

Het selecteren van aspirant-agenten

Evaluatie van de selectieprocedure voor
agenten op niveau 2 tot en met 4

Erasmus Universiteit Rotterdam



Oktober 2013

Dimitri van der Linden

Marise Born

Liza Phielix

Lisanne Touw

Dit onderzoek is uitgevoerd in opdracht van het WODC, afdeling Extern Wetenschappelijke Betrekkingen, Ministerie van Veiligheid en Justitie.

©2013, WODC, Ministerie van Veiligheid en Justitie, Auteursrechten voorbehouden

Samenvatting

Elk jaar melden zich bij de politieacademie honderden kandidaten die solliciteren naar de functie van agent. Voordat de kandidaten aangenomen worden als aspirant-agenten, doorlopen ze een uitvoerige selectieprocedure waarin hun geschiktheid voor politiewerk wordt getest. Het huidige project had als doel de selectieprocedure van de politieacademie te evalueren. Eén van de redenen voor deze evaluatie was om na te gaan of de eisen en bijbehorende selectie-instrumenten nog voldoen aan de nieuwste wetenschappelijke inzichten in de personeelspsychologie.

Om de selectieprocedure te evalueren werd gebruik gemaakt van verschillende bronnen van informatie. Eén daarvan betrof interviews met medewerkers van de politieorganisatie. Een tweede bron werd gevormd door selectiegegevens van alle kandidaten vanaf 2001 en opleidingsgegevens van aspirant-agenten vanaf 2005. Daarnaast waren er rapporten beschikbaar van eerder onderzoek door de politieacademie.

Ten eerste werd ten behoeve van het bepalen van de effectiviteit van de selectieprocedure het criteriumprofiel geanalyseerd van politiewerk op niveau 2 tot en met 4. De vraag werd gesteld welke eigenschappen iemand tot een goede agent maken. De conclusie was dat bij de politieorganisatie het criteriumprofiel reeds tamelijk goed is uitgewerkt. De uitvoerige reeks van wettelijke aanstellingseisen vertoont sterke overlap met de eigenschappen van politiewerk zoals deze worden genoemd in de wetenschappelijke literatuur en in erkende bronnen van criteriuminformatie zoals de website van het O*NET. Mogelijk valt de lijst van criteria voor politiewerk nog enigszins uit te breiden met beroepsinteressematen.

Om de effectiviteit van de selectieprocedure te toetsen, is er ten tweede empirisch onderzoek uitgevoerd naar de zogeheten voorspellende waarde van verschillende selectie-instrumenten. Eén van de voornaamste conclusies die kon worden getrokken, is dat er geen goede criteriuminformatie aanwezig is. Dat wil zeggen, er zijn geen goede gegevens aanwezig die aangeven hoe goed een politieagent functioneert in de praktijk. Om die reden is er in het huidige project gebruikt gemaakt van opleidingsgegevens als criteriummaten, te weten i) of aspirant-agenten de proeven van bekwaamheid tijdens de opleiding goed hebben afgerond, en ii) of aspirant-agenten de opleiding vroegtijdig, zonder diploma, hebben verlaten (uitval).

Een algemene bevinding van het empirisch onderzoek was dat geen van de gebruikte selectie-instrumenten een sterke voorspellende waarde liet zien voor de opleidingscriteria. De cognitieve capaciteitentest vertoonde bijvoorbeeld geen voorspellende waarde voor opleidingsprestatie of uitval. De voorspellende waarden van persoonlijkheid en competenties, zoals gemeten in het selectiegesprek en de praktijkproef, waren vaak niet significant. Indien er wel significant voorspellende waarden werden gevonden waren de effectgroottes over het algemeen laag.

Deze relatief lage voorspellende waarden moeten echter met voorzichtigheid worden geïnterpreteerd. De belangrijkste reden hiervoor is dat er niet zonder meer kan worden

aangenomen dat deze waarden een correcte weergave zijn van de daadwerkelijk voorspellende waarden voor politiewerk. Er kon namelijk niet worden vastgesteld in hoeverre de gehanteerde opleidingscriteria valide waren. In navolging hierop is de voornaamste aanbeveling in het huidige project dan ook dat de politie gebaat zou zijn bij het verzamelen van betrouwbare en valide informatie over het functioneren van de agenten. Zulke informatie is essentieel voor een solide evaluatie van de selectieprocedure en kan tevens als basis dienen voor het evalueren van de effectiviteit van de opleiding tot agent.

Naast bovengenoemde algemene aanbeveling, bevat het rapport nog 41 conclusies en aanbevelingen die betrekking hebben op specifieke delen van de selectieprocedure. Zo wordt bijvoorbeeld vermeld dat de psychometrische kenmerken van de cognitieve capaciteitentest voorsnog onvoldoende vastgesteld konden worden. Er wordt ook vermeld dat aanpassingen aan selectienormen gepaard moeten gaan met zorgvuldig onderzoek naar de effecten hiervan op de samenstelling van het politiekorps. Bovendien wordt beredeneerd dat relatief kleine aanpassingen aan de normen grote effecten kunnen hebben op de selectieratio. Het eindgesprek (het voormalige korpsgesprek) heeft in januari 2013 een grondige herziening ondergaan. De effecten hiervan moeten goed worden geëvalueerd omdat dit gesprek een prominent effect heeft op de selectieratio. Voorheen vertoonde de uitval van kandidaten na dit gesprek grote verschillen tussen de korpsen. In de huidige opzet wordt uitgegaan van een uitval van omstreeks 12 procent in de landelijke selectie.

In één van de laatste hoofdstukken van het rapport zijn enkele nieuwe wetenschappelijke ontwikkelingen besproken. Het betreft onder andere de algemene factor van persoonlijkheid en nieuwe inzichten in emotionele intelligentie en in de duistere kanten van persoonlijkheid. De politieacademie zou na kunnen gaan of deze concepten een toekomstige rol zouden kunnen spelen in het selecteren van aspirant-agenten.

