



Rapport onderzoek

Evaluatie van het selectieproces
voor de rio-opleiding

Dr. A.S.M. Niessen
Prof. dr. R.R. Meijer

Colofon

Raad voor de rechtspraak
Postbus 90613
2509 LP Den Haag

Deze publicatie verschijnt in het kader van het wetenschappelijk onderzoeksprogramma van de Raad voor de rechtspraak. Uitgave daarvan betekent niet dat de inhoud het standpunt van de Raad voor de rechtspraak weergeeft.

Onderzoek in opdracht van de Raad voor de rechtspraak.

Begeleidingscommissie

Mr. G. Tangenberg (voorzitter SSR)
Prof. dr. M.Ph. Born (Erasmus Universiteit Rotterdam)
Dr. H.G.E. Verberk (wetenschappelijk adviseur Rvdr)
Mr. H. Vetter (rechter Rechtbank Den Haag)
Mr. A. Olland (teamvoorzitter Gerechtshof Den Haag)
Mr. F. van der Grinten (NOVA)
Drs. E. Bremer (bestuurssecretaris SSR)
A.J.B. Verheijen MA (coördinerend beleidsmedewerker)

De integrale tekst van dit rapport is gratis te downloaden van: www.rechtspraak.nl/Organisatie/Raad-Voor-De-Rechtspraak
Rubriek: wetenschappelijk onderzoek
dr. A. S. M. (Susan) Niessen & prof. dr. R. R. (Rob) Meijer
Heymansinstituut voor Psychologisch Onderzoek
Rijksuniversiteit Groningen
april 2019

Uitgever

Sdu Uitgevers bv, Den Haag

Vormgeving

Corps, Den Haag

Opmaak binnenwerk

LINE UP boek en media bv, Groningen

De verantwoordelijkheid voor de inhoud berust bij de Rijksuniversiteit Groningen. Het gebruik van cijfers en/of teksten als toelichting of ondersteuning in artikelen, scripties en boeken is toegestaan mits de bron duidelijk wordt vermeld. Vermenigvuldigen en/of openbaarmaking in welke vorm ook, alsmede opslag in een retrieval system, is uitsluitend toegestaan na schriftelijke toestemming van de Rijksuniversiteit Groningen. Rijksuniversiteit Groningen aanvaardt geen aansprakelijkheid voor drukfouten en/of andere onvolkomenheden. The responsibility for the contents of this report lies with Rijksuniversiteit Groningen. Quoting numbers or text in papers, essays and books is permitted only when the source is clearly mentioned. No part of this publication may be copied and/or published in any form or by any means, or stored in a retrieval system, without the prior written permission of Rijksuniversiteit Groningen. Rijksuniversiteit Groningen does not accept responsibility for printing errors and/or other imperfections.

In opdracht van de Raad voor de rechtspraak

Rapport onderzoek

Evaluatie van het selectieproces
voor de rio-opleiding

Inhoudsopgave

Samenvatting	9
1 Inleiding	24
1.1 Doelstelling en aanleiding	24
1.2 Context en beschrijving van de selectieprocedure	25
1.3 Werkwijze	27
1.3.1 Interviews	27
1.3.2 Vragenlijsten voormalig deelnemers	28
1.3.3 Data selectieprocedure en rio-volgsysteem	28
1.3.4 Begrippenlijst	29
2 De selectiecommissie	31
2.1 Totstandkoming selectiecommissie	31
2.2 Training selectiecommissie	32
2.3 Ervaringen met het functioneren van de selectiecommissie	34
3 Organisatie van de selectieprocedure	35
3.1 Secretariële ondersteuning	35
3.2 Doorlooptijd selectieprocedure	35
4 Analyse proces van selectie en selectiecriteria	37
4.1 Mening van betrokkenen over de selectieprocedure	37
4.2 Briefselectie	39
4.2.1 Controle formele vereisten	39
4.2.2 Andere criteria	39
4.2.3 Ervaring betrokkenen	40
4.2.4 Aanbevelingen	40

4.3	Analytische tests	41
4.3.1	Ervaringen van betrokkenen	42
4.3.2	Kwaliteit van de tests	43
4.3.3	Aanbevelingen	43
4.4	Voorgesprek	44
4.4.1	Procedure voorgesprek	44
4.4.2	Ervaringen leden LSR	45
4.4.3	Kwaliteit voorgesprekken	45
4.4.4	Aanbevelingen	48
4.5	Referenties	49
4.6	Assessment	50
4.6.1	Kwaliteit van de instrumenten	51
4.6.1.1	<i>Watson-Glaser</i>	52
4.6.1.2	<i>Vragenlijsten</i>	52
4.6.1.3	<i>Olympics game en Dilemmatest</i>	53
4.6.1.4	<i>Assessmentinterview</i>	54
4.6.2	Procedure van beoordelen in het assessment	54
4.6.3	Selectie op basis van de competenties uit het rechtersprofiel	56
4.6.4	Ervaringen leden LSR	58
4.7	Eindgesprek	60
4.7.1	Procedure eindgesprek	60
4.7.2	Ervaringen leden LSR	61
4.7.3	Kwaliteit eindgesprekken	62
4.7.4	Aanbevelingen	64
4.8	Lokale selectieprocedure	67
4.8.1	Opzet lokale selectieprocedures	67
4.8.2	Aansluiting landelijke en lokale selectie	68
4.8.3	Plaatsing na landelijk 'groen licht'	69
4.8.4	Aanbevelingen	70
4.9	Bijdrage selectie aan ontwikkelpunten in de opleiding	72
4.10	Cultuurwaardevrij selecteren	73
5	Meningen van deelnemers aan de selectieprocedure	76
5.1	Tevredenheid over de selectieprocedure als geheel	77
5.1.1	Gepastheid inhoud procedure	78
5.1.2	Ontwikkelpunten opleiding	78

5.2	Meningen over onderdelen van de selectieprocedure	78
5.2.1	Briefselectie	82
5.2.2	Analytische tests	82
5.2.3	Voorgesprek	83
5.2.4	Assessment	83
5.2.4.1	<i>Persoonlijkheidsvragenlijsten</i>	83
5.2.4.2	<i>Dilemmatest</i>	83
5.2.4.3	<i>Olympics game</i>	84
5.2.4.4	<i>Watson-Glaser test</i>	84
5.2.4.5	<i>Assessmentinterview</i>	84
5.2.5	Eindgesprek	85
5.2.6	Lokale selectie	85
5.3	Verschillen naar leeftijd en migratieachtergrond	85
5.3.1	Leeftijd	85
5.3.2	Migratieachtergrond	86
5.4	Conclusie en aanbevelingen	87
6	Analyses op basis van het rio-volgsysteem	89
6.1	Data en analyse	89
6.1.1	Selectiegegevens	89
6.1.2	Gegevens beoordelingen rio-opleiding	90
6.1.3	Steekproefgrootte	93
6.1.4	Analyse	94
6.2	Overzicht predictieve validiteit selectieonderdelen	95
6.3	Analytische tests	96
6.4	Andere cognitief georiënteerde tests	96
6.5	Voorgesprek	100
6.6	Assessment	102
6.7	Eindgesprek	104
6.8	Aanvullende analyses	104
6.8.1	De relatie tussen leeftijd en scores op de analytische tests	104
6.8.2	Herkansen van de analytische tests	107
6.9	Conclusies en aanbevelingen	108
6.9.1	Beperkingen van het onderzoek	108
6.9.2	Mogelijke verklaringen	109
6.9.3	Conclusie	110

6.9.4	Aanbevelingen analytische tests	110
6.9.4.1	<i>Controverse rondom analytische tests</i>	112
6.9.5	Aanbevelingen voorgesprek	114
6.9.6	Aanbevelingen assessment	114
6.9.7	Aanbevelingen eindgesprek	115
7	Toelichting op terugkerende aanbevelingen	117
7.1.	Structureren van interviews	117
7.2	Statistische combinatie van gegevens	118
7.3	Tot slot	119
	Referenties	121
	Bijlagen	127
Bijlage 1:	Specifieke onderzoeksvragen uit de Startnotitie evaluatie selectieproces rio's	127
Bijlage 2:	Competentiematrix assessment	130
Bijlage 3:	Wegingen van de verschillende test- en schaalscores uit het assessment voor iedere competentie	131
Bijlage 4:	Tabellen	132

Samenvatting

Inleiding

In het kader van een veelomvattend vernieuwingstraject rondom de werving, selectie en opleiding van rechters is de selectieprocedure voor de rio-opleiding in 2013 onderworpen aan een wetenschappelijk onderzoek (Born, 2013). Uit dit onderzoek bleek dat veel aspecten in de toenmalige wijze van selectie zorgvuldig verliepen en dat betrouwbare en valide selectie-instrumenten werden ingezet. De voorspellende waarde van de verschillende onderdelen van de selectieprocedure kon toen echter niet worden vastgesteld. Bovendien is de selectieprocedure sindsdien op verschillende punten veranderd. Het doel van dit evaluatieonderzoek naar het selectieproces van rio's is om te beoordelen in hoeverre de doelstellingen die aan het vernieuwde selectieproces zijn verbonden worden gerealiseerd en om aanbevelingen voor verbeteringen op te stellen. Onderwerpen die zijn onderzocht zijn ervaringen en meningen van betrokkenen vanuit de Landelijke Selectiecommissie Rechters (LSR), de lokale gerechten en de opleiding over de selectieprocedure, ervaringen en meningen van oud-kandidaten met de selectieprocedure, de kwaliteit van de verschillende onderdelen van de selectieprocedure en de voorspellende waarde (predictieve validiteit) van de verschillende onderdelen van de selectieprocedure.

Methode

Bovenstaande onderwerpen zijn onderzocht aan de hand van interviews met verschillende betrokkenen vanuit de LSR, het Studiecentrum Rechtspleging (SSR) en de Raad voor de rechtspraak, een vragenlijst afgenomen bij voormalig deelnemers aan de selectieprocedure, documenten aangeleverd door de LSR, SSR en het assessmentbureau LTP en een analyse van gegevens over scores en beoordelingen uit de selectie-

procedure en beoordelingen uit de rio-opleiding. De resultaten van de statistische analyses die zijn uitgevoerd om de voorspellende waarde en de interbeoordelaarsbetrouwbaarheid van de verschillende selectie-onderdelen in kaart te brengen zijn echter onnauwkeurig, gezien de kleine steekproeven en het grote aantal ontbrekende gegevens. De conclusies over deze onderwerpen worden dan ook met veel voorzichtigheid getrokken. Hieronder bespreken we de belangrijkste conclusies.

De selectiecommissie & organisatie

De LSR bestaat uit zes interne leden vanuit de zittende magistratuur en zes externe leden afkomstig uit diverse sectoren. De leden worden benoemd door de Raad voor de rechtspraak. Daarnaast zijn er zes plaatsvervangend leden, waarvan drie intern en drie extern lid. Het functioneren van de LSR wordt positief geëvalueerd door de leden, met name door de zorgvuldigheid en betrokkenheid van de leden en het Presidium. De samenwerking tussen interne en externe leden wordt als prettig ervaren. Kandidaat-leden van de LSR worden geworven via het plaatsen van vacatures. Aan de hand van sollicitatiebrieven en het cv worden kandidaten uitgenodigd voor een gesprek waarin onder andere gelet wordt op de eigenschappen aangeraden in het rapport Born. In hoeverre dat ervoor zorgt dat de leden van de LSR daadwerkelijk over de gewenste eigenschappen beschikken is echter niet te controleren. Om te waarborgen dat leden van de LSR de over gewenste eigenschappen en vaardigheden beschikken bevelen we aan om de interviews met kandidaat-leden te structureren. Ook kan overwogen worden om een assessment aan de procedure toe te voegen waarin eigenschappen die samenhangen met het maken van accurate selectiebeslissingen, met name: empathisch vermogen, sociale vaardigheden, cognitieve capaciteiten en dispositional reasoning, eventueel in de vorm van een test (De Kock, Lievens & Born, 2018).

Voor nieuwe leden van de LSR wordt een kennismakingsdag georganiseerd en voor alle leden is er een jaarlijkse 'terugkomdag', waarop aandacht is voor training en oefening van interviewvaardigheden en voor discussie en reflectie. De kennismakingsdagen en de terugkomdagen wordt positief gewaardeerd. Het zou volgens de leden wenselijk zijn als er nog meer ruimte was voor training, oefening en feedback. Ook is het type training dat aangeboden wordt niet de meeste effectieve variant. Net als Born (2013) raden we aan om een frame-of-reference training aan te bieden aan alle selecteurs, waarin veel aandacht besteed wordt aan het bereiken van een gemeenschappelijke conceptualisatie van de verschillende competenties en eigenschappen die beoordeeld worden in de interviews (zie paragraaf 2.2).

Wat de organisatie van de procedure betreft waren er weinig knelpunten. De gewenste doorlooptijd van de selectieprocedure wordt gehaald en de leden zijn tevreden met de beschikbare ondersteuning. Het belangrijkste knelpunt wat betreft de organisatie was de beschikbaarheid van goed functionerende digitale middelen ten behoeve van de organisatie en registratie van de procedure. Daaraan wordt momenteel gewerkt.

De selectieprocedure

De selectieprocedure van rio's bestaat uit een briefselectie, twee analytische tests, een voorgesprek, een assessment en een eindgesprek. Vervolgens gaan succesvolle kandidaten door naar de selectie bij de lokale gerechten.

De geïnterviewde leden van de LSR en het Presidium ervaren de procedure unaniem als zorgvuldig en volledig, maar zien ook dat het een zware procedure is voor de kandidaten. De hoogte van de eisen die gehanteerd moet worden is een punt van aandacht onder alle geïnterviewde betrokkenen, vooral omtrent de grensscores die gehanteerd worden bij de analytische tests. Enkelen vermoeden bovendien dat de hoge eisen geschikte kandidaten af zou kunnen schrikken. Anderen merken op dat de uitval uit de rio-opleiding is afgenomen sinds de nieuwe selectieprocedure is ingevoerd. Of de verminderde uitval veroorzaakt wordt door de veranderingen in de selectieprocedure is echter niet vast te stellen. Ook de veranderingen in de opleiding, die ongeveer gelijktijdig hebben plaatsgevonden, zouden daaraan bij kunnen dragen.

Briefselectie

Tijdens de briefselectie worden de formele vereisten gecontroleerd en wordt er op zaken als cv, inhoudelijke expertise en motivatie gescreend. De leden van de LSR en de oud-kandidaten zien de briefselectie als een noodzakelijk onderdeel van de procedure. Het was niet mogelijk om de betrouwbaarheid of validiteit van de briefselectieprocedure te onderzoeken. Uit gesprekken met betrokkenen bleek dat deze procedure niet gestandaardiseerd is en er geen expliciete definities van de formele vereisten rondom het voldoen aan minimaal twee jaar relevante juridische werkervaring en voldoende maatschappelijke ervaring gehanteerd worden. Ook worden er aanvullende zaken, zoals motivatie en werkervaring beoordeeld, die niet geëxpliciteerd of geformaliseerd zijn. We bevelen aan om de criteria die gehanteerd worden bij de briefselectie te expliciteren, de briefselectie te structureren en om daarbij gebruik te maken van een beoordelingsformulier.

Analytische tests

De analytische tests zijn de Matrix test, met als doel het abstract redeneervermogen te meten, en de Test voor Verbale Relaties en Analogieën (TVRA), met als doel het verbaal redeneervermogen te meten. Kandidaten moeten op dit onderdeel bepaalde scores halen om de procedure te mogen voortzetten. Over dit onderdeel is binnen de rechtspraak veel discussie; bij sommigen bestaat het beeld dat veel geschikte kandidaten afvallen bij deze stap. De meeste leden van de LSR waren van mening dat de analytische tests een geschikt onderdeel van de procedure zijn, vooral gezien de wetenschappelijke onderbouwing voor dit soort tests. Oud-kandidaten waren gematigd positief over het opnemen van deze tests, maar velen waren wel van mening dat de gehanteerde norm te streng is en dat ook andere zaken, zoals het cv of een gesprek, meegenomen zouden moeten worden in de beslissing om de procedure voort te zetten met kandidaten.

Een analyse van gegevens uit de selectieprocedure en de opleiding liet zien dat de analytische tests de sterkst voorspellende onderdelen voor beoordelingen in de rio-opleiding uit de selectieprocedure lijken te zijn, waarbij de TVRA de grootste predictieve validiteit had. De Watson-Glaser test voor kritisch redeneervermogen (nu gebruikt in het assessment) liet een grotere voorspellende waarde zien dan de Matrix test en werd ook positiever beoordeeld door oud-kandidaten. Daarom bevelen we aan om de TVRA te handhaven en om te overwegen om de Matrix test te vervangen door de Watson-Glaser test. Over de kwaliteit van de analytische tests was niet veel informatie beschikbaar. We bevelen aan om tests te gebruiken waarvan de kwaliteit uitvoerig is onderzocht en die bij voorkeur zijn beoordeeld door de Commissie Testaangelegenheden Nederland (COTAN).

Voorgesprek

Tijdens het voorgesprek spreken de kandidaten met drie personen, waaronder een afgevaardigde van een lokaal gerecht, over verschillende vooraf bepaalde onderwerpen. Ook is er ruimte om te spreken over andere zaken, zoals zaken die in het cv opvallen. Het uitgangspunt bij de voorgesprekken is dat kandidaten bij twijfel door mogen naar de volgende ronde van de selectieprocedure. Daarom is dit onderdeel volgens de leden van de LSR het minst noodzakelijke onderdeel van de procedure. Wel zijn zij erg positief over de betrokkenheid van de lokale gerechten en de wederkerige opzet van deze gesprekken. De oud-kandidaten waren ook positief over het

voorgesprek. Op basis van een analyse van de beschikbare gegevens lijken de beoordelingen uit de voorgesprekken beperkt betrouwbaar¹ te zijn en lijkt de voorspellende waarde van het voorgesprek erg klein te zijn (zie paragraaf 4.4.3 en 6.5). Gezien de beperkingen van de beschikbare gegevens zijn deze conclusies echter niet met stelligheid te trekken. Aangezien er ook nog een eindgesprek gevoerd wordt kan worden overwogen om dit onderdeel te schrappen. Als het voorgesprek gehandhaafd blijft raden we aan om de gesprekken sterker te structureren, zodat de betrouwbaarheid en voorspellende waarde vergroot worden.

Assessment

Het assessment bestaat uit een interview met een assessmentpsycholoog, verschillende vragenlijsten, een Dilemmatest, de Watson-Glaser test voor kritisch redeneervermogen en een spelopdracht bedoeld om de informatieverwerkingssnelheid te meten. Op basis van deze onderdelen wordt iedere kandidaat beoordeeld op de competenties uit het rechtersprofiel². Oud-kandidaten waren gematigd positief over de meeste onderdelen van het assessment, maar merkten op dat de vragenlijsten gemakkelijk sociaal-wenselijk ingevuld konden worden en dat de relevantie van de spelopdracht niet duidelijk was. Leden van de LSR waren positief over het assessment en gaven aan dat er in de besluitvorming veel gewicht wordt toegekend aan het assessmentrapport, zoals ook de doelstelling was. De voorspellende waarde van de competentiebeoordelingen uit het assessmentrapport voor de beoordelingen uit de rio-opleiding lijkt op basis van de beschikbare gegevens echter klein en/of negatief te zijn, hoewel deze conclusie, gezien de beperkingen van de gegevens waarop deze analyses gebaseerd zijn, slechts met grote voorzichtigheid getrokken kan worden. Er waren daarnaast weinig gegevens beschikbaar over de kwaliteit van de gebruikte tests en instrumenten. We bevelen aan om het assessment in te richten als een 'assessment center', bestaande uit opdrachten en rollenspellen die in inhoud en in vorm overeenkomen met belangrijke taken die in (het begin van) de rio-opleiding vervuld moeten worden. Een dergelijke invulling draagt waarschijnlijk ook bij aan de doelstelling om cultuurwaardevrijer te selecteren, aangezien dit soort assessments vaak tot kleinere verschillen in scores op basis van etnische achtergrond leidt (De Soete et al., 2013). Als het assessment in de huidige vorm gehandhaafd blijft bevelen we aan om tests en instrumenten te gebruiken waarvan de kwaliteit uitvoerig is onderzocht en die bij voorkeur beoordeeld zijn door de COTAN. Ook raden we aan om, gezien de

1 Er lijkt sprake te zijn van beperkte overeenstemming tussen verschillende beoordelaars.

2 Regievoering, analytisch vermogen, flexibiliteit, juridische kennis, leervermogen, luisteren, omgevingsbewustzijn, overredingskracht, samenwerken, zelfbewustzijn, zelfvertrouwen & authenticiteit, besluitvaardigheid.

mogelijkheid tot sociaal-wenselijk antwoorden, zelfrapportagevragenlijsten niet, of alleen voor 'select-out'³ beslissingen te gebruiken. Daarnaast raden we aan om het assessmentinterview te structureren en om beoordelingen tot stand te laten komen volgens een beslisregel, waarbij de verschillende onderdelen expliciete gewichten krijgen.

Eindgesprek

Het eindgesprek bestaat uit gesprekken met drie 'teams' bestaande uit twee personen, die ieder een eigen agenda hanteren. Daarnaast is er ruimte om ook andere onderwerpen te bespreken, bijvoorbeeld resultaten uit het assessmentrapport. Na afloop van de eindgesprekken wordt besloten of de kandidaten 'groen licht' krijgen en door mogen naar de lokale selectieprocedure. Volgens de leden van de LSR is dit het meest waardevolle onderdeel van de selectieprocedure, vooral vanwege de zorgvuldige procedure rondom de beraadslagingen en de belangrijke onderwerpen die in deze gesprekken aan bod komen. Ook de oud-kandidaten waren positief over de eindgesprekken. Op basis van de beschikbare gegevens varieert de betrouwbaarheid⁴ van de beoordelingen die voortkomen uit het eindgesprek van onvoldoende tot redelijk en lijkt de voorspellende waarde zeer klein te zijn. Die conclusies zijn echter, gezien de beperkingen van de beschikbare gegevens, niet met stelligheid te trekken. Om de betrouwbaarheid en voorspellende waarde van het eindgesprek te vergroten bevelen we aan om het eindgesprek verder te structureren. Daarnaast bevelen we aan om de combinatie van beoordelingen en gegevens uit het assessmentrapport en de eindgesprekken te combineren volgens een beslisregel, waarbij de verschillende onderdelen expliciete gewichten krijgen.

Lokale selectieprocedure

Na de landelijke selectieprocedure vindt er nog een selectieronde plaats bij de lokale gerechten. Dit levert met regelmaat kandidaten op die landelijk 'groen licht' hebben, maar lokaal geen rio-plaats weten te verwerven. Om dit aantal terug te dringen kan worden overwogen om de lokale selectieprocedure te laten vervallen. De lokale gerechten kunnen dan nog wel bij de briefselectie en de voorgesprekken betrokken worden. De opzet van de lokale selectieprocedures verschillen per gerecht, maar lijken weinig gestructureerd te zijn. Ook wordt, afhankelijk van het gerecht, op verschillende zaken beoordeeld, waarvan sommige nog niet getoetst zijn in de landelijke procedure

3 Bij select-out beslissingen worden alleen kandidaten met zeer lage scores op de vragenlijsten afgewezen.

4 De mate van overeenstemming tussen verschillende beoordelaars.

en gerelateerd zijn aan de 'fit' met het gerecht (bijvoorbeeld binding met de regio en inhoudelijke expertise), terwijl andere zaken soms overlap vertonen met de landelijke procedure (bijvoorbeeld motivatie voor het rechterschap en samenwerken). Als de lokale selectie gehandhaafd blijft bevelen we aan om expliciete geoorloofde afwijzingsgronden af te spreken en de lokale selectieprocedures te structureren.

Beschikbaarheid gegevens

De gegevens uit de selectie en de opleiding die we hebben kunnen gebruiken voor de statistische analyses hadden veel beperkingen. Met name de kleine steekproeven en het grote aantal ontbrekende gegevens speelden daarbij een rol zorgden ervoor dat het niet mogelijk is om stellige conclusies te trekken over de voorspellende waarde van de verschillende selectie-onderdelen op basis van de statistische analyses die zijn uitgevoerd aan de hand van deze gegevens. We bevelen met klem aan om in de nabije toekomst een completer en handzamer systeem voor het bewaren en opslaan van gegevens te realiseren. Een vergelijkbaar onderzoek zou dan na enkele jaren herhaald kunnen worden op basis van nieuwe gegevens van betere kwaliteit en kwantiteit.

Overzicht van aanbevelingen

Hieronder volgt een beknopt overzicht van de aanbevelingen, waarbij we verwijzen naar de secties van het rapport waar de toelichtingen en onderbouwingen te vinden zijn.

Selectie leden LSR

Op basis van gesprekken met het Presidium van de LSR en de wetenschappelijke literatuur bevelen we aan om:

- De selectiegesprekken met kandidaat-leden van de LSR te structureren, aangezien dit tot een hogere voorspellende waarde en accuratere selectiebeslissingen leidt (zie paragraaf 2.1).
- Te overwegen om in de selectie van kandidaat-leden aandacht besteden aan kenmerken die positief samenhangen met het maken van accurate selectiebeslissingen, met name: empathisch vermogen, sociale vaardigheden, cognitieve capaciteiten en dispositional reasoning, eventueel in de vorm van een test (zie paragraaf 2.1).

Training leden selectiecommissie

Op basis van gesprekken met het Presidium en leden van de LSR bevelen we aan om:

- Nieuwe leden meer mogelijkheden te bieden om zich te verdiepen in het beroep van rechter (externe leden) en op de selectieprocedure, bijvoorbeeld door zittingen bij te wonen en deel te nemen aan verschillende tests en assessments uit de selectieprocedure (zie paragraaf 2.2).
- Op kennismakingsdagen en terugkomdagen voor leden van de LSR meer tijd aan oefenen, rollenspelen en casussen te besteden, aangezien de leden hier behoefte aan hebben (zie paragraaf 2.2).

Op basis van de wetenschappelijke literatuur bevelen we aan om:

- Een frame-of-reference training te (laten) ontwikkelen en aan te bieden aan selecteurs, aangezien dit de meest effectieve vorm van interviewtraining is. In een dergelijke training wordt veel aandacht besteed aan een gemeenschappelijke conceptualisatie van de verschillende competenties en eigenschappen die beoordeeld worden in interviews (zie paragraaf 2.2).

Organisatie selectieprocedure

Op basis van gesprekken met het Presidium, leden van de LSR, het secretariaat van de LSR en de reacties op de vragenlijst ingevuld door oud-kandidaten bevelen we aan dat:

- De kwaliteit en beschikbaarheid van digitale middelen voor het organiseren en registreren van de selectieprocedure en de doorlooptijd van de selectieprocedure ook in de toekomst een aandachtspunt moeten blijven, zeker in het geval van veranderingen in de procedure. Er lijken momenteel echter geen dringende aandachtspunten te zijn op dit gebied (zie hoofdstuk 3).

Briefselectie

Op basis van gesprekken met leden van het Presidium, leden van de LSR en leden van lokale selectiecommissies bevelen we aan om:

- Te expliciteren wat er verstaan wordt onder relevante juridische werkervaring en wanneer er voldaan is aan voldoende maatschappelijke ervaring. Daar worden momenteel geen formele definities voor gehanteerd (zie paragraaf 4.2.1).
- Ook expliciete definities te hanteren van andere zaken waarop kandidaten tijdens de briefselectie beoordeeld worden (zie paragraaf 4.2.2).
- Deze definities op te nemen in een beoordelingsformulier dat gebruikt kan worden voor de briefselectie, zodat de briefselectie gestructureerd kan worden uitgevoerd (zie paragraaf 4.2.4).

Analytische tests

Op basis van gesprekken met betrokkenen bij de selectieprocedure, documenten aangeleverd door LTP, reacties op de vragenlijst ingevuld door oud-kandidaten, de wetenschappelijke literatuur en de resultaten van de analyse van de predictieve validiteit, bevelen we aan om:

- De Taalgevoeltest in de toekomst niet meer op te nemen in de selectieprocedure, aangezien deze onvoldoende betrouwbare scores oplevert (zie paragraaf 4.3.2). Indien het testen van taalgevoel wenselijk wordt gevonden kan een ander instrument van voldoende kwaliteit worden gezocht.
- Tests te gebruiken die in voldoende mate zijn onderzocht op psychometrische kwaliteit (betrouwbaarheid en validiteit) en waarvan de resultaten laten zien dat de tests aan de wetenschappelijke standaard voldoen (zie paragraaf 4.3.3).
- Bij voorkeur tests te gebruiken die op relevante punten minstens als 'voldoende' zijn beoordeeld door de COTAN (zie paragraaf 4.3.3).
- Wat betreft de gebruikte tests:
 - De TVRA of een gelijksoortig instrument te handhaven, aangezien de scores op deze test samenhangen met beoordelingen uit de rio-opleiding (zie paragraaf 6.3).
 - Bij het gebruik van meerdere tests te overwegen om de tests compensatorisch te scoren, zodat een lage score gecompenseerd kan worden door een hoge score op een andere test, om het aantal vals-negatieve beslissingen terug te dringen (zie paragraaf 6.9.4).
 - Te overwegen om de Matrix test te vervangen door de Watson-Glaser test, of een ander instrument van goede kwaliteit, dat kritisch redeneervermogen meet, aangezien de scores op deze test samenhangen met beoordelingen uit

- de rio-opleiding en positief werd gewaardeerd door kandidaten (zie paragraaf 5.2 en 6.4).
- Het gebruik van 'job-knowledge' tests als alternatief voor de analytische tests te overwegen, aangezien een dergelijke test waarschijnlijk een goede voorspellende waarde en positieve reacties van kandidaten zal opleveren (zie paragraaf 6.9.4).
 - Zeer voorzichtig om te gaan met uitzonderingen op de gehanteerde grensscores op basis van leeftijd of ervaring. Het verband tussen leeftijd en de scores op de analytische tests is terecht een punt van aandacht. Wat de oorzaak van dit verschil is, is echter niet duidelijk en het verband tussen het aantal jaren relevante werkervaring en latere baanprestaties is gering (zie paragraaf 6.8.1).
 - De mogelijkheid tot herkansing niet meer aan te bieden, aangezien dit een negatief effect heeft op de voorspellende waarde van de test scores en het aantal vals-positieve beslissingen doet toenemen (zie paragraaf 6.8.2).
 - Indien het toch wenselijk wordt gevonden om een herkansing aan te bieden, kandidaten alle onderdelen te laten herkansen en de gemiddelde scores over de pogingen te gebruiken in de besluitvorming. Dit zal het aantal vals-positieve beslissingen als gevolg van herkansen verminderen (zie paragraaf 6.8.2).

Voorgesprek

Op basis van gesprekken met betrokkenen bij de selectieprocedure, reacties op de vragenlijst ingevuld door oud-kandidaten, de wetenschappelijke literatuur en de resultaten van de analyse van de predictieve validiteit, bevelen we aan om:

- Te overwegen om het voorgesprek te laten vervallen, aangezien het geen doorslaggevend onderdeel is en de betrouwbaarheid en de voorspellende waarde van de beoordelingen uit het voorgesprek gering lijken (zie paragraaf 6.5 en 6.9.5).
- Indien het voorgesprek gehandhaafd blijft:
 - Ook lokale selecteurs te trainen in interviewvaardigheden (zie paragraaf 4.4.2).
 - Erop toe te zien dat de beoordelingsformulieren volledig ingevuld worden (zie paragraaf 4.4.2).
 - De voorgesprekken sterker te structureren. Dat leidt waarschijnlijk tot een hogere betrouwbaarheid en een hogere voorspellende waarde (zie paragraaf 4.4, 6.9.5 en 7.1).

- Ook bij het voorgesprek gebruik te maken van een statistische methode van gegevenscombinatie en besluitvorming. Dit zal de voorspellende waarde waarschijnlijk vergroten (zie paragraaf 6.9.5 en 7.2).

Referenties

Uit gesprekken met leden van de begeleidingscommissie bleek dat de procedure rondom het opvragen van ambtshalve referenties niet is geformaliseerd en als onvoldoende transparant wordt ervaren. Daarom bevelen we aan om:

- De procedure rondom het opvragen van ambtshalve referenties te formaliseren en te structureren door aan referenten duidelijk te maken wat voor informatie de LSR wel en niet wenst te ontvangen (zie paragraaf 4.5).

Assessment

Op basis van gesprekken met betrokkenen bij de selectieprocedure, documenten aangeleverd door LTP, reacties op de vragenlijst ingevuld door oud-kandidaten, de wetenschappelijke literatuur en de resultaten van de analyse van de predictieve validiteit, bevelen we aan om:

- De oordelen die voortvloeien uit het assessment volgens statistische gegevenscombinatie te bepalen, aangezien dit de voorspellende waarde waarschijnlijk zal vergroten (zie paragraaf 4.6.2 en 7.2).
- Vragenlijsten op basis van zelfrapportage niet, of alleen voor 'select-out' beslissingen te gebruiken, gezien de mogelijkheid tot sociaal-wenselijk antwoorden en de lage predictieve validiteit in selectiesituaties (zie paragraaf 4.6.1.2).
- Het assessment anders in te richten:
 - Het is het overwegen waard om het assessment in te richten als 'assessment center', bestaande uit opdrachten en rollenspellen die in inhoud en in vorm overeenkomen met belangrijke taken die in (het begin van) de rio-opleiding vervuld moeten worden. Dit zal waarschijnlijk een goede predictieve validiteit en positieve reacties van kandidaten opleveren, maar zal ook tot hogere kosten leiden (zie paragraaf 4.6.3 en 6.9.6).
 - Het assessmentinterview te laten vervallen (zie paragraaf 6.9.6).

- Als het assessment in de huidige of in een vergelijkbare vorm gehandhaafd blijft, bevelen we aan om:
 - De beoordelingen op competentieniveau in de selectie los te laten, aangezien de validiteit van de gehanteerde competentiestructuur onduidelijk is (zie paragraaf 4.6.3).
 - Het assessmentinterview te structureren en te scoren (zie paragraaf 4.6.2 en 7.1).
 - Bij handhaving van de Dilemmatest en een simulatie-opdracht zoals de 'Olympics game', raden we aan om varianten te ontwikkelen die inhoudelijk aansluiten bij het rechterschap. Dit zal de voorspellende waarde en de reacties van kandidaten waarschijnlijk positief beïnvloeden (zie paragraaf 4.6.1.3).
 - Tests te gebruiken die in voldoende mate zijn onderzocht op psychometrische kwaliteit (betrouwbaarheid en validiteit) en waarvan de resultaten laten zien dat de tests aan de wetenschappelijke standaard voldoen (zie paragraaf 4.6.1).
 - Bij voorkeur tests te gebruiken die op relevante punten minstens als 'voldoende' zijn beoordeeld door de COTAN, zie paragraaf 4.6.1).

Eindgesprek

Op basis van gesprekken met betrokkenen bij de selectieprocedure, reacties op de vragenlijst ingevuld door oud-kandidaten, de wetenschappelijke literatuur en de resultaten van de analyse van de predictieve validiteit, bevelen we aan om:

- Het eindgesprek verder te structureren, aangezien dit de betrouwbaarheid en de voorspellende waarde waarschijnlijk zal verhogen (zie paragraaf 4.7, 6.7 en 7.2).
- De besluitvorming volgens een statistische procedure te laten verlopen, waarbij vooraf wordt gespecificeerd hoe zwaar ieder thema en iedere score gewogen moet worden. Dat geldt dan zowel voor de scores of beoordelingen uit het assessment als uit het eindgesprek. Dit zal de voorspellende waarde waarschijnlijk verhogen (zie paragraaf 4.7.4, 6.9.7 en 7.2).
- Om draagvlak te creëren voor een statistische procedure is het belangrijk om de onderbouwing daarvoor duidelijk naar belanghebbenden te communiceren en kan overwogen worden om ze te betrekken bij het opstellen van beslisregels en wegen van de informatie die gebruikt wordt in de besluitvorming (zie paragraaf 4.7.2 en 7.2).
- Een eventuele beoordeling van het assessmentrapport en het voeren van de eindgesprekken niet door dezelfde personen te laten doen, zodat de informatie

afkomstig uit beide onderdelen onafhankelijk van elkaar kan worden meegenomen (zie paragraaf 6.9.7).

Lokale selectieprocedure

Op basis van gesprekken met betrokkenen bij de selectieprocedure en de wetenschappelijke literatuur, bevelen we het volgende aan:

- Het komt voor dat kandidaten die succesvol door de procedure van de LSR zijn gekomen lokaal geen rio-plaats verwerven. Om het aantal succesvolle kandidaten dat lokaal geen rio-plaats kan vinden terug te dringen, kan overwogen worden om evenveel rio's te selecteren als er plaatsen zijn (zie paragraaf 4.8.3).
- De lokale selectie kan dan ofwel:
 - Komen te vervallen en worden vervangen door plaatsing.
 - Al tijdens de briefselectie en/of het voorgesprek plaatsvinden.

(Welk alternatief de voorkeur heeft is een beleidskeuze)

- Indien de lokale selectie gehandhaafd blijft bevelen we aan om:
 - De lokale selectieprocedure te structureren. Dat lijkt op basis van de gevoerde gesprekken nu niet het geval te zijn (zie paragraaf 4.8).
 - Expliciet geoorloofde afwijzingsgronden af te spreken (zie paragraaf 4.8).
 - Te zorgen voor meer bekendheid bij de lokale gerechten over de landelijke procedure, om het vertrouwen in de procedure gehanteerd door de LSR te vergroten (zie paragraaf 4.8).
 - De doelstelling van de lokale selectie duidelijk te communiceren naar kandidaten (zie paragraaf 4.8).

