht van de scores van de afzonderlijke onderdelen van het toetsingsformulier (zie paragraaf 5.2). De totaalscore kwam dus niet tot stand door een optelsom van de scores van de afzonderlijke onderdelen gedeeld door het totaal aantal onderdelen, maar door reflectie op en prioritering van de bevindingen. Op zichzelf is dan ook denkbaar dat een onderdeel in de ene zaak de doorslag heeft gegeven in de totaalscore, maar in een andere zaak van ondergeschikte betekenis is geacht in de totaalscore. In de praktijk zijn uit de bevindingen van de koppels echter wel enkele opvallende punten af te leiden.

Allereerst is ten aanzien van de 46 van 54 zaken (85%) met een gemiddelde totaalscore van een voldoende of meer, de juridische beoordeling in 45 gevallen (98%) door ieder van de koppelgenoten op een voldoende of meer gewaardeerd en in 1 geval (2%) gemiddeld op een voldoende. Verder is in 39 van die 46 zaken met een totaalscore van een voldoende of meer (85%) het onderdeel motivering en beslissing gescoord op een voldoende of meer. In 4 van die 46 zaken (8%) kwam de een tot een voldoende en de ander tot een onvoldoende voor het onderdeel motivering, maar was het gemiddelde steeds voldoende. In 3 van de 46 zaken met een totaalscore van een voldoende of meer (7%) is de motivering onvoldoende geoordeeld. Een en ander kan worden gezien als een aanwijzing dat de bevindingen met betrekking tot de juridische beoordeling en in iets mindere mate met betrekking tot de motivering van de beslissing van groot gewicht zijn geacht voor de totaalscore van een voldoende of meer.

Omgekeerd laat zich dat laatste echter niet zeggen voor de zaken waarin de totaalscore onvoldoende was. Deze zaken laten namelijk een zeer wisselend beeld zien wat betreft het gewicht van het oordeel op de afzonderlijke onderdelen. In twee zaken hebben de feitenvaststelling, de juridische beoordeling en de motivering bij beide koppelgenoten een onvoldoende gescoord. In een derde zaak is de juridische beoordeling en de motivering met een onvoldoende gewaardeerd. In een vierde zaak heeft het koppel

gemiddeld alleen voor de feitenvaststelling een onvoldoende gegeven. In de vijfde zaak werd alleen de motivering met een onvoldoende gewaardeerd. In een zesde zaak zijn de motivering en de aanvaardbaarheid onvoldoende geacht door het koppel. In een zevende zaak is alleen de juridische beoordeling met een onvoldoende gewaardeerd. In een achtste zaak zijn de juridische beoordeling en de inkleding van het vonnis voldoende gescoord en de overige onderdelen onvoldoende.

In de zaken die een onvoldoende totaalscore hebben gekregen, is dus geen lijn te ontdekken in de rol die het soortelijk gewicht van de onderdelen heeft gespeeld. Het testen van het toetsingsformulier heeft aldus geen aanknopingspunt opgeleverd om te veronderstellen dat vonnissen met een onvoldoende totaalscore een structurele tekortkoming op een bepaald onderdeel zouden vertonen.

Hieruit vloeien wel aandachtspunten voort met betrekking tot de methode van het onderzoek en de bruikbaarheid van het toetsingsformulier. Het is immers niet bekend waardoor het wisselende beeld is veroorzaakt. Houdt het wisselende beeld (onder meer) verband met de onderzoeksopzet – die gericht was op het testen van het toetsingsformulier en niet op toetsing van de kwaliteit van civiele vonnissen in een sociaal wetenschappelijk verantwoorde steekproef – dan kan een toekomstig gebruik van het formulier in een andere onderzoeksopzet uitwijzen of het beeld stabiliseert in toekomstig te beoordelen zaken. In die andere onderzoeksopzet verdient in ieder geval ook de methode hernieuwde aandacht. Ieder oordeel over een vonnis is in zekere mate subjectief. Het is van belang zo goed mogelijk inzichtelijk te maken in hoeverre de toetsing tot geobjectiveerde uitkomsten leidt.

6.5.5 Het gebruik van opmerkingen op het toetsingsformulier

Door de leden is een verschillend gebruik gemaakt van de mogelijkheid om een toelichting te geven op een cijfer. Sommigen hebben in hoofdzaak volstaan met het toekennen van cijfermatige waarderungen, terwijl anderen min of meer uitvoerige beschouwingen hebben gewijd aan de onderbouwing van hun waardering. Deze beschouwingen zijn veelal bijzonder waardevol, omdat zij de achterliggende gedachte van de toetsing waarneembaar en dus ook controleerbaar maken. Dit wil intussen niet zeggen dat, ingeval niet gebruik is gemaakt van de mogelijkheid opmerkingen te maken, de toetsing minder gewicht heeft. Met kennis van het dossier is in de meeste gevallen ook zonder nadere toelichting in te zien waarom een bepaalde waardering is gegeven. Het verschillend gebruik van de mogelijkheid om een toelichting te geven, maakt wel dat het lastiger is om verklaringen te vinden voor verschillen in toetsingsresultaten en daarmee om de mate van objectiviteit van de toetsing inzichtelijk te maken. Ook het tijdsbeslag van de toetsing kan verschillen naar gelang er meer of minder toelichting wordt gegeven.