Alhoewel het gebrek aan goede criteriuminformatie eenduidige uitspraken over de effectiviteit van de selectieprocedure moeilijk maken, moet worden vermeld dat de politieacademie een grondige werkwijze hanteert. De huidige selectieonderdelen zijn over het algemeen ontworpen aan de hand van recente wetenschappelijke inzichten in personeelsselectie. Bovendien heerst er bij de afdeling werving- en selectie een open cultuur die erop gericht is om continue de kwaliteit van de selectie te monitoren en te verbeteren.

Summary

Each year, the Dutch police academy receives hundreds of applications from candidates who want to become a police officer. Before candidates are hired however, they need to go through a selection procedure in which their suitability for police work is tested. The goal of the present project was to evaluate the quality of the selection procedure at the police academy. One of the important aims of this evaluation was to examine whether the requirements in the selection procedure are still in accordance with current scientific insights in the area of personnel selection and with changed societal demands.

In order to evaluate the selection procedure, various sources of information were available. One source of information was formed by interviews held with employees from the police organization. Another source included the selection data of all applicants from 2001 until 2012, and training outcome measure from applicants who were hired between 2005 and 2012. In addition, several research reports were available that were provided by the police academy.

The first step in the evaluation of the selection procedure consisted of analyzing the so-called criterion profile for a police officer. This theme relates to the question which personal characteristics are necessary to conduct adequate police work. It could be concluded that in the police organization, the criterion profile for police officers is well developed. The extensive list of requirements for police work shows substantial overlap with the police officer requirements as described in the scientific literature and in acknowledged sources such as O*Net. Possibly, the list of requirements may be extended by measures that assess the occupational interests of the candidates.

To assess the effectiveness of the selection procedure, empirical analyses were conducted that tested the predictive validity of the selection instruments. One of the major conclusions from the empirical tests is that there is a lack of clear and valid criterion-information. Specifically, no clear information was available that provided direct insight into the performance of police officers in practice. Due to this limitation, the present project used the following training outcomes as criteria, i) the performance of the trainee on the tests during the police academy, and ii) whether the trainee dropped out of the training.

One general finding from the empirical research was that none of the selection instruments that were evaluated showed strong predictive validity. For example, the cognitive capacity test could not predict the training outcomes. The majority of the personality and competence measures also did not show significant predictive validity. In those cases, where significant predictive validity was found, effect sizes often were small.

The low predictive validity found in the present project should be interpreted with great caution. One of the main reasons for this, is that it cannot automatically be assumed that the reported validities reflect the true validities. The quality of the criterion-information was too unclear to draw strong conclusions.

In line with the above findings, one of the main recommendations in this project is that the police organization would benefit from clear and valid criterion-information regarding the

performance of police officers. Such information is essential in any evaluation of a selection procedure and can also be helpful in evaluating the effectiveness of the police training.

Besides the general recommendation presented above, the report contains 41 conclusions and recommendations that relate to specific parts of the selection procedure. One of these conclusions is that the psychometric properties of the cognitive capacity test could not yet be established. Furthermore, it is argued that it is imperative to monitor the effect of any adjustments in norm scores on the composition of the police force. Moreover, it is argued that relatively small adjustments in the norms on selection instruments can have large effects on the selection ratio.

The selection interviews conducted by representatives of the police force that candidates have to go through in the final stage of the selection procedure, has gone through a major revision in January 2013. The effects of this revision have to be carefully monitored as the final selection interview has a prominent impact on the selection ratio. Previously, the dropout rate of candidates in this selection interview varied widely among the different police districts. In its present form, however, it is expected that the dropout rate in this selection interview will be approximately 12 percent.

In one of the last chapters of the report, several new scientific developments are discussed. Specifically, information is provided about the general factor of personality, emotional intelligence, and the dark side of personality. The police may wish to examine the possibility of using these new psychological construct in the selection procedure for police officers.

Although the lack of clear criterion-information prevents conclusive statements about the effectiveness of the selection procedure, it has to be noted that the police academy has adopted a thorough approach in developing the selection procedure. Specifically, the majority of instruments and procedures used are in line with recent scientific developments in personnel selection. Moreover, the department of recruitment and selection at the police academy has an organizational culture supporting the continuous monitoring and improvement of the quality of the selection procedure.

Inhoudsopgave

Samenvatting	3
Summary	5
Voorwoord	9
Deel 1: Achtergrondinformatie	11
Hoofdstuk 1: Inleiding	12
Hoofdstuk 2: De selectieprocedure van aspirant-agenten bij de Nederlandse politie	17
Deel 2: Criteriavaststelling	25
Hoofdstuk 3: Wat kenmerkt een goede agent? De criteria van politiewerk	26
Deel 3 Evaluatie selectieprocedure	41
Hoofdstuk 4: Methodeverantwoording	42
Selectieonderdeel A: Cognitieve capaciteit, het fysiek-motorisch onderzoek en de taaltoets	49
Hoofdstuk 5: Cognitieve capaciteit	51
Hoofdstuk 6: Het fysiek-motorisch onderzoek	65
Hoofdstuk 7: De taaltoets	75
Selectieonderdeel B: Persoonlijkheid en competenties	81
Hoofdstuk 8: Persoonlijkheid	80
Hoofdstuk 9: Competenties in het eindgesprek en de praktijkopdracht	89
Hoofdstuk 10: Het eindgesprek/korpsgesprek	99
Hoofdstuk 11: Discussie en conclusies selectieonderdelen	105
Hoofdstuk 12: Norm en cut-off scores	117
Deel 4: Proces- en beleidsmatige aspecten van de selectie	123
Hoofdstuk 13: Proces- en beleidsmatig aspecten van de selectieprocedure	124
Deel 5: Nieuwe constructen-instrumenten in de selectieprocedure?	127
Hoofdstuk 14: Nieuwe constructen in de selectieprocedure?	130
Hoofdstuk 15: Overzicht conclusies/aanbevelingen	143
Referenties	155
Bijlagen	163

Voorwoord

Voor u ligt het rapport waarin onze evaluatie van de selectieprocedure bij de Nederlandse politie is beschreven. De evaluatie is uitgevoerd naar aanleiding van de startnotitie van het Wetenschappelijk Onderzoek- en Documentatie Centrum (WODC) van het Ministerie van Veiligheid en Justitie (2012). Hoofdverantwoordelijken in het project waren professor Marise Born, Hoogleraar Personeelspsychologie en dr. Dimitri van der Linden, Universitair hoofddocent. Beiden zijn werkzaam bij het Instituut voor Psychologie aan de Erasmus Universiteit Rotterdam. Ze zijn werkzaam bij de sectie Arbeids- en Organisationspsychologie aldaar.