Cultuurwaardevrijer selecteren

Of de doelstelling om zo cultuurwaardevrij mogelijk te selecteren wordt behaald kan niet worden vastgesteld op basis van de beschikbare informatie. Op basis van gesprekken met betrokkenen bij de selectieprocedure, reacties op de vragenlijst ingevuld door oud-kandidaten en de wetenschappelijke literatuur bevelen we aan om:

- Indien mogelijk gebruik te maken van simulaties, zoals een assessment center of work samples. Op deze assessments worden vaak kleinere verschillen in scores

gevonden tussen personen met en zonder migratieachtergrond, ze hebben een hoge predictieve validiteit en worden positief gewaardeerd door kandidaten (zie paragraaf 4.10).

- Indien mogelijk analytische tests te gebruiken die zijn ontworpen om zo cultuurwaardevrij mogelijk te meten en die ook op dit punt onderzocht zijn (zie paragraaf 4.10)
- Aan de cultuurwaardevrijheid van de gehanteerde tests en assessments blijvend aandacht te besteden. Momenteel lijken er weinig kandidaten met een migratieachtergrond te solliciteren voor een rio-plaats. Als het aandeel kandidaten met een migratieachtergrond toeneemt zullen er wellicht meer knelpunten wat betreft cultuurwaardevrij selecteren aan het licht komen (zie paragraaf 4.10).

Acceptatie van de procedure door kandidaten

De kandidaten waren over de meeste onderdelen van de procedure gematigd positief. Op basis van gesprekken met betrokkenen bij de selectieprocedure, reacties op de vragenlijst ingevuld door oud-kandidaten en de wetenschappelijke literatuur bevelen we aan om:

- Informatie te verstrekken aan kandidaten over de reden om de procedure op een bepaalde manier in te richten. Dit kan de acceptatie vergroten (zie paragraaf 5.4).
- Waar mogelijk selectie-instrumenten te gebruiken die inhoudelijk goed aansluiten bij de rio-opleiding en de functie van rechter, zoals een 'assessment center'. Deze instrumenten zorgen waarschijnlijk voor een grotere tevredenheid van kandidaten (en hebben een goede predictieve validiteit), maar zijn waarschijnlijk kostbaarder dan het huidige assessment (zie paragraaf 5.4).

Beoordelingen in de rio-opleiding

Op basis van gesprekken met betrokkenen bij de selectieprocedure, documenten en gegevens aangeleverd door de LSR en de wetenschappelijke literatuur bevelen we aan om:

- Indien mogelijk beoordelingen niet per taakniveau te laten uitvoeren, maar een algemene beoordeling te hanteren en het gewenste prestatieniveau aan te passen aan de mate van vordering in de opleiding. Voor evaluatieonderzoek is het

- noodzakelijk dat beoordelingen uit de opleiding- samen geanalyseerd kunnen worden. Ook voor het in kaart brengen van de ontwikkeling van rio's is het inzichtelijker als verschillende beoordelingen met elkaar vergeleken kunnen worden (zie paragraaf 6.1.2).
- De beoordelingslijst zo in te richten dat iedere vraag onder één competentie geschaard kan worden. De huidige structuur heeft waarschijnlijk tot gevolg dat de competentiescores niet empirisch van elkaar onderscheiden kunnen worden (zie paragraaf 6.1.2).
 - Een weging toe te passen, zodat competenties die op basis van een groter aantal vragen worden beoordeeld niet langer (onbedoeld) een groter gewicht krijgen in de totaalscore (zie paragraaf 6.12).

Beschikbaarheid van gegevens

Op basis van de gegevens die zijn aangeleverd door de LSR en op het proces dat hieraan voorafging, bevelen we aan om:

- In de nabije toekomst een beter, completer en handzamer systeem voor het bewaren en opslaan van gegevens te realiseren. Dit maakt het mogelijk om in de toekomst evaluatieonderzoek uit te voeren op basis van meer en meer complete gegevens, wat tot nauwkeurigere conclusies zal leiden (zie paragraaf 6.1).

Vervolgonderzoek

- Gezien de beperkingen van de gegevens die in het huidige onderzoek gebruikt konden worden voor statistische en psychometrische analyses, bevelen we sterk aan om dit onderzoek, na aanpassingen van de selectieprocedure en het realiseren van een verbeterd systeem voor het bijhouden van gegevens, nog eens te herhalen (zie paragraaf 6.1 en 7.3).

Inleiding

1.1 Doelstelling en aanleiding

De wijze van selectie van rechters en raadsheren zoals die sinds 2014 plaatsvindt maakt onderdeel uit van het meer omvattende vernieuwingstraject van werving, selectie en opleiding van rechters. De aanpassing van het selectieproces moet worden bezien in het kader van de actualisatie van het rechtersprofiel, de afschaffing van de raio-opleiding en de start van de nieuwe rio-opleiding.

In dit kader zijn de (oude) selectie-instrumenten en selectienormen onderworpen aan een wetenschappelijk onderzoek (Born, 2013). Uit dit onderzoek bleek dat veel aspecten in de toenmalige wijze van selectie zorgvuldig verliepen en dat betrouwbare en valide selectie-instrumenten werden ingezet. De voorspellende waarde van de verschillende onderdelen van de selectieprocedure kon toen echter niet worden vastgesteld. Op basis van aanbevelingen uit het rapport Born zijn inmiddels gegevens verzameld om dit mogelijk te maken en heeft de raad voor de rechtspraak een opdracht tot vervolgonderzoek uitgeschreven. Het doel van dit evaluatieonderzoek naar het selectieproces van rio's is om te beoordelen in hoeverre de doelstellingen die aan het vernieuwde selectieproces zijn verbonden worden gerealiseerd en om aanbevelingen voor verbeteringen op te stellen.

De hoofdvraag die de Raad van de rechtspraak voor dit onderzoek heeft geformuleerd is:

In hoeverre slaagt de selectieprocedure erin om de doelstellingen die bij de introductie ervan zijn geformuleerd daadwerkelijk te realiseren en wat zijn daarbij de belangrijkste succesfactoren en verbeterpunten?

De vorengenoemde doelstellingen luiden:

1. De nieuwe selectieprocedure dient zó te zijn vormgegeven dat zoveel mogelijk rio's de opleiding met succes afronden;
2. De nieuwe selectieprocedure dient rekening te houden met een zo kort mogelijke doorlooptijd van de selectie, zonder iets af te doen aan de kwaliteit;
3. De aansluiting van het landelijke op het lokale proces dient zo te zijn vormgegeven dat zo min mogelijk overlap in de inhoud van het selectieproces ontstaat;
4. De betrouwbaarheid van de verschillende in te zetten selectie-instrumenten waar mogelijk te vergroten;
5. Het aantal kandidaten dat ná landelijk groen licht alsnog lokaal geen plaats kan vinden terug te dringen;
6. De nieuwe selectie-instrumenten zo cultuurwaardenvrij te maken als mogelijk – de diversiteit van de geselecteerde groep rio's dient zoveel mogelijk een afspiegeling te zijn van de maatschappij;
7. Meer profijt te halen uit het duurste selectie-instrument (het assessment).

Als overkoepelende doelstelling geldt dat de nieuwe selectieprocedure de kwaliteit van de Rechtspraak bevordert.

Aansluitend bij deze doelstellingen heeft de Raad voor de rechtspraak een aantal specifieke onderzoeksvragen geformuleerd (zie Bijlage 1). Deze vragen worden beantwoord door verschillende kwalitatieve en kwantitatieve gegevens te verzamelen door middel van interviews, vragenlijsten, documentanalyse en een analyse van de data uit het beschikbare rio-volgsysteem. De onderwerpen die in dit rapport worden belicht zijn tot stand gekomen op basis van de vragen vermeld in de *Startnotitie evaluatie selectieproces rio's*, bevindingen en aanbevelingen uit het rapport Born (2013) en de suggesties en aanbevelingen van de begeleidingscommissie, bestaande uit: Gerard Tangenberg, Else Bremer, Antonie Verheijen, Suzan Verberk, Hans Vetter, Annette Olland, Fenneke van der Grinten en Marise Born.

1.2 Context en beschrijving van de selectieprocedure

De rio-opleiding wordt gevolgd bij één van de gerechten en bij het Studiecentrum Rechtspleging (SSR). De duur en de inhoud van de opleiding hangt af van de werker-
varing en van de leerbehoefte van de rio en het gerecht. Juristen met werkervaring tot zes jaar volgen de volledige opleiding van 4 jaar, juristen met minstens vijf jaar ervaring volgen een verkorte opleiding variërend van 15 maanden tot drie jaar. De

opleiding bestaat uit een voorfase en een hoofdfase, waarin leren op de werkplek centraal staat. De rio wordt aangesteld als rechter-plaatsvervanger en doet begeleid dan wel zelfstandig rechterlijk werk.

De selectieprocedure bestaat uit de volgende onderdelen:

- Briefselectie Toetsing van formele vereisten en beslissing voortzetting procedure door de lid-secretarissen van het Presidium van de LSR, het secretariaat van de LSR, en een vertegenwoordiger van het gerecht waarbij gesolliciteerd is.
- Analytische tests Tests met als doel de beheersing van een minimumnorm verbaal redeneervermogen en abstract redeneervermogen vast te stellen, afgenomen en ontwikkeld door LTP.
- Voorgesprek Gesprek met een intern lid van de LSR, een vertegenwoordiger van een gerecht waar gesolliciteerd is en een HRM-adviseur, waarin de motivatie, overtuigingskracht, mondelinge uitdrukingsvaardigheid en het beeld van de functie worden getoetst.
- Referenties Kandidaten geven op het sollicitatieformulier twee referenten op. Deze referenties worden na afloop van het voorgesprek opgevraagd.
Bij kandidaten die werkzaam zijn of waren als advocaat of notaris, of die werkzaam zijn of waren bij een gerecht of het Openbaar Ministerie, worden ambtshalve referenties opgevraagd.
- Assessment Gesprek met een adviseur van het LTP, een aantal persoonlijkheidsvragenlijsten en een aantal tests, bedoeld om de competenties behorend bij het rechtersprofiel in kaart te brengen. Het assessment wordt geheel uitgevoerd door LTP en resulteert in een rapportage en een advies.
- Eindgesprek Drie gesprekken met steeds twee leden van de LSR (een intern en extern lid), om inzicht te krijgen in persoonlijke kwaliteiten zoals samenwerken, maatschappelijke oriëntatie en visie op het rechterschap.
- Lokaal selectie-gesprek Gesprek met de lokale selectiecommissie met nadruk op lokale aspecten: Past iemand in het team wat betreft collegialiteit en samenwerking?

De selectie voor de rio-opleiding wordt uitgevoerd door de Landelijke Selectiecommissie Rechters (LSR). Deze commissie bestaat uit zes interne leden vanuit de zittende magistratuur en zes externe leden afkomstig uit diverse sectoren. De leden worden benoemd door de Raad voor de rechtspraak. Daarnaast zijn er zes plaatsvervangend leden, waarvan drie intern en drie extern lid.

Na de openstelling van een rio-vacature en eventueel een oriënterend gesprek bij het gerecht volgt een getrapte selectieprocedure. Met uitzondering van de referentiecheck en het assessment kan na elke stap door de LSR besloten worden om de procedure niet voort te zetten met de kandidaat. De analytische tests en het assessment zijn ontwikkeld en worden afgenomen door LTP, een psychologisch advies- en assessmentbureau. Bij de briefselectie en het voorgesprek is bovendien een afgevaardigde van een gerecht waar is gesolliciteerd betrokken.

1.3 Werkwijze

De onderzoeksvragen zijn beantwoord aan de hand van interviews met verschillende betrokkenen vanuit de LSR, SSR en de Raad voor de rechtspraak, een vragenlijst afgenomen bij voormalig deelnemers aan de selectieprocedure en een analyse van gegevens over scores en beoordelingen uit de selectieprocedure en beoordelingen uit de rio-opleiding. Dit onderzoek is goedgekeurd door de Ethische Commissie Psychologie van de Rijksuniversiteit Groningen. Hieronder volgt een korte uiteenzetting van de werkwijzen die zijn gebruikt. Ook hebben we voor dit onderzoek gebruik gemaakt van documenten aangeleverd door de LSR en LTP en van de wetenschappelijke literatuur. In dit evaluatieonderzoek borduren we voort op de bevindingen en aanbevelingen uit het rapport Born (2013). Wij zullen regelmatig naar dit rapport verwijzen.

1.3.1 Interviews

Om vragen te beantwoorden over de ervaringen met, de meningen over en de uitvoering van de selectieprocedure zijn gesprekken gevoerd met 15 personen, waarvan vier leden van de LSR (twee intern en twee extern), de toenmalig voorzitter en beide lid-secretarissen van het Presidium van de LSR, een gerechtsbrede opleidingscoördinator, het afdelingshoofd HRM en organisatieontwikkeling van de Raad voor de rechtspraak, de coördinator van het secretariaat van de LSR, de voorzitter van het college van bestuur van SSR en vier leden van lokale selectiecommissies. De geïnterviewde

leden van de LSR en de lokale selectiecommissies zijn willekeurig gekozen, rekening houdend met een spreiding op basis van regio en ervaring. De gesprekken duurden ongeveer een uur en hadden een semigestructureerd karakter, waarbij gebruik werd gemaakt van een vooraf opgestelde vragenlijst, maar ook ruimte was voor aanvulling en eigen inbreng. De gesprekken zijn met toestemming van de geïnterviewden opgenomen en getranscribeerd. Vervolgens zijn de transcripten geanalyseerd volgens thematische analyse (Braun & Clarke, 2006) in het programma Atlas.Ti.

1.3.2 Vragenlijsten voormalig deelnemers

Om de ervaringen en meningen van oud-kandidaten in kaart te brengen is een vragenlijst opgesteld op basis van de organisational justice theory (Steiner & Gilliland, 1996). De uitnodiging voor deelname aan dit deelonderzoek is uitgezet bij geselecteerde- en niet geselecteerde oud-kandidaten (respectievelijk via het rio-portaal en via e-mail). In hoofdstuk 5 wordt de werkwijze in meer detail beschreven.

1.3.3 Data selectieprocedure en rio-volgsysteem

Om uitspraken te kunnen doen over de effectiviteit van de verschillende onderdelen van de selectieprocedure is het van belang om empirische relaties tussen scores op de verschillende onderdelen en latere criteriumprestaties (prestaties in de rio-opleiding) te onderzoeken. Daarvoor is gebruik gemaakt van gegevens over prestaties en beoordelingen in de selectieprocedure van de LSR, aangeleverd door de LSR en LTP en van beoordelingen uit de rio-opleiding, aangeleverd door SSR. De gegevens zijn aan elkaar gekoppeld door de LSR en volledig geanonimiseerd aan de onderzoekers geleverd. Deze gegevens zijn in de eerste plaats gebruikt om de voorspellende waarde (de predictieve validiteit) van de verschillende onderdelen van de selectieprocedure in kaart te brengen. Daarnaast zijn de gegevens ook gebruikt om uitspraken te kunnen doen over de kwaliteit van de beoordelingen uit de rio-opleiding (het criterium) en voor analyses omtrent andere aandachtspunten die tijdens de uitvoering van het project aan het licht zijn gekomen.

1.3.4 Begrippenlijst

In dit onderzoek worden verschillende psychometrische en statistische begrippen gebruikt. In het overzicht hieronder worden steeds terugkerende begrippen toegelicht.

Begrippenlijst	
Voorspeller	De scores of beoordelingen gebruikt om een voorspelling te doen over toekomstige prestaties. In deze context: de scores en beoordelingen uit de selectieprocedure.
Criterium	De te voorspellen prestaties. In deze context: de prestaties tijdens de rio-opleiding.
Construct	De eigenschap of vaardigheid die men met een test, vragenlijst of selectie-instrument beoogt te meten.
Betrouwbaarheid	• De mate waarin een testscore of een beoordeling vrij is van meetfout; de nauwkeurigheid van de meting. • Meestal uitgedrukt als Cronbach's alfa (α). • Interpretatie^a: > .80: voldoende > .90: goed
Interbeoordelaarsbetrouwbaarheid	• De mate waarin een beoordeling onafhankelijk is van de persoon die de beoordeling uitvoert. • Uitgedrukt als de intra-klasse correlatie (ICC) • Interpretatie^b: < .50: onvoldoende .50-.75: redelijk .75-.90: goed > .90: excellent

Begrippenlijst

Predictieve validiteit	• De voorspellende waarde van een voorspeller voor een bepaald criterium. • Uitgedrukt als een correlatie tussen de scores op de voorspeller en de score op het criterium (r, waarden tussen -1 en 1). • Negatieve waarden drukken een negatief verband uit, waarden dichtbij 0 laten zien dat er geen verband is, positieve waarden drukken een positief verband uit. • Interpretatie^c: <.20: klein verband .20-.50: middelgroot verband > .50:groot verband
Incrementele validiteit	• De toegevoegde waarde van het gebruik van een voorspeller bovenop een andere voorspeller, voor de predictieve validiteit. • Uitgedrukt als de toename in verklaarde variantie (ΔR^2)
Constructvaliditeit	De mate waarin een test (of interview) erin slaagt datgene te meten wat men beoogt te meten.
95% betrouwbaarheidsinterval (BHI)	Geeft de nauwkeurigheid van een bepaalde schatting weer (bijv. van een correlatie, een gemiddelde, of een betrouwbaarheidscoëfficiënt).
Klinische beoordeling	Een manier van gegevenscombinatie waarbij informatie 'in het hoofd' van de selecteur(s) gewogen en gecombineerd wordt.
Statistische beoordeling	Een manier van gegevenscombinatie waarbij informatie volgens een beslisregel of formule gewogen en gecombineerd wordt.

^a Evers, Lucassen, Meijer, & Sijtsma, 2010. ^bKoo & Li, 2016. ^c Cohen, 1998.

De selectiecommissie

In dit hoofdstuk behandelen we de totstandkoming van de LSR, de training die leden van de LSR krijgen en de ervaringen die de leden hebben met het functioneren van de LSR. Dit hoofdstuk is tot stand gekomen op basis van een analyse van documenten over de selectieprocedure en interviews met verschillende leden van de LSR en van het Presidium van de LSR.

2.1 Totstandkoming selectiecommissie

De leden van de LSR worden geworven via een open vacature. Voor rechterlijk leden verschijnen vacature op het intranet van de Raad voor de rechtspraak, voor externe leden worden advertenties geplaatst op het internet en in de krant. Geïnteresseerden schrijven vervolgens een sollicitatiebrief. Uit de brieven wordt een selectie gemaakt door de voorzitter van de LSR, leden van de Raad voor de rechtspraak en een extern persoon. Met de geselecteerde kandidaten wordt vervolgens een gesprek gevoerd. Een lid van de LSR stelde voor om in het vervolg ook een (scheidend) lid te laten plaatsnemen in de selectiecommissie voor de leden van de LSR.

Er wordt bij de selectie van selecteurs specifiek gelet op affiniteit met selecteren en opleiden. Daarnaast wordt er voor interne leden gekeken of er een spreiding is over de verschillende regio's en gerechten. De reden om externe leden op te nemen in de LSR is om een ander en breder perspectief te bieden dan de die van interne leden. Daarom wordt er voor externe leden geselecteerd op verscheidenheid in professionele achtergrond. Daarnaast is de doelstelling om ook de culturele diversiteit van de leden van de LSR te vergroten en wordt er aandacht besteed aan de man-vrouw verhouding.

In het rapport Born (2013) werd aanbevolen om bij het selecteren van selecteurs aandacht te besteden aan empathisch vermogen en sociale vaardigheden. Uit

gesprekken met leden van het Presidium van de LSR blijkt dat zij de indruk hebben dat hier in de gesprekken met kandidaat-selecteurs op wordt gelet. Een geïnterviewde merkte op dat het de vraag is of een gesprek de beste manier is om de aanwezigheid van deze wenselijke vaardigheden te waarborgen. Een van het Presidium van de LSR opperde dat het wellicht een goed idee is om een assessment in de selectieprocedure voor leden van de LSR op te nemen, maar dat dit de selectieprocedure voor kandidaat-leden wel zou verzwaren.

Zoals ook in het rapport Born (2013) aan de orde is gesteld, zorgt het structureren van interviews voor een verhoogde validiteit. De interviews met kandidaat-leden zijn momenteel niet sterk gestructureerd. Aanbevolen wordt om de interviews met kandidaat-leden te structureren en een scoringlijst te gebruiken waarop de wenselijke eigenschappen en vaardigheden expliciet vermeld worden. Daarnaast kan worden overwogen of het wenselijk en mogelijk is om een assessment aan de procedure toe te voegen die de specifieke competenties en eigenschappen zoals aanbevolen in het rapport Born (2013) in kaart brengen. Uit recenter en ander onderzoek blijkt bovendien dat cognitieve capaciteiten en *dispositional reasoning* (kennis hebben van de invloed van persoonlijkheidskenmerken en situaties op de uiting van persoonlijkheidskenmerken in termen van gedrag (De Kock, Lievens & Born, 2018), samenhangen met het maken van accuratere beslissingen op basis van gestructureerde interviews met kandidaten. Aangezien het voeren van interviews met kandidaten de kerntaak van leden van de LSR is, zouden kandidaat-leden van de LSR op deze capaciteiten getest kunnen worden.

AANBEVELINGEN 1 – SELECTIE LEDEN LSR

- Het structureren van selectiegesprekken met kandidaat-leden.
- Overwogen kan worden om bij de selectie van kandidaat-leden aandacht besteden aan empathisch vermogen, sociale vaardigheden, cognitieve capaciteiten en *dispositional reasoning*, eventueel in de vorm van een test.

2.2 Training selectiecommissie

Voor nieuwe leden van de LSR wordt een kennismakingsdag georganiseerd waarin informatie wordt verstrekt over de selectieprocedure en een training wordt verzorgd over interviewtechnieken op basis van de STAR-methode. Dit is een vorm van situationeel interviewen waarbij gevraagd wordt naar *situaties*, de eigen *taak* daarin, de

uitgevoerde acties en het resultaat daarvan. Deze training wordt verzorgd door LTP. De selecteurs gaven aan dat zij deze trainingen als nuttig hebben ervaren. Eén extern lid was van mening dat het wel wat weinig was, zeker voor nieuwe leden die nog niet erg veel ervaring hebben met het voeren van selectiegesprekken. Er had meer ruimte mogen zijn voor oefening en rollenspel en meer aandacht besteed kunnen worden aan de specifieke thema's die met kandidaten besproken worden in de gesprekken, bijvoorbeeld in de vorm van een casus. Sommige leden geven aan dat zij op eigen initiatief andere activiteiten hebben ondernomen die zij belangrijk vonden om de functie van selecteur goed te kunnen uitvoeren. Voorbeelden zijn het bijwonen van zittingen (externe leden) of het deelnemen aan tests van het LTP. Een geïnterviewde oppert het idee om alle selecteurs de analytische tests en wellicht ook delen van het assessment eens te laten doen, om meer inzicht te krijgen in de selectie-instrumenten op basis waarvan beoordelingen en beslissingen gemaakt worden.

Voor zittende leden van de LSR wordt jaarlijks een terugkomdag georganiseerd waarop vaardigheden worden opgefrist en onderwerpen die aandacht of discussie verdienen gezamenlijk worden besproken. Op die dag kunnen leden van de LSR input leveren aan elkaar en aan het Presidium over onderwerpen die volgens hen aandacht verdienen. De terugkomdagen worden positief gewaardeerd door de geïnterviewde leden. Als verbeterpunten worden genoemd dat er meer ruimte mag zijn voor het oefenen en opfrissen van interviewvaardigheden, de uitgangspunten van de selectie, de gehanteerde begrippen en de wijze van toetsing. Op de laatste terugkomdag is er een training geweest over interviewvaardigheden, waarbij selecteurs zelf casussen konden aandrazen, in rollenspellen konden oefenen en feedback kregen. Een dergelijke opzet zou iedere terugkomdag een vast onderdeel kunnen zijn. In het rapport Born (2013) en in het rapport Stevens (2010) werd aanbevolen om het aantal selecteurs terug te brengen zodat de selecteurs snel veel ervaring kunnen opdoen. Die aanbeveling is opgevolgd; voorheen bestond de LSR uit 70 leden en inmiddels uit 12 leden. Bovendien gaf Born de aanbeveling om selecteurs een frame-of-reference training te laten volgen, omdat dit type training volgens de wetenschappelijke literatuur het meeste effectief is (Woehr & Huffcutt, 1994). Geen van de geïnterviewden heeft aangegeven dat dit type training is aangeboden. Wel is er een interview-training volgens de STAR-methode gegeven en wordt er aandacht besteed aan het voorkomen van beoordelingsfouten zoals het halo-effect. Dit laatste komt overeen met rater-error training. Dit type training is minder effectief in het bevorderen van goede beoordeling dan frame-of-reference training (Woehr & Huffcutt, 1994). In een frame-of-reference training wordt vooral veel aandacht besteed aan het bereiken van

een gemeenschappelijke conceptualisatie van de verschillende competenties en eigenschappen die beoordeeld worden in de interviews. Een dergelijke training moet dus op maat gemaakt worden en aansluiten bij de eigenschappen en competenties die in de betreffende interviews beoordeeld worden. We bevelen aan om in samenwerking met een assessmentbureau een dergelijke training te ontwikkelen, aan te bieden aan nieuwe en zittende leden en deze training een vast onderdeel van de initiële training en vervolgtraining te maken.

AANBEVELINGEN 2 – TRAINING SELECTIECOMMISSIE

- Bied nieuwe leden veel mogelijkheden om zich te oriënteren op het beroep van rechter (externe leden) en op de selectieprocedure, bijvoorbeeld door zittingen bij te wonen en deel te nemen aan verschillende tests uit de selectieprocedure.
- Besteed op kennismakingsdagen en terugkomdagen meer tijd aan oefenen, rollenspelen en casussen.
- Bied een frame-of-reference training aan aan selecteurs.

2.3 Ervaringen met het functioneren van de selectiecommissie

De geïnterviewde leden waren het er allen over eens dat de LSR als geheel goed functioneert. Veel geïnterviewden benoemen de grote zorgvuldigheid en betrokkenheid van het Presidium en de leden en roemen de inspanningen die verricht worden om het selectieproces steeds te evalueren en te verbeteren. De samenwerking tussen interne en externe leden wordt door beide partijen als prettig en belangrijk ervaren. Een intern lid benoemt dat de externe leden een goede spiegel voorhouden over de verwachtingen die de maatschappij heeft van rechters. De externe leden voelen zich gewaardeerd en gelijkwaardig behandeld door de interne leden en vinden de samenwerking prettig. Als aandachtspunt noemde één lid dat de aansturing vanuit het Presidium vaak top-down gebeurt. Deze persoon zou het op prijs zou stellen als de leden wat meer betrokken worden bij het bepalen van de koers van de LSR en de besluitvorming. Een ander lid was van mening dat de selectieprocedure erg gericht is op het beoordelen van geschiktheid. Deze persoon zou graag zien dat er meer wordt gezocht naar uitzonderlijkheid en de bijdrage die een kandidaat kan leveren aan de rechtspraak.

Organisatie van de selectieprocedure

In dit hoofdstuk behandelen we de meningen van de leden van de LSR over de secretariële ondersteuning van de LSR en eventuele knelpunten rondom het halen van de gewenste doorlooptijd (de duur van de selectieprocedure). Dit hoofdstuk is tot stand gekomen op basis van een analyse van documenten over de selectieprocedure en interviews met leden van de LSR, leden van het Presidium en het secretariaat van de LSR.

3.1 Secretariële ondersteuning

De geïnterviewde leden en de leden van het Presidium zijn allen tevreden over de secretariële ondersteuning, soms met de kanttekening 'gegeven die tijd en middelen die voorhanden zijn'. Knelpunten zitten met name in de digitale middelen en de automatisering van processen. Dat zorgt voor verminderde efficiëntie van werken en moeilijkheden in het bijhouden en veilig bewaren van gegevens. Hier wordt momenteel aan gewerkt en de verwachting is dat de geplande aanpassingen verbetering zullen opleveren. Daarnaast zijn er geen aandachtspunten over de secretariële ondersteuning genoemd.

3.2 Doorlooptijd selectieprocedure

In het verleden was de gemiddelde doorlooptijd van de selectieprocedure 3 tot 4 maanden. In de rapporten Born (2013) en Stevens (2010) is aanbevolen om die doorlooptijd in te korten. De huidige streeftijd is 11 weken. Deze doorlooptijd wordt vooraf naar kandidaten gecommuniceerd. Uit gesprekken met het secretariaat van de LSR blijkt dat de huidige doorlooptijd is in de regel ook 11 weken bedraagt en dat

het soms zelf lukt om de procedure in 10 weken af te ronden. Het lukt goed om de doorlooptijd te halen omdat de planning met alle betrokkenen vanuit de LSR al een jaar van tevoren wordt gemaakt. Eén keer is de doorlooptijd vanwege ongunstige schoolvakanties met een week verlengd. In de huidige vorm van de procedure lijkt het niet mogelijk om de doorlooptijd van de selectieprocedure verder te verkorten. De doorlooptijd van de procedure is dus naar tevredenheid ingekort.

AANBEVELING 3 – ORGANISATIE SELECTIEPROCEDURE

- Er lijken geen dringende aandachtspunten te zijn het gebied van de organisatie van de selectieprocedure. De kwaliteit en beschikbaarheid van digitale middelen en de doorlooptijd zullen ook in de toekomst, zeker in het geval van veranderingen in de procedure, aandachtspunten moeten blijven.

Analyse proces van selectie en selectiecriteria

Dit hoofdstuk is tot stand gekomen op basis van een analyse van documenten over de procedure aangeleverd door de LSR, documenten over de gebruikte tests en het assessment aangeleverd door LTP, de gevoerde interviews en de wetenschappelijke literatuur. De verschillende stappen in de selectieprocedure worden afzonderlijk besproken. Daarin komen de volgende aspecten aan bod: de gevolgde procedure, de betrouwbaarheid en constructvaliditeit van de tests en beoordelingen, de ervaringen van betrokkenen bij de selectie en daaruit voortvloeiende aanbevelingen.

4.1 Meningen van betrokkenen over de selectieprocedure

De geïnterviewde leden van de LSR en het Presidium ervaren de procedure unaniem als zeer zorgvuldig. De geïnterviewden realiseren zich ook dat het een zware procedure is voor de kandidaten. Dat wordt echter door de meesten niet als negatief punt gezien, een zware procedure waarin hoge eisen worden gesteld aan de kandidaten is ook op zijn plaats voor dit beroep. Wel twijfelt een aantal betrokkenen of er in de procedure niet te hoge eisen gesteld worden, vooral omdat zij met regelmaat personen zien stranden waarvan het beeld bestaat dat zij zeer geschikt zijn. Bij het bepalen van hoogte van de eisen moet een afweging gemaakt worden over de ernst van het toelaten van ongeschikte kandidaten (vals-positieven) ten opzichte van de ernst van het afwijzen van geschikte kandidaten (vals-negatieven). Deze selectiefouten kunnen geen van beide geheel voorkomen worden. We merken op dat een procedure met hoge eisen en hoge grensscores er vooral voor zal zorgen dat het aantal vals-positieve beslissingen wordt teruggedrongen, tegen de prijs van een toename in vals-negatieve beslissingen. Andersom zal een procedure met minder hoge eisen ervoor zorgen dat het aantal vals-negatieve beslissingen daalt, ten koste van een toename in vals-positieve beslissingen. Welk van deze twee typen beslissingsfouten onwenselijker

zijn, is beleidskeuze die door de Raad van de rechtspraak gemaakt zal moeten worden.

Het is een beleidskeuze of de doelstelling van de selectieprocedure vooral moet zijn om het aantal vals-positieven of het aantal vals-negatieve beslissingen tot een minimum te beperken. Een betrokkene bij de opleiding was van mening dat vals-negatieve beslissingen onwenselijker zijn dan vals-positieve beslissingen. Die vals-positieve kandidaten worden tijdens de opleiding wel gecorrigeerd. Een andere geïnterviewde merkte juist als positief punt op dat de uitval in de rio-opleiding aanzienlijk is verminderd en dat de nieuwe procedure van selectie, met de daarbij horende hoge eisen, daar vermoedelijk een rol in speelt. Een grote kans op uitval tijdens de opleiding is volgens deze persoon onwenselijk, omdat dit de stap van een andere carrière naar het rechterschap een risicovolle onderneming maakt. Dat komt de aantrekkelijkheid van het rechterschap niet ten goede. Tegelijkertijd noemen enkele anderen ook dat de huidige opleiding en de hoge eisen in de selectieprocedure die daaraan voorafgaat veel kandidaten met voornamelijk academische werkervaring lijkt aan te trekken. De zwaarwegendheid van analytische vaardigheden in de selectieprocedure zou daaraan bij kunnen dragen.

De meeste geïnterviewden waren van mening dat de procedure volledig is wat betreft de eigenschappen en competenties die men in huis moet hebben. Enkel waren van mening dat een aantal aspecten nog toegevoegd zouden kunnen worden, zoals schrijfvaardigheid, juridische kennis en zittingsvaardigheden. Ook de geïnterviewde betrokkenen vanuit de opleiding zijn van mening dat goede juridische vaardigheden essentieel zijn voor goed rechterschap. Die worden nu niet getoetst in de selectieprocedure. Desondanks vallen er volgens betrokkenen bij de opleiding en de lokale gerechten momenteel weinig rio's uit door een gebrek aan inhoudelijke vakbeheersing. Juridische vaardigheden worden ook genoemd als iets wat door 'training on the job' tot stand komt. In de selectie gaat het erom die kandidaten te kiezen die de nodige inhoudelijke vaardigheden kunnen ontwikkelen. Een ander punt, genoemd door twee geïnterviewden, is of indrukken uit gesprekken, die nu veel gebruikt worden als basis voor selectiebeslissingen, de beste basis vormen om kandidaten te beoordelen op de gewenste competenties en vaardigheden. Beiden vragen zich af of in ieder geval sommige competenties niet beter in de vorm van een rollenspel, opdracht, of werksimulatie gemeten kunnen worden.

4.2 Briefselectie

De belangrijkste functie van de briefselectie is de controle op de formele vereisten. Enkele van deze vereisten zijn wettelijk vastgelegd, zoals het bezit van de Nederlandse nationaliteit en een Mastergraad in de Rechten. Andere vereisten zijn minimaal twee jaar relevante juridische werkervaring (opgedaan na afstuderen) en voldoende maatschappelijke ervaring. Deze vereisten worden gecontroleerd door het secretariaat van de LSR, de lid-secretarissen van de LSR en een vertegenwoordiger van het gerecht waarbij de kandidaat gesolliciteerd heeft. Zij ontvangen alle dossiers digitaal en komen vervolgens fysiek of via videoconferentie bij elkaar om de dossiers één voor één te bespreken.

4.2.1 Controle formele vereisten

De leden van het Presidium van de LSR geven aan dat er nooit afgeweken wordt van de formele vereisten. De beoordeling van de twee jaar juridische werkervaring en maatschappelijke ervaring is echter niet geformaliseerd of gestandaardiseerd. Wel is er volgens de geïnterviewde betrokkenen vanuit de LSR informeel overeenstemming over wanneer er voldaan wordt aan deze eisen. Enkele geïnterviewden die niet betrokken zijn bij de briefselectie gaven echter aan dat er wel eens twijfels bestaan over het voldoen aan minimaal twee jaar relevante juridische werkervaring. Die twijfels hebben vooral betrekking op kandidaten die vooral academische werkervaring hebben en soms alleen promotieonderzoek hebben uitgevoerd. De vraag is of dit voldoende relevant is en een voldoende basis biedt om de rio-opleiding te beginnen. Aanbevolen wordt om de gehanteerde definities van relevante juridische werkervaring en voldoende maatschappelijke ervaring en de onderbouwing van die definities te expliciteren en te communiceren naar de lokale gerechten, de opleiders en het opleidingsinstituut.

4.2.2 Andere criteria

Tijdens de briefselectie wordt niet alleen gekeken naar het voldoen aan de formele vereisten, maar ook naar de werkervaring, cijfers behaald tijdens de studie, de motivatie, de manier waarop iemand schrijft en de inhoudelijke expertise. Waar op wordt gelet is deels afhankelijk van de vacature en het profiel waarnaar het betreffende gerecht op zoek is. Daarbij kijken de lid-secretarissen met een andere bril naar de

kandidaat dan de afgevaardigde van het gerecht. Vanuit de LSR wordt gekeken naar algemene geschiktheid, terwijl het gerecht kijkt naar geschiktheid voor de specifieke vacature die daar openstaat. De eisen waaraan een kandidaat moet voldoen om door te gaan naar de volgende stap in de selectieprocedure is hierbij mede afhankelijk van het aantal kandidaten voor een vacature; wanneer er veel kandidaten zijn voor een vacature worden vaak hogere eisen gesteld.

4.2.3 Ervaring betrokkenen

De betrokkenen vanuit de LSR zijn tevreden met de huidige gang van zaken en merken op dat die verbeterd is in vergelijking met het verleden. Eerder zag men de dossiers pas op de dag dat er overleg gepleegd werd, nu is er veel meer tijd om zich van tevoren goed in te lezen. Een lid van het Presidium van de LSR merkt op dat de briefselectie wellicht nog wat zorgvuldiger zou kunnen wat betreft de gehanteerde criteria en vereisten. Daar komt wel bij dat er vaak veel dossiers beoordeeld moeten worden in relatief korte tijd. Een zorgvuldigere procedure zou ook de tijdsinvestering verhoogen. Ook merken verschillende leden op dat het digitale sollicitatieformulier soms niet naar behoren functioneert, wat erg vervelend is voor kandidaten.