Een en ander afwegend meent de Commissie dat bij een praktische toepassing van deze toetsingsmethode het maken van opmerkingen mogelijk moet worden gemaakt. De Commissie vindt dat opmerkingen bij de totaalscore niet kunnen worden gemist, maar ziet geen aanleiding voor een verplichting tot het maken van opmerkingen bij de afzonderlijke onderdelen en aandachtspunten. Het wordt aan de professionaliteit van de gebruiker van het formulier gelaten om te beoordelen waar een "score" van een motivering of toelichting zou moeten worden voorzien. Bij het eindoordeel over het vonnis wordt dus wel steeds een toelichting verlangd.

6.5.6 De schaal van 1 tot 5

Voor zover de koppelgenoten niet tot dezelfde totaalscore zijn gekomen, liggen de verschillen grotendeels binnen enge grenzen. Bij versie 1 was er in één geval een verschil van 2 (4-2) en in zeven gevallen een verschil van 1 (5-4, 4-3 en 3-2). Bij versie 2 waren er vier gevallen met een verschil van 1 (5-4 en 4-3). Opvallend is verder dat het cijfer 5 (uitstekend) in de totaalscore relatief weinig is gebruikt en het cijfer 1 (slecht) geheel niet. Het is kennelijk moeilijk, ook met behulp van het toetsingskader, om te bepalen wanneer een vonnis nu echt als uitstekend of juist als slecht moet worden beschouwd.

Bovendien laat een schaal van 1 tot 5 de nodige ruimte om van mening te verschillen over het gewicht van een afzonderlijk onderdeel in de totaalscore. Een en ander roept de vraag op of de schaal ook kleiner zou kunnen zijn, bijvoorbeeld een schaal van voldoende en onvoldoende, of een schaal van onvoldoende, voldoende en goed. Naar verwachting zal echter elke schaal (1-5, 1-2, 1-3 of bijvoorbeeld 1-10) voor- en nadelen hebben. De Commissie beveelt aan de ervaringen in dit onderzoek te betrekken in de keuze van de schaal in een eventueel vervolgonderzoek.

6.5.7 Korte conclusie

Bij het testen van het toetsingskader met behulp van de 54 zaken is gebleken dat het mogelijk is een werkbaar toetsingskader te ontwikkelen waarvan de inhoud wordt gedragen door de Commissieleden. Met dit kader kan een beeld van de kwaliteit van een vonnis worden verkregen. Het betreft zowel een totaalbeeld als een beeld op onderdelen die de kwaliteit van een civiel vonnis in belangrijke mate bepalen. Hoewel de koppelgenoten in aanzienlijke mate tot dezelfde uitkomsten zijn gekomen, kunnen de bevindingen desondanks nog uiteen lopen.

6.5.8 Vervolgonderzoek?

Dat de bevindingen toch nog uiteen kunnen lopen, roept de vraag op wat de oorzaken van de gevonden verschillen zijn. Komt het door de ontwikkelde methode, door opvattingen van de leden van de Commissie, door het aangeleverde onderzoeksmateriaal, doordat professionals nu eenmaal gemotiveerd van mening kunnen verschillen, en dergelijke? De Commissie heeft getracht bij de bespreking en beschouwing van haar bevindingen tevens zo goed mogelijk eventuele aanknopingspunten te belichten voor het zoeken naar verklaringen voor de gevonden verschillen. De Commissie kan deze kwestie echter niet nader onderzoeken en de onderzoeksopzet leent zich daarvoor ook niet. Het doel van het onderzoek was immers het toetsingskader te ontwikkelen en te testen. Zoals gezegd is de verzameling getoetste vonnissen in dit onderzoek te gering om uit de testresultaten iets in algemene zin te kunnen afleiden over de kwaliteit van civiele vonnissen in handelszaken met verweer.

Wanneer het vinden van verklaringen voor verschillen in de bevindingen van groot belang wordt geacht, zal een vervolgonderzoek nodig zijn. De Commissie acht het van belang dat dan in de onderzoeksopzet ook wordt betrokken hoeveel zaken moeten worden onderzocht om, anders dan in dit onderzoek is beoogd en mogelijk was, wetenschappelijk verantwoord in algemene zin conclusies te kunnen trekken over de kwaliteit van civiele vonnissen van de betrokken rechtbanken in handelszaken met verweer. De Commissie verwacht dat bij de vorming van die onderzoeksopzet bovendien niet alleen het aantal te onderzoeken zaken van belang zal zijn, maar bijvoorbeeld ook hypothesevorming over mogelijke oorzaken van de gevonden verschillen, de onderzoekbaarheid daarvan, of de vraagstelling en de methoden van het vervolgonderzoek. De Commissieleden zijn graag bereid om hierover mee te denken indien daaraan behoefte bestaat.