Het onderzoek zou niet mogelijk zijn geweest zonder de hulp van een groot aantal betrokkenen. Allereerst willen we onze dank betuigen aan de medewerkers van afdeling Werving en Selectie van de Politieacademie. Het slagen van het huidige project was voor een groot deel afhankelijk van de informatie en gegevens die aan ons zijn verstrekt door de Academie. De afdeling Werving en Selectie speelde een belangrijke rol hierin. De medewerkers van deze afdeling waren altijd bereid om onze vragen te beantwoorden en ons de data te verstrekken waar we om vroegen.

Uiteraard willen we ook graag alle betrokkenen bedanken die we mochten interviewen. De inzichten die de geïnterviewden ons gaven waren zeer waardevol. De begeleidingscommissie (zie Bijlage 7 voor lijst van commissieleden) in het huidige project willen we graag bedanken voor hun inzet en waardevolle tips die hebben bijgedragen aan de kwaliteit van het onderzoek en het rapport. Tot slot willen we het WODC, en met name Olivier Hendriks bedanken om ons in de gelegenheid te stellen de selectieprocedure bij de politie onder de loep te nemen.

Rotterdam, 2013

Dimitri van der Linden
Marise Born
Liza Phielix
Lisanne Touw

Deel 1

Achtergrondinformatie

Hoofdstuk 1

Inleiding

1. Inleiding

De politieorganisatie besteedt veel aandacht aan het selecteren van nieuw personeel. Kandidaten die zich aanmelden doorlopen een aantal goed afgebakende stappen waarin hun geschiktheid voor politiewerk wordt getest. Ondanks deze grondige aanpak is er een aantal redenen om de selectieprocedure bij de politie te evalueren (Zie Startnotitie WODC, 2012).

Een eerste reden is te onderzoeken of de huidige aanstellingseisen voldoen in het licht van de veranderingen in de maatschappij die zich de afgelopen tien jaar hebben voorgedaan en aan de nieuwe wetenschappelijke inzichten over personeelsselectie.

Een tweede relevante ontwikkeling is de invoering van de Nationale Politie vanaf januari 2013, waardoor de aard van de organisatie verandert. Dit werd gezien als een belangrijke gelegenheid om de selectieprocedure van de politie opnieuw onder de loep te nemen (Startnotitie 2012, 2224, WODC). De invoering van de Nationale Politie zou immers effect kunnen hebben op de eisen die worden gesteld aan agenten. Deze mogelijkheid wordt expliciet benoemd in twee recente rapporten van de politie waarin wordt gespeculeerd over de manier waarop het politiewerk zich in de toekomst zou kunnen ontwikkelen (e.g., 'De toekomst begin vandaag', 2012, Directie HRM Nationale Politie).

Twee andere redenen om de selectieprocedure opnieuw te evalueren zijn i) de arbeidspopulatie vergrijsd terwijl initiële politiefuncties juist vaak door jongeren vervuld worden, ii) het is minder vanzelfsprekend geworden om lang bij dezelfde werkgever te blijven terwijl dit voor politiepersoneel juist kenmerkend is.

Naast bovengenoemde organisatorische en arbeidsmarktveranderingen kregen bovendien enkele aspecten van politiewerk recent aandacht in de pers. Een voorbeeld daarvan is de psychische belastbaarheid van agenten (zie bijvoorbeeld: Van der Velden, Bosmans, & Brekveld, 2012). Een eerder onderzoek suggereerde dat relatief veel agenten last hebben van verminderde psychische weerbaarheid met alle gevolgen van dien (Andersson, Elfvers & Felix, 2011). Alhoewel deze aanname momenteel verder onderzocht wordt en er inmiddels ook al enkele maatregelen zijn getroffen om de psychische weerbaarheid bij agenten te verhogen, is het zinvol om ook aanvullende aandacht te besteden aan het meten van psychische weerbaarheid in de selectieprocedure.

In de startnotitie van het WODC (2012) wordt verwezen naar geluiden dat er in de jaren negentig teveel nadruk zou zijn gelegd op de sociale aspecten van politiewerk terwijl het toepassen van noodzakelijk geweld onderbelicht is gebleven. Gezien de aard van het werk moeten agenten echter bereid en in staat zijn om geweld toe te passen indien dat nodig is. Het zou daarom wenselijk zijn om na te gaan of in de selectie een voorspelling kan worden gedaan over de mate waarin een aspirant-agent in staat is om geweld te gebruiken indien dat nodig is.

Bovenstaande ontwikkelingen liggen ten grondslag aan het huidige rapport. Het onderzoek dat hierin wordt beschreven is uitgevoerd door dr. Dimitri van der Linden en Prof. dr. Marise Born

van het Instituut voor Psychologie aan de Erasmus Universiteit Rotterdam. Het onderzoek bestond uit de volgende onderdelen. Ten eerste zijn er interviews gehouden met betrokken personen bij de politieorganisatie (zie bijlage 1 voor overzicht). Diverse functionarissen van de afdeling Werving en Selectie bij de politie zijn geïnterviewd, maar ook personen uit andere relevante gebieden van de organisatie, zoals beleidsmakers, docenten aan de Politieacademie en executieve politiefunctionarissen.

Naast de interviews is gebruik gemaakt van de wetenschappelijke literatuur over selectiepsychologie. Dit betreft onder andere nieuwe ontwikkelingen op het gebied van persoonlijkheid, intelligentie en competenties. Deze literatuur is gebruikt ter interpretatie van de empirische analyses van de selectieonderdelen en ook bij de beschrijving van nieuwe constructen en/of instrumenten in de selectieprocedure (zie bijvoorbeeld hoofdstuk 14).