4.2.4 Aanbevelingen

Het was niet mogelijk om de betrouwbaarheid en validiteit van de briefselectie te onderzoeken op basis van empirische gegevens. Om het proces van de briefselectie zo betrouwbaar en valide mogelijk te laten verlopen zonder een onevenredige grote tijdsinvestering van de selecteurs te vragen, bevelen we aan om definities van de formele vereisten en de aanvullende zaken zoals cijfers en cv te expliciteren en op te nemen in een beoordelingsformulier voor de briefselectie, zoals ook aanbevolen in het rapport Born (2013). Dat kan in samenwerking met personen die vanuit de lokale gerechten aan de briefselectie deelnemen, zodat ook de doelstellingen en belangrijke aandachtspunten vanuit de lokale gerechten op het formulier vertegenwoordigd zijn. Ook kan er een 'open' deel aan het formulier toegevoegd worden, zodat de lokale gerechten daar vooraf aandachtspunten kunnen invullen die passen bij de specifieke vacature of de huidige behoeften van het gerecht.

AANBEVELINGEN 4 – STRUCTUREREN BRIEFSELECTIE

- We bevelen aan om te expliciteren wat er verstaan wordt onder relevante juridische werkervaring en wanneer er voldaan is aan voldoende maatschappelijke ervaring.
- Expliciete definities van de formele vereisten en van andere zaken waarop men tijdens de briefselectie wil letten kunnen worden opgenomen in een beoordelingsformulier voor de briefselectie.

4.3 Analytische tests

Na de briefselectie worden de kandidaten uitgenodigd voor deelname aan de analytische tests. Dit zijn de Matrix test, met als doel het abstract redeneervermogen te meten, en de Test voor Verbale Relaties en Analogieën (TVRA), met als doel het verbaal redeneervermogen te meten. Daarnaast wordt ook de Taalgevoeltest afgenomen, met als doel kennis van en gevoel voor zinsbouw in de Nederlandse taal te meten. De score op de Taalgevoeltest wordt echter niet meegenomen in de voortzittingsbeslissing die volgt op deze stap. De tests worden afgenomen door LTP. De scores worden weergegeven aan de hand van normscores op een schaal van 1 tot 19, genormeerd op een normgroep bestaande uit personen met WO-opleidingsniveau. De voorzettingsbeslissing wordt volgens de regels in onderstaande tabel genomen. Als hieraan niet wordt voldaan valt de kandidaat af.

Scores	Beslissing
Beide scores > 10	Door naar voorgesprek
Eén score > 13, één score tussen 8-10	Door naar voorgesprek
Eén score > 10, één score tussen 8-10	Herkansing

Voor deze gang van zaken is gekozen naar aanleiding van aanbevelingen uit het rapport Born (2013). Uit een veelvoud aan wetenschappelijk onderzoek is gebleken dat cognitieve capaciteiten één van de beste voorspellers voor latere baanprestaties zijn (Schmidt & Hunter, 1998) en een sterker verband met baanprestaties tonen naarmate de complexiteit van de baan toeneemt (Kuncel & Hezlett, 2010). De analytische tests

hebben, op basis van de wetenschappelijke literatuur, waarschijnlijk de hoogste predictieve validiteit van alle onderdelen uit de selectieprocedure.

4.3.1 Ervaringen van betrokkenen

Over de aanwezigheid en vooral de zwaarwegendheid van de analytische tests in de procedure bestaat controverse (Ahsmann, 2018; Hofhuis, 2018). Deze controverse wordt door veel geïnterviewden benoemd. Toch vinden de meeste geïnterviewden het een goede zaak dat deze tests deel uitmaken van de selectieprocedure. Redenen daarvoor zijn de wetenschappelijke onderbouwing voor de validiteit van dit soort tests, de objectiviteit van het instrument en het beschouwen van analytische vaardigheden als een basisvoorwaarde voor goed rechterschap. Ook de efficiëntie van tests werd een enkele keer als voordeel genoemd; het is erg kostbaar om alle kandidaten naar een volledig assessment te laten gaan. Wel twijfelden enkele betrokkenen aan de harde grens die bij dit onderdeel gehanteerd wordt. Enkel vroege zich af of het niet beter zou zijn om de scores op de tests te bezien samen met andere relevante informatie, zoals ervaring en motivatie. De voornaamste reden voor deze twijfels is het afvallen van kandidaten die door henzelf of door collega's erg geschikt werden geacht, waaronder kandidaten die eerder naar tevredenheid als rechter-plaatsvervanger hebben gefungeerd. De controverse rondom dit onderdeel lijkt dus met name betrekking te hebben op de gehanteerde harde grensscores en niet zozeer op het opnemen van analytische tests in de procedure.

Een belangrijk aandachtspunt dat door veel betrokkenen genoemd is, is het effect van leeftijd op de kans om de tests met goed gevolg af te leggen. Oudere kandidaten behalen lagere scores op de analytische tests. De betrokkenen vinden dat hier aandacht aan besteed moet worden, maar dat dit verschil niet per definitie problematisch hoeft te zijn. Ook oudere kandidaten moeten over de nodige kwaliteiten beschikken. Bovendien bestaat de indruk dat dit verschil deels verklaard kan worden door een gebrek aan oefening en wellicht zelfs een onderschatting van de zwaarte van de procedure bij oudere, meer ervaren kandidaten. Een ander genoemd aandachtspunt is dat de strenge grens die op dit onderdeel wordt gehanteerd wellicht geschikte personen afschrikt om te solliciteren. Ook de mogelijkheid tot herkansing is als aandachtspunt genoemd. De indruk bestaat dat er dan wel erg vaak veel beter wordt gescoord na herkansing. Ten slotte vragen enkelen zich af of de specifieke tests die gebruikt worden in de procedure wel voldoen aan de wetenschappelijke standaard. Deze aandachtspunten zijn niet opgenomen in het oorspronkelijk onderzoeks-

plan, maar we besteden er hier en in nadere analyses waar mogelijk aandacht aan (zie paragraaf 6.8).

4.3.2 Kwaliteit van de tests

Uit rapportages verstrekt door LTP blijkt dat de scores op de Matrix test en de TVRA voldoende betrouwbaar zijn (Matrix: $\alpha = .81$, TVRA: $\alpha = .87$). Verder onderzoek naar de constructvaliditeit en de predictieve validiteit van de testscores ontbreekt grotendeels. Correlaties met scores op andere capaciteitentests (de zogenoemde convergente validiteit) zijn onderzocht, maar alleen op basis van tests ontwikkeld door LTP die andersoortige capaciteiten beoogden te meten. De normen voor de tests worden bepaald aan de hand van personen met WO-opleidingsniveau die aan de tests deelnemen. Het is niet duidelijk hoe representatief deze normgroep is. De scores op de Taalgevoeltest zijn onvoldoende betrouwbaar ($\alpha = .69$). Er zijn weinig gegevens over de validiteit van deze test, maar de correlaties met scores op andere capaciteitentests waren klein. In de documentatie van LTP wordt ook aangegeven dat deze test slechts een signaalfunctie mag hebben en als input voor verder assessments en gesprekken kan dienen.

4.3.3 Aanbevelingen

De betrouwbaarheid van de Matrix test en de TVRA is voldoende, hoewel tests met een betrouwbaarheid hoger dan .90 de voorkeur zouden hebben, gezien de harde grensscore die in deze stap gehanteerd wordt. Verder is er weinig informatie beschikbaar over de psychometrische kwaliteit (betrouwbaarheid en validiteit) van de tests. De tests zijn niet ter beoordeling aangeboden aan de Commissie Testaangelegenheden Nederland (COTAN). De COTAN beoordeelt tests en vragenlijsten die in Nederland worden gebruikt op verschillende relevante criteria. Voor een beoordeling dient de uitgever van de test zelf de benodigde informatie en evidentie aan te leveren. We bevelen aan dat er uitgebreider onderzoek plaatsvindt naar de psychometrische kwaliteit van de tests en dat de resultaten ter beoordeling aan de COTAN worden aangeboden, of dat er gebruik wordt gemaakt van tests die wel door de COTAN zijn beoordeeld en over een goede psychometrische kwaliteit beschikken. Daarnaast bevelen we aan om de huidige Taalgevoeltest in de toekomst niet meer in te zetten. Overwogen kan worden om een ander instrument dat taalvaardigheid meet in te zetten, waarvan de psychometrische kwaliteit goed onderbouwd is.

AANBEVELINGEN 5 – ANALYTISCHE TESTS

- We bevelen aan om de Taalgevoeltest in de toekomst niet meer op te nemen in de selectieprocedure.
- We raden aan om tests te gebruiken die in voldoende mate zijn onderzocht op psychometrische kwaliteit (betrouwbaarheid en validiteit) en waarvan de resultaten laten zien dat de tests aan de wetenschappelijke standaard voldoen.
- Bij voorkeur worden tests gebruikt die op relevante punten minstens als ‘voldoende’ zijn beoordeeld door de COTAN.

4.4 Voorgesprek

In het voorgesprek spreken de kandidaten ongeveer 30 minuten lang met een lid van de LSR, een afgevaardigde van een gerecht waar de kandidaat gesolliciteerd heeft en een HRM-adviseur. Voor de voorgesprekken is door de LSR een agenda opgesteld met de volgende thema's: *motivatie, overtuigingskracht, mondelinge uitdrukkingsvaardigheid* en het *beeld van de functie*. De leden van de commissie beschikken vooraf over het sollicitatieformulier en de cijferlijsten van de kandidaten, maar niet over de resultaten van de analytische tests. Het uitgangspunt bij de voorgesprekken is dat de kandidaat bij twijfel door mag naar het assessment en het eindgesprek.

4.4.1 Procedure voorgesprek

De geïnterviewde leden van de LSR stellen dat de agenda met onderwerpen die aan de orde moeten komen tijdens de voorgesprekken wordt gevolgd, hoewel de tijd die aan ieder onderwerp wordt besteed kan variëren per kandidaat. Wel komen er ook andere zaken aan de orde, bijvoorbeeld opvallende zaken uit het cv en maatschappelijke nevenactiviteiten. Gezien de doelstelling om te selecteren aan de hand van de competenties uit het rechtersprofiel valt bovendien op dat de punten op de agenda van het voorgesprek niet direct overeenkomen met die competenties. Na het voeren van een gesprek vullen de leden van de dagcommissie voorgesprek elk een beoordelingsformulier in met daarop de thema's waarop beoordeeld dient te worden en ruimte voor toelichting. Aan het einde van de dag volgt een overleg, waarin de leden van de dagcommissie eerst ieder naar hun oordeel gevraagd worden (ja, twijfel, nee) en dat oordeel vervolgens kunnen toelichten. Daarna wordt in overleg consensus

gezocht. In de meeste gevallen is consensus snel bereikt. De beoordelingen op de agendapunten zijn volgens de geïnterviewden het meest doorslaggevend voor de voortzettingsbeslissing. Hoe hoog men de lat legt bij het voorgesprek is volgens een geïnterviewde deels afhankelijk van het aantal kandidaten voor een vacature. Dit is wel in mindere mate het geval sinds de kandidaten één voorgesprek voeren, in plaats van een gesprek voor iedere vacature waarop zij hebben gesolliciteerd, zoals in het verleden het geval was.

4.4.2 Ervaringen leden LSR

De geïnterviewde leden van de LSR geven aan dat zij het voorgesprek een goede en noodzakelijk filter vinden, waarbij het uitgangspunt van het voordeel van de twijfel positief wordt gewaardeerd. Wel wordt het voorgesprek gezien het 'voordeel van de twijfel' principe, vaak genoemd als het minst belangrijke onderdeel van de selectieprocedure. Toch zijn de leden van de LSR van mening dat dit onderdeel moet blijven bestaan. Gezien de efficiëntie van de procedure is het noodzakelijk om op dit moment in de procedure alvast een schifting te maken. Ook over de gehanteerde thema's en het beoordelingsformulier is men tevreden. Een geïnterviewde merkt op dat die formulieren niet altijd volledig ingevuld worden. Dit wordt genoemd als verbeterpunt, ook omdat de verslagen van het voorgesprek weer ingezien worden door de personen die de eindgesprekken voeren. Als positief punt wordt genoemd dat het voorgesprek een wederkerig karakter heeft, waarbij ook de kandidaat de geschiktheid van de functie kan toetsen. Een ander genoemd positief aspect is de verbinding die wordt gelegd tussen de landelijke en de lokale selectieprocedure. Dat is echter ook een aandachtspunt. Enkele leden van de LSR noemden dat de lokale afgevaardigden niet allemaal kijken naar algemene geschiktheid, maar soms te veel vanuit de behoeften en voorkeuren van het eigen gerecht kijken. Dat is niet de insteek. Ook zijn de lokale afgevaardigden meestal niet getraind in het voeren van selectiegesprekken. Dat zou wel wenselijk zijn.

4.4.3 Kwaliteit voorgesprekken

Aan de hand van gegevens over beoordelingen op de voorgesprekken (tussen de 86 en 103 kandidaten, afhankelijk van het aantal ontbrekende gegevens) is de interbeoordelaarsbetrouwbaarheid van de oordelen voortkomend uit de voorgesprekken in kaart gebracht. Deze gegevens waren afkomstig van kandidaten die zijn toegelaten

tot de rio-opleiding. De betrouwbaarheid van de totaalscore (berekend op basis van de gemiddelde score per beoordelaar) was onvoldoende en de betrouwbaarheid van het eindoordeel van het voorgesprek (ja, twijfel, nee) was gelijk aan 0 en was dus onvoldoende. De interbeoordelaarsbetrouwbaarheid voor de verschillende thema's was ook onvoldoende tot redelijk (zie de notitie onderaan Tabel 1). Dit betekent dat de beoordelingen van de verschillende beoordelaars op basis van het voorgesprek in slechts in beperkte mate met elkaar overeenkomen. Hierbij moet de kanttekening geplaatst worden dat de analyse alleen gebaseerd is op oordelen over kandidaten die tot de rio-opleiding zijn toegelaten, waardoor restrictie in range waarschijnlijk tot een onderschatting van de interbeoordelaarsbetrouwbaarheid heeft geleid. In het rapport Born (2013, p. 80) is een vergelijkbare analyse uitgevoerd op basis van gegevens van zowel afgewezen als toegelaten kandidaten. Die analyse resulteerde in wat hogere betrouwbaarheidscoëfficiënten van redelijke grootte. Dit zijn waarschijnlijk accuratere schattingen van de betrouwbaarheid dan we hier konden realiseren. Wel zijn de opzet en thema's die aan bod komen in het voorgesprek veranderd sinds het rapport Born is geschreven. Stellige uitspraken over de betrouwbaarheid van de beoordelingen uit het voorgesprek zijn op basis van deze resultaten niet te doen.

Tabel 1 Interbeoordelaarsbetrouwbaarheid van totaalscores en themascores uit het voorgesprek.

Onderdeel/score	ICC
Motivatie	.28 [< .01, .49]
Overtuigingskracht	.58 [.42, .71]
Mondelinge uitdrukkingsvaardigheid	.54 [.37, .68]
Beeldfunctie	.51 [.31, .65]
Eindoordeela	-.03 [-.05, -.01]
Totaalscore voorgesprek	.48 [.25, .64]

Notitie. ICC = intra-class correlatie, 95% betrouwbaarheidsinterval tussen haakjes. Interpretatie: < .50 = onvoldoende, 0.5 – 0.75 = redelijk, 0.75 – 0.9 = goed, > .90 = excellent (Koo & Li, 2016).

^a Krippendorff's alfa voor ordinale data i.p.v. ICC (3 ordinale categorieën).

In Tabel 2 staat de samenhang tussen de oordelen op de verschillende thema's die behandeld worden in de voorgesprekken. Aangezien de thema's inhoudelijk verschillen en afzonderlijke competenties en vaardigheden zouden moeten meten, zou de samenhang (correlaties) tussen de oordelen op de verschillende thema's laag moeten zijn. Dit geeft de zogenoemde divergente validiteit, ofwel de onderscheidbaarheid van de thema's weer. Zoals ook gevonden in het rapport Born (2013, p. 82) zijn de correlaties hoog. Op basis van de inhoud van de thema's zouden we bijvoorbeeld niet verwachten dat er een sterke samenhang bestaat tussen de beoordelingen op het *beeld van de functie* en *overtuigingskracht*, terwijl de resultaten in Tabel 2 hier wel een sterke samenhang tussen de beoordelingen op deze twee thema's laten zien. Het is waarschijnlijk dat er in de voorgesprekken een halo-effect optreedt; wanneer een beoordelaar een positieve indruk van een kandidaat heeft, zal de beoordelaar geneigd zijn om die kandidaat positief te beoordelen op alle thema's. Dit effect zorgt er waarschijnlijk voor dat de competenties die beoordeeld worden in het voorgesprek niet onafhankelijk en onderscheidbaar gemeten worden. De correlaties tussen de scores op de verschillende thema's en het eindoordeel zijn niet onderzocht omdat er alleen gegevens beschikbaar waren van kandidaten die zijn toegelaten tot die rio-opleiding. Er was daarom erg weinig spreiding in de eindoordelen (bijna allemaal positief). In het rapport Born zijn deze relaties wel onderzocht op basis van gegevens van toegelaten en afgewezen kandidaten.

Tabel 2 Divergente validiteit van de thema's beoordeeld binnen het voorgesprek.

Thema	Motivatie	Overtuigingskracht	Mondelinge uitdrukkingsv.
Overtuigingskracht	.67*		
Mondelinge uitdrukkingsv.	.59*	.71*	
Beeld functie	.60*	.47*	.48*

Notitie. $n = 341\text{--}333$ oordelen van individuele selecteurs. * $p < .05$.

4.4.4 Aanbevelingen

Een eerste aanbeveling is om ook de afgevaardigden van de lokale gerechten te trainen in het voeren van selectiegesprekken en hen goed voor te lichten over de doelstelling van het voorgesprek. Ook is het noodzakelijk dat erop wordt toegezien dat de beoordelingsformulieren volledig worden ingevuld, zowel voor de kwaliteit van de besluitvorming en de bruikbaarheid van de verslagen later in de procedure, als voor de het gebruik van de gegevens in evaluatieonderzoek. In de gegevens die gebruikt konden worden voor de betrouwbaarheidsanalyses ontbraken veel beoordelingen.

Hoewel stellige uitspraken over de betrouwbaarheid van de oordelen voortkomend uit het voorgesprek niet gedaan kunnen worden, lijkt er in ieder geval ruimte te zijn voor verbetering. Een aandachtspunt daarbij is, zoals ook geformuleerd is in de Startnotie, de uniformiteit van de procedure, ofwel de mate van structuur en standaardisatie. Hoewel de onderwerpen van de agenda's gevolgd worden, wordt er ook over andere onderwerpen gevraagd die niet voor iedere kandidaat dezelfde zijn. In de literatuur worden vier niveaus van standaardisatie onderscheiden (niveau 4 is het hoogst mogelijke niveau), waarbij onderscheid gemaakt wordt in de structuur van het interviewproces en de structuur van de scoring (Huffcutt & Arthur, 1994). Op basis van de opzet van het voorgesprek (gedefinieerde onderwerpen met voorbeeldvragen; niveau 2) en de scoring (beoordelingen op een eenvoudige schaal naar onderwerp; niveau 2) heeft het voorgesprek een structuurniveau van 2. Een hogere mate van structuur verhoogt de betrouwbaarheid van interviewscores (Huffcutt et al., 2013). De wijze waarop een hogere mate van structuur bereikt kan worden zal verderop in dit document besproken worden.

AANBEVELINGEN 6 – STRUCTUUR VOORGESPREK

- Het zou bevorderlijk zijn voor het selectieproces als ook lokale selecteurs getraind worden in interviewvaardigheden.
- Het is belangrijk om erop toe te zien dat de beoordelingsformulieren volledig ingevuld worden.
- We bevelen aan om de voorgesprekken sterker te structureren.

4.5 Referenties

Als na afloop van het voorgesprek wordt besloten om de procedure met een kandidaat voort te zetten worden referenties opgevraagd bij de referenten die de kandidaten hebben opgegeven op het sollicitatieformulier. Deze referenties worden in de besluitvorming volgend op het assessment en het eindgesprek meegenomen. De referenties worden voor het eindgesprek met de kandidaat gedeeld en tijdens het eindgesprek met de kandidaat besproken. Voor kandidaten die werkzaam zijn of zijn geweest als advocaat of notaris, of werkzaam zijn of zijn geweest bij een gerecht of het Openbaar Ministerie, worden ambtshalve referenties opgevraagd. Omdat dit vaak veel tijd in beslag neemt worden deze referenties vaak al voor het voorgesprek opgevraagd. Voor advocaten gebeurt dat bij de president van de rechtbank en de deken van de Orde van Advocaten, voor notarissen bij de voorzitter van de Kamer van het Notariaat. Voor (voormalig) medewerkers bij een gerecht of het Openbaar Ministerie gebeurt dat bij de president van het gerecht of de hoofdofficier van justitie. Deze referenties worden opgevraagd om na te gaan of kandidaten te maken hebben gehad met berispingen of tuchtrechtspraak. Het resultaat wordt altijd voorgelegd aan en besproken met de kandidaat.

Uit gesprekken met leden van de begeleidingscommissie bleek dat het opvragen van ambtshalve referenties niet is geformaliseerd en dat deze referenties soms meer dan de gevraagde informatie opleveren. Ook bestaat de indruk dat de procedure van het opvragen van referenties niet transparant genoeg is voor de kandidaten. We bevelen aan om de procedure rondom het opvragen van ambtshalve referenties te formaliseren en te structureren, door te expliciteren welke informatie gewenst is. Deze geformaliseerde procedure kan ook aan kandidaten duidelijk gemaakt worden. Aangezien er geen gegevens over de referenties worden bijgehouden die in aanmerking komen voor empirische analyse, besteden we hier in de rest van dit rapport geen aandacht aan.

AANBEVELING 7 – REFERENTIES

- We bevelen aan om de procedure rondom het opvragen van ambtshalve referenties te formaliseren en te structureren door aan referenten duidelijk te maken wat voor informatie de LSR wel en niet wenst te ontvangen.

4.6 Assessment

Het assessment bestaat uit verschillende tests en vragenlijsten en een interview met een assessmentpsycholoog, afgenomen door LTP. Aan de hand van de informatie die hieruit wordt verkregen beoordeelt de psycholoog van LTP iedere kandidaat op de 11 competenties uit het rechtersprofiel (beneden, op, of boven niveau) en op integriteit. Op basis van die beoordelingen wordt een positief of negatief advies uitgebracht. De bevindingen worden weergegeven in een assessmentrapport, bestaande uit een kwantitatief en een kwalitatief beschrijvend deel. Er volgt geen voortzettingsbeslissing na het assessment, dat volgt pas als de kandidaten ook het eindgesprek hebben gevoerd. Wel kan de kandidaat ervoor kiezen om de procedure niet voort te zetten en het rapport niet naar de LSR te verzenden. De verschillende onderdelen van het assessment worden hieronder beschreven. Ook de scores op de analytische tests worden meegenomen in de competentiebeoordelingen en het assessmentrapport.

Onderdeel	Beschrijving
Watson-Glaser	Test met als doel <i>kritisch redeneervermogen</i> te meten. Ingekocht door LTP.
Big 5+ Persoonlijkheidsvragenlijst	Zelfrapportagevragenlijst. 18 verschillende subschalen; facetten van de Big Five en <i>materialisme, trots, oprechtheid, inschikkelijkheid</i> en <i>veranderingsgerichtheid</i> . (Zie Bijlage 3) Ontwikkeld door LTP.
Werkstijlenvragenlijst	Zelfrapportagevragenlijst. 8 subschalen over <i>motivatie, ontwikkelingsgerichtheid, faalangst, doortastendheid, doelgerichtheid</i> en <i>risico nemen</i> (zie Bijlage 3). Ontwikkeld door LTP.
Drijfverenvragenlijst	Zelfrapportagevragenlijst. 4 subschalen over de drijfveren <i>prestatie, macht, autonomie</i> en <i>affiliatie</i> . Ontwikkeld door LTP.

Onderdeel	Beschrijving
Beïnvloedingsstijlen-vragenlijst	Zelfrapportagevragenlijst. 4 subschalen met de beïnvloedingsstijlen <i>manipulatief, rechtdoorzee, assertief en berekenend</i> . Ontwikkeld door LTP.
Hogan Development Survey	Zelfrapportagevragenlijst. 11 <i>dark side</i> persoonlijkheidskenmerken die een negatief effect op prestaties zouden hebben als ze in sterke mate aanwezig zijn, maar onderbenut worden als ze in zwakke mate aanwezig zijn. Ingekocht door LTP.
Dilemmatest	Situational judgment test (SJT) waarbij kandidaten aangeven wat zij zouden doen en welke overwegingen daarbij een rol spelen, in situaties met dilemma's op moreel en ethisch gebied. De test is ontwikkeld om inzicht te krijgen in de aard van iemands <i>morele kompas</i> . Ontwikkeld door LTP.
Olympics game	Digitaal spel/simulatie-opdracht waarin de meest geschikte locatie voor de Olympische Spelen moet worden gezocht aan de hand van veel informatie. De aanpak is daarbij het belangrijkste, niet het resultaat. Bedoeld om <i>informatieverwerkingssnelheid</i> te meten. Ontwikkeld door LTP.
Interview	Interview van 1,5 uur met een psycholoog van LTP. Er is geen vaste structuur, maar wel onderwerpen die altijd besproken worden. Het interview wordt niet gescoord.

4.6.1 Kwaliteit van de instrumenten

Het was niet mogelijk om de interbeoordelaarsbetrouwbaarheid van de competentie-oordelen en het assessmentadvies in kaart te brengen, aangezien deze door één persoon beoordeeld worden. Daarom gaan we hieronder alleen in op de kwaliteit van

de verschillende onderdelen. Geen van de gebruikte tests en instrumenten is ter beoordeling voorgelegd aan de COTAN.

4.6.1.1 Watson-Glaser

Er was geen informatie beschikbaar over de psychometrische kwaliteit van de Nederlandstalige Watson-Glaser test, maar er zijn wel veel onderzoeksgegevens bekend uit de Verenigde Staten (Watson & Glaser, 2010). Daaruit blijkt dat de Watson-Glaser voldoende betrouwbaar is ($\alpha \approx .80$) en dat de constructvaliditeit in orde is, onder andere blijkend uit resultaten van factor analyses en uit sterke correlaties met scores op andere capaciteitentests. Ook is er eerder evidentie gevonden voor middelgrote predictieve validiteit voor werkprestaties (Watson & Glaser, 2010).

4.6.1.2 Vragenlijsten

Op basis van rapportages over de vragenlijsten die ontwikkeld zijn door LTP blijkt dat de betrouwbaarheid van de schalen op de verschillende vragenlijsten variëren van matig tot goed ($\alpha = .75-.94$). De samenstelling van de groepen gebruikt om de norm-scores te bepalen worden niet beschreven. Onderzoek naar de constructvaliditeit en de predictieve validiteit van de vragenlijstcores is beperkt. De theoretische onderbouwing van de Big 5+ schalen is aanwezig voor de gebruikelijke Big Five schalen, maar onduidelijk voor de 'plusschalen' *materialisme*, *trots*, *oprechtheid*, *inschikkelijkheid* en *veranderingsgerichtheid*. Voor deze vragenlijst wordt door LTP vermeld dat de veronderstelde structuur van de schalen die worden gebruikt in de rio-selectieprocedure niet werd gevonden in een factoranalyse. Wat verder opvalt is dat er voor de Big 5+ vragenlijst aparte normen gehanteerd worden voor mannen en vrouwen, omdat er verschillen in scores gevonden zijn (in welke richting is niet vermeld). Dit heeft tot gevolg dat mannen en vrouwen met dezelfde testscore verschillende normscores krijgen. Het is de vraag of het in deze context wenselijk is, aangezien dit tot verschillende beslissingen kan leiden voor mannen en vrouwen met dezelfde testcores. Verder lieten de resultaten van een factoranalyse op de Drijfverenvragenlijst niet de verwachte resultaten zien en ontbreekt een theoretische ondersteuning voor de schalen van de Werkstijlenvragenlijst. Een gedetailleerde toelichting op deze bevindingen ontbrak.

Er was geen informatie beschikbaar over de kwaliteit van de Nederlandstalige versie van de Hogan Development Survey (HDS), maar uit onderzoek op basis van gegevens verzameld in de V.S. door de testuitgever (Hogan Press, 2014) blijkt dat de schaalcores onvoldoende betrouwbaar zijn voor het maken van selectiebeslissingen ($\alpha = .63-.77$).

Uit twee recente overzichtsstudies blijkt bovendien dat scores op de schalen van de HDS slechts kleine verbanden hadden met succesvol en ethisch leiderschapsgedrag; de schalen met betrekking op vermijdend gedrag wanneer men geconfronteerd wordt met stress (o.a. *voorzichtig*, *gereserveerd* en *skeptisch*) laten significante maar kleine verbanden zien met succesvol leiderschap en de schaal *gewaagd* heeft een klein verband met ethisch leiderschap (Gaddis & Foster, 2015; Nei, Foster, Ness, & Nei, 2018).

Ook merken we op dat er discussie bestaat rondom het gebruik van zelfrapportagevragenlijsten in selectieprocedures. In de wetenschappelijke literatuur wordt vaak vermeld dat scores op dit soort vragenlijsten bruikbare niveaus van predictieve en incrementele validiteit hebben (Schmidt & Hunter, 1998). Deze resultaten zijn echter voornamelijk gebaseerd op vragenlijstscores verkregen in een onderzoekscontext. Er is erg weinig onderzoek gedaan waarbij de scores verkregen zijn in selectiesituaties. Wel is gebleken dat personen in staat zijn om dit soort vragenlijsten sociaal wenselijk in te vullen (Viswesvaran & Ones, 1999) en dat men dit in ook doet in selectiesituaties (Birkeland et al., 2006). Ook zijn er studies die uitwijzen dat het sociaal-wenselijk invullen van dit soort vragenlijsten de predictieve validiteit aantast (Anglim et al., 2018; Niessen et al., 2017; Peterson et al., 2011). Op basis van de beschikbare studies bleek de predictieve validiteit van dit soort vragenlijsten in de context van selectie erg klein te zijn (Anglim et al., 2018; Niessen et al., 2017; Peterson et al., 2011; van Iddekinge et al., 2012). Uit onderzoek bleek echter ook dat de predictieve validiteit van zelfrapportagevragenlijsten afgenomen in een gesimuleerde selectiesituatie wel van een bruikbare grootte was voor personen die erg lage scores behaalden op de vragenlijst (Mueller-Hanson et al., 2003). Daarom raden we aan om zelfrapportagevragenlijsten niet, of slechts voor 'select-out' doeleinden te gebruiken. Bij select-out beslissingen worden alleen kandidaten die erg laag scoren op de vragenlijsten afgewezen.

4.6.1.3 Olympics game en Dilemmatest

Er was geen informatie beschikbaar over de betrouwbaarheid en validiteit van de Dilemmatest (de situational judgment test) en de Olympics game (een simulatie-opdracht). De Dilemmatest bestaat momenteel uit beschrijvingen van ethische en morele dilemma's waarbij kandidaten moeten aangeven hoe ze in dat geval zouden handelen en welke overwegingen daarbij een rol spelen. De overwegingen die daarbij worden aangemerkt zijn bepalend voor de uitslag. De inhoud van de test is gebaseerd op het model van morele ontwikkeling van Kohlberg. Er wordt gewerkt aan een versie waarbij de inhoud van de test specifiek gericht is op de functie van rechter. We

verwachten dat dit de validiteit zal verhogen en positieve reacties van kandidaten zal opleveren. Eenzelfde inhoudelijk relevante invulling zou daarom ook overwogen kunnen worden voor de Olympics game simulatie-opdracht. Verder is het op basis van de beschikbare informatie niet mogelijk om uitspraken te doen over de kwaliteit van deze tests. We bevelen aan om onderzoek uit te laten voeren naar de betrouwbaarheid en validiteit van deze tests, of om instrumenten te gebruiken waarvan deze aspecten al onderzocht zijn en van goede kwaliteit bleken.

4.6.1.4 *Assessmentinterview*

Het assessmentinterview is een interview met een psycholoog van LTP en duurt anderhalf uur. Er is geen vast format voor het interview, maar er zijn wel vaste onderwerpen die altijd aan bod komen, zoals de uitslag en aanpak van de Dilemmatest. Vaak wordt in het interview aanvullend daarop gevraagd naar een moreel of ethisch dilemma en hoe de kandidaat daarmee is omgegaan. Ook de uitslagen op de persoonlijkheidsvragenlijsten worden vaak besproken, hoewel die niet altijd al voorhanden zijn tijdens het gesprek. Het assessmentinterview wordt niet gescoord. Daarmee heeft het assessmentinterview een structuurniveau van 1, het laagste niveau (Huffcutt & Arthur, 1994). In het kader van de uniformiteit van de procedure en de betrouwbaarheid en validiteit, raden we aan om dit onderdeel te structureren en om de interviews te scoren (zie hoofdstuk 7 voor gedetailleerde aanbevelingen).

4.6.2 Procedure van beoordelen in het assessment

Op basis van de informatie verkregen uit de verschillende onderdelen van het assessment worden beoordelingen op competenties gegeven en volgt een assessmentadvies. De wijze waarop de informatie uit de verschillende bronnen wordt gecombineerd is van belang voor de predictieve validiteit. Daarin kan onderscheid gemaakt worden tussen een *klinische* en een *statistische manier* van gegevenscombinatie. Bij klinische gegevenscombinatie gebeurt het wegen en combineren van informatie ‘in het hoofd’ van de expert, bij de statistische methode gebeurt dat volgens een vooraf opgestelde beslisregel of formule (bijvoorbeeld een (gewogen) som of gemiddelde). Veelvuldig wetenschappelijk onderzoek heeft aangetoond dat de statistische methode even goede en vaak zelfs betere voorspellingen en beslissingen oplevert dan de klinische methode. Voor het voorspellen van werkprestaties levert de statistische methode gemiddeld een 50% betere voorspelling op dan de klinische methode (Kuncel et al., 2013).

In de toelichting van het assessmentrapport wordt het volgende vermeld over de manier van gegevenscombinatie: "De uitkomsten van de verschillende onderdelen zijn met behulp van een 'evidence based' scoreprogramma gecombineerd tot een totaaloordeel. De verschillende onderdelen en het interview hebben binnen dit programma verschillende gewichten." In het Handboek Landelijke Selectiecommissie Rechters staat hierover ook: "Het assessmentrapport is een rapport op basis van statistische predictie, waardoor de conclusie uit het rapport in belangrijke mate gebaseerd is op statistische gegevens (evidence based)". Deze beschrijvingen impliceren dat er een statistische manier van scoren wordt toegepast om tot de beoordelingen en een advies te komen. Uit gesprekken met LTP en documenten die door LTP zijn aangeleverd bleek dat dit niet het geval is. In Bijlage 2 staat een schema met daarin vermeld welke onderdelen medebepalend zijn voor het oordeel op iedere competentie en in Bijlage 3 staat gespecificeerd welke scores worden meegenomen en hoe zwaar, om tot het automatisch gegenereerde deel van het assessmentrapport te komen. Het assessmentinterview wordt meegewogen in alle competentieoordelen, met uitzondering van *analytisch vermogen*. Aangezien het interview niet wordt gescoord, is het niet mogelijk een beslisregel of formule op de verzamelde informatie te toe passen om tot een oordeel te komen. De beoordelingen op de verschillende competenties komen dus tot stand volgens de klinische methode, waarbij Bijlage 2 als leidraad gebruikt wordt. Ook het assessmentadvies (positief of negatief) komt tot stand op basis van een klinisch oordeel. Daarbij zijn er wel een aantal informele richtlijnen. Wanneer twee of meer competenties als 'beneden niveau' worden beoordeeld volgt vaak een negatief advies. De competenties worden voor het advies niet allemaal even zwaar gewogen; de ontwikkelbaarheid van de competenties wordt sterk meegenomen en regievoering en besluitvaardigheid wegen zwaarder, terwijl samenwerken en omgevingsbewustzijn bijvoorbeeld minder zwaar wegen. Aan de hand van deze gegevens is het dus niet geheel duidelijk hoe de competentieoordelen precies tot stand komen. Wel kan op basis van deze informatie worden opgemaakt dat het huidige assessment kan worden gedefinieerd als een 'individueel assessment'. Een individueel assessment bestaat uit meerdere assessmentmethoden en instrumenten en de daaruit verkregen informatie wordt volgens de klinische methode gecombineerd. Uit een recente overzichtsstudie (Morris et al., 2015) bleek dat individuele assessments gemiddeld een middelgrote predictieve validiteit ($r = .30$) hadden en dat die validiteit groter was wanneer tests voor cognitieve capaciteiten waren opgenomen, maar kleiner wanneer een interview deel uitmaakte van het assessment. Als de scores op onderdelen waaruit een individueel assessment meestal

bestaat worden gecombineerd volgens de statistische methode is de predictieve validiteit aanzienlijk hoger ($r = .63$).