Tot slot vestigt de Commissie er de aandacht op dat een bespreking van de resultaten met de individuele rechter (nog) geen deel uitmaakte van haar onderzoek. Dat valt wel te overwegen in een vervolgonderzoek, wanneer een statistisch verantwoorde steekproef wordt gedaan. Een bespreking van de resultaten met de betrokken rechter zal in sommige gevallen mogelijk een ander licht werpen op gevonden verschillen, al was het maar omdat een derde professional zich dan mengt in het debat over de kwaliteit van het getoetste vonnis.

7 Het toetsingskader getoetst door drie hoven

Naast het onderzoek door de Commissie hebben de hoven te Amsterdam, Arnhem en 's-Gravenhage de formulieren toegepast op zaken van de betrokken rechtbanken, die zij in hoger beroep hebben

behandeld. Het gaat daarbij niet om dezelfde vonnissen die werden beoordeeld door de Commissie. Het onderzoek van de hoven heeft niet plaatsgevonden onder verantwoordelijkheid van de Commissie (zie hoofdstuk 2).

De hoven hebben gebruik gemaakt van de door de Commissie ontwikkelde toetsingsformulieren (versies 1 en 2). Er heeft verder – hoewel dit door de voorzitter van de Commissie wel is aangeboden – geen overleg plaatsgevonden tussen de Commissie en de beoordelaars van de hoven. De Commissie heeft geen inzicht in de wijze waarop vonnissen in hoger beroep zijn geselecteerd.

De hoven hebben in totaal 12 vonnissen onderworpen aan toetsing aan de hand van versie 1 en 22 vonnissen aan de hand van versie 2 van het toetsingsformulier. Onder de getoetste vonnissen was één kantonzaak, vier kort gedingen en één in eerste aanleg meervoudig afgedane zaak.

De vraag aan de hoven was of het toetsingskader zich leent voor toepassing door de hoven. Zij hebben hun testresultaten aangeleverd aan de Commissie en de vraag bevestigend beantwoord. De hoven hebben per vonnis één toetsingsformulier ingevuld. Dit is soms door meerdere of alle leden van de kamer gedaan, maar soms ook door één van de leden van de kamer. In dat laatste geval is niet duidelijk of er overleg heeft plaatsgevonden over de toetsing. De hoven hebben niet per raadsheer een toetsingsformulier ingevuld. De Commissie heeft verder geen inzicht in de wijze waarop de hoven tot de bevinding zijn gekomen dat het toetsingskader zich leent voor toepassing door de hoven.

De hoven hebben eveneens gebruik gemaakt van de mogelijkheid om opmerkingen bij de toetsingsresultaten te maken. Hieruit blijkt dat degenen die de toetsing bij de hoven hebben verricht, de beoordeling van de kwaliteit van een vonnis duidelijk plaatsen in het kader van het hoger beroep daartegen, de argumenten die partijen tijdens het hoger beroep aanvoeren en hoe het hof de zaak uiteindelijk zelf heeft beoordeeld. Dit is een ander perspectief dan is gehanteerd bij de toetsing door de Commissieleden.

7.1 Korte conclusie

De indruk bestaat dat de hoven, ofschoon zij niet betrokken zijn geweest bij de ontwikkeling van de toetsingsformulieren, er eveneens in geslaagd zijn de formulieren adequaat te gebruiken en daarmee inzicht te geven in de kwaliteit van de vonnissen. Er zijn verder geen opvallende verschillen gesignaleerd in de wijze waarop de hoven de formulieren hebben gebruikt en de wijze waarop de Commissie dat heeft gedaan. Uit de opmerkingen die de hoven op de toetsingsformulieren hebben gemaakt, kan wel worden opgemaakt dat zij een ander referentiekader hebben. De hoven hebben immers behalve het dossier, ook kennis kunnen nemen van de stellingen die partijen in hoger beroep hebben ingenomen en hun eventuele onvrede met het vonnis in eerste aanleg. De Commissie heeft niet kunnen vaststellen dat dit de toetsingsresultaten van de hoven heeft gekleurd.

8 Bruikbaarheid van het toetsingskader

8.1 Mogelijke toepassingen

8.1.1 Inleiding

Het toetsingskader is een kwaliteitsinstrument en kan als zodanig op vele manieren worden gebruikt. Op de mogelijke toepassingen wordt hierna nader ingegaan. De Commissie beveelt aan bij de keuze voor toepassingen tot uitgangspunten te nemen dat het leereffect van het toetsingskader optimaal wordt benut

en dat het toetsingskader dient bij te dragen aan de accumulatie van kennis en ervaring rondom het civiele vonnis.

8.1.2 Toepassing in het kader van visitatie

In deze paragraaf wordt nader ingegaan op de toepassing van het toetsingskader in het kader van visitatie. Die toepassing ligt voor de hand omdat de wens tot inhoudelijke toetsing van de kwaliteit van rechterlijke uitspraken vorm heeft gekregen in het voorstel van de Projectgroep Visitatieprotocol 2009, dat heeft geleid tot de instelling van de Commissie en de aan haar verleende opdracht.