Een substantieel deel van het huidige onderzoek bestond uit kwantitatieve analyses van de selectieonderdelen teneinde de psychometrische kenmerken vast te stellen van de selectieonderdelen en, indien mogelijk, het toetsen van hun toegevoegde waarde in de selectieprocedure (e.g., hun voorspellende waarde). Op grond van de startnotitie van het WODC (2012) en het onderzoeksvoorstel van het instituut voor Psychologie aan de Erasmus Universiteit Rotterdam komen de volgende onderzoeksvragen aan bod:

A) Wat kenmerkt een goed agent?

Dit betreft de volgende vragen:

- Welke eigenschappen en competenties moet een agent bezitten (of ontwikkelen) om politiewerk adequaat uit te voeren? Het huidige rapport richt zich bij de beantwoording van deze vraag specifiek op executief politiewerk op de niveaus twee tot en met vier.
- Welke van deze eigenschappen/competenties zijn stabiel en welke kunnen nog goed worden ontwikkeld tijdens de opleiding?

B) Wat is de waarde van de verschillende selectieonderdelen en de meetinstrumenten die worden gebruikt?

Deze vraag kan worden onderverdeeld in de volgende deelvragen:

- Hoe valide en betrouwbaar zijn de meetinstrumenten in de selectieprocedure? Daarbij moet worden gedacht aan de cognitieve capaciteitentesten, het fysiek-motorisch onderzoek, de taaltoets, persoonlijkheidstesten, het selectiegesprek en de praktijkopdrachten.
- Wat is de voorspellende waarde van meetinstrumenten in de selectieprocedure?
- Zijn de stappen in de selectieprocedure in overeenstemming met huidige wetenschappelijke inzichten in de personeelsselectie?

C) Afhankelijk van de antwoorden op bovenstaande vragen, worden de volgende vervolgvragen gesteld:

- Zijn er op grond van de evaluatie van de selectieprocedure aanpassingen nodig in het gebruik van, of de normering van de selectie-instrumenten? Zo ja, welke en wat zijn de verwachte financiële gevolgen van de veranderingen?
- Is het op grond van de uitkomsten van de evaluatie mogelijk om aanbevelingen te doen voor het verminderen van de uitval uit de opleiding? Zo ja, welke aanbevelingen en wat zijn de mogelijke financiële consequenties?
- Is het op grond van de evaluatie mogelijk om aanbevelingen te doen voor een kostenreductie van de selectieprocedure? Zo ja, welke?

1.1 Leeswijzer

In dit rapport worden de onderzoeksuitkomsten per hoofdstuk behandeld. Ten eerste wordt een overzicht gegeven van de selectieprocedure bij de politie (Hoofdstuk 2). Dit hoofdstuk beschrijft de stappen die kandidaten moeten doorlopen voordat zij worden aangenomen als aspirant-agent. Informatie over instroom en uitval in elke stap selectie bepaalt uiteindelijk de zogeheten selectieratio, oftewel het percentage kandidaten dat uiteindelijk wordt toegelaten tot de opleiding.

Vervolgens komen de criteria van politiewerk aan bod (Hoofdstuk 3). In dit hoofdstuk gaat het om de vraag welke eigenschappen iemand tot een goede agent maken. Het gaat hier om een cruciaal onderdeel in de evaluatie van elke selectieprocedure. Om na te gaan of een selectieprocedure adequaat is dient immers informatie beschikbaar te zijn over wat men wilt voorspellen. Deze zogeheten criteria zijn richtinggevend in de selectie. Indien deze criteria fundamentele fouten bevatten zou personeel op de verkeerde kenmerken geselecteerd worden.

Hoofdstuk 4 geeft een verantwoording voor de methoden die we hebben gebruikt in het project. De hoofdstukken 5 tot en met 10 beschrijven vervolgens de uitkomsten van de analyses. Alhoewel deze hoofdstukken enige aanbevelingen en discussiepunten bevatten, is er nog een aanvullend hoofdstuk dat de uitkomsten van de analyses gezamenlijk bespreekt en de aanbevelingen en de punten van discussie samenvat (Hoofdstuk 11).

Aangezien hoofdstukken 5 tot en met 10 specifieke onderdelen van de selectie (persoonlijkheidstests, taaltoets etc.) bespreken, zijn er nog enkele hoofdstukken toegevoegd die ingaan op vraagstukken die de gehele selectieprocedure betreffen. Dit zijn de zogeheten norm-en cutoff scores en de politieke en beleidsmatige aspecten van de selectieprocedure bij de politie.

Hoofdstuk 14 behandelt vervolgens enkele relatief recente concepten uit de psychologie die mogelijk gebruikt kunnen worden in de toekomstige selectieprocedure bij de politie.

1.2 De conclusies en aanbevelingen

Door het rapport heen trekken we conclusies waaruit aanbevelingen volgen. Beiden worden beschreven in aparte delen van de tekst, onder het label '**Conclusie**'. De conclusies en aanbevelingen in het huidige rapport worden gegeven op die locaties in het rapport waar ze het meest relevant zijn. Dat wil zeggen dat ze logisch volgen uit de voorgaande tekst. Voor de

duidelijkheid zijn alle conclusies en aanbevelingen nog eens op een rij gezet in het laatste hoofdstuk (15).

Hoofdstuk 2
**De selectie van aspirant-agenten bij de
Nederlandse politie**

2 De selectie bij de politie

Het selecteren van aspirant-agenten vindt plaats op basis van procedure- en beoordelingskaders die zijn opgesteld in 2002 en sindsdien herhaaldelijk zijn aangepast. Verschillende partijen die betrokken zijn bij de politieorganisatie hebben destijds samen met relevante experts vastgesteld wat de aanstellingseisen moeten zijn voor personeel bij de politie. Deze eisen zijn wettelijk vastgesteld en openbaar (zie: <http://wetten.overheid.nl/BWBR0014138>). De selectieprocedure bij de Politieacademie is afgestemd op deze wettelijke aanstellingseisen. Dit hoofdstuk beschrijft de selectieprocedure.