We bevelen aan om de oordelen die voortvloeien uit het assessment via statistische gegevenscombinatie te bepalen en om te expliciteren wanneer er sprake is van prestaties beneden, op, of boven niveau. Hiervoor is het noodzakelijk om alle onderdelen van het assessment te scoren. Daarnaast moet ook geëxpliciteerd worden wanneer een positief of negatief advies gegeven wordt. Deze methode van beoordelen leidt waarschijnlijk tot een hogere predictieve validiteit (Kuncel et al., 2013; Morris et al., 2015). Het scoren van alle onderdelen maakt bovendien toekomstig empirisch validatieonderzoek mogelijk.

AANBEVELINGEN 8 – INHOUD ASSESSMENT IN HUIDIGE OPZET

- We bevelen aan om de oordelen die voortvloeien uit het assessment volgens statistische gegevens combinatie te bepalen.
- We bevelen aan om het assessmentinterview te structureren en te scoren.
- We raden aan om tests te gebruiken die in voldoende mate zijn onderzocht op psychometrische kwaliteit (betrouwbaarheid en validiteit) en waarvan de resultaten laten zien dat de tests aan de wetenschappelijke standaard voldoen. Deze informatie ontbreekt voor sommige onderdelen volledig.
- Bij voorkeur worden tests gebruikt die op relevante punten minstens als 'voldoende' zijn beoordeeld door de COTAN.
- We raden aan om vragenlijsten op basis van zelfrapportage niet, of slecht voor 'select-out' doeleinden te gebruiken.
- Bij handhaving van de SJT en een simulatie-opdracht zoals de 'Olympics game' raden we aan om varianten te ontwikkelen die inhoudelijk aansluiten bij het rechterschap.

4.6.3 Selectie op basis van de competenties uit het rechtersprofiel

De doelstelling van de LSR is om de selectieprocedure zo in te richten dat deze inhoudelijk aansluit bij het competentieprofiel van rechters. Alleen het assessment is expliciet op die manier ingericht. De enige competentie uit het rechtersprofiel die niet wordt getoetst in de selectieprocedure is '*juridische vaardigheden*'. In het assessment worden de kandidaten op alle andere competenties uit het rechtersprofiel beoordeeld. Deze opzet is tot stand gekomen op verzoek van de LSR. De theoretische

en empirische onderbouwing van het meten van de competenties volgens het schema weergegeven in Bijlage 3 is echter een aandachtspunt. Het is de vraag of het assessment er in de huidige opzet in slaagt om de gewenste competenties valide in kaart te brengen. Er is bijvoorbeeld geen empirische evidentie voorhanden dat scores op de schalen *nuchterheid*, *taakgerelateerde- en relationele stressbestendigheid* en *optimisme* gezamenlijk de competentie *zelfvertrouwen en authenticiteit* meten. Voor de andere competenties geldt hetzelfde. Voor de meeste competenties uit het rechtersprofiel zijn, voor zover bij ons bekend, geen gevalideerde instrumenten voorhanden. Dat de meeste gehanteerde competenties specifiek zijn voor het rechterschap speelt daarin waarschijnlijk een grote rol. Het beoordelen van de kandidaten op ieder van deze competenties is dus moeilijk.

Hoewel uit onderzoek is gebleken dat het bevorderlijk is om de inhoud van selectieprocedures aan te laten sluiten bij de competenties die uitmaken van het te voorspellen criterium (Hoffman et al., 2015), hoeft dit niet te betekenen dat de gewenste competenties expliciet gemeten worden in de selectieprocedure (Bartram, 2005). Als de huidige competentiestructuur gehandhaafd blijft, wordt aanbevolen om de beoordelingen op de competenties in de selectieprocedure in expliciete zin los te laten. In plaats daarvan kunnen eigenschappen, vaardigheden en kenmerken gedefinieerd worden die van belang worden geacht voor het succesvol voltooien van de rio-opleiding. De scores op de instrumenten die in het assessment worden gebruikt worden dan rechtstreekt gebruikt in de besluitvorming, zonder ze eerst als basis voor competentiebeoordelingen te gebruiken. Een dergelijke aanpak is theoretisch en empirisch waarschijnlijk beter te verdedigen. Zoals ook opgemerkt in het rapport Born (2013) is het gehanteerde competentieprofiel bovendien niet gevalideerd. Vaak blijkt dat de verschillende gehanteerde competenties empirisch niet van elkaar te onderscheiden zijn. Zoals ook door Born (2013) aangeraden kunnen de 'Grote Acht' competenties als alternatief, gevalideerd competentiemodel gebruikt worden. Voor deze competentiestructuur is er wel een theoretische en empirische fundering voorhanden, zowel op het gebied van criteriummaten als voor geschikte voorspellers (Bartram, 2005).

Een alternatieve invulling van het assessment dat inhoudelijk dichterbij de functie van rechter ligt is het gebruik van simulaties, in de vorm van rollenspellen en/of opdrachten. Het assessment zou in dat geval een 'assessment center' worden. Uit de wetenschappelijke literatuur blijkt dat beoordelingen op basis van assessment center opdrachten goede voorspellers zijn voor werkprestaties. In groepen die relatief

homogeen¹ zijn wat betreft cognitieve capaciteiten en wanneer een breed gedefinieerd en niet strikt cognitief criterium voorspeld moet worden, zoals hier het geval is, zijn assessment center scores zelfs betere voorspellers dan scores op cognitieve capaciteitentests (Sackett et al., 2017). In assessment centers worden kandidaten beoordeeld op meerdere oefeningen (bijvoorbeeld in-basket tests, rollenspellen, presentaties en groepsopdrachten), door meerdere beoordelaars (Kuncel & Sackett, 2014). Vaak zijn de opdrachten in vorm en inhoud ontworpen om aan te sluiten bij de baan. Wel zal het een kostbare aangelegenheid zijn om een dergelijk assessment center te ontwikkelen, te evalueren en af te nemen.

AANBEVELINGEN 9 – ALTERNATIEVE INVULLING ASSESSMENT

- We bevelen aan om te overwegen om het assessment in te richten als 'assessment center', bestaande uit opdrachten en rollenspellen die in inhoud en in vorm overeenkomen met belangrijke taken die in (het begin van) de rio-opleiding vervuld moeten worden. Een dergelijk assessment center zou in samenwerking met assessmentpsychologen ontwikkeld moeten worden, om de betrouwbaarheid en validiteit te waarborgen.
- Als de huidige opzet van het assessment gehandhaafd blijft, bevelen we aan om de beoordelingen op competentieniveau in de selectie los te laten, aangezien de validiteit van de competentiestructuur onduidelijk is.

4.6.4 Ervaringen leden LSR

Een belangrijke doelstelling van de LSR was om het gewicht dat in de besluitvorming wordt toegekend aan het assessmentrapport te vergroten. Deze doelstelling is tot stand gekomen naar aanleiding van aanbevelingen uit het rapport Born (2013). Uit een analyse uit dit rapport bleek dat er erg weinig gewicht werd gegeven aan het advies dat voortvloeide uit het assessment, terwijl dit een bruikbaar² en kostbaar instrument is. Omdat we alleen beschikten over gegevens van succesvolle kandidaten, die zelden een negatief assessmentoordeel ontvingen, was het niet mogelijk om de analyse uit het rapport Born te herhalen. Ook heeft Born (2013) aanbevolen om het gewicht van het assessment te vergroten door de besluitvorming naar aanleiding van het assessment en het eindgesprek te laten plaatsvinden volgens een statistische

1 Dat wil zeggen, aanmerkelijk homogener dan de algemene beroepsbevolking.

2 We merken op dat het assessment toen nog een rollenspel en een 'fact finding' opdracht bevatte en daarom meer kenmerken van een assessment center had. Deze aanbeveling was daarop gebaseerd.

methode van gegevenscombinatie. Uit gesprekken met het Presidium en leden van de LSR blijkt dat het assessment meer gewicht heeft gekregen in de beslissing, hoewel niet volgens een statistische procedure. Alle geïnterviewden geven aan dat het assessmentrapport zwaar weegt in de eindbeslissing. Hoewel sommigen het moeilijk vonden om hier een percentage aan te verbinden, gaven zij aan dat er in een minderheid van de gevallen wordt afgeweken van het advies gegeven door LTP. Die afwijking komt vaker voor in de vorm van een negatieve beslissing naast een positief assessmentadvies dan andersom. Aanpassingen aan de invulling van de assessmentrapporten hebben hier volgens de betrokkenen aan bijgedragen. De rapportage is statistischer geworden door de opname van een tabel met competentiebeoordelingen en de beschrijvingen zijn concreter, waardoor ze gemakkelijker te interpreteren zijn voor de leden van de LSR. Ook heeft het geholpen dat de selecteurs de assessmentrapporten nu voor de eindgesprekken zien, in plaats van achteraf. Hierdoor biedt het assessmentrapport ook input voor de eindgesprekken en wordt er vaak om toelichting op opvallende uitkomsten gevraagd.

De geïnterviewde leden van de LSR vinden het assessment een zeer waardevol onderdeel van de procedure waarin veel belangrijke competenties getoetst worden. Welke delen van het assessmentrapport het meest inzichtelijk wordt gevonden verschilt, sommige selecteurs hebben de voorkeur voor het kwantitatieve deel en anderen voor de kwalitatieve beschrijving. Als aandachtspunt is genoemd dat er, ondanks de verbeteringen, volgens de meeste betrokkenen nog steeds behoefte is aan een verdere concretisering van de beschrijvende delen van de assessmentrapporten. Het is niet altijd duidelijk op welke informatie (het interview, vragenlijstscores, testscores) de kwalitatieve beschrijvingen van de kandidaten gebaseerd zijn. Ook komen de gehanteerde definities van competenties niet altijd overeen. Een voorbeeld is *regievoering*, wat LTP volgens een geïnterviewde definieert als delegeren en organiseren, terwijl de LSR regievoering tijdens een zitting bedoelt. Daarnaast geeft LTP soms voorwaardelijke adviezen, bijvoorbeeld 'positief, indien extra aandacht wordt geschonken aan samenwerken'. Enkele leden van de LSR gaven aan dat die voorwaardelijkheid voor hen niet erg nuttig is, aangezien zij er niet op kunnen toezien dat het aandachtspunt wordt aangepakt. We bevelen aan om de gewenste definities en formuleringen van adviezen expliciet met LTP te blijven afstemmen. De leden zijn erg positief over de bereikbaarheid van LTP. Wanneer er vragen ontstaan over een assessmentrapport van een kandidaat zijn zij altijd bereikbaar voor toelichting, die meestal naar tevredenheid gegeven wordt. Andere aandachtspunten die de geïnterviewden noemden waren de mogelijkheid tot sociaal-wenselijk antwoorden op de vragenlijst-

ten en het ontbreken van rollenspellen of een inhoudelijke opdracht, zoals de casus die eerder bij de selectie van raio's gebruikt werd.

4.7 Eindgesprek

In het eindgesprek spreken de kandidaten met drie 'teams' bestaande uit twee leden van de LSR. Ieder team bestaat uit een intern en een extern lid en heeft een eigen agenda. Ieder gesprek duurt ongeveer 40 minuten. De leden van de dagcommissie eindgesprekken (de drie teams) beschikken vooraf over het assessmentrapport, het sollicitatieformulier, het cv, cijferlijsten, de formulieren van het voorgesprek en opgevraagde referenties. In de eindgesprekken komen de volgende thema's aan bod:

- Agenda 1:
 - Maatschappelijke nevenactiviteiten
 - Samenwerken
 - Evenwichtigheid
- Agenda 2:
 - Visie op maatschappelijke ontwikkelingen
 - Besluitvaardigheid
 - Regievoering
- Agenda 3:
 - Visie op rechterschap
 - Inlevingsvermogen
 - Flexibiliteit

4.7.1 Procedure eindgesprek

Aan de vorm van de eindgesprekken en de daaropvolgende beraadslagingen is veel aandacht besteed. De agenda's worden strikt gevolgd en alle onderwerpen die besproken moeten worden komen aan bod, hoewel niet alle kandidaten woordelijk dezelfde vragen voorgelegd krijgen. Ook komen er daarnaast zaken aan bod die niet op de agenda staan, zoals opvallende zaken uit het assessmentrapport of uit het cv. Na ieder gesprek vult iedere selecteur een beoordelingsformulier in met daarop de gehanteerde thema's en ruimte voor toelichting. De selecteurs bespreken de kandidaten onderling niet voordat de beraadslagingen beginnen. Tijdens de beraadslagingen vervult een lid-secretaris van de LSR de rol van voorzitter. De lid-secretaris heeft zelf

niet deelgenomen aan de gesprekken en ziet erop toe dat de beraadslagingen gestructureerd verlopen en dat iedereen aan het woord wordt gelaten. Aan het begin van de beraadslagingen geeft ieder lid van de dagcommissie kort een voorlopig oordeel over de kandidaat (negatief, twijfel negatief, twijfel positief, positief). Vervolgens krijgt iedereen de gelegenheid om het oordeel toe te lichten. Daarbij wordt er zorg voor gedragen dat de oordelen onderbouwd worden op basis van concreet gedrag van de kandidaat en niet op algemene indrukken. Daarna is er ruimte voor overleg, vragen en discussie. Pas als iedereen vindt dat alles is gezegd en besproken vindt stemming plaats. Die stemming gebeurt sinds kort besloten, om onderlinge beïnvloeding te voorkomen en de objectiviteit te waarborgen. Daarbij wordt de regel gehanteerd dat een kandidaat geselecteerd wordt als tenminste vier van de zes leden van de dagcommissie een positief oordeel geven.

Sommige thema's die in de eindgesprekken aan de orde komen, komen overeen met de competenties uit het rechtersprofiel (*samenwerken, besluitvaardigheid, regievoering, flexibiliteit*), maar andere niet. Het is onduidelijk waarom juist deze competenties in het eindgesprek nogmaals aan bod komen (de competenties zijn ook beoordeeld in het assessment) en waarop de keuze voor de andere thema's is gebaseerd. *Maatschappelijke nevenactiviteiten* en *de visie op maatschappelijke ontwikkelingen* zijn wel gerelateerd aan de functieomschrijving in het rechtersprofiel, maar *evenwichtigheid, de visie op het rechterschap* en *inlevingsvermogen* komen daar niet direct in terug. Gezien de doelstelling om te selecteren aan de hand van de competenties uit het rechtersprofiel is dit opvallend.

4.7.2 Ervaringen leden LSR

De geïnterviewde leden van de LSR noemden het eindgesprek unaniem als belangrijkste onderdeel van de selectieprocedure. Redenen daarvoor waren de zorgvuldigheid waarmee de procedure is vormgegeven en de mogelijkheid om daardoor een goed gemotiveerd en onafhankelijk oordeel te geven, het beoordelen door zes verschillende selecteurs die allen getraind zijn in het voeren van de gesprekken en erg gemotiveerd zijn om tot adequate beslissing te komen, en de belangrijke thema's die tijdens de eindgesprekken besproken worden. Over de formulieren die bij de eindgesprekken gebruikt worden is men tevreden. Eén lid van de LSR stelt voor om ook een veld toe te voegen waarop een onderbouwing van de beoordeling gegeven moet worden ('waar bleek dat uit?'), om verder te voorkomen dat men afgaat op algemene indrukken. Een ander aandachtspunt dat werd genoemd is dat niet alle selecteurs

even 'streng' lijken te beoordelen en dat er geen expliciete, vaste standaard gehanteerd wordt. Ook bestaat de indruk dat de beoordelingen en beslissingen onbewust worden geïkt op de kandidaten die die dag aan de beurt zijn geweest. Er wordt dan een soort relatieve ranking toegepast, terwijl dat niet de bedoeling is.

Wat ook opvalt is dat de antwoorden op de vraag welke competenties, eigenschappen of kenmerken het belangrijkste zijn en het zwaarst wegen, behoorlijk uiteenlopen. *Regievoering, inlevingsvermogen, analytisch vermogen en evenwichtigheid* werden door twee personen genoemd. Andere antwoorden waren de *visie op de maatschappij en het rechterschap, taalvaardigheid, persoonlijk overwicht, interactie, integriteit en zelfreflectie*. Kortom, niet aan alle competenties en thema's wordt evenveel gewicht toegekend en aan welke competenties en thema's het meeste gewicht wordt toegekend verschilt per selecteur. Daarnaast valt op dat enkele van de genoemde competenties niet in het rechtersprofiel of de agenda van de eindgesprekken voorkomen. Ook merkt een lid van de LSR op dat er voor sommige thema's beperkt consensus is over wat een goed of wenselijk antwoord is, bijvoorbeeld voor de visie op het rechterschap en op de maatschappij. Deze ogenschijnlijke verschillen in de weging en definitie van verschillende competenties is niet bevorderlijk voor de uniformiteit van de procedure.

4.7.3 Kwaliteit eindgesprekken

Aan de hand van gegevens over beoordelingen op de eindgesprekken (tussen de 87 en 107 kandidaten, afhankelijk van het aantal ontbrekende gegevens) is de interbeoordelaarsbetrouwbaarheid van de oordelen voortkomend uit de eindgesprekken in kaart gebracht. Deze gegevens waren afkomstig van kandidaten die zijn toegelaten tot de rio-opleiding. De resultaten staan in Tabel 3.

De betrouwbaarheid van het eindoordeel van het eindgesprek (niet door, twijfel, wel door) was ongeveer gelijk aan 0 en dus onvoldoende. De interbeoordelaarsbetrouwbaarheid voor de verschillende thema's waren onvoldoende tot redelijk (zie de notitie onderaan Tabel 3). Dit betekent dat de beoordelingen van de verschillende beoordelaars op basis van het eindgesprek niet of slechts in beperkte mate met elkaar overeenkomen. Omdat in de analyse alleen oordelen meegenomen konden worden van kandidaten die tot de rio-opleiding zijn toegelaten, heeft restrictie in range waarschijnlijk tot een onderschatting van de interbeoordelaarsbetrouwbaarheid geleid. In een vergelijkbare analyse op basis van gegevens van afgewezen en aangenomen

Tabel 3 Interbeoordelaarsbetrouwbaarheid van totaalscores en themascores uit het eindgesprek.

Thema/score	ICC
Agenda 1	
Maatschappelijke nevenactiviteiten	.64 [.45, .76]
Samenwerken	.43 [.14, .62]
Evenwichtigheid	.24 [-.14, .50]
Eindoordeel selecteurs thema 1 ^a	.25 [< .01, .51]
Totaalscore thema 1	.41 [.09, .61]
Agenda 2	
Visie op maatschappelijke ontwikkelingen	.57 [.37, .71]
Besluitvaardigheid	.53 [.30, .69]
Regievoering	.50 [.26, .66]
Eindoordeel thema 2 ^a	.44 [.18, .66]
Totaalscore thema 2	.61 [.41, .74]
Agenda 3	
Visie op rechterschap	.51 [.29, .67]
Inlevingsvermogen	.56 [.35, .70]
Flexibiliteit	.38 [.09, .58]
Eindoordeel thema 3 ^a	.24 [-.01, .46]
Totaalscore thema 3	.50 [.25, .67]
Totaalscore	
Eindoordeel selecteurs ^a	.06 [.01, .12]

Notitie. ICC = intra-class correlatie, 95% betrouwbaarheidsinterval tussen haakjes. Interpretatie: < .50 = onvoldoende, 0.5 – 0.75 = redelijk, 0.75 – 0.9 = goed, > .90 = excellent (Koo & Li, 2016).

^a Krippendorff's alfa voor ordinale data i.p.v. ICC (3 ordinale categorieën).

kandidaten in het rapport Born (2013, p. 80) werden aanzienlijk hogere betrouwbaarheidscoefficienten gevonden. Die resultaten geven waarschijnlijk een accurater beeld van de betrouwbaarheid. Daarbij moet wel worden opgemerkt dat de inhoud van het eindgesprek sindsdien is gewijzigd. Het is mogelijk dat de nu gehanteerde thema's meer of minder betrouwbaar beoordeeld kunnen worden in vergelijking met de eerder gehanteerde thema's, bijvoorbeeld door verschillen in de mate van explicitering en overeenstemming over de definities van de thema's. Stellige uitspraken over de betrouwbaarheid van de beoordelingen uit het eindgesprek zijn op basis van deze resultaten niet te doen.

In Tabel 4 staan de correlaties tussen de oordelen op de verschillend thema's die behandeld worden in de eindgesprekken. Aangezien de thema's inhoudelijk verschillen en afzonderlijke competenties en vaardigheden zouden moeten meten, verwachten we wederom dat de correlaties tussen de scores op de verschillend thema's klein zijn. Dit geeft een beeld van de divergente validiteit, ofwel de onderscheidbaarheid van de thema's. Voor scores op thema's die binnen hetzelfde gesprek en dus door dezelfde beoordelaars beoordeeld werden, waren die correlaties echter hoog. De correlaties tussen scores op thema's die niet door de dezelfde beoordelaars beoordeeld werden waren, zoals verwacht, klein. Dit duidt erop dat ook in de eindgesprekken een halo-effect optreedt, zoals ook gevonden in eerdere analyses in het rapport Born (2013, p. 81).

4.7.4 Aanbevelingen

Hoewel stellige uitspraken over de betrouwbaarheid van de oordelen voortkomend uit de eindgesprekken niet gedaan kunnen worden, lijkt er ook hier ruimte te zijn voor verbetering. De uniformiteit is daarbij wederom een aandachtspunt. Hoewel de onderwerpen van de agenda's gevolgd worden, wordt er ook over andere onderwerpen gevraagd die niet voor iedere kandidaat hetzelfde zijn. Van de vier niveaus van standaardisatie die in de literatuur worden onderscheiden (Huffcutt & Arthur, 1994), heeft het eindgesprek daarmee op procesniveau een standaardisatieniveau van 2. De scoring heeft ook een standaardisatieniveau van 2 (beoordelingen op een eenvoudige schaal, naar onderwerp). Omdat een hogere mate van structuur de betrouwbaarheid en predictieve validiteit van interviewscores verhoogt (Huffcutt et al., 2014), raden we aan om de eindgesprekken verder te structureren. De wijze waarop een hogere mate van structuur bereikt kan worden zal verderop in dit document besproken worden.

Tabel 4 Divergente validiteit van de thema's beoordeeld binnen het eindgesprek.

Criterium	Score								
	Maatsch. nevenact.	Samen- werken	Even- wichtig.	Visie maatsch. ontw.	Besluit- vaardigh.	Regie- voering	Visie rechter- schap	Inlevings- vermogen	Flexi- biliteit
Thema 1									
Maatschappelijke nevenact.									
Samenwerken	.42*								
Evenwichtigheid	.27*	.55*							
Thema 2									
Visie maatsch. ontw.	.04	.09	.09						
Besluitv.	-.05	.12	.29*	.40*					
Regievoering	-.02	.07	.19*	.46*	.61*				
Thema 3									
Visie rechterschap	-.03	.10	.11	.15*	-.03	.12			
Inlevingsvermogen	-.02	.08	.11	-.01	.04	.07	.55*		
Flexibiliteit	-.05	.01	.10	.13	.12	.13	.53*	.45*	
Totaalscores									
Eindoordeel selecteurs ^a	.20*	.30*	.18*	.41*	.40*	.25*	.39*	.38*	.36*
Gemiddeld eindoordeel ^b	.14	.16	.23*	.26*	.26*	.17	.23*	.12	.19

Notitie. $n = 223-200$ oordelen van individuele selecteurs. ^a Samenhang met eindoordeel van de selecteren op het bijbehorende thema, beoordelingen in de 'twijfel' categorie zijn buiten beschouwing gelaten, het eindoordeel was in ongeveer 6% van de gevallen negatief, $n = 195$. ^b Dit is de correlatie tussen het gemiddelde eindoordeel o.b.v. alle selecteurs en het gemiddelde oordeel op de subthema's o.b.v. twee selecteurs, $n = 123$. * Statistisch significant verband ($p > .05$).

Verder kan de beoordeling volgend op het eindgesprek, net als het assessment, opgevat worden als een 'individueel assessment', waarbij scores en beoordelingen afkomstig van verschillende instrumenten volgens de klinische methode worden gecombineerd (zie paragraaf 4.5.2). In dit geval gebeurt dat door meerdere beoordelaars. Dit is het geval omdat de besluitvorming gebaseerd is op zowel de indrukken uit het eindgesprek als het assessmentrapport en daarnaast waarschijnlijk ook beïnvloed wordt door de andere informatie uit de dossiers. Beoordelingen op basis van 'individuele assessments' hebben een aanzienlijk lagere predictieve validiteit dan beoordelingen op basis van statistische combinatie van de verzamelde informatie (Morris et al., 2015). Als de besluitvorming na afloop van de procedure op een statistische manier plaatsvindt hoeven er geen beraadslagingen meer plaats te vinden. De huidige manier van beraadslagen is zorgvuldig opgezet om onderlinge beïnvloeding te voorkomen en objectiviteit te waarborgen. Dat is echter niet wat er gebeurt; er vindt op een gecontroleerde manier onderlinge beïnvloeding plaats. Uit onderzoek is gebleken dat beoordelingen die op basis van dit soort gecontroleerde beraadslagingen tot stand komen toch vaak lager zijn dan de predictieve validiteit op basis van een statistische combinatie van beoordelingen en scores (Dilchert & Ones, 2009; Highhouse & Brooks, 2017). Het is dus waarschijnlijk effectiever en efficiënter om de beraadslagingen te vervangen door een statistische manier van het combineren van informatie. Dat maakt het ook meteen noodzakelijk om de wegingen die worden toegekend aan verschillende competenties, eigenschappen en kenmerken te expliciteren, wat voor meer uniformiteit zal zorgen.

Voorstellen tot statistische besluitvormingsprocedures stuiten vaak op weerstand, bijvoorbeeld omdat beoordelaars het als een aantasting van de autonomie zien (Nolan & Highhouse, 2014), of omdat zulke procedures als onpersoonlijk worden gezien (Dawes, 1979). Daarom hebben we met enkele leden van de LSR besproken wat zij vinden van een dergelijke procedure. Het merendeel staat hiervoor open als blijkt dat dit betere beslissingen en beoordelingen oplevert. Eén extern lid van de LSR vindt een statistische procedure afdoen aan het eigen beoordelingsvermogen en hecht er waarde aan om te horen wat de ervaringen van andere selecteurs waren. Ook wordt genoemd dat verschillende competenties en thema's verschillende wegingen moeten krijgen. Een lid van het Presidium vindt dat indrukken en 'gezond verstand', hoe feilbaar ook, een onderdeel moeten blijven. We merken op dat dit ook binnen een statistische procedure kan, mits die indrukken gecombineerd worden volgens een vooraf opgestelde regel of formule. Een lid van het presidium van de LSR was van mening dat de autonomie om te beoordelen ook in een statistische

procedure gewaarborgd is, ook al worden de beoordelingen volgens een statistische regel gecombineerd. Het lijkt erop dat de leden en het Presidium in principe open staan voor het gebruik van statistische beoordelingsprocedures. Mocht dit worden ingevoerd, dan is het voor het draagvlak onder selecteurs en andere belanghebbenden van belang om de onderbouwing van die keuze duidelijk te communiceren. Ook kan het betrekken van de leden bij het bepalen van de wegen van verschillende thema's, eigenschappen en vaardigheden het draagvlak vergroten (Nolan & Highhouse, 2014).

AANBEVELINGEN 10 – VORM EN BESLUITVORMING EINDGESPREK

- We bevelen aan om ook het eindgesprek verder te structureren.
- Daarnaast wordt aanbevolen om de besluitvorming volgens een statistische procedure te laten verlopen, waarbij vooraf wordt gespecificeerd hoe zwaar ieder thema en iedere score gewogen moet worden. De beraadslagingen zijn dan niet meer nodig.
- Om draagvlak te creëren voor een dergelijke procedure is het belangrijk om de onderbouwing voor een dergelijke procedure duidelijk naar belanghebbend te communiceren en om kan overwogen worden om ze te betrekken bij het opstellen van beslisseregels en wegen.

4.8 Lokale selectieprocedure

Na het succesvol doorlopen van de landelijke selectieprocedure volgt de selectie bij de lokale gerechten. De lokale selectie heeft als doelstelling de 'fit' met het gerecht te beoordelen. Het uitgangspunt is dat de geschiktheid voor de rio-opleiding door de LSR is vastgesteld en in de lokale procedures niet meer wordt getoetst.

4.8.1 Opzet lokale selectieprocedures

De lokale selectiecommissie bestaat bij sommige gerechten uitsluitend uit rechters, bijvoorbeeld uit verschillende rechtsgebieden, terwijl bij andere gerechten ook juridisch medewerkers en personen van de griffie deel uitmaken van de commissie, om een afspiegeling van het gerecht te vormen. Bij de gerechten waarvan we een lid van de selectiecommissie gesproken hebben worden de selectiegesprekken altijd met meerdere personen gevoerd. Sommige lokale procedures bestaan uit een enkel ge-

spreek waaraan verschillende personen deelnemen, andere bestaan uit twee gesprekken met verschillende teams. Uiteindelijk hebben de lokale selectiecommissies een adviserende rol naar het gerechtshof. Alle geïnterviewden gaven aan dat het advies in de meeste gevallen door het bestuur gevolgd wordt.

Bij sommige gerechten wordt er een taakverdeling afgesproken voor de gesprekken en/of is er een lijst met te bespreken onderwerpen. Bij geen van de lokale gerechten worden scoreformulieren gebruikt bij de interviews. Wel wordt er bij sommige gerechten aan het eind expliciet een ranking van kandidaten opgesteld. Eén lid van een lokale selectiecommissie merkte op dat de procedures bij de verschillende gerechten niet erg uniform zijn wat betreft het aantal gesprekken, de doelstelling, de samenstelling van de selectiecommissie en de aanpak van de selectiegesprekken. De geïnterviewde vindt het wenselijk om dit sterker te uniformeren. Enkel merken op dat de doelstelling van de lokale selectiegesprekken ook niet altijd duidelijk is voor kandidaten. Sommige kandidaten verwachten dat het gaat om een kennismakingsgesprek, terwijl het gerecht het ziet als een sollicitatiegesprek waarin de geschiktheid als rechter bij een specifiek gerecht getoetst wordt.

4.8.2 Aansluiting landelijke en lokale selectie

De leden van de lokale selectiecommissies stellen dat zij de lokale selectie benaderen vanuit het uitgangspunt dat in principe alle kandidaten geschikt zijn bevonden. De doelstelling van de lokale selectie om de geschiktheid voor de specifieke rechtbank in kaart te brengen. Meestal is de reden voor het afwijzen van kandidaten dan ook de 'fit' met de rechtbank. Incidenteel zien de lokale selecteurs een kandidaat die zij niet geschikt achten. De reden daarvoor ligt meestal op het gebied van communicatieve vaardigheden. Toch noemt ook iemand dat het niet geheel duidelijk is of de lokale procedure wel als 'selectiemoment' aangemerkt moet worden en dat verdere afstemming van de doelstelling van de landelijke- en lokale procedure wenselijk is. In de lokale gesprekken wordt op uiteenlopende zaken getoetst, zoals binding met de regio, match met de rechtbank op basis van persoonlijkheid, klik met de kandidaat, inhoudelijke expertise (afhankelijk van de behoefte van het gerecht), maar ook behoeften wat betreft diversiteit (man-vrouw verhoudingen, culturele achtergrond en persoonlijkheid). Ook werden zaken genoemd die al in de landelijke procedure getoetst zijn, zoals communicatieve vaardigheden, samenwerken, analytisch vermogen, de motivatie om rechter te willen worden, ontwikkelbaarheid, nevenactiviteiten en maatschappelijke betrokkenheid. Drie leden van een lokale selectiecommissie gaven

aan dat delen van de landelijke procedure nog wel eens dunnetjes worden overgegaan, bij het eigen of bij andere gerechten. Eén lid van een lokale selectiecommissie was van mening dat enige overlap onvermijdelijk is wanneer de lokale gerechten nog inspraak hebben na afronding van de landelijke procedure. Een andere betrokkene vond dit expliciet onwenselijk, vooral omdat de lokale procedures duidelijk minder zorgvuldig en betrouwbaar zijn dan de procedure bij de LSR. Zorgen voor voldoende bekendheid met de landelijke procedure kan volgens deze persoon helpen bij het voorkomen van overlap. Twee van de geïnterviewde leden gaven aan redelijk goed op de hoogte te zijn van de landelijke procedure en de verschillende stappen die daarbij horen. Zij gaven aan uitleg te hebben ontvangen vanuit de LSR, bijvoorbeeld op een dagdeel georganiseerd voor lokale selecteurs. Twee andere leden van lokale selectiecommissies gaven aan de grote lijnen van de landelijke procedure enigszins te kennen, maar niet van de precieze details en onderdelen op de hoogte te zijn.

4.8.3 Plaatsing na landelijk 'groen licht'

Het komt met regelmaat voor dat kandidaten die na de landelijke selectieprocedure 'groen licht' hebben gekregen geen rio-plaats weten te verwerven bij een lokaal gerecht. Als een kandidaat niet wordt aangenomen bij het gerecht of de gerechten waar in eerste instantie is gesolliciteerd, kunnen zij nog één jaar op andere rio-vacatures solliciteren. Zij hoeven dan alleen deel te nemen aan de briefselectie (rio's met ruime werkervaring) of de briefselectie en het voorgesprek (rio's met beperkte werkervaring) en gaan vervolgens door naar de lokale selectie. De lokale selectiecommissies weten in dat geval dat het om een kandidaat gaat die al eerder 'groen licht' heeft gekregen, maar nog geen plaats heeft gevonden. Een doelstelling is om het aantal landelijk succesvolle kandidaten die lokaal geen plaats kunnen vinden terug te brengen. Of dit is gelukt kon aan de hand van de beschikbare gegevens niet beantwoord worden.

De geïnterviewde leden van lokale selectiecommissies herkenden het probleem van kandidaten die 'groen licht' hebben gekregen van de LSR, maar lokaal geen rio-plaats weten te vinden. Volgens hen wordt het mede veroorzaakt doordat er regelmatig meerdere kandidaten met groen licht overblijven voor één vacature. Vooral bij gerechten in en rondom de Randstad is dat het geval. Dat is echter niet het enige mechanisme. Het komt ook voor dat een gerecht er voor kiest om een vacature niet te vervullen, terwijl er wel een kandidaat met 'groen licht' beschikbaar is. Eén lid van een lokale selectiecommissie vindt dat dit eigenlijk niet zou moeten gebeuren, terwijl

een ander lid denkt dat dit onwenselijk, maar ook onvermijdelijk is. Een derde lokaal lid oppert dat het misschien beter is om rio's in overleg te plaatsen en de lokale gerechten minder autonomie te geven. Verschillende betrokkenen bij lokale gerechten en de LSR merken bovendien op dat de lokale selectie niet het enige inspraakmoment is, maar dat de gerechten ook al betrokken zijn bij de briefselectie en het voor-gesprek. Enkele leden van lokale selectiecommissies gaven aan met regelmaat bedenkingen te hebben bij kandidaten die eerder bij andere gerechten niet zijn geselecteerd. Dat komt met name omdat het gerecht waar vervolgens wordt gesolliciteerd blijkbaar niet de eerste keus van de kandidaat was. Daaruit blijkt ook een mindere binding met de regio (vooral bij gerechten buiten de Randstad), wat op praktisch en professioneel gebied problemen zou kunnen geven. Hoe groot dit probleem momenteel is en of het aantal kandidaten dat na landelijk groen licht geen lokale plaats kunnen vinden is teruggedrongen, is dus niet te beantwoorden. Duidelijk is wel dat het probleem nog steeds bestaat en volgens verschillende betrokkenen aandacht verdient. Vooral voor de betreffende kandidaten is dit een erg vervelende situatie.

4.8.4 Aanbevelingen

Het lijkt wenselijk om vanuit de LSR de doelstelling van de lokale selectiegesprekken alvast duidelijker richting de kandidaten te communiceren en er vervolgens gezamenlijk zorg voor te dragen dat de lokale selectieprocedures zich ook tot die doelstellingen beperken. Ook bevelen we aan dat de lokale selectiecommissies tot meer uniforme selectieprocedures komen, waarin rekening wordt gehouden met kennis over best practices, zoals het structureren van interviews, het scoren van reacties en het gebruik van beslisregels. De lokale selectiegesprekken zijn momenteel niet of slechts beperkt gestructureerd (niveau 1, Huffcutt & Arthur, 1994). De LSR en een assessmentbureau zouden daarin kunnen assisteren en adviseren.

Het probleem van de kandidaten met 'groen licht' die geen plaats kunnen vinden kan worden ingeperkt door hetzelfde aantal rio's te selecteren als er rio-plaatsen zijn. Dat is echter moeilijk te realiseren zonder de autonomie van de lokale gerechten in te perken. Ook komt op deze manier de scheiding tussen het beoordelen van geschiktheid voor de functie (person-job fit, de LSR) en de geschiktheid voor de organisatie (person-organisation fit, het lokale gerecht) in het gedrang. Daarbij merken we op dat deze scheiding in de huidige opzet ook niet strikt gehandhaafd wordt; ook nu wordt tijdens de briefselectie en de voorgesprekken al in zekere mate op 'fit' gescreend.

Daarbij is het ook de vraag of het uiteindelijke oordeel van de LSR in termen van 'geschikt' of 'ongeschikt' te verdedigen is. Een onderscheid in de mate van geschiktheid is theoretisch en empirisch waarschijnlijk beter te verdedigen. Kandidaten worden dan afgewezen omdat andere kandidaten geschikter werden bevonden, niet omdat zij ongeschikt zijn bevonden.