De ervaringen van de Commissie met de toetsing van een beperkt aantal (54) vonnissen van drie rechtbanken in het kader van de ontwikkeling van het toetsingskader hebben haar tot de conclusie geleid dat het ontwikkelde toetsingsformulier (versie 3) met de daarbij behorende toelichting een bruikbaar en geschikt instrument lijkt te zijn voor het toetsen van de kwaliteit van een rechtbankvonnis in een handelszaak met verweer. Het beperkte aantal getoetste vonnissen geeft wel aanleiding tot het maken van een voorbehoud, zoals is toegelicht aan het slot van paragraaf 6.5.4.

Een ander voorbehoud betreft de kwestie of het instrument bruikbaar is in het kader van visitatie. In overleg met de Stuurgroep werd gesignaleerd dat een van de uitkomsten van de pilot zou kunnen zijn dat de toetsing van een vonnis te tijdrovend is en daardoor (in de ontwikkelde vorm) niet bruikbaar is voor visitatie. Is het wel binnen redelijke tijd mogelijk een beeld te krijgen over de kwaliteit van een vonnis, dan zou een soortgelijke groep de toetsing moeten kunnen uitvoeren. Hierop wordt thans ingegaan.

Ofschoon de Commissie, als gezegd, van oordeel is dat met het toetsingskader een redelijk betrouwbaar oordeel over de kwaliteit van een vonnis kan worden gegeven, kan een praktisch bezwaar zijn dat de toetsing een aanzienlijk tijdsbeslag meebrengt. Toetsing van een vonnis met het dossier door een koppel van twee beoordelaars heeft in dit onderzoek naar schatting gemiddeld ongeveer 13 uur per zaak gekost. Daarbij moet wel worden bedacht dat bij de keuze voor toetsing in koppels heeft meegespeeld dat het ging om de ontwikkeling van het toetsingskader. Naast de toetsing in een zaak hielden de koppelgenoten zich bezig met de beoordeling van het toetsingskader. Niettemin zal bij een toekomstig representatief onderzoek het tijdsbeslag niet gering zijn. Volledigheidshalve wordt hierbij opgemerkt dat het tijdsbeslag ook anders kan uitvallen indien het onderzoek tevens vonnissen in familiezaken en vonnissen van de sector kanton zou omvatten.

In dit verband rijst nog een organisatorische vraag, namelijk of een vaste toetsingscommissie die bij visitatie kan worden ingezet als een subcommissie van de Visitatiecommissie, dan wel een ad hoc-commissie voor elke visitatie wordt ingesteld. In het laatste geval zal rekening moeten worden gehouden met de noodzaak voor de leden van de commissie zich de methode van toetsing eigen te maken en met aanloopproblemen. Overigens zou deze aanpak ook van invloed kunnen zijn op de mate waarin de resultaten van verschillende toetsingen onderling vergelijkbaar zijn. Een tussenoplossing is een commissie waarin de leden gedurende een bepaald aantal jaren zitting hebben met een schema van aftreden waarin de continuïteit is gewaarborgd.

De Commissie heeft onder ogen gezien of, ingeval het hiervoor geschatte tijdsbeslag te bezwaarlijk wordt geacht, varianten denkbaar zijn in de inrichting van de toetsing. De varianten die de Commissie onder ogen heeft gezien, zijn de volgende:

1. Overwogen zou kunnen worden uit te gaan van een nog kleiner aantal te toetsen vonnissen. Het gebruik van het toetsingskader in een visitatie zou in de opvatting van de Commissie echter

- moeten plaatsvinden in de vorm van een wetenschappelijk verantwoorde steekproef. Naar het oordeel van de Commissie verdient deze variant dan ook geen aanbeveling.
2. Ongeveer een halvering van de voor toetsing benodigde tijd zou kunnen worden bereikt door de toetsing per vonnis niet door twee, maar door één persoon te doen geschieden. De Commissie is hiervan geen voorstander. Zelfstandige toetsing door twee personen, gevolgd door overleg tussen beiden, draagt volgens de Commissie bij aan consistentie in en objectivering van de toetsing. Aan toetsing door één persoon zijn naar het oordeel van de Commissie in wezen dezelfde bezwaren verbonden als aan unus-rechtspraak zonder meeleezer. Voorts verdient het aanbeveling juist bij de toetsing van vonnissen die zijn geweest door een alleen rechtsprekende rechter, te vermijden dat slechts één ander daarover een oordeel geeft.
 3. Volledigheidshalve merkt de Commissie op dat een aanzienlijke tijdsbesparing zou kunnen worden bereikt, indien de toetsing zou geschieden uitsluitend aan de hand van het vonnis, dus zonder het dossier. Het behoeft geen betoog dat een zodanige toetsing niet goed mogelijk is en dus weinig waarde heeft. Hiermee wil overigens niet gezegd zijn dat de omvang van het dossier niet van invloed kan zijn op de vraag welke vonnissen worden betrokken in de toetsing.