2.1 De selectieprocedure bij de politie

Jaarlijks melden zich bij de politie voldoende kandidaten aan om aan de vraag naar aspirant-agenten te voldoen. Het aanbod van kandidaten fluctueert wel per jaar en per regio en is afhankelijk van de mate waarin de politie zich bezighoudt met wervingsactiviteiten. In 2011 meldden zich bijvoorbeeld zeer veel kandidaten aan omdat er destijds veel grote wervingscampagnes zijn gehouden.

Kandidaten kunnen zich via een website aanmelden door middel van een sollicitatieformulier waarop zij informatie over hun achtergrond invullen (demografische kenmerken, opleiding) en een korte motivatie (= < 80 woorden) geven over hun reden om bij de politie te willen werken. Op dat moment start de selectieprocedure waarvan de verschillende stappen in Figuur 2.1 zijn weergegeven. Deze stappen worden in het nu volgende toegelicht.

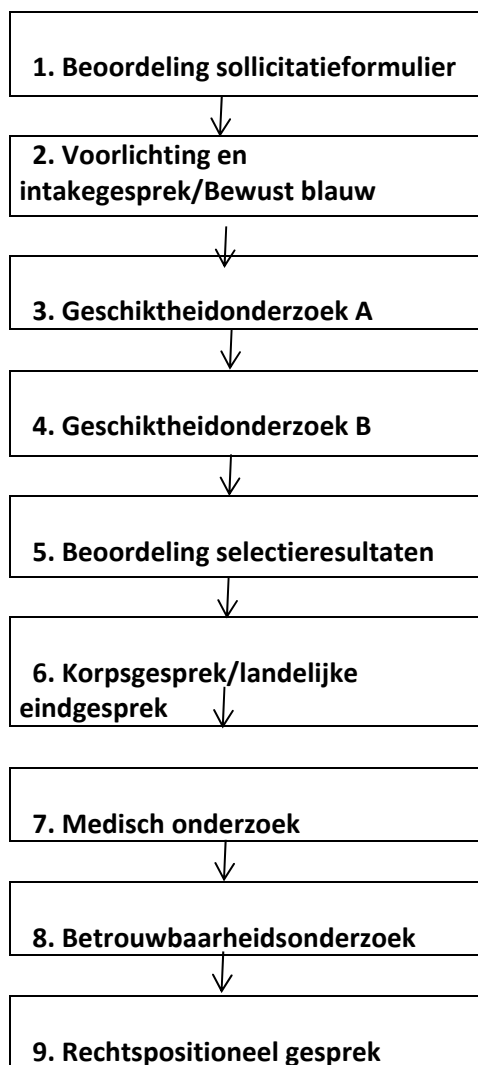
2.1.1 Beoordeling Sollicitatieformulier

Bij de beoordeling van het sollicitatieformulier worden de vereisten gecontroleerd om toe te mogen treden tot de verdere selectieprocedure. Er wordt gecontroleerd of de kandidaat het juiste opleidingsniveau heeft. Daarnaast vindt er een globale evaluatie plaats van de motivatie voor politiewerk en van het gehanteerde taalgebruik. Indien er duidelijk sprake is van onwenselijke motivatie wordt de kandidaat niet uitgenodigd. Een voorbeeld hiervan is een sterke motivatie om geweld of macht te mogen gebruiken. Onwenselijk taalgebruik betreft uitspraken waarin onder andere duidelijk zichtbaar is dat de kandidaat gebrek aan respect heeft voor bepaalde groepen in de samenleving.

2.1.2 Voorlichting en intakegesprek

Indien de informatie op het sollicitatieformulier voldoet, wordt de kandidaat uitgenodigd voor een voorlichting waarbij deze tevens een intakegesprek van 15 minuten krijgt met een selectiefunctionaris van de politie. In dit gesprek wordt verder gevraagd naar de motivatie voor politiewerk en maakt de selectiefunctionaris een inschatting van opvallende uiterlijke kenmerken zoals tatoeages, piercings, haardracht etc. Politiefunctionarissen dienen zich namelijk te onthouden van uitingen en versierselen die afbreuk doen aan een gezagsuitstralende, neutrale en veilige houding. Verder moet afstand worden gedaan van zichtbare accessoires die op enige wijze letsel voor de politiefunctionaris kunnen opleveren. Recent is de voorlichtings- en intakefase aangepast en wordt de kandidaat ook nadrukkelijker gewezen op welke aspecten er zitten aan het werken bij de politie. Het gaat hierbij zowel om de

positieve als de negatieve kanten van de politieopleiding en politiewerk. Het idee hierachter is dat de kandidaat op grond van deze informatie beter zou kunnen inschatten of de politieopleiding en politiewerk iets voor hem/haar is. De nieuwe vorm van de voorlichtings-intakefase worden aangeduid met de term 'Bewust Blauw'. Bij een positieve uitkomst in deze fase wordt de kandidaat uitgenodigd om deel te nemen aan deel A van de selectieprocedure.



Figuur 2.1. Schematische weergave van de huidige selectieprocedure bij de politie

2.1.3 Deel A

Deel A in de selectieprocedure bestaat uit 3 onderdelen: Een cognitieve capaciteitentoets, een fysiek-motorisch onderzoek en een taaltoets. Bij elk onderdeel zijn minimumscores (grenswaarden) gesteld die elkaar niet kunnen compenseren. Indien de kandidaat aan elk van de drie grenswaarden voldoet kan deze doorgaan naar deel B van de selectieprocedure.

2.1.4 Deel B

Deel B richt zich op de persoonlijkheid en de competenties van de kandidaat. Deel B bestaat, net als deel A uit drie onderdelen. In dit geval gaat het om: Een persoonlijkheidsvragenlijst, een selectiegesprek van een uur met een psycholoog van de afdeling Werving en Selectie en een praktijkopdracht van 15 minuten met een acteur. Ook in deel B zijn er grenswaarden geformuleerd van de scores die een kandidaat minimaal moet behalen om te slagen voor een onderdeel. Deze grenswaarden gelden echter niet voor alle aspecten van deel B. Zo zijn er twee competenties waarvoor geen formele minimumscores zijn geformuleerd. Dit zijn zelfinzicht en communicatieve vaardigheden. De reden voor het achterwege laten van deze minimumeisen is dat wordt aangenomen dat de competenties nog goed ontwikkelbaar zijn tijdens de opleiding.