Een mogelijkheid waarbij de lokale gerechten nog enige inspraak hebben is om de beslissing over de bereidheid van het lokale gerecht om een kandidaat aan te nemen te verplaatsen en al na de voorgesprekken te laten plaatsvinden. Er kunnen dan afspraken worden gemaakt over geoorloofde gronden (bijvoorbeeld binding met de regio of de organisatie en inhoudelijke expertise) om een kandidaat niet aan te willen nemen. Aangezien kandidaten nu slechts één voorgesprek voeren, ook als zij op meerdere vacatures hebben gesolliciteerd, betekent dit wel dat die gang van zaken herzien moet worden, wat de organisatie van de voorgesprekken complexer maakt en de doorlooptijd van de selectieprocedure waarschijnlijk zou verlengen.

Een andere mogelijkheid is het afschaffen van de lokale selectie en in plaats daarvan rio's bij gerechten te plaatsen. De 'person-organisation fit' zou dan geheel niet meer beoordeeld worden, of slechts in beperkte mate tijdens de briefselectie. Er is relatief weinig onderzoek gedaan naar de predictieve validiteit van 'person-organisation fit' en de waarde ervan in selectiesituaties is een punt van discussie in de wetenschappelijke literatuur. De voorspellende waarde voor werkprestaties lijkt klein te zijn, maar er is wel samenhang gevonden met betrokkenheid bij de organisatie en baantevredenheid (Arthur et al., 2006). Deze verbanden met betrokkenheid en tevredenheid kunnen een reden zijn om de 'fit' met het gerecht een rol te laten spelen in de selectieprocedure. Het is echter de vraag of het wenselijk is dat de 'fit' met het gerecht een dermate grote rol speelt bij het verwerven van een rio-plaats. Het antwoord op deze vraag is een beleidskeuze en afhankelijk van de uitkomstmaten waaraan waarde wordt gehecht en in welke mate.

Mocht de lokale selectie gehandhaafd blijven, dan wordt aanbevolen om de lokale gesprekken te structureren en daarbij te letten op vooraf gespecificeerde eigenschappen en kernwaarden die tijdens de interviews gescoord worden. Uit onderzoek blijkt dat er accuratere schattingen van fit bereikt worden als selecteurs letten op expliciet geformuleerde waarden en kenmerken die relevant zijn voor de fit met de organisatie (Parson et al., 1999).

AANBEVELINGEN 11 – LOKALE SELECTIEPROCEDURE

- Om het aantal succesvolle kandidaten dat lokaal geen rio-plek kunnen vinden terug te dringen kan overwogen worden om evenveel rio's te selecteren als er plaatsen zijn.
- De lokale selectie kan dan:
 - komen te vervallen en worden vervangen door plaatsing;
 - al tijdens de briefselectie en/of het voorgesprek plaatsvinden (welk alternatief de voorkeur heeft, is een beleidskeuze).
- Indien de lokale selectie gehandhaafd blijft bevelen we aan om:
 - de lokale selectieprocedure te structureren;
 - expliciet geoorloofde afwijzingsgronden af te spreken;
 - te zorgen voor meer bekendheid over de landelijke procedure, om het vertrouwen te vergroten;
 - de doelstelling van de lokale selectie duidelijk te communiceren naar kandidaten.

4.9 Bijdrage selectie aan ontwikkelpunten in de opleiding

Een vraag vanuit de Startnotie was of de selectieprocedure zinvolle input levert voor het bepalen van ontwikkelpunten waaraan de kandidaat tijdens de opleiding kan werken. Dit punt is voorgelegd aan personen betrokken bij de selectie en de opleiding en aan oud-deelnemers aan de selectieprocedure (zie ook paragraaf 5.5.2). De geïnterviewden waarmee dit punt is besproken hadden hier uiteenlopende meningen over. Als input voor de opleiding ontvangt de intakecommissie een bijlage van het assessmentrapport, waarin ontwikkelpunten worden opgenomen. De hoeveelheid ontwikkelpunten verschilt per persoon. Eén geïnterviewde geeft aan dat hier niet zo veel mee gebeurt en dat de ontwikkelpunten vaak wat vaag geformuleerd zijn, zodat het lastig is om concrete aanknopingspunten te bieden. Het ontvangen van het hele assessmentrapport, of zelfs ook notities van de selecteurs vanuit de LSR, zou veel rijkere informatie opleveren. Dat gebeurt nu alleen op vrijwillige basis, als de rio in kwestie deze punten zelf mondeling of schriftelijk meedeelt. De indruk bestaat dat rio's zich soms te kwetsbaar voelen om deze informatie met de intakecommissie of de opleiders te delen.

De algemene tendens lijkt dat er in de selectieprocedure veel informatie wordt verzameld die relevant kan zijn voor de opleiding. Toch is het bewaken van de privacy van

de rio's ook belangrijk en dienen de selectie en de opleiding verschillende doelen. Gezien de opvattingen van de geïnterviewden lijken er geen dringende redenen te zijn om de selectie in sterkere mate als ontwikkelingsinstrument in te richten. Een afstemming met LTP over de formulering en definiëring van ontwikkelpunten en de aansluiting op en mogelijkheden in de opleiding zou de waarde van informatie uit de selectieprocedure voor de ontwikkeling in de opleiding wellicht wat kunnen vergroten.

4.10 Cultuurwaardevrij selecteren

Een doelstelling van de selectieprocedure is het bijdragen aan en vergroten van de diversiteit van de geselecteerde groep rio's, zodat de geselecteerde groep rio's zoveel mogelijk een representatieve afspiegeling van de maatschappij vormt. Daarom is het streven om selectie-instrumenten te gebruiken die zo cultuurwaardevrij zijn als mogelijk. Hierbij merken we op dat geheel cultuurwaardevrije tests en assessments niet bestaan. Er waren geen gegevens beschikbaar die empirisch onderzoek naar de cultuurwaardevrijheid van de gehanteerde instrumenten mogelijk maakten. In het rapport Born (2013, p. 50) is al een uitgebreide beschrijving van de literatuur over dit onderwerp opgenomen. Samenvattend laten instrumenten die sterk cognitief georiënteerd zijn vaak de grootste verschillen in scores zien tussen personen met en zonder migratieachtergrond, waarbij personen met een migratieachtergrond lagere scores behalen (Dahlke & Sackett, 2017). Deze verschillen rechtvaardigen echter niet zonder meer de conclusie dat het gebruik van cognitief georiënteerde instrumenten een onterecht nadeel opleveren voor personen met migratieachtergrond. Daarvan wordt pas gesproken als er sprake is van onderpredictie, wat inhoudt dat de voorspellingen van latere prestaties op basis van de testcores consistent te laag uitvallen voor personen met een migratieachtergrond. Uit onderzoek blijkt dat dit in de regel niet het geval is; voorspellingen van prestaties van personen met een migratieachtergrond op basis van cognitieve tests zijn vaak zelfs iets hoger dan de daadwerkelijke latere prestaties (Berry, 2015). Daarbij moet wel de kanttekening worden geplaatst dat deze onderzoeken voornamelijk in de Verenigde Staten zijn uitgevoerd. Een paar in Nederland uitgevoerde studies lieten vergelijkbare resultaten zien (te Nijenhuis & van der Flier, 1999).

Zelfs als cognitief georiënteerde tests geen 'bias' vertonen richting personen met een migratieachtergrond, zorgen de lagere scores van personen met een migratieachtergrond ervoor dat het gebruik van deze tests de diversiteit op dit vlak nadelig beïn-

vloedt. Cognitief georiënteerde instrumenten zijn echter ook vaak de beste voorspelers voor werkprestaties (Schmidt & Hunter, 1998), wat zorgt voor diversiteits-validiteitsdilemma. Voor de selectie van rio-kandidaten zou dit bijvoorbeeld kunnen inhouden dat de instrumenten met de beste voorspellende waarde ook de meest nadelige effecten op diversiteit hebben. In de literatuur worden verschillende aanbevelingen gegeven om de scoreverschillen te reduceren zonder de validiteit aan te tasten. De meeste effectieve manieren zijn: het gebruik van tests voor logisch redeneren als alternatief voor traditionele tests voor cognitieve capaciteiten, het gebruik van simulatie-assessments zoals *assessment centers* of *work samples*, het aanpassen van de wegingen van verschillende onderdelen zodat onderdelen met kleinere scoreverschillen zwaarder worden gewogen, het reduceren van het belang van verbale vaardigheden en het inzetten van instrumenten die positief beoordeeld worden door kandidaten (De Soete, Lievens, & Druart, 2013).

Wat betreft de analytische tests die in de rio-selectieprocedure gebruikt worden is in het ontwerp van de vragen getracht om deze zo cultuurwaardenvrij mogelijk te maken, door enkel hoogfrequente woorden en geen gezegdes of spreekwoorden te gebruiken in de tests. Er zijn echter geen empirische gegevens beschikbaar die het mogelijk maken om uitspraken te doen over de effectiviteit van deze inspanningen. Ook hebben leden van het Presidium van de LSR en medewerkers van LTP gesprekken gevoerd met oud-kandidaten met een migratieachtergrond, om knelpunten in de selectie op dit gebied te achterhalen. Naar aanleiding van die gesprekken kwam naar voren dat de wenselijkheid en interpretatie van sommige eigenschappen die met de persoonlijkheidsvragenlijsten uit het assessment werden gemeten, zoals assertiviteit, een aandachtspunt waren. Dat is opvallend, aangezien juist op dit type instrument vaak slechts kleine verschillen in scores worden gevonden (Ployhart & Holtz, 2008). Sindsdien wordt er extra aandacht besteed aan de invloed van culturele achtergrond op de uitslagen op de persoonlijkheidsvragenlijsten. Hierover wordt bijvoorbeeld doorgevraagd tijdens de assessmentinterviews.

De gesproken betrokkenen bij de selectieprocedure onderschrijven het doel om de diversiteit onder rechters waar mogelijk te vergroten en herkennen dat deze doelstelling nog niet behaald is. De meeste geïnterviewden stellen dat er een grote bereidwilligheid bestaat om de culturele diversiteit onder rechters te vergroten en dat kandidaten met een niet-Nederlandse achtergrond zowel vanuit de LSR als vanuit de lokale gerechten wat sneller het voordeel van de twijfel krijgen. De meeste betrokkenen hebben echter de indruk dat het gebrek aan diversiteit vooral aan de werving ligt. Er zijn weinig kandidaten met een migratieachtergrond die solliciteren voor een rio-

plaats. Een reden daarvoor zou juist de geringe culturele diversiteit binnen de rechtspraak kunnen zijn. Volgens de meeste betrokkenen lijkt uitval tijdens de selectieprocedure nu geen hoofdoorzaak van de geringe diversiteit te zijn, hoewel dit gezien het kleine aantal kandidaten met een migratieachtergrond lastig is in te schatten. Wel merken enkelen op dat als kandidaten met een migratieachtergrond solliciteren, zij vaker uit lijken te vallen op taalvaardigheid. Dat geldt zowel voor kandidaten met een westerse- en een niet-westerse migratieachtergrond. De meeste geïnterviewden vinden taalvaardigheid echter een zeer belangrijke vaardigheid voor rechters en vinden het niet wenselijk om op dit terrein de lat te verlagen. Slechts één extern lid van de LSR vroeg zich af of de hoge eisen op het gebied van taalvaardigheid noodzakelijk zijn.

Om de diversiteit van de rio's te vergoten lijkt een aangepaste werving een optie. In het rapport Born (2013) zijn al aanbevelingen gedaan voor effectiever diversiteitsbeleid op het gebied van werving. Daarnaast bevelen we aan om ook eventuele knelpunten in de selectie in de gaten te blijven houden. Gezien het klein aantal kandidaten met een migratieachtergrond komen bestaande knelpunten nu wellicht niet voldoende aan het licht. Daarnaast bevelen we aan om onderzoek te laten verrichten naar de cultuurwaardevrijheid van de analytische tests, of om tests te gebruiken die op dit punt al onderzocht zijn, omdat op deze tests de grootste verschillen in scores worden gevonden.

AANBEVELINGEN 12 – CULTUURWAARDEVRIJ SELECTEREN

- Een mogelijkheid is het gebruik van simulaties, zoals assessment centers of work samples. Op deze assessments worden vaak kleinere verschillen in scores gevonden en ze worden positief gewaardeerd door kandidaten.
- We bevelen aan om, indien mogelijk, analytische tests te gebruiken die zijn ontworpen om zo cultuurwaardevrij mogelijk te meten en die ook op dit punt onderzocht zijn.
- Aan de cultuurwaardevrijheid van de gehanteerde tests en assessments zal blijvend aandacht besteed moeten worden. Als het aandeel sollicitanten met een migratieachtergrond toeneemt zullen er wellicht meer knelpunten wat betreft cultuurwaardevrij selecteren aan het licht komen.

Meningen van deelnemers aan de selectieprocedure

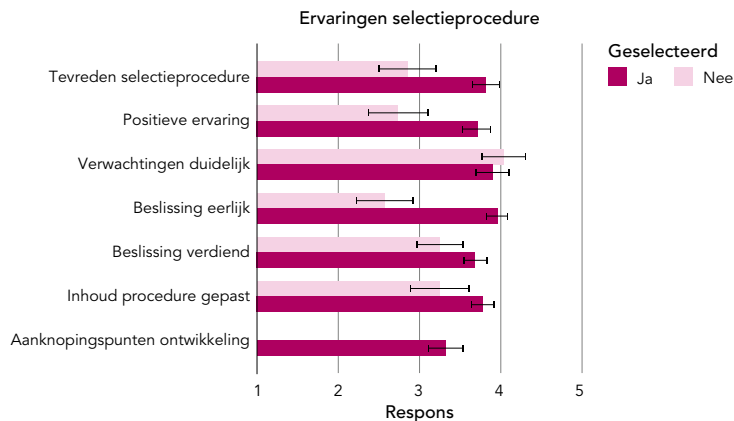
In dit hoofdstuk bespreken we de ervaringen en meningen van deelnemers aan de selectieprocedure. Om de ervaringen en meningen van voormalig deelnemers aan de selectieprocedure in kaart te brengen is een vragenlijst opgesteld op basis van de organisational justice theory (Steiner & Gilliland, 1996). Geselecteerde- en niet geselecteerde oud-kandidaten zijn uitgenodigd om aan dit deelonderzoek deel te nemen (respectievelijk via het rio-portaal en via e-mail). De vragenlijst is ingevuld door 153 respondenten, waarvan 116 respondenten voldoende vragen hebben beantwoord en na 2012 aan de selectieprocedure hebben deelgenomen. Van deze respondenten had 32% zich aangemeld voor een plaats voor rio's met beperkte werkervaring, had 68% zich aangemeld voor een plaats voor rio's met ruime werkervaring en was 65% toegelaten tot de rio-opleiding. Verder was 45% van de respondenten man en waren de meeste respondenten tussen de 30 en 49 jaar oud. Zes-en-negentig procent van de respondenten was in Nederland geboren, 8 procent had ten minste één ouder die in een ander westers land is geboren, en 7 procent had ten minste één ouder die in een ander niet-westers land is geboren. Gemiddeld hadden de respondenten 11 jaar relevante werkervaring ten tijde van hun sollicitatie ($SD = 6.2$, variërend van 2 tot 28 jaar). Alle respondenten hebben deelgenomen aan de briefselectie, 99% heeft deelgenomen aan de analytische test en 90% heeft deelgenomen aan het voorgesprek, het assessment, en het eindgesprek. Oud-kandidaten die in een vroeg stadium van de procedure zijn afgefallen waren dus ondervertegenwoordigd.

Alle respondenten hebben vragen over de selectieprocedure als geheel gekregen, vragen over de verschillende onderdelen zijn alleen voorgelegd aan respondenten die aan dat onderdeel hebben deelgenomen. Vragen over de lokale selectieprocedure zijn alleen voorgelegd aan respondenten die aangaven te zijn geselecteerd. Niet alle respondenten hebben alle vragen die aan hen zijn voorgelegd beantwoord (10%). In de analyse is 'pairwise deletion' toegepast bij respondenten met ontbrekende

gegevens, wat inhoud dat niet alle respondenten met ontbrekende gegevens buiten beschouwing zijn gelaten, maar dat voor iedere variabele en ieder onderzocht verband alle beschikbare gegevens zijn gebruikt. De steekproefgrootte verschilt hierdoor dus per variabele en per analyse ($n = 67 - 108$).

5.1 Tevredenheid over de selectieprocedure als geheel

In Figuur 1 staan de resultaten over de tevredenheid over de selectieprocedure als geheel weergegeven voor geselecteerde en niet-geselecteerde kandidaten. Zoals verwacht waren de geselecteerde kandidaten over het algemeen tevredener dan de niet-geselecteerde kandidaten; de geselecteerde kandidaten waren op alle punten gematigd positief. De niet-geselecteerde kandidaten vonden vooral de selectiebeslissingen niet erg eerlijk en vonden de deelname aan de procedure geen positieve ervaring. Dit kan verklaard worden aan de hand van self-serving bias; personen die minder goed presteren zijn sterker geneigd om dat resultaat aan hiaten of oneerlijkheden in de procedure toe te schrijven (Ployhart & Harold, 2004; Schmitt et al., 2004). Beide groepen oud-kandidaten geven aan dat de verwachtingen vooraf duidelijk waren.



Figuur 1. Gemiddelden en 95% betrouwbaarheidsintervallen voor reacties op vragen over ervaringen met de selectieprocedure voor de rio-opleiding.

5.1.1 Gepastheid inhoud procedure

Respondenten die aangaven dat zij de inhoud van de procedure niet gepast vonden (13%) is gevraagd om een toelichting te geven. Veertien respondenten hebben dat gedaan (waarvan 11 niet zijn geselecteerd). De meeste opmerkingen hadden betrekking op de Matrix test, die men te zwaarwegend, niet relevant, of te moeilijk vond. Verder gaven verschillende respondenten aan dat ze graag hadden gezien dat het functioneren in een relevante baan of juridische kennis ook onderdeel uitmaakte van de procedure. Een paar respondenten vonden de procedure niet efficiënt gezien de overlap in of de volgorde van onderdelen. Ook gaven een paar respondenten aan de sfeer tijdens het eindgesprek niet prettig te vinden (*negatief, onprettig, schofferend, respectloos*). Tenslotte werd de subjectiviteit van het oordeel op basis van het assessment en het eindgesprek een aantal keren als negatief punt benoemd.

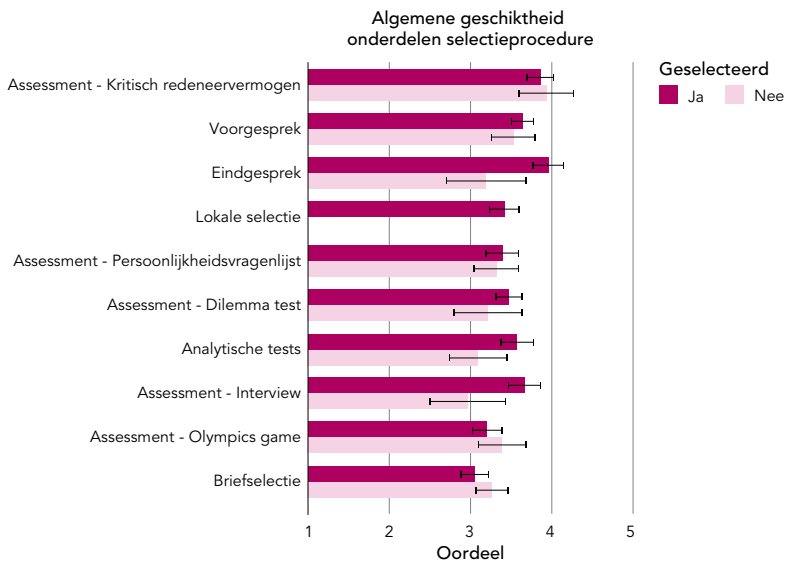
5.1.2 Ontwikkelpunten opleiding

Geselecteerde respondenten konden ook een toelichting geven op de mate waarin zij vonden dat de procedure aanknopingspunten bood voor hun ontwikkeling als rio. Negenenvijftig respondenten hebben dat gedaan. Ongeveer de helft van de respondenten gaf aan dat vooral het assessment die aanknopingspunten geboden had. Daarvan vonden twee respondenten dat er erg zwaar aan de ontwikkelpunten uit het assessmentrapport werd getild in de opleiding. Daarnaast gaven een aantal respondenten aan dat er geen ontwikkelpunten in het assessment rapport stonden, dat de ontwikkelpunten niet concreet genoeg of niet bruikbaar waren, of dat de ontwikkelpunten niet juist of niet passend waren. De meeste respondenten leken het bieden van ontwikkelpunten voor de opleiding geen hoofddoel van de selectie te vinden en enkelen respondenten gaven expliciet aan dat de opleiding zelf hier geschikter voor is. Wel stelden sommige respondenten dat de selectieprocedure aanzette tot zelfreflectie over de keuze voor het rechterschap en de eigen sterke en zwakke punten.

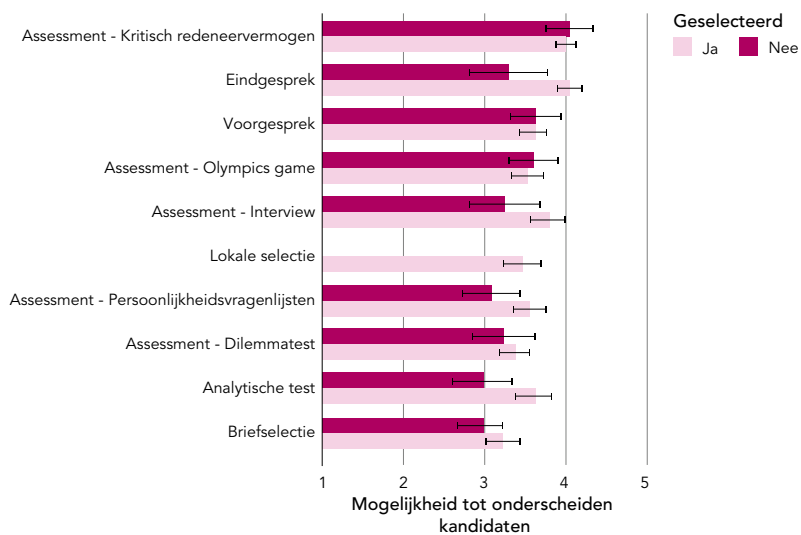
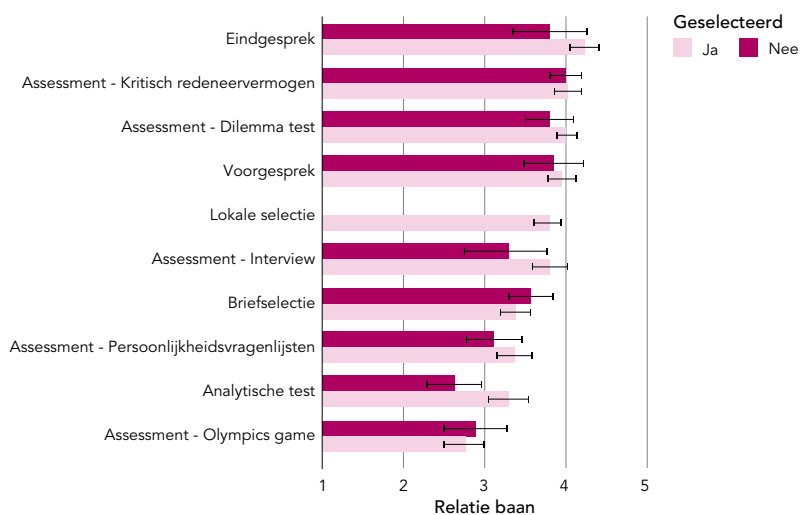
5.2 Meningen over onderdelen van de selectieprocedure

De algemene geschiktheid van de verschillende onderdelen van de selectieprocedure is gebaseerd op een vraag over de effectiviteit en een vraag over de eerlijkheid van dat onderdeel (Steiner & Gilliland, 1996). In Figuur 2 staan de reacties op vragen over de geschiktheid van de verschillende onderdelen van de selectieprocedure weerge-

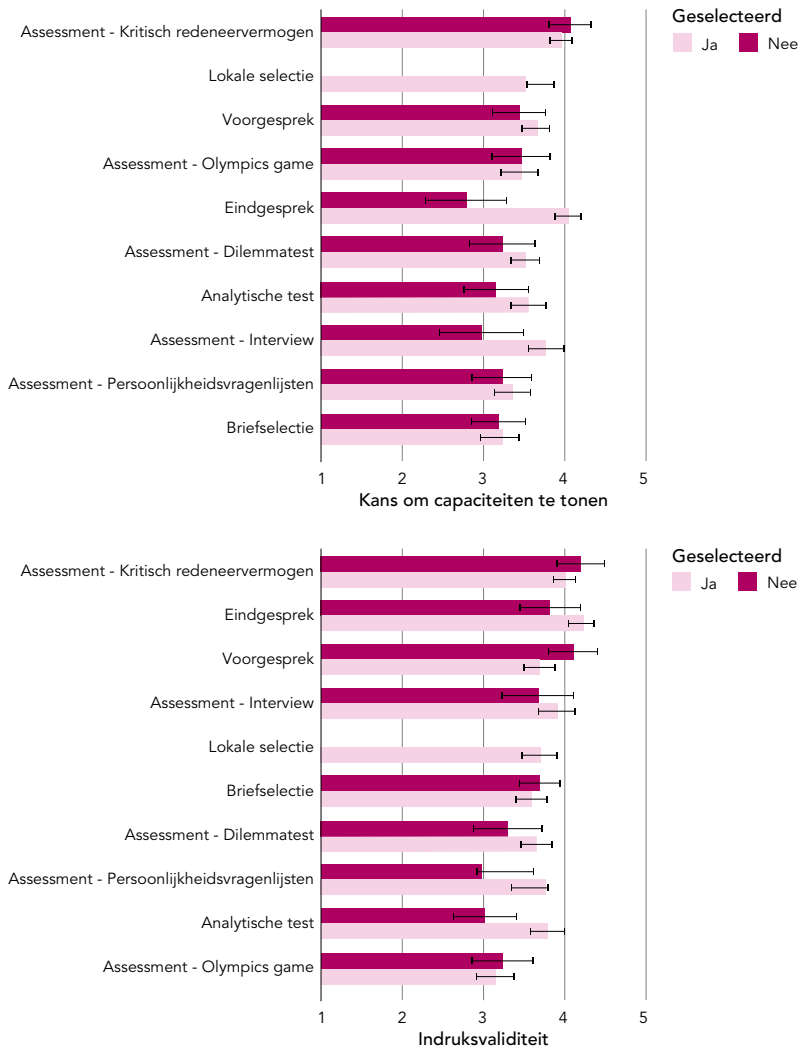
geven. In Figuur 3 staan de oordelen over de verschillende onderdelen op vier relevante procedurele rechtvaardigheidscriteria (de relatie met de baan, de mogelijkheid om zich te onderscheiden van andere kandidaten, de ogenschijnlijke validiteit en de kans om vaardigheden en kwaliteiten te laten zien), ieder gemeten met één vraag (Bauer et al., 2001; Steiner & Gilliland, 1996). Daarnaast konden de respondenten over elk onderdeel een tekstuele toelichting geven. Die toelichtingen zijn gecodeerd en geanalyseerd aan de hand van thematische analyse (Braun & Clarke, 2006). Hieronder worden de verschillende onderdelen van de selectieprocedure aan de hand van deze resultaten en de tekstuele opmerkingen van de respondenten besproken.



Figuur 2. Gemiddelden en 95% betrouwbaarheidsintervallen voor algemene geschiktheid van de verschillende onderdelen van de selectieprocedure voor de rio-opleiding.



Meningen van deelnemers aan de selectieprocedure



Figuur 3. Gemiddelden en 95% betrouwbaarheidsintervallen voor de procedurele rechtvaardigheidscriteria van de verschillende onderdelen van de selectieprocedure voor de rio-opleiding.

5.2.1 Briefselectie

De briefselectie werd van alle onderdelen het laagst beoordeeld op algemene geschiktheid en scoorde laag op de mogelijkheid om de eigen capaciteiten te tonen en om zich te onderscheiden van andere kandidaten. Dat is een opvallend resultaat, aangezien het cv over het algemeen als een geschikte methode wordt beoordeeld door sollicitanten (Anderson et al., 2010). Van de 38 respondenten die een tekstuele toelichting hebben gegeven merkten ongeveer de helft op dat ze de beoordelingscriteria als onduidelijk en subjectief hebben ervaren. Enkelen vonden het daarbij opvallend dat zij bij het een gerecht wel door de briefselectie kwamen, maar bij een ander gerecht niet. Een aantal respondenten stelde dat ze de briefselectie een beperkt maar noodzakelijk onderdeel van de selectie vonden. Verschillende respondenten hadden commentaar op het online systeem dat gebruikt wordt voor de briefselectie, met name op het functioneren daarvan ('viel uit') en op de omslachtigheid en de grote hoeveelheid informatie die gevraagd wordt. Andere thema's waren het gebrek aan eigen inbreng door het gestandaardiseerde format, de zwaarwegendheid van de maatschappelijke nevenactiviteiten (zowel het gebrek aan relevantie daarvan als de ongelijkheid in kansen om daaraan te voldoen), de negatieve invloed van de briefselectie op diversiteit, (door het talige aspect ervan) en het voordeel van het hebben van een bepaalde achtergrond (door te weten hoe zich te presenteren).

5.2.2 Analytische tests

De analytische tests werden als redelijk geschikt beoordeeld door de respondenten, wat in lijn is met eerdere onderzoeksresultaten (Anderson et al., 2010). De oordelen op de procedurele rechtvaardigheidsprincipes waren ook gematigd positief, maar, op het kunnen tonen van de eigen capaciteiten na, laag ten opzichte van de andere onderdelen. De oordelen van de niet-geselecteerde respondenten waren steeds wat lager dan die van de geselecteerde respondenten. Dit beeld komt ook terug in de tekstuele opmerkingen die 43 respondenten gemaakt hebben. De meeste opmerkingen hadden betrekking op de norm die gehanteerd wordt; veel respondenten vonden het opnemen van dergelijke test te verdedigen, maar de strenge grensscores die op bij dit onderdeel gehanteerd worden niet. Zij zouden graag zien dat ook andere zaken, zoals werkervaring, cv, een gesprek, of een relevante schrijfpodracht zouden worden meegenomen in de beslissing. Een aantal respondenten trokken de predictieve validiteit van de scores op de analytische tests in twijfel, waarvan de helft speci-

fiek de Matrix test noemde. Ook uitten enkele respondenten hun zorgen over de oefenbaarheid van de tests, wat oneerlijkheid in de hand zou kunnen werken. Daarnaast trokken enkele respondenten de betrouwbaarheid van de tests in twijfel, zij zagen de uitslag als het resultaat van een momentopname. Daarnaast stelden sommige respondenten juist ook dat het goed was om deze tests als onderdeel op te nemen, mist ze voorspellende waarde hebben voor het functioneren als rio of rechter.

5.2.3 Voorgesprek

Het voorgesprek werd als een geschikt onderdeel van de procedure gezien en is ook goed beoordeeld op de procedurele rechtvaardigheidsprincipes. Van de 28 respondenten die aanvullende opmerkingen hebben gemaakt stelden enkelen dat ze het gesprek als prettig en waardevol hebben ervaren en werd het contact met iemand van het gerecht ook als positief punt genoemd. Een aantal respondenten stelde dat ze het gesprek wel erg kort en inhoudelijk beknopt vonden, waarvan sommigen specifiek noemden dat er weinig ruimte was voor eigen inbreng. Opvallend was dat alleen een paar respondenten die wel door de procedure heen zijn gekomen de sfeer in het gesprek als negatief hebben ervaren.

5.2.4 Assessment

5.2.4.1 Persoonlijkheidsvragenlijsten

De persoonlijkheidsvragenlijsten die gebruikt worden in het assessment werden als redelijk geschikt beoordeeld, wat overeenkomt met eerdere onderzoeksresultaten (Anderson et al., 2010). In de tekstuele opmerkingen gaven verschillende respondenten aan dat het erg gemakkelijk is om deze vragenlijsten sociaal-wenselijk in te vullen. Sommige respondenten trokken de validiteit en betrouwbaarheid van de vragenlijsten in twijfel ("afhankelijk van de interpretatie van de vragen", "momentopname", "mist nuance").

5.2.4.2 Dilemmatest

De Dilemmatest werd als redelijk geschikt beoordeeld en ook als relevant voor de functie gezien. In de tekstuele opmerkingen kwam naar voren dat men dit inhoudelijk een zeer geschikt onderdeel vond. Een aantal respondenten gaven aan dit onderdeel liever in een gesprek te zien, zodat zij hun keuzes en afwegingen konden toelichten.

Ook gaven verschillende respondenten aan dat het erg gemakkelijk was om sociaal-wenselijk te antwoorden op dit onderdeel.

5.2.4.3 *Olympics game*

De Olympics game uit het assessment werd als redelijk geschikt beoordeeld, maar wel minder geschikt dan de andere onderdelen. Ook werd dit onderdeel relatief laag beoordeeld op de relatie met de baan en ogenschijnlijke validiteit. In de tekstuele opmerkingen gaven verschillende respondenten aan dat ze het niet duidelijk vonden waarom dit een onderdeel was, wat ermee getoetst werd en waarom dat relevant was. Andere respondenten waren van mening dat een dergelijke simulatieopdracht goed en nuttig kon zijn, maar dat de inhoud en de context aangepast zou moeten worden op de functie.

5.2.4.4 *Watson-Glaser test*

Deze tests voor kritisch redeneervermogen werd door de respondenten als meest geschikt beoordeeld en scoorde ook hoog op de procedurele rechtvaardigheidsprincipes. Ook in de tekstuele opmerkingen benadrukten de respondenten dat ze dit onderdeel als zeer relevant ervaren hebben; de relatie met de functie van rechter was erg duidelijk. Eén respondent (zonder migratieachtergrond) gaf aan deze test gezien de taligheid niet geschikt te vinden om het redeneervermogen van kandidaten die Nederlands niet als moedertaal spreken in kaart te brengen.

5.2.4.5 *Assessmentinterview*

Interviews worden vaak zeer positief gewaardeerd op algemene geschiktheid en op de procedurele rechtvaardigheidsprincipes (Anderson et al., 2010). Het is daarom opvallend dat dit interview slechts gematigd positief beoordeeld werd. Uit de tekstuele opmerkingen blijkt dat veel respondenten dit onderdeel als erg subjectief hebben ervaren. Zij noemden het voeren van een gesprek met slechts één persoon als negatief punt en enkelen vonden dat de toon en inhoud van de gesprekken niet consistent was, zoals zij hebben vernomen na met andere kandidaten te spreken. Enkele respondenten gaven aan liever met een rechter te spreken en twijfels te hebben aan het inzicht van de assessmentpsycholoog in de functie van rechter.

5.2.5 Eindgesprek

Het eindgesprek werd positief gewaardeerd op geschiktheid en op de procedurele rechtvaardigheidsprincipes. Opvallend is dat de verschillen tussen respondenten die wel en niet geselecteerd zijn op dit onderdeel het grootst waren. Een mogelijke verklaring daarvoor is dat dit het laatste onderdeel van de procedure is, waardoor het hieruit volgende oordeel beslissend is geweest voor de niet-geselecteerde kandidaten die hieraan hebben deelgenomen. Enkele respondenten gaven aan de sfeer onprettig te hebben gevonden en de indruk te hebben gehad dat dit met opzet werd gedaan “om je te testen”. Enkele anderen vonden het opmerkelijk dat er weinig aandacht was voor inhoudelijke kennis en kunde. Ook vonden sommige respondenten het onprettig dat het assessmentrapport een grote rol speelde in de eindgesprekken.

5.2.6 Lokale selectie

De lokale selectie is redelijk positief beoordeeld op geschiktheid en op de procedure rechtvaardigheidsprincipes. Van de 24 respondenten die een tekstuele toelichting gaven, vond ongeveer de helft dat de lokale procedure om ‘fit’ bij het gerecht zou moeten gaan en dat dit ook het geval was. Een paar respondenten vonden de lokale procedure omslachtig en overbodig na de landelijke procedure en sommigen gaven aan dat de landelijke procedure nog eens dunnetjes werd overgedaan, terwijl een paar andere respondenten het lokale gesprek juist erg kort vonden en ervoeren als een formaliteit. Enkele respondenten die bij de lokale selectie zijn afgewezen gaven aan dat zij de indruk kregen dat eerdere afwijzing een stigma opleverde, terwijl de afwijzing betrekking zou moeten hebben op ‘fit’ bij het gerecht, niet op geschiktheid. Ook vonden sommigen van hen dat er wel erg veel kandidaten voor een enkele plek waren bij de lokale selectie.