De Commissie is al met al van oordeel dat geen van deze varianten voldoet en zij worden dus niet aanbevolen. Naar het oordeel van de Commissie moet worden aanvaard dat, wil men komen tot een verantwoorde toetsing van de kwaliteit van vonnissen, daarmee een aanzienlijke investering van tijd is gemoeid.

8.1.3 Toetsing door hoven

De Commissie heeft nog de vraag onder ogen gezien of het zinvol zou zijn, op het voetspoor van de toetsing door de hoven in het onderhavige onderzoek, de hoven te verzoeken de toetsing in het kader van visitatie met toepassing van het door de Commissie ontwikkelde toetsingskader op zich te nemen. Het voordeel zou zijn dat dan geen extra tijd besteed hoeft te worden aan de bestudering van het dossier, omdat het hof het vonnis toch al beoordeelt en dat de kwaliteitstoetsing in het verlengde ligt van die beoordeling.

Niettemin zijn er, nog afgezien van de bereidheid van de hoven een extra taak op zich te nemen, aan deze oplossing bezwaren verbonden. De beslissing van een partij in hoger beroep te gaan, kan op een veelheid van factoren berusten, die niet noodzakelijkerwijze tot eenzelfde verzameling vonnissen leiden als een aselechte steekproef. Voorts kan een kwaliteitsbeoordeling door een hof wel in het verlengde liggen van een beslissing op het hoger beroep, maar deze beoordelingen zijn, zoals blijkt uit hoofdstuk 7 hiervoor, niet identiek. Bepaald niet uitgesloten is dat het oordeel van het hof over de kwaliteit wordt gekleurd door zijn beslissing op het appel. Ook kan per definitie geen sprake zijn van anonimiteit van de rechter die het beroepen vonnis heeft gewezen.

8.1.4 Andere toepassingsmogelijkheden

Denkbaar is dat het ontwikkelde toetsingskader op zichzelf een uitstraling krijgt naar de beroepsgroep en op die wijze bijdraagt aan de verbetering van de kwaliteit van vonnissen. In concreto zou dit bijvoorbeeld ook vorm kunnen krijgen doordat het door de Commissie ontwikkelde toetsingskader een functie vervult als "checklist" bij het meelesen van een vonnis. Ook kan worden gedacht aan gebruik van het toetsingskader tijdens de opleiding tot rechter of secretaris, aan opname van het toetsingskader in het Handboek civiel of aan gebruik als ijkpunt tijdens reflectiebijeenkomsten van rechtbanken en hoven, respectievelijke hoven en Hoge Raad. Verspreiding van het toetsingskader in de praktijk zal wellicht nog andere toepassingsmogelijkheden aan het licht brengen.

8.1.5 Toetsing van arresten van hoven

De onderhavige pilot heeft uitsluitend betrekking op toetsing van rechtbankvonnissen. Indien ook een toetsing van de kwaliteit van hofarresten zou worden overwogen, ligt het naar het oordeel van de Commissie voor de hand daarvoor een overeenkomstige opzet als voor de toetsing van rechtbankvonnissen te kiezen, zij het uiteraard met een nog te ontwikkelen op de appelrechtspraak toegespitst toetsingskader.

8.2 Organisatorische aspecten

8.2.1 Samenstelling onderzoekcommissie

De samenstelling van de Commissie werd voorafgaand aan de pilot vastgelegd (zie hoofdstuk 3). De Commissie heeft zich de vraag gesteld hoe de samenstelling van een toetsingscommissie er in de toekomst uit zou moeten zien. De conclusie van de Commissie is dat de toetsing ook in de toekomst zou moeten plaatsvinden door gezaghebbende juristen. Het lijkt niet zonder meer noodzakelijk dat deze juristen in hun dagelijks werk direct betrokken zijn bij de totstandkoming van civiele vonnissen.

In de toelichting bij het toetsingskader is al gesignaleerd dat niet elk vonnis zonder meer voor een leek begrijpelijk is. Mede hierin kan een aanleiding worden gezien leden uit de advocatuur bij de toetsing te betrekken. Zij zijn niet alleen jurist, maar kunnen daarnaast inschatten of een vonnis valt uit te leggen aan de betrokken partijen. De Commissie heeft verder gemerkt dat de leden uit de advocatuur een genuanceerde kijk hebben op de processtukken en de invloed daarvan op het verloop van de procedure. De Commissie beveelt aan ook in het vervolg leden uit de advocatuur bij de toetsing te betrekken.

Het is verder nuttig gebleken om bij de toetsing ook leden uit de rechtbanken te betrekken. Zij hebben recente ervaring met de vonnispraktijk en bijvoorbeeld ook zicht op de gang van zaken tijdens de comparitie. De gang van zaken ter comparitie heeft de Commissie in dit onderzoek slechts kunnen vaststellen aan de hand van het proces-verbaal. De invloed van de comparitie op de oordeelsvorming van de rechter kon niet worden getoetst. Denkbaar is dat niet in het proces-verbaal gerelateerde zaken, zoals kleuring van feiten en uitleg van stellingen, van betekenis zijn geweest voor het maken van het vonnis. De invloed hiervan op de toetsingsbevindingen kon niet worden achterhaald. Daarmee wordt het belangrijker dat voldoende leden in de Commissie ervaring hebben met het verloop van comparities na antwoord.