2.1.5 Beoordeling selectieresultaten

Tijdens deze stap moet de selectiefunctionaris op grond van de verschillende bronnen van informatie uit deel B een eindadvies formuleren over de geschiktheid van de kandidaat. De selectiefunctionaris beslist op grond van de beschikbare informatie of de kandidaat doorgaat naar het korpsgesprek. Hiertoe gaat deze functionaris na of de minimumscores zijn behaald op de betreffende onderdelen, en weegt hij of zij daarnaast af of de scores op de competenties waar geen minimumeis voor is gesteld elkaar voldoende compenseren om de kandidaat door te laten.

2.1.6 Korps/eindgesprek

Wanneer de uitkomst van deel B positief is krijgt de kandidaat een korpsgesprek (voor 2013). Dit gesprek heeft als doel om de passendheid van de kandidaat voor het korps te bepalen. Er bleken in het korpsgesprek nog relatief veel kandidaten afgewezen te worden. Er werd vaak aanvullend getest op verschillende competenties, maar de mate waarin dit gebeurde en de gehanteerde methoden en percentages toegelaten kandidaten verschilden sterk per korps. Daarom is vanaf 2013 de status van het korpsgesprek veranderd en heeft het korps een andere rolverdeling gekregen in het aannemen of afwijzen van een kandidaat.

2.1.7 Medisch en betrouwbaarheidsonderzoek en rechtspositioneel gesprek

Wanneer een kandidaat een positieve uitkomst krijgt na het korpsgesprek volgt een medische keuring en een betrouwbaarheidsonderzoek (oftewel een antecedentenkeuring). Als beide goed worden afgerond volgt het rechtspositioneel gesprek waarna het contract kan worden opgesteld. Vanaf dat moment is de kandidaat aspirant-agent en kan deze met de opleiding beginnen. De huidige evaluatie richt zich op stappen 1 tot en met 6 in de selectieprocedure (zie Figuur 2.1). De medische keuring, het betrouwbaarheidsonderzoek en het rechtspositioneel gesprek vallen dan ook buiten de scope van de huidige evaluatie.

2.2 Uitval en selectieratio's

Om inzicht te verkrijgen in de selectieprocedure is het relevant om na te gaan hoe groot de uitval van de kandidaten is per stap. Hierbij is het belangrijk om onderscheid te maken in de situatie vóór augustus 2011 en daarna. In augustus 2011 is er een aantal maatregelen vanuit de politiek ingevoerd, gestuurd door een behoefte aan meer agenten. Eén van die maatregelen omvatte een gestandaardiseerde rol voor het korpsgesprek in de selectieprocedure.

In de tabellen hieronder zijn de uitvalpercentages weergegeven voor (Tabel 2.1) en na (Tabel 2.2) augustus 2011 volgens het rapport van Taris (november 2012). Doordat een aantal kandidaten zich nog in bepaalde fasen van de selectie bevonden, kan er in deze tabellen enige discrepantie zichtbaar zijn in de grootte van de groep (*N*).

Tabel 2.1 Doorstroom en uitval per selectiestap in de periode voor invoering maatregelen (t/m 31 juli 2011) zoals weergegeven in he rapport van de Politieacademie (Taris, 2012)

	<i>N</i>	% uitval
1. Beoordeling sollicitatieformulier	30400	39.8
2. Voorlichting en intakegesprek	14479	35.5
3. Geschiktheidsonderzoek A	9333	43.3
4. Geschiktheidsonderzoek B+ Beoordeling selectieresultaten	5290	68.8
5. Korpsgesprek	1651	27.3
Aangesteld	1370	

Noot: *N* geeft het aantal kandidaten weer dat doorstroomt naar de volgende stap

Tabel 2.2. Doorstroom en uitval per selectiestap in de periode na invoering maatregelen (na 1 augustus 2011) zoals weergegeven in he rapport van de Politieacademie (Taris, 2012)

	<i>N</i>	% uitval
1. Beoordeling sollicitatieformulier	25120	30.8
2. Voorlichting en intakegesprek	13462	31.7
3. Geschiktheidsonderzoek A	9128	38.4
4. Geschiktheidsonderzoek B + Beoordeling selectieresultaten	5446	58.3
5. Korpsgesprek	2268	16.7
Aangesteld	1213	

Noot: *N* geeft het aantal kandidaten weer dat doorstroomt naar de volgende stap

Volgens de interviews en enkele rapporten van de Politieacademie is de selectieratio van het centrale testgedeelte (stappen 1 t/m 5) na de invoering van de maatregelen verhoogd. Voor augustus 2011 werd ongeveer 1 op 20 kandidaten aangenomen, terwijl vanaf augustus 2011 ongeveer 1 op 10 is aangenomen¹. De uitvalspercentages in de stappen in Tabel 2.2 zijn inderdaad lager dan in Tabel 2.1. Doordat er de afgelopen jaren verschillende veranderingen in de selectieprocedure hebben plaatsgevonden kunnen de percentages van voor augustus 2011 echter niet zonder meer worden vergeleken met de percentages na augustus 2011. Door een aantal oorzaken (zoals kandidaten die nog in de procedure zitten) komen de specifieke getallen in Tabellen 2.1 en 2.2 tevens niet overeen met de hierboven genoemde selectieratio's van 1 op 10 en 1 op 20.

Of men nu uitgaat van de selectieratio's zoals deze uit de interviews naar voren komen en in verschillende rapporten worden genoemd (1 op 10 en 1 of 20), of uitgaat van de informatie in de Tabellen 2.1 en 2.2, de conclusie blijft dat er na juli 2011 lagere uitvalspercentages zijn.