5.3 Verschillen naar leeftijd en migratieachtergrond

5.3.1 Leeftijd

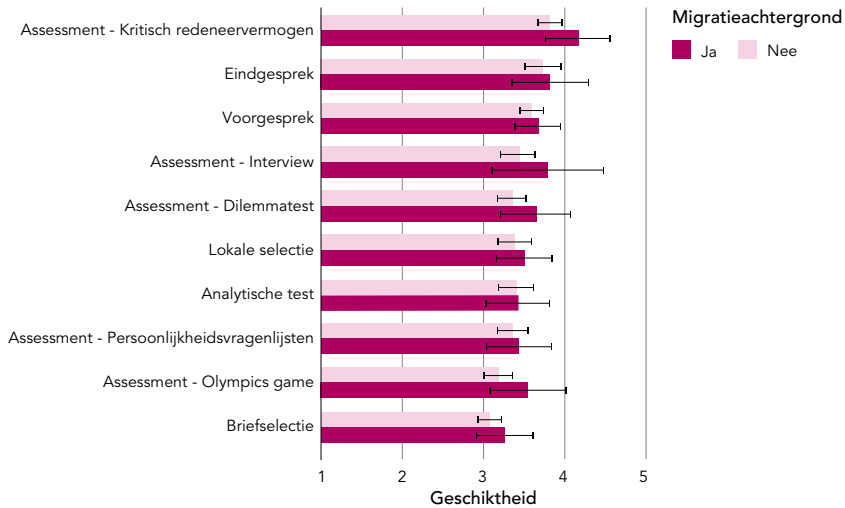
Gezien de bevindingen vanuit LTP en de wetenschappelijke literatuur (Fischer et al., 2017; Klein et al., 2015) dat oudere kandidaten lager scoren op de analytische test, zouden ook de meningen over deze tests kunnen verschillen tussen kandidaten van

verschillende leeftijden. Tests die fluïde intelligentie meten, zoals de Matrix test, laten verschillen in scores zien ten nadele van oudere kandidaten. Voor tests die gekristalliseerde intelligentie meten is dat niet of in mindere mate het geval (Klein et al., 2015). Een vaak gehanteerde leeftijdsgrens in deze context is die van 40 jaar of ouder. Respondenten van 40 jaar ($n = 37$) of ouder vonden de analytische tests zoals verwacht minder geschikt dan respondenten jonger dan 40 jaar ($n = 71$). Een t-toets voor het verschil in de beoordeelde geschiktheid van de analytische tests tussen deze twee groepen leverde een statistisch significant verschil op ($t_{(106)} = 2.20, p = .03$). Het verschil in beoordeelde geschiktheid was middelgroot ($d = 0.45, 95\% \text{ BHI } [0.85, 0.04]$).

5.3.2 Migratieachtergrond

De verschillende onderdelen van de selectieprocedure laten waarschijnlijk in verschillende mate scoreverschillen zien tussen personen met of zonder migratieachtergrond. (Ployhart & Holtz, 2008). Dat zou ook de percepties over geschiktheid van de verschillende onderdelen kunnen beïnvloeden. Daarom hebben we de verschillen in de beoordeelde geschiktheid van de verschillende onderdelen vergeleken voor personen met ($n = 18$) en zonder migratieachtergrond ($n = 97$). Gezien het kleine aantal respondenten met een migratieachtergrond is hierbij geen uitsplitsing gemaakt naar een westerse en niet-westerse migratieachtergrond. Wel is er eerst gekeken of de respondenten met of zonder migratieachtergrond verschillen in hoe vaak ze zijn geselecteerd. Daarin vonden we in deze steekproef geen wezenlijk verschil (respectievelijk 61% en 65%). De beoordeelde geschiktheid per onderdeel staat weergegeven in Figuur 4. Tegen de verwachtingen in is te zien dat de respondenten met een migratieachtergrond de meeste onderdelen als even geschikt of zelfs als iets geschikter beoordeelde dan de respondenten zonder migratieachtergrond. Een ANOVA voor herhaalde metingen resulteerde in een niet-significant verschil in beoordeelde geschiktheid op basis van migratieachtergrond ($F_{(1,59)} = 3.18, p = .08, \eta_p^2 = .05$) en een niet-significant interactie-effect tussen onderdeel en migratieachtergrond ($F_{(7.96, 469.85)} = 0.41, p = .60, \eta_p^2 = .01$). Er zijn dus geen verschillen gevonden in de beoordeelde geschiktheid van de selectie-onderdelen op basis van migratieachtergrond.

Meningen van deelnemers aan de selectieprocedure



Figuur 4. Gemiddelden en 95% betrouwbaarheidsintervallen voor de algemene geschiktheid van de verschillende onderdelen van de selectieprocedure voor de rio-opleiding, voor personen met en zonder migratieachtergrond.

5.4 Conclusie en aanbevelingen

Over de gesprekken bij de LSR waren de respondenten overwegend positief. Het minst positief is geoordeeld over de 'Olympics game' in het assessment en de briefselectie. De test voor kritisch redeneervermogen in het assessment werd het positiefst beoordeeld. Respondenten ouder dan 40 jaar beoordeelden de analytische tests als minder geschikt dan respondenten jonger dan 40 jaar. Mogelijke oorzaken hiervoor zijn een verschil in verwachte prestaties (Oostrom & De Soete, 2016), maar het kan ook zijn dat oudere kandidaten, die in de regel over meer relevante werkervaring zullen beschikken, vinden dat ervaring relevanter is dan analytisch vermogen zoals op deze manier getest. Verder vonden we geen verschillen in de mate van beoordeelde geschiktheid van de onderdelen op basis van migratieachtergrond. Ten slotte blijkt dat vooral instrumenten die inhoudelijk goed aansluiten bij de functie van rechter (de

Watson-Glaser test voor kritisch redeneervermogen en de gesprekken bij de LSR), positief beoordeeld worden, zoals ook gevonden wordt in de literatuur (Lievens & De Soete, 2012).

Deze resultaten laten weinig onverwachte resultaten of grote redenen tot zorg zien, maar tonen wel enkele aandachtspunten. Het doel van iedere stap en ieder onderdeel van de selectieprocedure, ook van de lokale selectie en de verschillende onderdelen van het assessment, kunnen duidelijker naar de kandidaten gecommuniceerd worden. Dat geldt ook voor de gestelde eisen en regels. Ook bevelen we aan om een document op te stellen voor kandidaten met een verantwoording en onderbouwing van de opzet van de procedure. Dit kan de acceptatie vergroten. De mogelijkheid tot sociaal-wenselijk antwoorden op de persoonlijkheidsvragenlijsten en de Dilemmatest is een aandachtspunt dat ook al naar voren kwam in de gesprekken met betrokkenen bij de selectieprocedure en volgens de literatuur een serieus probleem voor de validiteit kan opleveren. De opmerkingen die kandidaten hierover maakten bevestigen dit beeld en onze aanbeveling om dit type instrument niet te gebruiken en selectieprocedures.

AANBEVELINGEN 13 – MENINGEN EN ERVARINGEN VAN KANDIDATEN

- Om de acceptatie van de gebruikte procedures te vergroten kan informatie worden verstrekt aan kandidaten over de reden om de procedure op een bepaalde manier in te richten.
- Het gebruiken van selectie-instrumenten die inhoudelijk goed aansluiten bij de rio-opleiding en de functie van rechter, zorgen waarschijnlijk voor een grotere tevredenheid van kandidaten (en hebben een goede predictieve validiteit).

Analyses op basis van het rio-volgsysteem

In dit hoofdstuk behandelen we de relaties tussen scores en beoordelingen uit de selectieprocedure en beoordelingen uit de rio-opleiding; de predictieve validiteit. Om uitspraken te kunnen doen over de effectiviteit van de verschillende onderdelen van de selectieprocedure is het van belang om empirische relaties tussen scores op de verschillende onderdelen en latere criteriumprestaties (prestaties in de rio-opleiding) te onderzoeken. Voor deze analyses zijn gegevens over kandidaten gebruikt waarvan informatie over prestaties tijdens de selectieprocedure en beoordeling(en) van prestaties tijdens de rio-opleiding (deels) beschikbaar waren. De gegevens zijn aangeleverd door de LSR, SSR en LTP. Ook bespreken we aan de hand van deze gegevens het verband tussen leeftijd en scores op de analytische tests en het verschil in scores tussen de eerste kans en een herkansing op de analytische tests.

6.1 Data en analyse

6.1.1 Selectiegegevens

Om de predictieve validiteit van de verschillende onderdelen te kunnen analyseren zijn gegevens aangeleverd over de behaalde scores op de analytische tests, de beoordelingen op basis van het voorgesprek (per selecteur, per thema en eindbeoordelingen), de competentiebeoordelingen (beneden, op, of boven niveau) en het advies uit het assessmentrapport, scores op verschillende tests en vragenlijsten uit het assessment en de beoordelingen op basis van het eindgesprek (per selecteur, per thema en eindbeoordelingen).

6.1.2 Gegevens beoordelingen rio-opleiding

Ook de beschikbaarheid en kwaliteit van beoordelingsgegevens uit de rio-opleiding zijn van belang om conclusies te kunnen trekken over de predictieve validiteit van de verschillende onderdelen van de selectieprocedure. De beoordelingen in de opleiding worden gedaan door leden van de beoordelingscommissie. Zij beoordelen het portfolio van de rio aan de hand van een beoordelingslijst.

In de beoordelingen die gebruikt zijn voor dit onderzoek zijn twee verschillende versies van de scoringslijst gebruikt. Versie 1 van de lijst bestond uit 101 vragen en versie 2 (de huidige lijst) bestond uit 75 vragen. Iedere vraag behoort tot een bepaalde competentie uit het rechtersprofiel. Om de scores op beide versies te kunnen gebruiken zijn alleen vragen gebruikt die in beide versies voorkomen. Er waren 69 vragen die in beide versies voorkwamen. In de tabel hieronder staat het aantal vragen per competentie weergegeven. Dezelfde vraag kan bij meerdere competentiescores mee worden geteld.

Tabel 5 Aantal vragen per competentie in de beoordelingslijst.

Competentie	Aantal beoordelingspunten
Regievoering	15
Analytisch vermogen	20
Omgevingsbewustzijn	9
Luisteren	10
Overredingskracht	11
Samenwerken	9
Zelfvertrouwen & authenticiteit	10
Flexibiliteit	5
Zelfbewustzijn	7
Leervermogen en zelfreflectie	4
Besluitvaardigheid	7
Juridische kennis	6
Totaalscore	69

Op basis van deze vragen is een totaalscore berekend (het percentage behaalde punten ten opzichte van het maximaal te behalen punten). Dit komt overeen met de wijze waarop totaalscores worden berekend op de beoordelingslijst. Op deze wijze leveren sommige competenties echter een onevenredig grote bijdrage aan de totaalscore, omdat het aantal vragen behorend bij de verschillende competenties verschilt. Daarom is ook een gewogen totaalscore berekend, waarbij de score op iedere competentie even sterk bijdraagt aan de totaalscore.

De scoringslijst kan op vier verschillende taakniveaus (A, B, C en D, waarbij A het startniveau is en D het eindniveau) worden ingevuld. De scores op de vragen die volgens verschillende taakniveaus zijn ingevuld zijn daarom niet vergelijkbaar. Het gehanteerde taakniveau was daarnaast in een groot percentage van de aangeleverde beoordelingen (55%) onbekend. Als voor de beoordelingen waarvan het taakniveau wel bekend was afzonderlijke analyses voor ieder niveau zouden worden uitgevoerd, zou dat erg kleine steekproeven opleveren. Daarom is ervoor gekozen om alleen beoordelingen te gebruiken die op taakniveau B of C zijn uitgevoerd. De omschrijvingen van deze taakniveaus leken relatief weinig te verschillen en deze keuze leverde de groots mogelijke steekproeven op.

De interbeoordelaarsbetrouwbaarheid van de resulterende competentiescores en totaalscores staan in Tabel 6. De interbeoordelaarsbetrouwbaarheid van de totaalscores en van de meeste competentiescores was redelijk, de interbeoordelaarsbetrouwbaarheid voor de competenties *omgevingsbewustzijn* en *samenwerken* waren onvoldoende. Dit betekent dat de beoordelingen van de verschillende beoordelaars in de opleiding redelijk met elkaar overeenkomen. Een lage betrouwbaarheid van de criteriummaten heeft een onderdrukkend effect op de predictieve validiteit. Bij de interpretatie van de resultaten over de predictieve validiteit, die verderop in dit hoofdstuk besproken worden, moeten we dus rekening houden met de suboptimale betrouwbaarheid van de beoordelingen uit de rio-opleiding.

Er is geen analyse uitgevoerd om de constructvaliditeit, ofwel de onderscheidbaarheid van de verschillende competenties te onderzoeken. Aangezien ongeveer de helft van de vragen uit de beoordelingslijst bij meerdere competentiescores wordt meegeteld, zal namelijk dit per definitie niet het geval zijn. We bevelen aan om de beoordelingslijst zo in te richten dat iedere vraag onder één competentie geschaard kan worden.

Tabel 6 Interbeoordelaarsbetrouwbaarheid van beoordelingen in de rio-opleiding.

Score	ICC
Totaalscore	.71 [.58, .81]
Gewogen totaalscore	.66 [.50, .77]
Competentiescores	
Regievoering	.72 [.58, .81]
Analytisch vermogen	.78 [.67, .85]
Omgevingsbewustzijn	.36 [.07, .58]
Luisteren	.64 [.48, .76]
Overredingskracht	.65 [.49, .77]
Samenwerken	.42 [.15, .61]
Zelfvertrouwen & authenticiteit	.58 [.39, .72]
Flexibiliteit	.73 [.61, .82]
Zelfbewustzijn	.50 [.27, .67]
Leervermogen en zelfreflectie	.50 [.27, .66]
Besluitvaardigheid	.67 [.51, .78]
Juridische kennis	.55 [.34, .70]

Notitie. ICC = intra-class correlatie, 95% betrouwbaarheidsinterval tussen haakjes. $n = 77$. Interpretatie: $< .50$ = onvoldoende, $0.5 - 0.75$ = redelijk, $0.75 - 0.9$ = goed, $> .90$ = excellent (Koo & Li, 2016).

AANBEVELINGEN 14 – BEOORDELINGEN IN DE RIO-OPLEIDING

- Voor evaluatieonderzoek is het noodzakelijk dat beoordelingen uit de opleiding samen geanalyseerd kunnen worden. Het zou voor dit doel erg handig zijn als de beoordelingen niet per taakniveau gedaan worden. In plaats daarvan kan bijvoorbeeld de cesuur worden aangepast aan het niveau.
- Ook voor het in kaart brengen van de ontwikkeling van rio's is het inzichtelijker als verschillende beoordelingen met elkaar vergeleken kunnen worden.
- Veel vragen op de beoordelingslijst worden onder meerdere competenties geschaard. Dat heeft waarschijnlijk tot gevolg dat de competentiescores niet empirisch van elkaar onderscheiden kunnen worden. We bevelen aan om de beoordelingslijst zo in te richten dat iedere vraag onder één competentie geschaard kan worden.
- Op de huidige manier van beoordelen hebben competenties die met meer vragen worden beoordeeld een zwaarder gewicht in de totaalscore. We raden aan om deze manier van beoordelen te herzien.

6.1.3 Steekproefgrootte

De hierboven beschreven selectie van beoordelingsgegevens leverde een steekproef op van 77 personen op die tussen april 2014 en april 2017 aan de selectie voor de rio-opleiding hebben deelgenomen, waarvan gegevens over de prestaties tijdens de selectieprocedure (deels) beschikbaar waren en die aan de rio-opleiding zijn begonnen en een beoordeling op taakniveau B of C hebben gekregen. Van deze personen was 30% man, volgde 46% de rio-opleiding voor personen met ruime werkervaring, volgende 33% de rio-opleiding voor personen met beperkte werkervaring en was voor 21% de variant van de rio-opleiding niet bekend. De gemiddelde leeftijd was 38 jaar. De gebruikte beoordeling was voor 70% de eerste beoordeling, voor 29% de tweede beoordeling en voor 1% de derde beoordeling in de rio-opleiding. Voor 8% was de gebruikte beoordeling de eindbeoordeling van de opleiding. Het aantal personen waarvan zowel beoordelingsgegevens als selectiescores bekend waren varieerde tussen 65 en 37, afhankelijk van het selectieonderdeel. Gezien het kleine aantal personen waarvan zowel de scores op de persoonlijkheidsvragenlijsten uit het assessment als beoordelingsgegevens beschikbaar waren, hebben we de verbanden tussen de scores op de persoonlijkheidsvragenlijsten en de beoordelingen uit de rio-opleiding niet opgenomen in het rapport.

AANBEVELING 15 – BESCHIKBAARHEID VAN GEGEVENS

- Het is erg belangrijk om in de nabije toekomst een beter, completer en handzamer systeem voor het bewaren en opslaan van gegevens te realiseren. Dit maakt het mogelijk om in de toekomst evaluatieonderzoek uit te voeren op basis van meer en meer complete gegevens, wat tot nauwkeurigere conclusies zal leiden.

6.1.4 Analyse

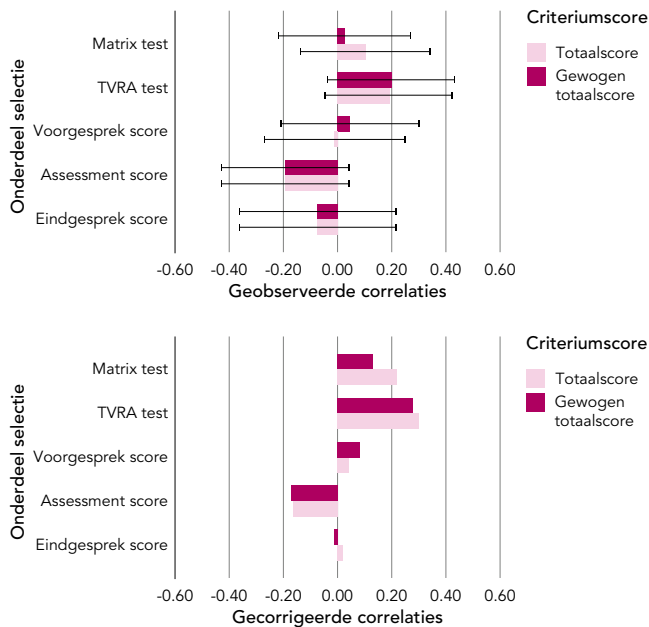
Om het verband tussen de scores op de selectieonderdelen en beoordelingen in de opleiding te onderzoeken zijn correlaties tussen deze variabelen berekend. Gezien de kleine steekproefgrootten waarop deze correlaties gebaseerd zijn, vermelden we waar mogelijk ook de 95% betrouwbaarheidsintervallen rondom de correlaties in de tabellen. Dit geeft de mate van (on)nauwkeurigheid van de correlaties aan. De kleine steekproefgrootten zorgen ervoor dat de schattingen van verbanden tussen variabelen minder nauwkeurig zullen zijn en dat er alleen bij relatief sterke verbanden sprake zal zijn van statistische significantie. We zullen bij de interpretatie van de gegevens vooral uitgaan van de sterkte van de verbanden. De resultaten zullen met voorzichtigheid geïnterpreteerd moeten worden.

Wanneer we, zoals hier het geval is, correlaties berekenen op basis van gegevens van rio's die geselecteerd zijn, zorgt dat voor *restrictie in range*. Restrictie in range zorgt voor een onderschatting van de correlatie tussen de scores op de selectie-instrumenten en prestaties in de rio-opleiding. Daarom hebben we de correlaties daarvoor gecorrigeerd¹. Statistische significantie en 95% betrouwbaarheidsintervallen van de correlaties zijn bepaald voor het toepassen van de correctie. De mate van (on)nauwkeurigheid wordt dus alleen voor de ongecorrigeerde (geobserveerde) correlaties getoond.

1 Voor de correctie voor range restrictie is gebruik gemaakt van het expectation maximization algoritme. Deze methode is het meest geschikt wanneer correlaties gebaseerd zijn op kleine steekproeven en wanneer het percentage geselecteerden klein is. Bovendien is deze methode geschikt voor getrapte selectieprocedures zoals deze en kunnen ook correlaties voor voorspellers waarvan geen scores van afwezen kandidaten bekend zijn gecorrigeerd worden (Fife et al., 2016). We beschikken over scores op de analytische tests van zowel geselecteerde als afwezen kandidaten ($n = 661$). Deze scores zijn gebruikt om ook selectievariabelen waarvan geen gegevens van afwezen kandidaten bekend zijn, zoals de scores op het voor- en eindgesprek, te corrigeren voor restrictie in range.

6.2 Overzicht predictieve validiteit selectieonderdelen

In Figuur 5 worden de correlaties tussen de belangrijkste onderdelen van de selectieprocedure en de totaalscores in de rio-opleiding weergegeven, voor en na correctie voor restrictie in range.



Figuur 5. Correlaties tussen onderdelen uit de selectieprocedure en totaalscores op beoordelingen in de opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 48-65$.

De scores op de Matrix test hadden lage correlaties met de beoordelingen uit de rio-opleiding en de correlaties tussen de scores op de TVRA en beoordelingen in de opleiding waren middelgroot. De correlaties tussen het gemiddelde eindoordeel op de voorgesprekken en de beoordelingen uit de rio-opleiding waren klein. Het gemiddelde van alle competentiebeoordelingen uit het assessmentrapport liet kleine nega-

tieve correlaties (verbanden) zien met de beoordelingen uit de rio-opleiding. Ten slotte waren de correlaties tussen het gemiddelde eindoordeel op de eindgesprekken en de beoordelingen uit de rio-opleiding nagenoeg gelijk aan nul. Op basis van deze resultaten lijken alleen de TVRA en wellicht de Matrix test een predictieve validiteit (voorspellend vermogen) van bruikbare grootte te hebben. Hieronder worden de resultaten voor de verschillende selectieonderdelen in meer detail uiteengezet.

6.3 Analytische tests

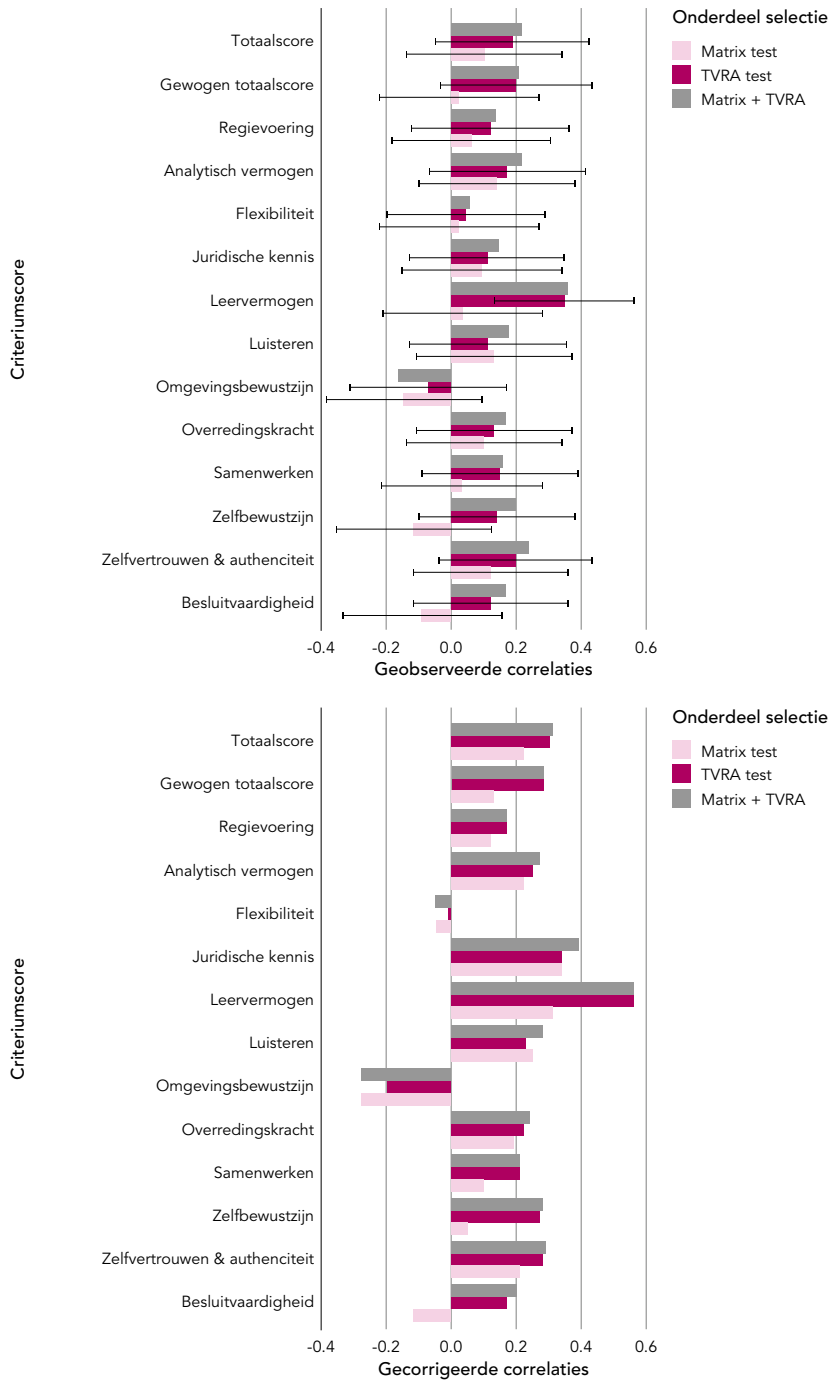
In Figuur 6 staan de correlaties tussen scores op de analytische tests en de totaal- en competentiescores behaald tijdens de rio-opleiding. Ook wordt de gezamenlijke correlatie (de correlatie als beide tests samen worden gebruikt) weergegeven. Meer gedetailleerde resultaten staan in Tabel B1 in Bijlage 4. De predictieve validiteit van de scores op de Matrix test is in de meeste gevallen klein, maar is middelgroot voor de competenties *juridische kennis* en *leervermogen & zelfreflectie*. De predictieve validiteit van de scores op de TVRA is voor de meeste competenties middelgroot en hoger dan die van de Matrix test. Gezien het sterk talige karakter van het rechtschap is dit niet verrassend.

In Figuur 6 is ook zichtbaar dat de predictieve validiteit van de Matrix test en de TVRA tezamen voor de totaalscores en de meeste competenties nauwelijks groter is dan de predictieve validiteit van de TVRA alleen. Dit houdt in dat de Matrix test weinig incrementele validiteit heeft bovenop de TVRA; voorspellingen op basis van de TVRA zijn ongeveer even goed als voorspellingen op basis van de TVRA en de Matrix test samen. Meer gedetailleerde resultaten over de incrementele validiteit van de Matrix test bovenop de TVRA staan in Tabel B1 in Bijlage 4.

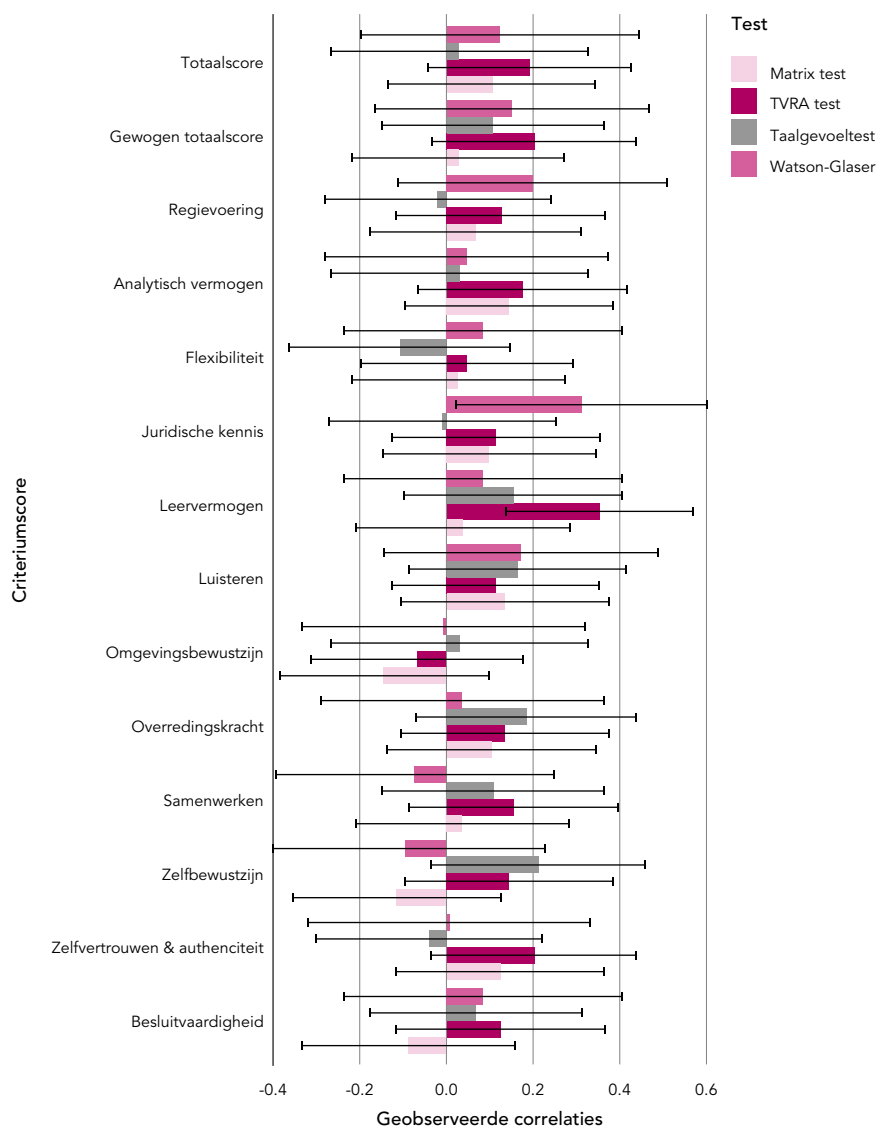
6.4 Andere cognitief georiënteerde tests

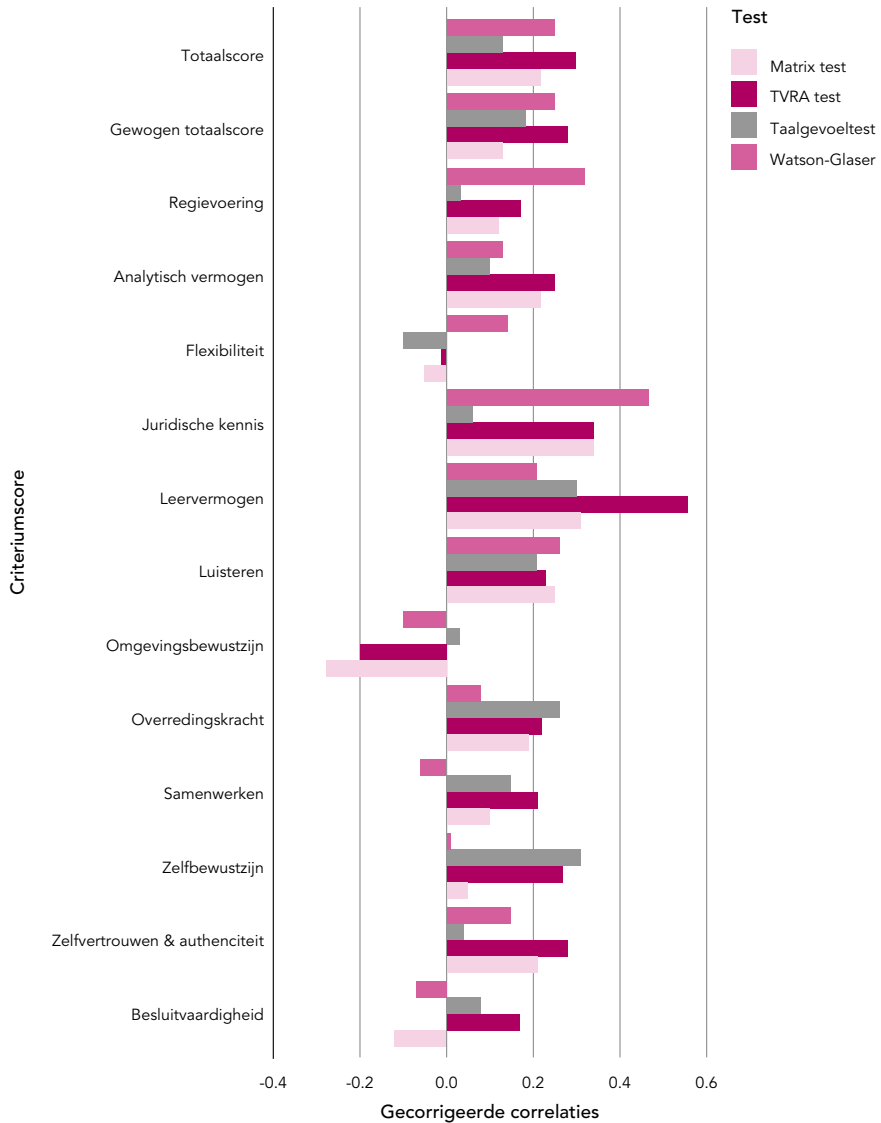
De Taalgevoeltest en de Watson-Glaser test voor kritisch redeneervermogen zijn twee andere cognitief georiënteerde tests die tijdens de selectieprocedure worden afgenomen. De scores op deze tests worden in de procedure wel meegewogen, maar er wordt geen harde ondergrens gehanteerd en de scores worden samen met andere zaken die tijdens het assessment naar voren komen beoordeeld. In Figuur 7 staan de correlaties tussen de scores op beide test en de verschillende beoordelingen tijdens de rio-opleiding. Ter vergelijking zijn ook de correlaties voor de Matrix test en de TVRA weergegeven. Meer gedetailleerde resultaten staan in Tabel B2 in Bijlage 4.

Analyses op basis van het rio-volgsysteem



Figuur 6. Correlaties tussen scores op de analytische tests en beoordelingen tijdens de rio-opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 65$.





Figuur 7. Correlaties tussen scores op de Taalgevoeltest, de Watson-Glaser test, de analytische tests en beoordelingen tijdens de rio-opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 65$ voor de analytische tests, $n = 58$ voor de Taalgevoeltest en $n = 37$ voor de Watson-Glaser test.

Voor beide varianten van de totaalscore is de predictieve validiteit van de Taalgevoeltest klein. Dat was ook het geval voor de competentiescores, met uitzondering van middelgrote correlaties met de scores voor *leervermogen*, *overredingskracht* en *zelfbewustzijn*. De scores op de Taalgevoeltest hadden geen incrementele validiteit bovenop de TVRA voor het voorspellen van de totaalscores². De correlaties tussen de scores op de Watson-Glaser test voor kritisch redeneervermogen en de beide totaalscores waren middelgroot. De predictieve validiteit was daarmee ongeveer even groot als die van de TVRA. De scores op de Watson-Glaser test lieten ook middelgrote correlaties zien met de competentiescores voor *regievoering*, *juridische kennis*, *leervermogen* en *luisteren*. De correlaties met de andere competentiescores waren klein. De scores op de Watson-Glaser test hadden echter geen incrementele validiteit bovenop de TVRA voor het voorspellen van de totaalscores³.

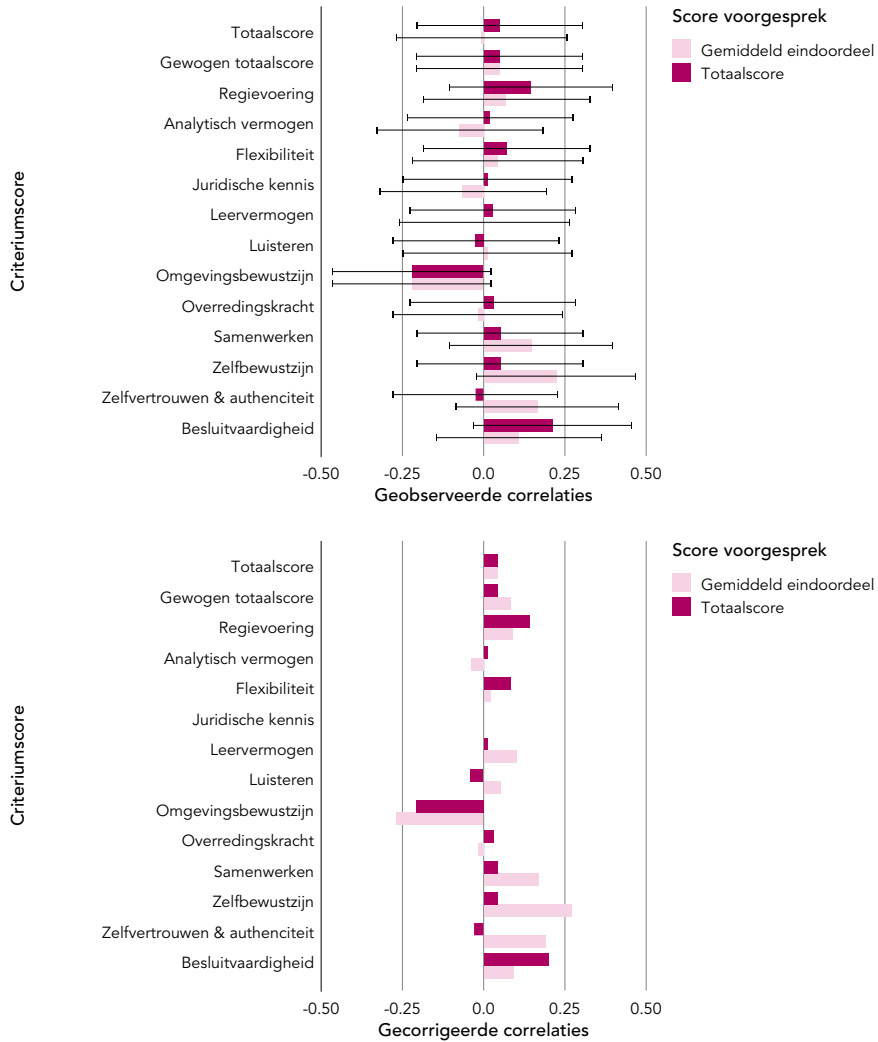
6.5 Voorgesprek

In Figuur 8 staan de correlaties tussen de beoordelingen gemaakt op basis van het voorgesprek en de beoordelingen tijdens de rio-opleiding. Het gemiddelde eindoordeel en de gemiddelde totaalscore op het voorgesprek zijn berekend als gemiddelde over de drie selecteurs die de voorgesprekken voeren. In Tabel B3 in Bijlage 4 staan ook de verbanden met de verschillende thema's die tijdens het voorgesprek beoordeeld worden. De correlaties van de totaalscores op het voorgesprek en het gemiddelde eindoordeel op het voorgesprek met de (gewogen) totaalscore in de rio-opleiding waren klein. Dat was ook het geval voor de meeste scores op de verschillende competenties beoordeeld in de rio-opleiding. De predictieve validiteit van het voorgesprek lijkt op basis van deze resultaten dus klein te zijn.