De Commissie beveelt aan ook in de toekomst meerdere leden uit de rechtbanken te laten deelnemen aan de beoordelingen. De Commissie heeft verder gemerkt dat de inbreng van ervaringen met de praktijk in hoger beroep en cassatie heeft bijgedragen aan de ontwikkeling van het toetsingskader. De Commissie kan zich voorstellen dat dergelijke ervaringen ook bij toekomstig gebruik van het toetsingskader waardevol kunnen zijn.

8.2.2 Selectie van de te toetsen vonnissen

Voor een doelmatige en verantwoorde toetsing is van belang dat deze betrekking heeft op een aselechte en voldoende omvangrijke steekproef uit de door de betrokken rechtbank gewezen vonnissen, anders dan in het kader van deze pilot het geval was. In het kader van een visitatie zou die selectie bijvoorbeeld als volgt kunnen worden georganiseerd.

Aan de te visiteren rechtbank wordt tijdig voor de visitatie verzocht alle (enkelvoudig gewezen) vonnissen uit een bepaalde periode, bijvoorbeeld twee maanden, tezamen met de dossiers en de producties afzonderlijk op te slaan. Uit deze verzameling wordt een blinde selectie tot het voor de toetsing vereiste aantal gemaakt aan de hand van de rolnummers (bijvoorbeeld de rolnummers

eindigend op 0 en 5). De in het dossier voorkomende namen van de rechter(s) worden onleesbaar gemaakt en vervangen door een letteraanduiding. De toetsingscommissie kan vonnissen in bagatelzaken en vonnissen die zij om andere redenen voor de beoordeling ongeschikt acht - bijvoorbeeld oververtegenwoordiging van een bepaalde rechter of de uitzonderlijke omvang van het dossier - uit de verzameling verwijderen en deze doen vervangen door andere op dezelfde wijze geselecteerde vonnissen.

8.2.3 Werkwijze

De Commissie is van mening dat het bij de toetsing essentieel is om kennis te kunnen nemen van het volledige rechtbankdossier, inclusief de producties. Zonder die producties is een goede toetsing van de kwaliteit van een vonnis – met name de beslissingen omtrent de bewijslast – niet mogelijk. Wanneer de toetsing van de kwaliteit van vonnissen in de toekomst regelmatig zal gaan plaatsvinden, is het van belang dat bij de dossiers die moeten worden beoordeeld de producties beschikbaar blijven.

De tijd die benodigd is voor de toetsing van een vonnis is aanzienlijk. De Commissie heeft geen mogelijkheden gezien om, zonder afbreuk aan de betrouwbaarheid van de resultaten te doen, in kortere tijd vonnissen te toetsen. In paragraaf 8.1.2 zijn verschillende andere methoden bekeken, maar geen van deze methoden wordt door de Commissie aanbevolen. De conclusie is dat een verantwoorde toetsing van de kwaliteit van vonnissen een aanzienlijke investering van tijd vergt.

Uit de in hoofdstuk 6.5 gepresenteerde resultaten blijkt dat de koppels het meestal eens waren over de totaalscore. Dit zou een argument kunnen zijn om bij toekomstige toetsingen niet meer in koppels te werken. Daar staat tegenover dat discussie tussen de koppels nuttig lijkt en ook tot afwijkende toetsingsresultaten kan leiden. De Commissie heeft geen alternatieven voor de toetsing door twee personen onderzocht en kan over de effecten van alternatieven niet adviseren. De Commissie beveelt aan ook in de toekomst steeds twee personen naar een vonnis te laten kijken.

In hoofdstuk 8 is weergegeven dat de toetsing onderdeel zou kunnen worden van de visitatie. De Commissie geeft de voorkeur aan toetsing door een vaste commissie in plaats van een commissie met een steeds wisselende samenstelling (zie paragraaf 8.1.2), waarbij gedacht kan worden aan benoemingen voor een bepaald aantal jaren met de mogelijkheid van een verlenging en een schema van aftreden dat de continuïteit waarborgt.

8.2.4 Rol van voorzitter en secretaris

Zoals hiervoor in paragraaf 5.3 is opgemerkt, hebben de voorzitter en de secretaris het niet tot hun taak gerekend de toetsing door de leden van de Commissie aan een herbeoordeling te onderwerpen. Zij hebben vooral een coördinerende en faciliterende rol gespeeld die onder meer bestond uit het beantwoorden van bij de leden gerezen vragen over bijvoorbeeld de wijze van invulling van de formulieren. Zij hebben ook de concepten opgesteld voor de door de Commissie vastgestelde toetsingsformulieren en de toelichtingen daarop.

Achteraf bezien kan de vraag rijzen of het niet de voorkeur zou hebben verdiend dat ook de voorzitter en de secretaris in de koppels zouden zijn ingedeeld en aldus zouden hebben deelgenomen aan de daadwerkelijke toetsing van de vonnissen. Het voordeel hiervan is dat de voorzitter en secretaris zelf ervaringen hadden kunnen opdoen met het door hen opgestelde toetsingskader. Dit is een aspect om mee te wegen bij het bepalen van een eventueel vervolg.