¹ In 2013 wordt er slechts zeer beperkt nieuwe aspirant-agenten aangenomen op niveau 2 t/m 4.

Door de vele grote, maar ook subtiele veranderingen die recent in de selectieprocedure hebben plaatsgevonden, is het moeilijk vast te stellen waardoor de verhoogde selectieratio precies is veroorzaakt. Er kunnen slechts enkele mogelijke verklaringen worden gegeven.

Ten eerste is het mogelijk dat de kandidaten die zich na augustus 2011 hebben aangemeld bij voorbaat al meer geschikt waren voor het politiewerk. Er zijn, voor zover bekend, op het gebied van zelfselectie echter geen grote veranderingen ingevoerd in 2011 die er toe zouden kunnen leiden dat er sterk verbeterde zelfselectie heeft plaatsgevonden bij kandidaten. Hoewel deze verklaring niet volledig kan worden uitgesloten lijkt deze dus niet waarschijnlijk.

Een tweede mogelijkheid is dat de formele of informele eisen voor geschiktheid op geen enkele wijze zijn veranderd maar dat men meer geschikte kandidaten aan heeft genomen omdat er meer opleidingsplekken waren te vullen. Aan de bottom-up selectie zou in dit geval niets zijn veranderd (alle kandidaten behalen de gestelde ondergrens van geschiktheid) maar kan de top-down selectie zijn uitgebreid (e.g., niet alleen de beste 10% maar de beste 20% wordt aangenomen).

Een derde mogelijkheid is dat de beoordeling van geschiktheid zelf op enige wijze is versoepeld. Voor zover bekend is, zijn de formele aanstellingseisen bij de Politieacademie niet veranderd. Dit laat nog wel de mogelijkheid open dat de interpretaties van de formele eisen kunnen zijn versoepeld. In een rapport van de Politieacademie over toelatingspercentages wordt bijvoorbeeld het volgende vermeld: *“Ook zien we een beduidend lager uitvalspercentage bij het beoordelen van de selectieresultaten (deel B). Hier zien we een verbetering van bijna 20%.”* (Rapport Politieacademie, 12 april 2012, p. 3).

Indien deze de mogelijkheid inderdaad een rol speelt, is het belangrijk dat goed wordt geëvalueerd wat mogelijke gevolgen kunnen zijn op de samenstelling van het politiekorps. Een verhoging van de selectieratio hoeft overigens niet noodzakelijk een probleem te zijn wanneer de selectieprocedure in eerste instantie streng is geweest. In zo'n geval is de toename van onterecht aangenomen kandidaten (false positives) minimaal.

Welke (combinatie) van de bovenstaande mogelijkheden de juiste is om de hogere toelatingspercentages te verklaren kan niet worden vastgesteld. Het lijkt er bovendien op dat geen van de betrokken partijen uitsluitend kan geven over de redenen van de hogere selectieratio. Dit is een aandachtspunt want wanneer de selectieratio is verdubbeld is het belangrijk voor de politieorganisatie om goed vast te stellen wat deze verdubbeling heeft veroorzaakt. Zonder deze kennis wordt het moeilijker om de kwaliteit van de selectieprocedure en de mogelijke effecten van selectie op het politiekorps goed te monitoren.

Conclusie 1: Door beleidsveranderingen in 2011 is de uitval per selectieonderdeel veranderd. In elk onderdeel is het uitvalspercentage verminderd. Er kon geen definitieve verklaring worden gegeven voor de toegenomen selectieratio, ook niet door de betrokken partijen. Bovendien is vooralsnog onduidelijk wat voor effecten de veranderde selectieratio heeft op de kwaliteit van de aspirant-agenten die worden aangenomen.

Maatregelen die zichtbaar effect hebben op uitvalspercentages en/of de selectieratio's hebben bij een gelijkblijvende pool van kandidaten per definitie effect op het type personeel dat wordt aangenomen. In het kader van geïnformeerde beleidsvoering, stellen we voor dat er in dit soort gevallen wordt onderzocht wat de mogelijke lange –termijneffecten in de praktijk zijn van zulke selectieveranderingen. Goed opgezette cohortstudies waarin aspirant-agenten worden gevolgd vanaf de selectie tot en met het afronden van de opleiding op de politieacademie (en mogelijk nog daarna) kunnen hiertoe een goed middel zijn.

Ter illustratie van bovenstaande aanbeveling: We refereerden al eerder naar geluiden dat een te grote nadruk op de sociale aspecten van politiewerk in de jaren negentig mogelijk hebben geleid tot agenten die minder goed in staat zouden zijn om geweld te gebruiken (zie de startnotitie WODC, 2012). Dit is een voorbeeld van hoe veranderingen in de selectie van personeel mogelijke lange-termijn negatieve gevolgen kunnen hebben op de kwaliteit van politiewerk.

Deel 2

Criteriumvaststelling

Hoofdstuk 3
Wat kenmerkt een goede agent?
De criteria van politiewerk

3. Wat kenmerkt een goede agent? De criteria van politiewerk

Een belangrijk doel in het huidige project is na te gaan welke instrumenten en procedures de geschiktheid voor politiewerk het best voorspellen. De kwaliteit van de voorspellingen is echter sterk afhankelijk van de mate waarin we kunnen bepalen welke eigenschappen een agent moet bezitten om goed te functioneren. Kennis over deze eigenschappen bepaalt onder andere de keuze van de selectie-instrumenten en de normeringen daarvan.

Wanneer de kenmerken van goed politiewerk (de zogenaamde criteria) juist zijn vastgesteld wordt het streven naar een adequate selectie teruggebracht tot de vraag welke metingen deze criteria goed voorspellen. Die vraag valt dan empirisch te beantwoorden. Men kan van een reeks voorspellers immers statistisch vaststellen in hoeverre ze gerelateerd zijn aan de criteria. Zonder adequate criteria kan men slechts beperkt uitspraak doen over de kwaliteit van de selectieprocedure. Daar komt bij dat criteria zijn te beschouwen als *richtinggevers* in het selectieproces. Indien men de verkeerde criteria kiest is de kans groot dan men ook onjuiste voorspellers zal gebruiken en komt men dus niet dichterbij het doel van het vinden van goed personeel.