2 $\Delta R^2 \leq .01$ voor beide totaalscores, gecorrigeerd en ongecorrigeerd.

3 $(\Delta R^2 \leq .01$ voor beide totaalscores, gecorrigeerd en ongecorrigeerd).

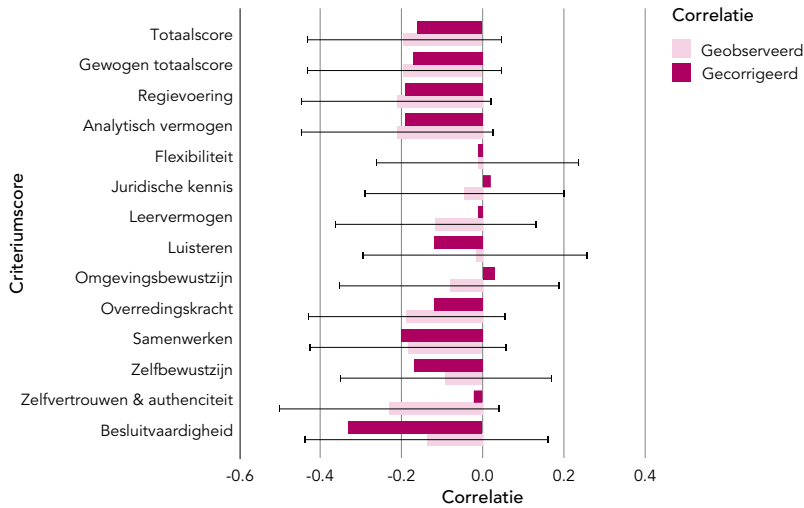
Analyses op basis van het rio-volgsysteem



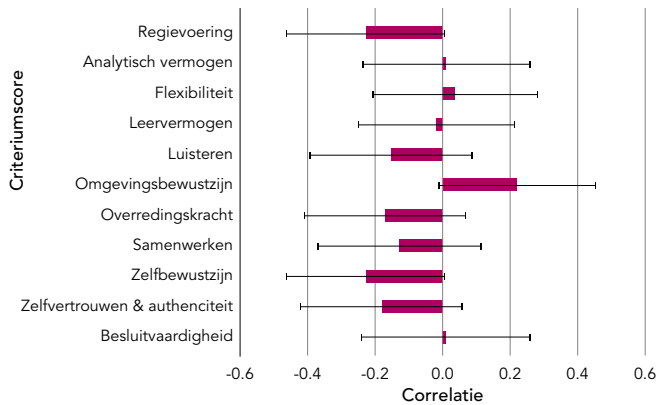
Figuur 8. Correlaties tussen scores op het voorgesprek en beoordelingen tijdens de rio-opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 58$.

6.6 Assessment

Beslissingen op basis van het assessment worden gemaakt na afloop van het eindgesprek, gebruikmakend van het assessmentrapport opgesteld door LTP, bestaand uit een kwantitatief en een kwalitatief beschrijvend gedeelte. Voor deze analyses kunnen we alleen het kwantitatieve gedeelte gebruiken. Dat gedeelte bestaat uit oordelen op de verschillende competenties uit het rechtersprofiel, zoals beoordeeld door een assessmentpsycholoog van LTP. Ook wordt er een advies gegeven (positief of negatief). Dit advies was voor de kandidaten waarvan gegevens beschikbaar waren bijna altijd positief, waardoor het niet mogelijk was om het verband tussen het LTP-advies en beoordelingen in de rio-opleiding te onderzoeken. Ook was het niet mogelijk om de competentiescores uit het assessment rapport op een statistische manier te reproduceren, aangezien de assessmentinterviews niet gescoord worden en de scores op verschillende assessment onderdelen (de Dilemmatest, de Hogan Development Survey en de Olympics game) niet bekend waren. Voor de scores op de persoonlijkheidsvragenlijsten was de steekproef met volledige gegevens bovendien dermate klein dat we de verbanden met beoordelingen uit de rio-opleiding niet rapporteren. De verbanden tussen het gemiddelde assessment oordeel (het gemiddelde over alle competentiebeoordelingen) en de beoordelingen uit de rio-opleiding staan weergegeven in Figuur 9. De correlaties tussen de beoordelingen op de verschillende competenties uit het assessment en de oordelen op dezelfde corresponderende competenties zoals beoordeeld in de rio-opleiding staan in Figuur 10. Meer gedetailleerde resultaten staan in Tabel B4 in Bijlage 4. Opvallend is dat, tegen de verwachtingen in, bijna alle correlaties tussen het gemiddelde assessment oordeel en de beoordelingen in de rio-opleiding dichtbij nul lagen of klein en negatief waren. Ook de correlaties tussen competentieoordelen uit het assessmentrapport en beoordelingen op dezelfde competenties uit de rio-opleiding lagen in de meeste gevallen dichtbij nul of waren klein en negatief. Eén van de mogelijke verklaring is de geringe spreiding in de oordelen uit het assessmentrapport. In de gebruikte gegevens werden de competenties zelden als 'beneden niveau' beoordeeld, waardoor er moeilijk onderscheid gemaakt kan worden tussen kandidaten. Hoewel de schattingen van de predictieve validiteit van de beoordelingen uit het assessmentrapport onnauwkeurig zijn (zie de 'error bars' in Figuur 9 en 10), lijken de beoordelingen uit het assessmentrapport geen predictieve validiteit te hebben voor de beoordelingen in de rio-opleiding.



Figuur 9. Correlaties tussen het gemiddelde assessment oordeel en beoordelingen tijdens de rio-opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 63$.



Figuur 10. Rangcorrelaties (Spearman's r) tussen oordelen op de competenties uit het assessmentrapport en oordelen over de corresponderende competenties uit de rio-opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 63$.

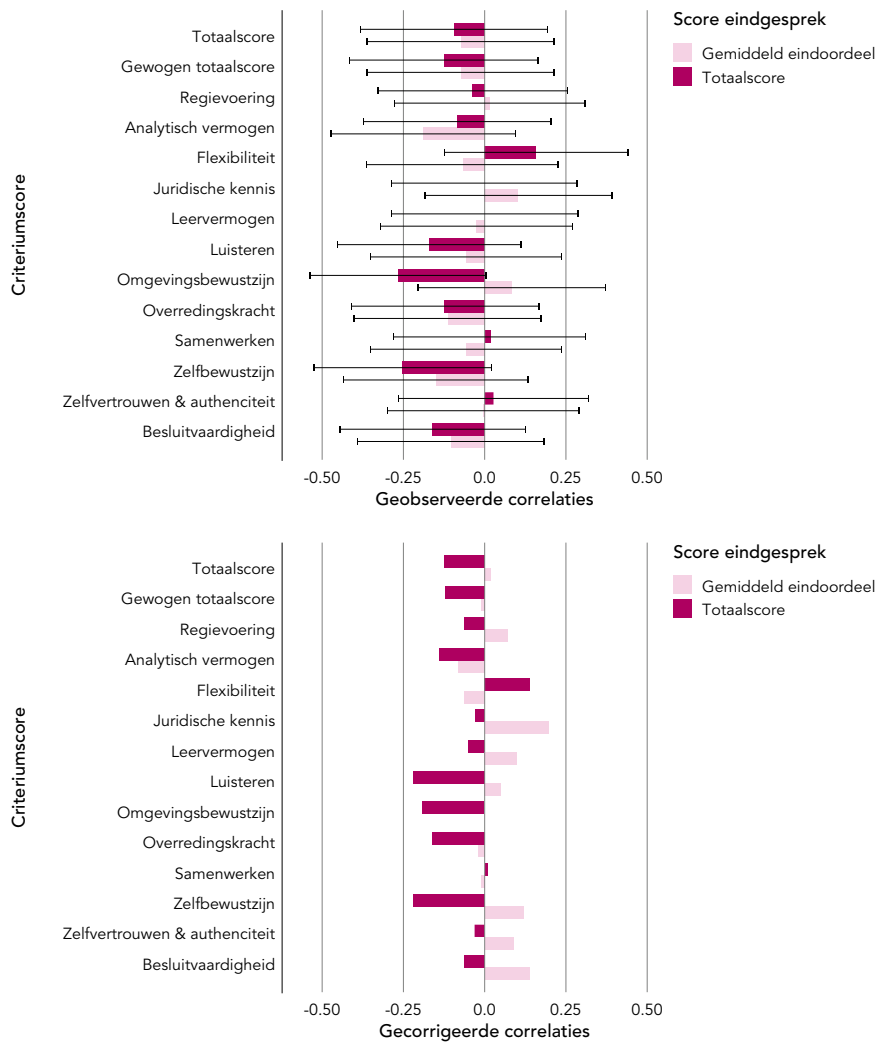
6.7 Eindgesprek

In Figuur 11 staan de correlaties tussen de beoordelingen gemaakt op basis van het eindgesprek en de beoordelingen uit de rio-opleiding. Het gemiddelde eindoordeel en de gemiddelde totaalscore zijn berekend als de gemiddelde beoordeling over de selecteurs die de eindgesprekken voeren. In Tabel B5 in Bijlage 4 staan ook de verbanden met de verschillende thema's die tijdens het eindgesprek beoordeeld worden. De correlaties tussen de totaalscore en het gemiddelde eindoordeel op het eindgesprek en de (gewogen) totaalscore in de rio-opleiding waren erg klein en dichtbij nul. Dat was ook het geval voor de scores op de verschillende competenties zoals beoordeeld in de rio-opleiding. De correlaties met de totaalscore op het eindgesprek waren bovendien vaak negatief. De predictieve validiteit van het eindgesprek lijkt op basis van deze resultaten erg klein te zijn.

6.8 Aanvullende analyses

6.8.1 De relatie tussen leeftijd en scores op de analytische tests

Gezien de signalen dat oudere kandidaten lagere scores behalen op de analytische tests dan jongere kandidaten, hebben we het verschil in scores tussen kandidaten jonger dan 40 jaar vergeleken met kandidaten van minstens 40 jaar oud. Uit recent onderzoek bleek dat kandidaten van 40 jaar en ouder vaak lager scoren op tests die zogenaamde 'fluide' cognitieve capaciteiten meten, zoals de Matrix test. Op tests voor 'gekrystalliseerde' cognitieve capaciteiten, zoals de TVRA, scoren oudere kandidaten juist vaak iets hoger (Fisher et al., 2013; Hough et al., 2001; Klein et al., 2015). Uit een analyse op basis van gegevens van 632 succesvolle en niet-succesvolle kandidaten die de analytische tests hebben gemaakt, bleek dat het verband tussen leeftijd en de testcores statistisch significant en middelgroot was voor beide tests (Matrix test: $r = -.32$, TVRA, $r = -.21$), waarbij oudere kandidaten lagere scores behaalden. Wanneer we het verschil bekijken tussen kandidaten van jonger dan 40 jaar oud en kandidaten van 40 jaar of ouder, vinden we een middelgroot verschil in scores voor de Matrix test (met $d = -.55$) en een klein verschil in scores voor de TVRA (met $d = -.31$). Het beeld dat oudere kandidaten minder hoge scores behalen op deze tests wordt door deze resultaten bevestigd. In overeenstemming met de literatuur is het verschil het grootste voor de scores op de Matrix test.



Figuur 11. Correlaties tussen scores op het eindgesprek en beoordelingen tijdens de rio-opleiding. De error bars geven het 95% betrouwbaarheidsinterval weer. $n = 45-48$.

Deze resultaten rechtvaardigen echter niet de conclusie dat de tests oudere kandidaten ten onrechte benadelen. Daarvan is wederom pas sprake als er onderpredictie zou optreden (een structureel onderschatting van toekomstige werkprestaties, zie ook paragraaf 4.9). Het was met de beschikbare gegevens echter niet mogelijk om te onderzoeken of dit het geval is, aangezien voor dergelijke analyses grote steekproeven gebruikt moeten worden. In de wetenschappelijke literatuur is nog geen onderzoek verricht naar differentiële predictie voor kandidaten uit verschillende leeftijdsgroepen. Verschillen in scores op basis van leeftijd worden daarnaast ook gevonden voor persoonlijkheidsvragenlijsten; jongere kandidaten scoren vaak hoger op extraversie, terwijl oudere kandidaten hoger scoren op conscientieusheid (Fisher et al., 2013; Hough et al., 2001). Ook uit rapportages van LTP blijkt dat er kleine verschillen naar leeftijd worden gevonden op schalen van sommige vragenlijsten. Oudere kandidaten scoren lager op *Doelgerichtheid* uit de Werkstijlenvragenlijst en op *Behoeftes aan prestatie en Behoeftes aan affiliatie* uit de Drijfverenvragenlijst.

Aangezien het niet mogelijk was om te toetsen of de scoreverschillen terecht of onterecht zijn, of om uitspraken te doen over de oorzaak van deze scoreverschillen, is het moeilijk om aanbevelingen te doen over hoe hier mee om te gaan. Het huidige beleid om voor zogenoemde 'gerede kandidaten' (kandidaten die over veel relevante kennis en ervaring beschikken en kunnen starten na een relatief kort opleidingstraject) een procedure te hanteren waarbij de analytische tests deel uitmaken van het assessment zonder dat strikte grensscores gehanteerd worden lijkt daarom een te verdedigende procedure. Om subjectiviteit te voorkomen is het belangrijk dat de voorwaarden om in aanmerking te komen voor deze procedure duidelijk zijn vastgelegd. Ook merkten verschillende leden van lokale selectiecommissies op dat dit soort kandidaten voor de rio-opleiding weinig voorkomen. Verder stoelt deze procedure op de aanname dat het beschikken over veel relevante werkervaring een erg goede en zelfs betere voorspeller is voor toekomstig functioneren dan de scores op de analytische tests. Uit de wetenschappelijke literatuur blijkt dat het aantal jaren werkervaring in hetzelfde beroep slechts een klein tot middelgroot verband vertoont met baanprestaties in complexe banen (McDaniel et al., 1988; Schmidt & Hunter, 1998) en dat dit verband kleiner wordt naarmate de jaren in werkervaring toenemen. Een recentere studie die nog niet is gepubliceerd vond bovendien slechts zeer kleine verbanden en adviseert om niet veel gewicht toe te kennen aan eerdere werkervaring (van Iddekinge et al., 2018). We adviseren daarom om de criteria om in aanmerking te komen voor een de gerede kandidaten procedure niet te ruimhartig te formuleren.

AANBEVELING 16 – LEEFTIJD EN DE ANALYTISCHE TESTS

- Het verband tussen leeftijd en de scores op de analytische tests is terecht een punt van aandacht. Zolang echter niet duidelijk is wat de oorzaak van deze verschillen is, bevelen we aan om zeer voorzichtig om te gaan met uitzonderingen.

6.8.2 Herkansen van de analytische tests

Een ander aandachtspunt dat in de gesprekken aan het licht kwam is de mogelijkheid om de analytische tests te herkansen als er onder de norm, maar binnen een bepaalde bandbreedte wordt gescoord. De herkansing houdt in dat er een andere versie van de test gemaakt wordt. De indruk bestaat dat herkansingen wel erg vaak in aanzienlijke hogere scores resulteren. Op basis van de beschikbare gegevens hebben we het verschil in scores tussen de eerste kans en de herkansing op beide tests geanalyseerd. In de beschikbare dataset waren er 29 kandidaten die de TVRA een tweede keer gemaakt hebben en 73 kandidaten die de Matrix test een tweede keer gemaakt hebben. Op de TVRA was de score na herkansing in 62% van de gevallen hoger dan de score op de eerste kans. Het verschil in scores tussen de eerste kans en de tweede kans was bovendien groot ($d = 0.82$). Op de Matrix test was de score in 79% van de gevallen hoger op de tweede kans dan op de eerste kans en was het verschil in scores ook groot ($d = 1.10$). Deze resultaten bevestigen het bestaande beeld dat herkansingen vaak aanzienlijk hogere scores opleveren. Daarbij moet wel opgemerkt worden dat kandidaten zelf kunnen beslissen of zij de test willen herkansen. Het is mogelijk dat kandidaten die terecht de indruk hebben dat zij, door verschillende denkbare omstandigheden, beter zouden kunnen presteren dan ze op de eerste kans gedaan hebben er vaker voor kiezen om gebruik te maken van de mogelijkheid tot herkansing. Zoals ook opgemerkt in het rapport Born (2013) blijkt echter uit onderzoek dat scores op cognitieve tests na een herkansing aanzienlijk minder valide zijn dan scores behaald tijdens de eerste kans (Lievens et al., 2007). We bevelen daarom aan om de mogelijkheid tot herkansen niet meer aan te bieden. In plaats daarvan zouden kandidaten in de gelegenheid gesteld moeten worden om het testmoment te verplaatsen als omstandigheden daarom vragen (bijvoorbeeld ziekte) en nadrukkelijk op die mogelijkheid gewezen worden. Als het toch wenselijk wordt gevonden om de mogelijkheid tot herkansing aan te bieden, raden we aan om de gemiddelde scores over de verschillende pogingen te gebruiken in de besluitvorming

en om (bij gebruik van meerdere tests) alle tests te laten herkansen. Dat leidt, in vergelijking met het gebruik van de laatst behaalde testscore en het herkansen van alleen het onderdeel waarop onvoldoende hoog is gescoord, tot minder vals-positieve beslissingen als gevolg van hertesten (Clauser & Wainer, 2016).

AANBEVELINGEN 17 – HERKANSEN VAN DE ANALYTISCHE TESTS

- We bevelen aan om de mogelijkheid tot herkansing van de analytische tests niet meer aan te bieden.
- Wanneer de organisatie het toch wenselijk acht om wel mogelijkheden tot herkansing aan te bieden, bevelen we aan om:
- Het gemiddelde resultaat over de pogingen te gebruiken in de besluitvorming.
- Bij het gebruik van meerdere tests, niet alleen de test waarop onvoldoende hoog is gescoord, maar alle tests te laten herkansen.

6.9 Conclusies en aanbevelingen

De correlaties tussen de scores en beoordelingen op de selectie-onderdelen en beoordelingen uit de rio-opleiding, zoals berekend op basis van de beschikbare gegevens, waren voor de meeste onderdelen klein. Alleen de scores op de TVRA, de Watson-Glaser test en, in mindere mate, de Matrix test (alleen voor de ongewogen totaalscore), resulteerden in een predictieve validiteit van praktisch bruikbare grootte.

6.9.1 Beperkingen van het onderzoek

De gegevens waarop de schattingen van de predictieve validiteit gebaseerd zijn hadden echter een aantal belangrijke beperkingen. Ten eerste zijn de schattingen door de kleine steekproefgrootten erg onnauwkeurig (te zien aan de 95% betrouwbaarheidsintervallen: de 'error bars' in de figuren de tabellen in Bijlage 4). Hoewel de betrouwbaarheid van de criteriumbeoordelingen (de beoordelingen in de rio-opleiding) redelijk was kan een suboptimale betrouwbaarheid een onderdrukkend effect op de correlaties hebben. Ook het samenvoegen van de taakniveaus B en C kan de resultaten beïnvloeden hebben. Bij de uitgevoerde analyses is de aanname dat een hogere score betekent dat de persoon in kwestie beter heeft gepresteerd dan een persoon met een lagere score. Die aanname is door het samenvoegen van taakniveaus misschien geschonden. Daar-

naast is het de vraag of de 77 rio's waarvan (een deel van) de selectiedata en beoordelingsgegevens op taakniveau B of C uit de opleiding beschikbaar waren, representatief zijn voor de gehele populatie rio's. Gezien deze kanttekeningen moeten de resultaten met voorzichtigheid geïnterpreteerd worden.

6.9.2 Mogelijke verklaringen

Duidelijk is dat de resultaten op basis van de hierboven beschreven analyses niet geheel overeenkomen met resultaten uit de wetenschappelijke literatuur. Op basis van eerder onderzoek zouden we een grote predictieve validiteit verwachten voor scores op de analytische tests (Kuncel & Hezlett, 2010; Schmidt & Hunter, 1998), terwijl we in bovenstaande analyses een middelgrote (TVRA) of kleine (Matrix test) predictieve validiteit vonden. Naast de beperkingen die hierboven zijn besproken, zijn daarvoor verschillen aanvullende verklaringen mogelijk. Ten eerste is de verwachte validiteit gebaseerd op de gemiddelde predictieve validiteit op basis van een groot aantal wetenschappelijke studies (meta-analyse). Bevindingen op basis van meta-analyses geven, bij gebrek aan resultaten op basis van gegevens verzameld in de eigen context, de best mogelijke schatting van de verwachte predictieve validiteit. De predictieve validiteit kan echter per situatie en context verschillen (Tett et al., 2017). In deze context hebben we bijvoorbeeld te maken met een selecte groep van kandidaten die allen minstens een Masterdiploma in de Rechten hebben behaald en dus waarschijnlijk bovengemiddeld intelligent zijn. Daardoor kan het moeilijker zijn om onderscheid te maken tussen kandidaten op basis van analytisch capaciteiten. De resultaten uit de wetenschappelijke literatuur zullen in de regel gebaseerd zijn op steekproeven die heterogener zijn wat betreft cognitieve capaciteiten (Sackett et al., 2017). Een andere mogelijke verklaring ligt in de kwaliteit van de gebruikte tests. Zoals ook opgemerkt in paragraaf 4.3 zijn er weinig gegevens bekend over de betrouwbaarheid en validiteit van deze tests.

Wat betreft de bevindingen voor de interviews (voorgesprek en eindgesprek) zou op basis van de literatuur een middelgrote predictieve validiteit verwacht worden, rekening houdend met de mate van structuur van de gesprekken (Huffcutt & Arthur, 1994; Huffcutt et al., 2014). De gevonden predictieve validiteit voor deze onderdelen was ongeveer gelijk aan nul en komt dus niet overeen met resultaten uit ander onderzoek. De suboptimale betrouwbaarheid van de interviewscores kan hierin een rol hebben gespeeld. De gevonden interbeoordelaarsbetrouwbaarheid voor de totaalscores op

het voorgesprek en de totaalscores op de thema's uit het eindgesprek waren echter vergelijkbaar met resultaten uit de wetenschappelijke literatuur (Huffcutt et al., 2013). Omdat het assessment uit veel verschillende onderdelen bestaat is het moeilijker om een verwachting over de predictieve validiteit te formuleren op basis van de literatuur. De predictieve validiteit van 'individuele assessments' leent zich hier het beste voor. Op basis daarvan zouden we voor het assessment een middelgrote predictieve validiteit verwachten (Morris et al., 2015). Ook hier vonden we echter een predictieve validiteit ongeveer gelijk aan nul, of zelfs kleine negatieve verbanden. Gezien de complexiteit van het assessment is het lastig om hier oorzaken voor aan te wijzen, maar mogelijkheden zijn onder andere de kwaliteit van de gebruikte tests en vragenlijsten, een gebrek aan structuur van het assessmentinterview, sociaal wenselijk antwoorden op de vragenlijsten, een geringe betrouwbaarheid en van de competentiebeoordelingen, een gebrek aan constructvaliditeit van de competentiebeoordelingen en het gebruik van de klinische methode om tot beoordelingen te komen. Een onderdeel uit het assessment dat wel een predictieve validiteit van bruikbare grootte liet zien was de Watson-Glaser test voor kritisch redeneervermogen. Dit komt overeen met eerdere bevindingen (Kuncel et al., 2014).

6.9.3 Conclusie

Op basis van de resultaten lijken de analytische tests, vooral de TVRA, en de Watson-Glaser test, de grootste bijdrage te kunnen leveren aan het selecteren van geschikte kandidaten voor de rio-opleiding. De predictieve validiteit van de beoordelingen uit het voorgesprek, de competentiebeoordelingen uit het assessmentrapport en de beoordelingen uit het eindgesprek lijken op basis van de huidige resultaten te klein om een praktische relevante bijdrage te leveren aan het selecteren van geschikte kandidaten. Gezien de kleine steekproeven, het ontbreken van veel gegevens en het samenvoegen van beoordelingen op verschillende niveaus, trekken we deze conclusies echter met grote voorzichtigheid. Deze analyses zullen in de toekomst herhaald moeten worden met meer en meer volledige gegevens.

6.9.4 Aanbevelingen analytische tests

Aangezien de scores op de TVRA een predictieve validiteit liet zien van bruikbare grootte, dit instrument de grootste predictieve validiteit lijkt te hebben van alle onderdelen van de procedure, en ook er op basis van de literatuur evidentie is voor de

predictieve validiteit van dit type test, adviseren we om de TVRA of een vergelijkbare test voor verbaal redeneervermogen te handhaven.

Aangezien de Matrix test een lage predictieve validiteit liet zien en geen incrementele validiteit lijkt te hebben bovenop de TVRA, kan ervoor worden gekozen om de Matrix test niet langer te gebruiken. De voortzettingsbeslissing is dan echter geheel afhankelijk van één testscore. In dat geval is het belangrijk dat deze testscore zeer betrouwbaar is. Het is voorstelbaar dat, zoals ook aangeraden in verschillende richtlijnen rondom psychologisch testgebruik, het de voorkeur heeft om deze beslissing op meer dan één testscore te baseren. Wanneer grensscores worden gehanteerd voor de afzonderlijke tests, zoals nu het geval is, leidt dit echter tot aanzienlijk grotere aantallen vals-negatieve beslissingen, vergeleken met het gebruik van één enkele testscore of een samengestelde score (waarbij een hoge score op de een test kan compenseren voor een lage score op een andere test; Clauser & Wainer, 2016). Of deze compensatorische manier van scores te verdedigen is hangt onder andere af van de constructen die met de tests gemeten worden. In het geval van het gebruik van verschillende tests die sterk gerelateerde constructen meten, zoals het geval is bij de huidige analytische tests, is compensatorisch scoren te verdedigen, terwijl dat minder voor de hand ligt wanneer inhoudelijk sterker verschillende constructen zouden worden gecombineerd. Wanneer er scores op meerdere tests gebruikt worden kan er dus, om het aantal vals-negatieve beslissingen te beperken, worden overwogen om de besluitvorming te baseren op een samengestelde, compensatorische score. Of er wordt gekozen voor het inzetten voor één of meerdere tests in deze fase van de selectieprocedure en of er wel of niet compensatorisch wordt gescoord, is een beleidskeuze. Het gebruik van meerdere instrumenten, waarbij sprake is van incrementele validiteit bovenop het gebruik van één instrument, heeft psychometrisch gezien de voorkeur en sluit aan bij de richtlijnen voor psychologisch testen. Beide keuzes kunnen empirisch gezien echter te verdedigen zijn.

Als de inzet van meerdere tests wenselijk wordt gevonden kan worden om in plaats van de Matrix test, de Watson-Glaser test of een vergelijkbaar instrument in te zetten. De scores op de Watson-Glaser test hadden ook geen incrementele validiteit bovenop de TVRA, maar hadden een grotere predictieve validiteit dan de Matrix test. Bovendien werd deze test positief ontvangen door kandidaten. Het inzetten van een dergelijk instrument in deze fase van de selectieprocedure zou de acceptatie van deze stap kunnen verhogen. Het gebruik van instrumenten die positief worden gewaardeerd door kandidaten is relevant, gezien de zorg dat met name de huidige analytische tests een afschrikwekkend effect hebben op potentieel geschikte kandida-

ten. We merken op dat dit beeld niet duidelijk naar voren kwam uit de resultaten van de vragenlijst afgenomen bij de oud-kandidaten. Deze steekproef bevatte echter wel alleen personen die daadwerkelijk hebben deelgenomen aan de procedure. Personen die wel interesse hadden in het rechterschap maar van sollicitatie hebben afgezien waren dus niet vertegenwoordigd. Een ander voordeel van deze verandering is dat het verband tussen leeftijd en de kans om deze stap met succes te voltooien waarschijnlijk af zal nemen. Verschillen in scores op basis van migratieachtergrond zullen echter waarschijnlijk iets toenemen.

Een alternatief voor het gebruik van analytische tests die waarschijnlijk tot een goede predictieve validiteit en positieve reacties van kandidaten zou leiden is het gebruik van een 'job-knowledge test', ofwel een kennistest (Schmidt & Hunter, 1998). Dat kan bijvoorbeeld naar Belgisch model (zie het rapport Born, 2013). Wel is het belangrijk om ervoor te zorgen dat het resultaat niet te sterk samenhangt met werkervaring, aangezien ook kandidaten met beperkte werkervaring in aanmerking moeten kunnen komen voor een rio-plaats. Er kan bijvoorbeeld voorbereidings- en oefenmateriaal aangeboden worden. Een test naar het model van het 'bar exam', die in verschillende landen voor toetreding tot de advocatuur wordt gebruikt is ook een mogelijkheid. Het ontwikkelen en evalueren van een dergelijk instrument zal wel kostbaar zijn. De kwaliteit, validiteit en betrouwbaarheid van een dergelijke test zal geëvalueerd moeten worden, de inhoud van de test moet niet bekend raken bij kandidaten en er moeten bij voorkeur meerdere versie van de test beschikbaar zijn.

AANBEVELINGEN 18 – ANALYTISCHE TESTS

- We bevelen aan om de TVRA of een gelijksoortig instrument te handhaven.
- Bij het gebruik van meerdere tests in deze fase kan worden overwogen om de tests compensatorisch te scoren, om het aantal vals-negatieve beslissingen terug te dringen.
- Overwogen kan worden om de Matrix test te vervangen door de Watson-Glaser test, of een vergelijkbaar instrument.
- Als alternatief voor de analytische tests kan het gebruik van een 'job-knowledge' test overwogen worden.

6.9.4.1 Controverse rondom analytische tests

Zoals eerder genoemd bestaat er controverse rondom het gebruik en de zwaarwegendheid van analytische tests in de selectieprocedure voor rio's (Ahsmann, 2018;

Hofhuis, 2018). Een deel van de kritiek heeft betrekking op het hanteren van harde grensscores, waardoor er een mechanische werkwijze wordt toegepast (Hofhuis, 2018) en er sprake is van het louter afvinken van resultaten (Ahsmann, 2018). Deze opvattingen over het toepassen van beslisregels zijn veelvoorkomend (Dawes, 1979; Nolan & Highhouse, 2014). Wanneer we aannamen dat het maken van zoveel mogelijk juiste selectiebeslissingen het doel is, leidt het afvinken van resultaten, ofwel het toepassen van een mechanische werkwijze (de statistische methode) juist vaak tot betere beslissingen, vergeleken met de klinische methode (Kuncel et al., 2013) waarbij 'rekening wordt gehouden met alle omstandigheden van het geval', zoals bijvoorbeeld Hofhuis (2018) voorstelt.

Een ander veelvoorkomend argument is gestoeld op de imperfect validiteit van dit soort tests (Ahsmann, 2018; Hofhuis, 2018). Het klopt dat deze tests geen absolute waarde hebben of foutloos zijn en dat er soms kandidaten die (ogenschijnlijk) geschikt zijn zullen worden afgewezen op basis van de testscores. Dat geldt echter voor ieder denkbaar selectie-instrument en voor iedere denkbare combinatie van instrumenten. De vraag moet daarom niet zijn of deze tests foutloze beslissingen opleveren, maar of er een alternatief is dat betere beslissingen oplevert. Vaakgenoemde alternatieven of toevoegingen als referenties, werkervaring en ongestructureerde interviews leveren geen meerwaarde op (Dana et al., 2013; Schmidt & Hunter, 1998). Toch zijn er wel zinvolle aanvullingen mogelijk, bijvoorbeeld door assessment centers of gestructureerde interviews toe te voegen aan een procedure waar analytische tests deel van uitmaken (Schmidt & Hunter, 1998; Sackett et al., 2017). De vraag is echter of het praktisch en financieel haalbaar is om alle kandidaten aan een dergelijk uitgebreide procedure te laten deelnemen. Als er dan, omwille van de haalbaarheid, een instrument ingezet moet worden om een belangrijke eerste 'schifting' te maken, dan zijn dit soort analytische tests daar bij uitstek geschikt voor, aangezien ze op zichzelfstaand een hoge predictieve validiteit opleveren voor baanprestaties en trainingsprestaties. De evidentie voor die conclusie is sterk en gebaseerd op honderden studies (Kuncel & Hezlett, 2010; Schmidt & Hunter, 1998). Bij gebrek aan empirisch onderzoek in de specifieke context van selectie voor de rio-opleiding zijn deze resultaten dan ook het meest bruikbare uitgangspunt voor de vormgeving van de selectieprocedure. Ook op basis van de statistische analyses uitgevoerd in het kader van dit onderzoek kunnen we voorzichtig concluderen dat de analytische test voor verbaal redeneervermogen, samen met de test voor kritisch redeneervermogen uit het assessment, de beste voorspellende waarde had van alle onderdelen uit de huidige selectieprocedure die we konden onderzoeken. Dit houdt niet in dat er, zoals hierboven al besproken is,

geen ruimte is voor verbetering aan de procedure rondom de analytische tests, maar wel dat het gebruik van analytische tests als stap in de selectieprocedure wetenschappelijk gezien te verdedigen is.

6.9.5 Aanbevelingen voorgesprek

De predictieve validiteit van het voorgesprek lijkt op basis van de beschikbare gegevens laag te zijn. Het is bovendien de vraag of het van toegevoegde waarde is dat er twee gesprekken deel uitmaken van de selectieprocedure. Er kan worden overwogen om het voorgesprek te laten vervallen. Wel zou dat zorgen voor een verminderde betrokkenheid van de lokale gerechten bij de procedure en zouden er meer kandidaten tot de volgende stap toegelaten moeten worden, of zouden er, om het voorgaande te voorkomen, hogere eisen aan de kandidaten moeten worden gesteld in de voorgaande stap van de procedure. Als het voorgesprek gehandhaafd blijft, bevelen we nogmaals aan om het voorgesprek verder te structureren en om ook de afgevaardigden van de lokale gerechten te trainen in het voeren van selectiegesprekken. Daarnaast bevelen we aan om de besluitvorming op basis van het voorgesprek op een statische manier te laten plaatsvinden (zie hoofdstuk 7).

AANBEVELINGEN 19 – VOORGESPREK

- Er kan overwogen worden om het voorgesprek te laten vervallen.
- We bevelen aan om ook bij het voorgesprek gebruik te maken van een statistische methode van gegevenscombinatie en besluitvorming (zie paragraaf 7.2).

6.9.6 Aanbevelingen assessment

De predictieve validiteit van de beoordelingen uit het assessment waren niet zoals verwacht. We herhalen onze aanbeveling om het assessment als 'assessment center' in te richten. Ook raden we aan om geen interview meer in het assessment op te nemen, aangezien de predictieve validiteit van 'individuele assessments' kleiner was wanneer ze een interview bevatten (Morris et al., 2015). Uit eerder onderzoek is ook gebleken dat informatie uit ongestructureerde interviews gebruikt in klinische besluitvorming tot een 'verdunding' van informatie kan leiden, wat een negatief effect heeft op de validiteit (Dana et al., 2013). Bovendien zijn interviews al voldoende vertegen-

woordigd in de andere stappen van de procedure. Wat verder opvalt is dat de informatie afkomstig uit het assessment in de huidige vorm twee keer beoordeeld wordt, door de assessmentpsycholoog van LTP en door de selecteurs van de LSR. We raden aan om deze dubbele klinische beoordeling te vervangen door een statistische manier van gegevenscombinatie. Dat geldt ook voor de beslissing die volgt op basis van het assessment en het eindgesprek. Deze statistische gegevenscombinatie kan gedaan worden door het assessmentbureau. In dat geval is de kwalitatieve beschrijving in het assessmentrapport overbodig. Een andere mogelijkheid is scoring van het assessmentrapport door selecteurs van LTP op basis van het assessmentrapport. De eerste variant zal echter waarschijnlijk de hoogste validiteit opleveren.

AANBEVELINGEN 20 – ASSESSMENT

- We bevelen aan om de oordelen die voortvloeien uit het assessment volgens statistische gegevens combinatie te bepalen.
- We bevelen aan om het assessmentinterview te laten vervallen.

6.9.7 Aanbevelingen eindgesprek

De predictieve validiteit van het eindgesprek lijkt op basis van de beschikbare gegevens niet van een bruikbare grootte te zijn. We bevelen nogmaals aan om het eindgesprek verder te structureren. Daarnaast bevelen we aan om de besluitvorming op basis van het voorgesprek op een statistische manier te laten plaatsvinden, waarbij de gewichten die worden toegekend aan scores en beoordelingen uit het eindgesprek en uit het assessment geëxpliciteerd moeten worden in een beslisregel. De beraadslagingen zijn dan niet meer nodig. In de huidige vorm kan de validiteit van de beoordelingen uit het eindgesprek bovendien niet geheel los worden gezien van de validiteit van het assessment, of zelfs van de validiteit van zaken zoals het cv, de cijferlijsten en de beoordelingen uit de voorgesprekken. De selecteurs die de eindgesprekken voeren beschikken immers vooraf over deze informatie en gebruiken het daarnaast als input voor de eindgesprekken. Indien er niet voor wordt gekozen om de beoordelingen via statistische gegevenscombinatie te laten plaatsvinden, bevelen we aan om een eventuele beoordeling van assessmentrapporten door andere personen te laten doen dan degenen die de eindgesprekken voeren met de kandidaat, om de onafhankelijkheid van het assessment en het eindgesprek te waarborgen en om te

voorkomen dat de informatie uit het assessment twee keer (klinisch) wordt beoordeeld.