8.2.5 Nadere onderzoeksvragen

De Commissie heeft gemerkt dat de indeling van het formulier van belangrijke invloed is op een efficiënte beoordeling. De gebruikers van de eerste versie van het formulier merkten op dat niet altijd duidelijk is waar zij een bepaald kritiekpunt moeten onderbrengen. Een verbetering van het formulier was gelegen in een gewijzigde vorm. Niet alleen de Commissie, maar ook de hoven hebben opgemerkt dat de tweede versie gebruiksvriendelijker is. De Commissie meent dat de derde versie een vorm heeft die een efficiënte toetsing mogelijk maakt. Dat sluit niet uit dat de vorm van het formulier verder zou kunnen worden verbeterd door een terzake deskundige. Denkbaar is bijvoorbeeld dat een geheel digitaal formulier direct kan worden gekoppeld aan een overzichtelijk rapportagemodel, zodat de resultaten van een toetsing sneller kunnen worden verwerkt.

Het aantal vonnissen dat in de pilot is onderzocht, is te beperkt geweest om verantwoorde conclusies te kunnen trekken over de kwaliteit van de betrokken rechtbanken en rechters. Voor een goede toetsing van een vonnis is veel tijd nodig. De voorzichtige inschatting (zie paragraaf 5.3) van de Commissie is dat voor de toetsing van één vonnis door twee beoordelaars ten minste 13 uur nodig is. Hierin is echter ook tijd begrepen die is besteed aan de ontwikkeling van het toetsingskader.

Tijdens de pilot is niet vastgesteld wat de omvang van de steekproef ten minste zou moeten zijn om wel uitspraken over de kwaliteit van de vonnissen van een rechtbank te kunnen doen. Dit zou onderdeel moeten worden van een eventueel vervolgonderzoek.

Uit de resultaten lijkt te volgen dat wanneer een vonnis niet voldoende werd gescoord, dit komt door een onvoldoende vaststelling van de feiten, de juridische onderbouwing of motivering van de beslissing. De inkleding van het vonnis wordt dan minder belangrijk. De invloed van de feiten op het kwaliteitsoordeel lijkt logisch, wanneer wordt bedacht dat de feiten steeds de basis van het vonnis vormen. Het zal niet eenvoudig zijn in een vervolgonderzoek de vaststelling van de feiten meer aandacht te geven, want daarvoor zal bijvoorbeeld met name de gang van zaken tijdens de comparitie van belang zijn. Het vaststellen van de betekenis van de comparitie voor het eindvonnis in dezelfde zaak vergt andersoortig onderzoek dan onderzoek gericht op de toetsing van vonnissen. De Commissie acht het echter wel van belang dat meer zicht ontstaat op bijvoorbeeld de volgende aspecten van de comparitie:

- Hebben partijen voldoende de gelegenheid gehad om tijdens de comparitie op elkaars stellingen te reageren? Wat is de rol van de advocaten hierin?
- Zijn er tijdens de comparitie nadere feiten gesteld of stellingen verder onderbouwd?
- Is er een juist en volledig proces-verbaal opgesteld?
- Zijn aan partijen de juiste vragen gesteld? Hebben zij adequate antwoorden gegeven?

Deze vragen konden in dit onderzoek niet of nauwelijks worden beantwoord, omdat - uiteraard - niet bekend is wat er feitelijk op de comparitie is gebeurd en wat daarvan in het proces-verbaal is neergelegd. Binnen de huidige opzet van het onderzoek kan dan ook niet worden beoordeeld in hoeverre de comparitie van invloed is geweest op het eindoordeel over een vonnis.

9 Conclusies en aanbevelingen

De Commissie heeft op grond van haar onderzoek de volgende conclusies en aanbevelingen geformuleerd. Daarbij wordt vooropgesteld dat het onderzoek gericht is geweest op het ontwikkelen van een methode om de kwaliteit van civiele vonnissen te toetsen en niet op het beoordelen van de kwaliteit van de onderzochte vonnissen.

1. Het is mogelijk gebleken een werkbaar toetsingskader te ontwikkelen waarvan de inhoud wordt gedragen door de Commissieleden. Met dit kader kan een beeld van de kwaliteit van een civiel vonnis in een handelszaak met verweer worden verkregen. Het betreft zowel een totaalbeeld als een beeld op onderdelen die de kwaliteit van een civiel vonnis in belangrijke mate bepalen.
2. De Commissie beveelt aan bij toekomstige toepassingen van het toetsingskader gebruik te maken van versie 3 van het toetsingsformulier en de toelichting (**bijlage 6**).
3. De toetsing van vonnissen door de hoven leert dat de door de Commissie ontworpen formulieren ook door anderen dan de leden van de Commissie adequaat kunnen worden gehanteerd.
4. Bij toekomstige toepassing van het toetsingskader in een visitatie gelden in ieder geval de volgende aandachtspunten:
 - a. In de onderzoeksopzet dient onder meer te worden betrokken hoeveel vonnissen moeten worden getoetst om wetenschappelijk verantwoorde uitspraken te kunnen doen over de kwaliteit van de vonnissen die worden betrokken in een onderzoek in het kader van visitatie.
 - b. De toetsing van vonnissen op de wijze die de Commissie heeft gehanteerd, vergt een aanzienlijke investering in tijd. De Commissie acht dit gelet op de gewenste kwaliteit van de toetsing onvermijdelijk. De Commissie ontraadt varianten op haar werkwijze die zijn ingegeven door de wens tijd te besparen en afbreuk doen aan de kwaliteit van de toetsing.
 - c. Voor het toetsen van de kwaliteit van vonnissen in het kader van een visitatie wordt bij voorkeur een vaste subcommissie van de Visitatiecommissie ingesteld. Deze commissie – hierna: de Subcommissie – dient te bestaan uit gezaghebbende juristen, onder wie in elk geval een aantal (oud-)rechters en een of meer (oud-)advocaten.
5. Bij het hanteren van de door de Commissie ontwikkelde methode in het kader van een visitatie dienen verder de volgende uitgangspunten in acht te worden genomen:
 - a. Door de betrokken rechtbank wordt een aselechte steekproef gedaan van een - volgens ter zake deskundig oordeel voldoende - aantal eindvonnissen. De steekproef omvat in ieder geval enkelvoudig gewezen vonnissen. Wanneer ook meervoudig gewezen vonnissen in de steekproef worden betrokken, verdient nader onderzoek of de verhouding tussen enkelvoudig en meervoudig gewezen vonnissen in de steekproef met het oog op de validiteit van de uitkomsten dient te stroken met de werkelijke verhouding tussen enkelvoudig en meervoudig gewezen vonnissen in de onderzochte periode in de betrokken rechtbank.
 - b. De dossiers van deze vonnissen bevatten tevens de overgelegde producties.
 - c. De dossiers worden zorgvuldig geanonimiseerd, in die zin dat de naam van de rechter wordt vervangen door een letteraanduiding; dit heeft niet alleen betrekking op het eindvonnis en eventuele tussenvonnissen, maar ook op andere dossierstukken waarin de naam van een rechter voorkomt.
 - d. Met het oog op de kwaliteit en de objectiviteit van de toetsing geschiedt de toetsing van elk vonnis door een koppel van twee personen, die in eerste instantie onafhankelijk van elkaar het toetsingsformulier invullen en daarna met elkaar overleg plegen. Deze koppels zijn van steeds wisselende samenstelling. Het

- verdient aanbeveling dat ook de voorzitter en de secretaris van de Subcommissie vonnissen toetsen. Ter bevordering van geobjectiveerde toetsing is het denkbaar dat van de koppelgenoten wordt gevraagd het overleg te gebruiken om het zo veel mogelijk eens te worden, of inzichtelijk te maken waarom zij het op bepaalde punten niet eens zijn geworden.
- e. De in een cijfer uitgedrukte waarderingen van de verschillende aspecten van het getoetste vonnis en van het totaal oordeel daarover kunnen van een toelichting worden voorzien. Het eindoordeel dient steeds van een toelichting te worden voorzien. Het wordt aan het professionele inzicht van de leden van de Subcommissie overgelaten wanneer en in welke mate een toelichting overigens dienstig is.
6. Indien de door de Commissie ontworpen methode in de praktijk van de visitatie zou worden gebruikt, zou de Subcommissie, hetzij vooraf hetzij werkendeweg, de door de Commissie in haar rapport geformuleerde vragen met betrekking tot het formulier, het aantal in het onderzoek te betrekken vonnissen en de (on)mogelijkheid van onderzoek met betrekking tot de comparitie nader onder ogen moeten zien.
 7. De Commissie acht de door haar ontwikkelde methode in beginsel ook bruikbaar voor de toetsing van de kwaliteit van civiele hofarresten, met dien verstande dat daarvoor een op de appelrechtspraak afgestemd toetsingskader moet worden ontworpen.
 8. Het verdient aanbeveling het toetsingskader te beschouwen als een instrument voor zowel kwaliteitstoetsing als kennisaccumulatie. Andere door de Commissie gesignaleerde mogelijke toepassingen van het toetsingskader zijn onder meer:
 - a. gebruik als checklist bij het meelesen van een vonnis;
 - b. gebruik tijdens de opleiding tot rechter of secretaris;
 - c. opname van het toetsingsformulier en de toelichting in het landelijke Handboek civiel van de rechtbanken.

11 Bijlagen

1. Samenstelling van de Commissie
2. Projectopdracht pilot Kwaliteitstoetsing rechterlijke uitspraken van 31 maart 2010
3. Werkwijze van de Commissie
4. Toetsingsformulier versie 1 met toelichting - 100603
5. Toetsingsformulier versie 2 met toelichting - 100703
6. Toetsingsformulier versie 3 met toelichting - 101205