AANBEVELINGEN 21 – EINDGESPREK

- We bevelen aan om de besluitvorming volgens een statistische procedure te laten verlopen, waarbij vooraf wordt gespecificeerd hoe zwaar ieder thema en iedere score gewogen moet worden (zie paragraaf 7.2). Dat geldt dan zowel voor de scores of beoordelingen uit het assessment als uit het eindgesprek.
- We bevelen aan om een eventuele beoordeling van het assessmentrapport en het voeren van de eindgesprekken niet door dezelfde personen te laten doen.

Toelichting op terugkerende aanbevelingen

In dit hoofdstuk bespreken we in meer detail hoe de terugkerende aanbevelingen over het structureren van interviews het maken van beslissingen op basis van statistische gegevenscombinatie uitgevoerd kunnen worden.

7.1. Structureren van interviews

We hebben over verschillende onderdelen van de selectieprocedure aanbevolen om ze verder te structureren. Hier zullen we uiteenzetten hoe dat gerealiseerd kan worden. In de wetenschappelijke literatuur worden vier structuurniveaus onderscheiden. Hoe hoger de mate van structuur, hoe hoger de betrouwbaarheid en de predictieve validiteit van beoordelingen voortvloeiend uit interviews (Huffcutt et al., 2013; 2014; Levashina et al., 2014). Dat verschil is echter het grootste tussen de laagste en de twee hoogste structuurniveaus.

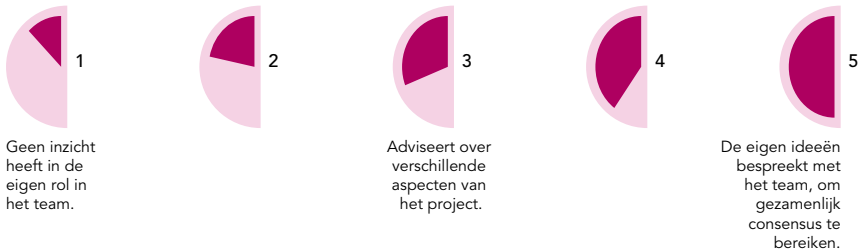
Een sterkere mate van structuur wordt bereikt door iedere kandidaat dezelfde vooraf opgestelde vragen te stellen met een beperkte mogelijkheid tot doorvragen (alleen ter verduidelijking), door de kandidaat op iedere gestelde vraag te beoordelen en/of te beoordelen op basis van een 'geankerde beoordelingsschaal'. Deze schalen geven specifieke beschrijvingen van wanneer er bijvoorbeeld sprake is van een onvoldoende, matige, of goede motivatie voor het rechterschap en vergroten de betrouwbaarheid van beoordelingen (Campion et al., 1997; Taylor & Small, 2002). Een voorbeeldvraag met een geankerde beoordelingsschaal om samenwerken te meten is:

Vraag:

Beschrijf een situatie waarin je in een team aan een project moest werken. Wat was je rol in dat team en wat droeg je bij in die rol?

Scoring:

Uit het antwoord blijkt dat de kandidaat in een vergelijkbare situatie waarschijnlijk:



* Voorbeeld afkomstig van de Society for Human Resource Management.

Andere zaken die bijdragen van de validiteit van interviewscores zijn het beperken van de toegang tot aanvullende informatie zoals testcores, het cv, eerdere interviews en referenties, de kandidaat pas aan het einde van het interview (na scoring) vragen te laten stellen, het maken van gedetailleerde notities, het inzetten van dezelfde interviewers voor alle kandidaten en de kandidaten niet onderling bespreken voordat wordt beoordeeld (Campion, 1997). Daarom raden we aan dat de selecteurs voor de gesprekken geen dossiers meer toegestuurd krijgen. Als de informatie uit de dossiers meegenomen moet worden kan die informatie onafhankelijk van de interviews beoordeeld worden. Dat zou ook op een gestructureerde manier moeten gebeuren. Voor het voorgesprek raden we aan om het wederkerige deel van het interview tot na het beoordelen te verplaatsen. Daarnaast zou de aanmoediging tot het maken van notities en toelichtingen en het beleid om de kandidaten niet onderling te bespreken gehandhaafd moeten blijven.

7.2 Statistische combinatie van gegevens

Omdat veel wetenschappelijk onderzoek heeft uitgewezen dat gegevenscombinatie volgens de statistische methode even goede en vaak zelfs betere voorspellingen en beslissingen oplevert dan volgens de klinische methode, raden we aan om de statistische methode van gegevenscombinatie te gebruiken bij alle stappen in de selectieprocedure. Voor het voorspellen van werkprestaties levert de statistische methode gemiddeld een 50% betere voorspelling op dat de klinische methode (Kuncel et al., 2013). De kern van statistische procedures is dat iedere kandidaat op dezelfde manier

beoordeeld wordt. Dat wil zeggen dat voor iedere kandidaat dezelfde informatie wordt gebruikt en dat die informatie voor iedere kandidaat dezelfde weging krijgt. De verklaring voor de betere resultaten die behaald worden met de statistische methode ligt vooral in de consistente toepassing van de wegingen; menselijke beoordelaars zijn daar erg inconsistent in (Dawes, 1979).

Een statistische procedure vraagt voor iedere stap waarop een voorzettingsbeslissing volgt om:

- Een explicitering van alle punten die in de beslissing moeten worden meegenomen.
- Een expliciete scoring of beoordeling van al die punten.
- Een afspraak over de weging van de verschillende punten.
 - Dit kan eenvoudig zijn, door alle punten op te tellen of een gemiddelde te nemen, waarbij ieder punt even zwaar weegt. Vaak leidt dit soort eenvoudige procedures tot valide beslissingen (Dawes, 1979; Dana & Dawes, 2004).
 - Ook kunnen de gewichten bepaald worden op basis van discussie met verschillende belanghebbenden, zoals de leden van de LSR en personen betrokken bij de opleiding.
- Een beslisregel die kan worden toegepast op het resultaat van de combinatie van de punten, waaruit wordt bepaald of de productie wordt voortgezet of niet.
 - Het is mogelijk om een grensscore vast te stellen waarboven men moet scoren om te worden doorgelaten.
 - Ook kan een vaststaand aantal personen met de hoogste scores worden doorgelaten.
 - Het is ook mogelijk om enkele 'kritieke punten' te hanteren waaraan in ieder geval voldaan moet worden.
 - Ook beslissingen rondom grensscores en 'kritieke punten' kunnen in samenwerking met belanghebbenden genomen worden.

7.3 Tot slot

We sluiten dit rapport af met de aanbeveling om dit onderzoek, na aanpassingen van de selectieprocedure en het realiseren van een goed functionerend datamanagementsysteem, nog eens te herhalen.

AANBEVELING 22 – VERVOLGONDERZOEK

- Gezien de gebreken wat betreft de kwantiteit en kwaliteit van de gegevens die in dit onderzoek gebruikt konden worden voor statistische en psychometrische analyses, bevelen we aan om dit onderzoek, na aanpassingen van de selectieprocedure en het realiseren van een beter systeem voor het bijhouden van gegevens, nog eens te herhalen.

Referenties

- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291-304. doi:10.1111/j.1468-2389.2010.00512.x
- Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education*, 23, 311-321. doi: 10.1007/s10459-017-9796-8
- Arthur, W., Jr., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology*, 91, 786-801. doi:10.1037/0021-9010.91.4.786
- Ahsmann, M. (2018). *Het nieuwe Opleidingshuis voor rechters: een fundamentele verbetering?*
- Bartram, D. (2005). The Great Eight competencies: A Criterion-centric approach to validation. *Journal of Applied Psychology*, 90, 1185-1203. doi:10.1037/0021-9010.90.6.1185
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, 54, 388-420. doi:10.1111/j.1744-6570.2001.tb00097.x
- Berry, C. M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior*, 2, 435-463. doi:10.1146/annurev-orgpsych-032414-111256

- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317–335. doi:10.1111/j.1468-2389.2006.00354.x.
- Born, M. Ph. (2013). *Rapport van het onderzoek selectie zittende magistratuur*. Rotterdam: Instituut voor Psychologie, Erasmus Universiteit Rotterdam
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101. doi:10.1191/1478088706qp063oa
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655–702. doi:10.1111/j.1744-6570.1997.tb00709.x
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates
- Clauser, A. L., & Wainer, H. (2016). A tale of two tests (and of two examinees). *Educational Measurement: Issues and Practice*, 35, 19–28. <https://doi.org/10.1111/emip.12110>
- Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology*. Advance online publication. doi:10.1037/apl0000235
- Dana, J., & Dawes, R. M. (2004). The Superiority of Simple Alternatives to Regression for Social Science Predictions. *Journal of Educational and Behavioral Statistics*, 29, 317–331. doi:10.3102/10769986029003317
- Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment & Decision Making*, 8, 512–520.
- Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571–582. doi:10.1037/0003-066X.34.7.571
- De Kock, F. S., Lievens, F., & Born, M. P. (2018). The profile of the 'good judge' in HRM: A systematic review and agenda for future research. *Human Resource Management Review*. doi:10.1016/j.hrmr.2018.09.003
- De Soete, B, Lievens, F., & Druart, C. (2013). Strategies for dealing with the diversity-validity dilemma in personnel selection: Where are we and where should we go? *Journal of Work and Organizational Psychology* 29, 3–12. doi: 10.5093/tr2013a2
- Dilchert, S., & Ones, D. S. (2009). Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection & Assessment*, 17, 254–270. doi:10.1111/j.1468-2389.2009.00468.x

- Evers, A., Lucasse, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. Utrecht: Nederlands Instituut voor Psychologen.
- Fife, D. (2016). *Selection: Correcting biased estimates under selection*. R package version 1.0. Retrieved from: <http://CRAN.R-project.org/package=selection>
- Fisher, G. G., Truxillo, D. M., Finkelstein, L. M., & Wallace, L. E. (2017). Age discrimination: Potential for adverse impact and differential prediction related to age. *Human Resource Management Review*, 27, 316–327. doi:10.1016/j.hrmr.2016.06.001
- Gaddis, B. H., & Foster, J. L. (2015). Meta-Analysis of dark side personality characteristics and critical work behaviors among leaders across the globe: Findings and implications for leadership development and executive coaching. *Applied Psychology: An International Review*, 64, 25–54. doi:10.1111/apps.12017
- Highhouse, S. & Brooks, M. E. (2017). Straight talk about selecting for upper management. In: *The Oxford Handbook of Talent Management*, D. G. Collings, K. Mellahi, and W.F. Cascio (Eds.). doi:10.1093/oxfordhb/9780198758273.013.24
- Hofhuis, H. (2018). De selectie voor rechters geanalyseerd Waarom wijst de Rechtspraak uitstekende kandidaten af? *Nederlands Juristenblad*, 20, 1431-1433.
- Hoffman, B. J., Kennedy, C. L., Lance, C. E., LoPilato, A. C., & Monahan, E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, 100, 1143–1168. doi:10.1037/a0038707
- Hogan Press (2014). *Hogan Development Survey: Technical Supplement*. Opgevraagd via: http://www.hoganassessments.com/sites/default/files/uploads/HDS_Subscale_Technical_Supplement.pdf
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9, 152-194. doi:10.1111/1468-2389.00171
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment Interview Reliability: New meta-analytic estimates by structure and format. *International Journal of Selection & Assessment*, 21, 264–276. doi:10.1111/ijsa.12036
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment*, 22, 297-309. doi:10.1111/ijsa.12078

- Klein, R. M., Dilchert, S., Ones, D. S., & Dages, K. D. (2015). Cognitive predictors and age-based adverse impact among business executives. *Journal of Applied Psychology, 100*, doi: 10.1037/a0038991
- Koo, T. K. & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155-63. doi:10.1016/j.jcm.2016.02.012
- Kuncel, N. R. & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19*, 339-345. doi:10.1177/0963721410389459
- Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology, 98*, 1060-1072. doi:10.1037/a0034156
- Kuncel, N. R., Rose, M., Ejiogu, K., & Yang, Z. (2014). Cognitive ability and socio-economic status relations with job performance. *Intelligence, 46*, 203-208. doi:10.1016/j.intell.2014.06.003
- Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment center construct validity problem (as we know it). *Journal of Applied Psychology, 99*, 38-47. doi:10.1037/a0034147
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241-293. doi:10.1111/peps.12052
- Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 383-410). New York, NY: Oxford University Press.
- Lievens, F., Reeve, C. L., & Heggstad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology, 92*, 1672-1682. doi: 10.1037/0021-9010.92.6.1672
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, 283-314. doi:10.1111/j.1744-6570.1988.tb02386.x
- Morris, S. B., Daisley, R. L., Wheeler, M., & Boyer, P. (2015). A meta-analysis of the relationship between individual assessments and job performance. *Journal of Applied Psychology, 100*, 5-20. doi:10.1037/a0036938.supp
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C., III. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355. doi:10.1037/0021-9010.88.2.348

- Nei, K. S., Foster, J. L., Ness, A. M., & Nei, D. S. (2018). Rule breakers and attention seekers: Personality predictors of integrity and accountability in leaders. *International Journal of Selection & Assessment*, 26, 17–26. doi:10.1111/ijsa.12201
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017b). Measuring noncognitive predictors in high-stakes contexts: The effect of self-presentation on selfreport instruments used in admission to higher education. *Personality and Individual Differences*, 106, 183–189. doi:10.1016/j.paid.2016.11.014
- Nolan, K. P., & Highhouse, S. (2014). Need for Autonomy and Resistance to Standardized Employee Selection Practices. *Human Performance*, 27, 328–346. doi:10.1080/08959285.2014.929691
- Ostrom, J. K., & De Soete, B. (2016). Ethnic differences in perceptions of cognitive ability tests: The explanatory role of self-serving attributions. *International Journal of Selection and Assessment*, 24, 14–23. doi:10.1111/ijsa.12125
- Parsons, C. K., Cable, D., & Wilkerson, J. M. (1999). Assessment of applicant work values through interviews: The impact of focus and functional relevance. *Journal of Occupational and Organizational Psychology*, 72, 561–566. doi:10.1348/096317999166842
- Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, 24, 270–290. doi:10.1080/08959285.2011.580808.
- Ployhart, R. E., & Harold, C. M. (2004). The applicant attribution-reaction theory (AART): An integrative theory of applicant attributional processing. *International Journal of Selection and Assessment*, 12, 84–98. doi:10.1111/j.0965-075X.2004.00266.x
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racio-ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153–172. doi:10.1111/j.1744-6570.2008.00109.x
- Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology*, 102, 1435–1447. doi:10.1037/apl0000236
- Schmidt, F. L. & Hunter, J. E. (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274. doi:10.1037/0033-2909.124.2.262

- Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., & Ramsay, L. J. (2004). The impact of justice and self-serving bias explanations of the perceived fairness of different types of selection tests. *International Journal of Selection and Assessment*, 12, 160–171. doi:10.1111/j.0965-075X.2004.00271.x
- Steiner, D. D., & Gilliland, S. W. (2001). Procedural justice in personnel selection: International and crosscultural perspectives. *International Journal of Selection and Assessment*, 9, 124–137. doi:10.1111/1468-2389.00169
- Stuurgroep Stevens (2010). *Rapport van de stuurgroep evaluatie werving, selectie en opleiding van rechters in opleiding*. Intern document Raad voor de rechtspraak.
- Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277–294. doi:10.1348/096317902320369712
- te Nijenhuis, J., & van der Flier, H. (1999). Bias research in The Netherlands: Review and implications. *European Journal of Psychological Assessment*, 15, 165–175. doi:10.1027//1015-5759.15.2.165
- Tett, R. P., Hundley, N. A., & Christiansen, N. D. (2017). Meta-analysis and the myth of generalizability. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 10, 421–456. doi:10.1017/iop.2017.26
- van Iddekinge, C. H., Arnold, J. D., Frieder, R. E., & Roth, P. L. (2018). *It's required, but is it job-related? A meta-analysis of the validity of prior work experience*. Academy of Management Annual Meeting Proceedings. doi: 10.5465/AMBPP.2018.278
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197–210. doi:10.1177/00131649921969802
- Watson, G. & Glaser, E.M. (2010). *Watson-Glaser II Critical Thinking Appraisal: Technical manual and user's guide*. San Antonio, U.S.: NCS Pearson, Inc.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organisational Psychology*, 67, 189–205.

Specifieke onderzoeksvragen uit de *Startnotitie evaluatie selectieproces rio's*

De selectiecommissie

1. Hoe is de selectie van de leden van de selectiecommissie tot stand gekomen? In hoeverre is daarbij rekening gehouden met de eigenschappen en vaardigheden die blijken uit het onderzoek van Born als wenselijk worden beschouwd (zoals empathie, goede sociale vaardigheden)?
2. Zijn de leden van de selectiecommissie opgeleid/ getraind voor hun werkzaamheden? In hoeverre is daarbij qua type training rekening gehouden met de aanbevelingen van Born?
3. Hoe ervaren de leden van de commissie het functioneren van de selectiecommissie (waaronder het werken met interne leden vanuit de Rechtspraak en externe leden van buiten de Rechtspraak)?
4. Hoe ervaren de leden van de selectiecommissie de ondersteuning zoals die wordt geboden door het secretariaat van de commissie?

De selectieprocedure

Het is van belang vast te stellen of het proces van selectie verloopt zoals dat bij aanvang van het nieuwe selectieproces was beoogd. Daartoe zijn de volgende vragen geformuleerd.

Selectiecriteria

1. Voldoen alle kandidaten die na de briefselectie doorgaan in het selectieproces aan de algemene selectie-eisen (minimaal twee jaar relevante juridische werkervaring en aantoonbare maatschappelijke ervaring) of wordt hiervan onder omstan-

digheden weleens vanaf geweken? Zo ja, wanneer is dat het geval en wat zijn hiervan de redenen?

2. Vindt selectie plaats op basis van het nieuwe rechtersprofiel? Vindt beoordeling plaats op basis van de competenties die ook gelden voor het rechtersprofiel? Op welke wijze wordt hieraan invulling gegeven?

Proces van selectie

1. Hoe zorgen de selecteurs ervoor dat de selectie op een zo uniform mogelijke wijze plaatsvindt (bijvoorbeeld door te werken met gestructureerde vragenlijsten)?
2. Wat is de doorlooptijd van de selectieprocedure en zou (met behoud van kwaliteit) nog versnelling mogelijk zijn?
3. Is het lokale proces van selectie goed afgestemd op het landelijke proces van selectie? Doet zich hier nog overlap voor? Voor zover die bestaat hoe zou die overlap dan kunnen worden teruggedrongen?
4. Hoe ervaren de kandidaten –de verschillende onderdelen van – het selectieproces (van het invullen van het digitale sollicitatieformulier tot en met het arbeidsvoorwaardengesprek)?

Werking van de selectie-instrumenten

1. Functioneren de gehanteerde selectie-instrumenten zoals beoogd: zijn zij cultuurwaardevrij en is het profijt van het assessment (het duurste selectie-instrument) vergroot?
2. Vormt de informatie uit het selectieproces zinvolle input voor het bepalen van de ontwikkelpunten van de kandidaten tijdens de opleiding?
3. Hoe ervaren de leden van de commissie het werken met het huidige selectie-instrumentarium? Hebben zij het idee dat het huidige instrumentarium hen inderdaad in staat stelt om de meeste geschikte kandidaten te selecteren?

Overall effecten van het selectieproces

1. Wat is de voorspellende waarde van de selectie-instrumenten? Komt de voorspelling over een competentie overeen met de feitelijke prestaties gedurende de rio-opleiding?

2. Slaagt de huidige selectieprocedure erin om die rio's te selecteren die daadwerkelijk de opleiding afronden?
3. Is het aantal kandidaten dat na landelijk groen licht lokaal alsnog geen plaats kan vinden teruggedrongen?
4. Wat zijn de kenmerken van de kandidaten die de selectieprocedure niet met succes hebben afgrond? Kunnen er basis van een vergelijking van de kandidaten die wel en die niet zijn geselecteerd conclusies worden verbonden aan de vraag of de selectieprocedure al dan niet te streng is?

Competentiematrix assessment

Competentiematrix selectieprocedure RIO's

Competenties & overige te meten factoren

Onderdelen	Luisteren	Analytisch vermogen	Besluitvaardigheid	Samenwerken	Omgevingsbewustzijn	Flexibiliteit	Leervermogen en zelfreflectie	Overredingskracht	Regievoering	Zelfvertrouwen en authenticiteit	Zelfbewustzijn	Integriteit	Kritisch redeneervermogen	Uitdrukkingsvaardigheid
Matrix		V					V							
TVRA		V					V							V
Taalgevoeltest														
Watson Glaser		V					V						V	
Vragenlijsten	V		V	V	V	V	V	V	V	V		V		
Interview	V		V	V	V	V	V	V	V	V	V	V		V
Hogan development survey			V									V		
Dilemma test					V							V		
Digitale case		V	V			V			V					

[illegible]

Scores geven de gewichten van de schalen aan in de competentie. Rode cellen tellen positief mee, groene cellen worden omgescoord en tellen dus negatief mee. De gele cellen worden afgekappt, tot ongeveer een sd bovengemiddeld telt de schaal positief mee, daarboven telt de schaal negatief mee.

Wegingen van de verschillende test- en schaalscores uit het assessment voor iedere competentie

[illegible]

Scores geven de gewichten van de schalen aan in de competentie. Rode cellen tellen positief mee, groene cellen worden omgescoord en tellen dus negatief mee. De gele cellen worden afgekap, tot ongeveer een sd bovengemiddeld telt de schaal positief mee, daarboven telt de schaal negatief mee.

Tabellen

Tabel B1. Correlaties tussen scores op de analytische tests en beoordelingen tijdens de rio-opleiding.

Criterium	Matrix		TVRA		TVRA + Matrix			
	r	r_c	r	r_c	R	ΔR^2	R_c	ΔR_c^2
Totaalscore	.11 [-.14, .34]	.22	.20 [-.05, .42]	.30	.22	.01	.31	.01
Gewogen totaalscore	.03 [-.22, .27]	.13	.21 [-.04, .43]	.28	.21	<.01	.28	<.01
Competentiescores								
Regievoering	.07 [-.18, .31]	.12	.13 [-.12, .36]	.17	.14	<.01	.17	<.01
Analytisch vermogen	.15 [-.10, .38]	.22	.18 [-.07, .41]	.25	.22	.02	.27	.01
Flexibiliteit	.03 [-.22, .27]	-.05	.05 [-.20, .29]	-.01	.06	<.01	.05	<.01
Juridische kennis	.10 [-.15, .34]	.34	.12 [-.13, .35]	.34	.15	.01	.39	.04
Leervermogen	.04 [-.21, .28]	.31	.36* [-.13, .56]	.56	.36*	<.01	.56	<.01
Luisteren	.14 [-.11, .37]	.25	.12 [-.13, .35]	.23	.18	.02	.28	.02
Omgevingsbewustzijn	-.15 [-.38, .10]	-.28	-.07 [-.31, .18]	-.20	.16	.02	.28	.04
Overredingskracht	.11 [-.14, .34]	.19	.14 [-.11, .37]	.22	.17	.01	.24	.01
Samenwerken	.04 [-.21, .28]	.10	.16 [-.09, .39]	.21	.16	<.01	.21	<.01
Zelfbewustzijn	-.12 [-.35, .13]	.05	.15 [-.10, .38]	.27	.20	.02	.28	.01
Zelfvertrouwen & authenticiteit	.13 [-.12, .36]	.21	.21 [-.04, .43]	.28	.24	.01	.29	.01
Besluitvaardigheid	-.09 [-.33, .16]	-.12	.13 [-.12, .36]	.17	.17	.01	.20	.03

Notitie. $n = 65$. 95% BHI tussen haakjes. * $p < .05$. r is de ongecorrigeerde correlatie, r_c is de correlatie gecorrigeerd voor restrictie in range, R is de ongecorrigeerde multiple correlatie voor de scores op de Matrix test en de TVRA tezamen, ΔR^2 is de ongecorrigeerde verklaarde variantie van de score op de Matrix test bovenop scores op de TVRA, R_c is de gecorrigeerde multiple correlatie voor de scores op de Matrix test en de TVRA tezamen, ΔR_c^2 is de gecorrigeerde verklaarde variantie van de score op de Matrix test bovenop scores op de TVRA.

Tabel B2. Correlaties tussen scores op de Taalgevoeltest, de Watson-Glaser test en beoordelingen tijdens de rio-opleiding.

Criterium	Taalgevoeltest		Watson-Glaser	
	r	r_c	r	r_c
Totaalscore	.06 [-.28, .31]	.13	.13 [-.20, .44]	.25
Gewogen totaalscore	.11 [-.15, .36]	.18	.16 [-.17, .46]	.25
Competentiescores				
Regievoering	-.02 [-.28, .24]	.03	.21 [-.12, .50]	.32
Analytisch vermogen	.06 [-.28, .31]	.10	.05 [-.28, .37]	.13
Flexibiliteit	-.11 [-.36, .15]	-.10	.09 [-.24, .40]	.14
Juridische kennis	-.01 [-.27, .25]	.06	.33* [.01, .59]	.47
Leervermogen	.16 [-.10, .40]	.30	.09 [-.24, .40]	.21
Luisteren	.17 [-.09, .41]	.21	.18 [-.15, .48]	.26
Omgevingsbewustzijn	.06 [-.28, .31]	.03	-.01 [-.33, .32]	-.10
Overredingskracht	.19 [-.07, .43]	.26	.04 [-.29, .36]	.08
Samenwerken	.11 [-.15, .36]	.15	-.08 [-.39, .25]	-.06
Zelfbewustzijn	.22 [-.04, .45]	.31	-.10 [-.41, .23]	.01
Zelfvertrouwen & authenticiteit	-.04 [-.30, .22]	.04	.01 [-.32, .33]	.15
Besluitvaardigheid	.07 [-.18, .31]	.08	.09 [-.24, .40]	-.07

Notitie. Taalgevoeltest: $n = 58$, Watson-Glaser: $n = 37$. 95% BHI tussen haakjes. * $p < .05$. r is de ongecorrigeerde correlatie, r_c is de correlatie gecorrigeerd voor restrictie in range.

Tabel B3. Correlaties tussen scores op het voorgesprek en beoordelingen tijdens de rio-opleiding.

Criterium	Score							
	Motivatie	Overtuigingskracht	Mondelinge uitdrukkingsv.	Beeldfunctie	Gem. eindoordeel	Totaalscore voorgesprek		
	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r_c</i>	<i>r</i>	<i>r_c</i>
Totaalscore	.06 [-.20, .31]	-.05 [-.30, .21]	.03 [.23, .29]	.19 [-.07, .43]	-.01 [-.27, .25]	.04	.05 [-.21, .30]	.04
Gewogen totaalscore	.04 [-.22, .29]	-.06 [-.29, .22]	.04 [-.22, .30]	.20 [-.06, .44]	.05 [-.21, .30]	.08	.05 [-.21, .30]	.04
Competentiescores								
Regievoering	.11 [-.15, .36]	.09 [-.17, .34]	.14 [-.12, .38]	.23 [-.03, .46]	.07 [-.19, .32]	.09	.15 [-.11, .39]	.14
Analytisch vermogen	.04 [-.22, .29]	-.10 [-.35, .16]	<.01 [-.26, .26]	.18 [-.08, .42]	-.08 [-.33, .18]	-.04	.02 [-.24, .27]	.01
Flexibiliteit	.11 [-.15, .36]	-.03 [-.28, .23]	.03 [.23, .29]	.16 [-.10, .40]	.04 [-.22, .30]	.02	.07 [-.19, .32]	.08
Juridische kennis	.07 [-.19, .32]	<.01 [-.26, .26]	.02 [-.24, .28]	-.01 [-.27, .25]	-.07 [-.32, .19]	<.01	.01 [-.25, .27]	<-.01
Leervermogen	-.02 [-.27, .24]	-.03 [-.28, .23]	.02 [-.24, .28]	.17 [-.09, .41]	<-.01 [-.26, .26]	.10	.03 [-.23, .28]	.01
Luisteren	.02 [-.24, .27]	-.09 [-.34, .17]	-.10 [-.35, .16]	.08 [-.18, .33]	.01 [-.25, .27]	.05	-.03 [-.28, .23]	-.04
Omgevingsbewustzijn	-.20 [-.43, .06]	-.26* [-.48, -.01]	-.18 [-.42, .08]	-.16 [-.40, .10]	-.23 [-.46, .03]	-.27	-.23 [-.46, .03]	-.21
Overredingskracht	.03 [-.23, .28]	-.04 [-.29, .22]	.02 [-.24, .28]	.13 [-.13, .38]	-.02 [-.28, .24]	-.02	.03 [-.23, .28]	.03
Samenwerken	-.03 [-.28, .23]	.02 [-.24, .27]	.02 [-.24, .28]	.19 [-.07, .43]	.15 [-.11, .39]	.17	.05 [-.21, .30]	.04
Zelfbewustzijn	<.01 [-.26, .26]	.01 [-.25, .27]	.05 [-.21, .30]	.22 [-.04, .45]	.23 [-.03, .46]	.27	.05 [-.21, .30]	.04
Zelfvertrouwen & authenticiteit	-.07 [-.32, .19]	-.02 [-.27, .24]	.05 [-.21, .30]	-.03 [-.29, .23]	.17 [-.09, .41]	.19	-.03 [-.28, .23]	-.03
Besluitvaardigheid	.19 [-.07, .43]	.07 [-.19, .32]	.21 [-.05, .44]	.31* [.06, .53]	.11 [-.15, .36]	.09	.22 [-.04, .45]	.20

Notitie. *n* = 58. 95% BHI tussen haakjes. * *p* < .05. *r* is de ongecorrigeerde correlatie, *r_c* is de correlatie gecorrigeerd voor restrictie in range.

Tabel B4. Correlaties tussen assessment oordelen en beoordelingen tijdens de rio-opleiding.

Criterium	Competentiebeoordeling assessment												
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	Gemiddelde competentie	
	r_s											r	r_c
Totaalscore	-.21 [-.44, .04]	-.01 [-.26, .24]	.03 [-.22, .28]	.07 [-.18, .31]	.01 [-.24, .26]	.06 [-.19, .30]	-.28* [-.49, .04]	-.03 [-.28, .22]	-.10 [-.34, .15]	-.32* [-.53, -.08]	-.16 [-.39, .09]	-.20 [-.43, .05]	-.16
Gewogen totaalscore	-.20 [-.43, .05]	-.09 [-.33, .16]	-.01 [-.26, .24]	.01 [-.24, .26]	.05 [-.20, .29]	.01 [-.24, .26]	-.32* [-.53, -.08]	-.09 [-.33, .16]	-.05 [-.29, .20]	-.22 [-.44, .03]	-.15 [-.38, .10]	-.20 [-.43, .05]	-.17
Competentiescores													
Regievoering	-.24 [-.46, .01]											-.22 [-.44, .03]	-.19
Analytisch vermogen		.01 [-.24, .26]										-.22 [-.44, .03]	-.19
Flexibiliteit			.04 [-.21, .28]									-.01 [-.26, .24]	-.01
Leervermogen & zelfreflectie				-.01 [-.26, .24]								-.05 [-.29, .20]	-.01
Luisteren					-.16 [-.39, .09]							-.16 [-.39, .09]	-.12
Omgevingsbewustzijn						.23 [-.02, .45]						.06 [-.19, .30]	.03
Overredingskracht							-.18 [-.41, .07]					-.16 [-.39, .09]	-.12
Samenwerken								-.13 [-.37, .12]				-.21 [-.44, .04]	-.20
Zelfbewustzijn									-.24 [-.46, .01]			-.18 [-.41, .07]	-.17
Zelfvertrouwen & authenticiteit										-.19 [-.42, .06]		-.05 [-.29, .20]	-.02
Besluitvaardigheid											.01 [-.24, .26]	-.34* [-.54, -.10]	-.33
Juridische kennis												-.04 [-.28, .21]	.02

Notitie. $n = 63$. 95% BHI tussen haakjes. * $p < .05$. r_s is Spearmans's rang correlatie (de beoordelingen op de competenties uit het assessment bestaan uit 3 ordinale categorieën). r is de ongecorrigeerde correlatie, r_c is de correlatie gecorrigeerd voor restrictie in range.

Tabel B5. Correlaties tussen scores op het eindgesprek en beoordelingen tijdens de rio-opleiding.

Criterium	Score											
	Totaalscore	Gemiddeld eindoordeel	Flexibiliteit	Inlevingsvermog.	Visie rechtschap	Regievoering	Bestuursvaardigheid	Visie maats. ontw.	Evenwichtigheid	Samenwerken	Maatsch nevenact.	
	r	r_c	r	r_c	r	r_c	r	r_c	r	r_c	r	r_c
Totaalscore	.09 [-.21, .37]	.02 [-.36, .22]	-.32*	-.08	-.18 [-.45, .12]	.01 [-.28, .30]	-.04 [-.33, .26]	.02 [-.28, .31]	.02 [-.28, .31]	-.03 [-.32, .27]	.09 [-.21, .37]	-.12 [-.38, .20]
Gewogen totaalscore	.14 [-.19, .42]	-.01 [-.36, .22]	-.36*	-.08	-.11 [-.39, .19]	-.03 [-.32, .27]	-.09 [-.37, .21]	-.02 [-.31, .28]	.01 [-.28, .30]	-.05 [-.34, .25]	.14 [-.19, .42]	-.12 [-.41, .17]
Competentie scores												
Regievoering	.16 [-.14, .43]	.07 [-.28, .31]	-.25 [-.51, .05]	.02	-.07 [-.36, .23]	-.01 [-.30, .28]	-.03 [-.32, .27]	.04 [-.26, .33]	.08 [-.22, .36]	-.06 [-.35, .24]	.16 [-.14, .43]	-.06 [-.33, .26]
Analytisch vermogen	.14 [-.19, .42]	-.08 [-.47, .10]	-.34*	-.20	-.11 [-.39, .19]	-.04 [-.33, .26]	-.10 [-.38, .20]	.01 [-.28, .30]	.02 [-.28, .31]	.01 [-.28, .30]	.14 [-.19, .42]	-.14 [-.37, .21]
Flexibiliteit	.17 [-.13, .44]	-.06 [-.36, .23]	-.07 [-.36, .23]	-.07	.11 [-.19, .39]	.01 [-.28, .30]	.15 [-.20, .38]	.05 [-.25, .34]	.21 [-.09, .47]	.04 [-.26, .33]	.17 [-.13, .44]	.14 [-.13, .44]
Juridische kennis	-.14 [-.42, .16]	.20 [-.19, .39]	-.05 [-.34, .25]	.11	.07 [-.23, .36]	-.10 [-.38, .20]	.10 [-.20, .38]	.08 [-.22, .36]	-.02 [-.31, .28]	-.05 [-.34, .25]	-.14 [-.42, .16]	-.03 [-.29, .29]
Leervermogen	.18 [-.12, .45]	.10 [-.32, .27]	-.31*	-.03	-.09 [-.37, .21]	-.14 [-.42, .16]	.05 [-.25, .34]	.14 [-.19, .42]	.05 [-.25, .34]	.12 [-.18, .40]	.18 [-.12, .45]	-.05 [-.29, .29]
Luisteren	-.02 [-.31, .28]	.05 [-.35, .24]	-.32*	-.06	-.15 [-.42, .15]	-.26 [-.51, .04]	.02 [-.28, .31]	-.09 [-.37, .21]	-.03 [-.32, .27]	-.06 [-.35, .24]	-.02 [-.31, .28]	-.22 [-.45, .12]
Omgevingsbewustzijn	-.08 [-.36, .22]	<.01 [-.21, .37]	-.22 [-.48, .08]	.09	-.11 [-.39, .19]	-.13 [-.41, .17]	-.08 [-.36, .22]	-.24 [-.50, .06]	-.21 [-.47, .09]	-.15 [-.42, .15]	-.08 [-.36, .22]	-.19 [-.53, .01]

Tabel B5. Correlaties tussen scores op het eindgesprek en beoordelingen tijdens de rio-opleiding.

Criterium	Score												
	Totaalscore	Gemiddeld eindoordeel	Flexibiliteit	Inlevingsvermog.	Visie rechterschap	Regievoering	Besluitvaardigheid	Visie maats. ontw.	Evenwichtigheid	Samenwerken	Maatsch nevenact.		
	r_c	r_c					r					r_c	
Overredingskracht	-.16	-.02	-.12	-.30*	-.18	-.22	.02	-.04	.05	-.01	-.09	.04	[-.26, .33]
Samenwerken	.01	-.01	-.06	-.16	.07	-.17	.03	.01	-.05	.14	<-.01	.12	[-.18, .40]
Zelfbewustzijn	-.22	-.12	-.16	-.30*	-.24	-.19	-.25	-.37*	-.13	-.18	-.03	.23	[-.07, .49]
Zelfvertrouwen & authenticiteit	-.03	.09	<.01	-.18	.05	-.06	-.03	-.14	<-.01	.15	-.02	.33*	[-.04, .57]
Besluitvaardigheid	-.06	-.14	-.11	-.30*	-.18	-.17	-.01	-.07	.02	-.05	-.07	-.03	[-.32, .27]

Notitie. $n = 45-48$. 95% BHI tussen haakjes. * $p < .05$. r is de ongecorrigeerde correlatie, r_c is de correlatie gecorrigeerd voor restrictie in range.