Het vaststellen van een zogeheten competentieprofiel van agenten blijkt overigens een relatief moeilijke opgave te zijn. Er wordt door betrokkenen, zoals de door ons geïnterviewden, op verschillende manieren gedacht over de kenmerken van een goede agent. Bovendien vermoedt men, blijkens de interviews, dat het politiewerk in de komende decennia zal veranderen en dat het werk daardoor andere eisen zal gaan stellen aan de agenten (e.g., Kwartiermakers Nationale Politie, 2012). Hierdoor zullen van agenten in de toekomst wellicht andere competenties verwacht worden dan in de huidige situatie.

3.1 Doel en procedure

In het huidige hoofdstuk zullen we ingaan op de criteria van politiewerk. We richten ons daarbij, vanwege de opdracht van het WODC, met name op de kenmerken van politieagenten op functieniveaus 2 tot en met 4. De informatiebronnen die we tot onze beschikking hadden waren: i) De interviews met medewerkers bij de politie, ii) achtergrondinformatie over politiewerk. Het gaat hierbij met name om beleidsstukken waarin functie- en competentieprofielen van politieagenten worden besproken, iii) de wetenschappelijke literatuur, met name met betrekking tot politiefuncties, iv) O*Net een website waarin gestandaardiseerde beroepsprofielen worden beschreven (<http://www.onetonline.org/>).

3.2 Leeswijzer

We beginnen met een beschrijving van de informatie uit deze verschillende bronnen (paragrafen 3.3 t/m 3.6). Vervolgens trekken we een aantal conclusies over de basiskenmerken van politiewerk en doen we een aantal aanbevelingen hieromtrent (paragraaf 3.7).

3.3 Politiewerk in een veranderde context

Recent zijn enkele rapporten uitgebracht waarin het functioneren van politieagenten in het verleden, heden en de toekomst is beschreven (Kwartiermakers HRM Nationale Politie, 2012; Ministerie van Veiligheid & Justitie, Directie Strategie, 2012).

Een eerste ontwikkeling die hierin wordt geschetst is dat de samenleving *complexer* wordt en dat daardoor de eisen die aan de politie worden gesteld ook complexer worden. Een voorbeeld is *internetcriminaliteit*. Een gevolg hiervan is dat de politie steeds meer *hoger opgeleid personeel met specifieke ICT-kennis* zou willen werven.

Een tweede toekomstbeeld dat in de rapporten wordt geschetst is dat steeds meer taken van de politie wellicht zullen worden *overgenomen door private bedrijven*. Dit betekent bijvoorbeeld dat bewaking van de veiligheid tijdens grote evenementen gedaan zal worden door de medewerkers van zulke bedrijven.

Ten derde, tot slot, wordt de vorming van de Nationale Politie vermeld als een grote verandering in de organisatie die met name invloed heeft op het bestuur van politie-eenheden en op de *informatie-uitwisseling* tussen verschillende afdelingen.

Een belangrijke vraag die in het kader van het huidige project moet worden gesteld is: In hoeverre hebben bovenstaande ontwikkelingen effect op het werk en daarmee op het profiel van de agenten op niveaus 2 tot en met 4?

De toekomstbeelden van politiewerk zoals in verschillende bronnen (rapporten, interviews) vermeld geven niet direct aanleiding om aan te nemen dat de kenmerken van agenten van niveau 2 t/m 4 substantieel zullen veranderen. Dat men in sommige gevallen krachtig moet optreden zal ook in de toekomst van belang blijven, met als verschil dat er een trend lijkt te zijn, aldus enkelen van de geïnterviewden, naar een toename in de mate van geweld (e.g., meer schietincidenten). In dit kader zal de *psychische weerbaarheid* en het vermogen om traumatische ervaringen te verwerken mogelijk nog belangrijker worden in de toekomst. Ook blijft van belang dat agenten de *sociale vaardigheden* hebben om met burgers te communiceren en hen te ondersteunen.

Kortom, men kan verwachten dat de basiskwaliteiten van agenten op straat niet wezenlijk zullen veranderen in de nabije toekomst. Dit idee werd ook ondersteund door sommigen van de geïnterviewden. Een docent aan de Politieacademie merkte bijvoorbeeld op dat het 'ouderwetse handwerk van vroeger' (contact hebben met de Nederlandse burger, sociale vaardigheden) hetzelfde is gebleven.

Het is een feit dat zelfs voor de agent op straat de digitalisering zal toenemen. Burgers zullen vaker aangifte doen via internet. Men kan zich ook voorstellen dat voor wijkagenten het contact met burgers voor een groter deel via social media zal verlopen. Echter, volgens de onderzoekers in het huidige project kunnen deze invloeden –zoals de toepassingen van nieuwe technologieën– gezien worden als algemene trends in de maatschappij die niet direct een grote of fundamentele invloed hebben op de eisen die worden gesteld aan de agent op straat. Men kan hierbij de vergelijking trekken met technologische veranderingen in de laatste eeuw zoals de intrede van radioapparatuur, de intrede van de computer, en de invoering van mobiele telefonie. Al deze ontwikkelingen hebben zeker invloed gehad op de aard van het politiewerk,

maar hebben ons inzien niet geleid tot fundamentele veranderingen in de eisen die aan politiepersoneel worden gesteld.

Hierop volgend zullen we in onderstaande paragrafen de eigenschappen van agenten bespreken vanuit de huidige situatie, er daarbij vanuit gaande dat deze eigenschappen ook in de nabije toekomst nog relevant zullen blijven.

Conclusie 2: Ondanks organisatorische (e.g., Nationale Politie) en technologische ontwikkelingen, zullen de eisen die aan uitvoerend politiepersoneel worden gesteld niet fundamenteel veranderen.

3.3.1 Kanttekening: Hoger politiepersoneel

Voordat we ingaan op de kenmerken van agenten op niveau 2 tot en met 4 willen we eerst enkele opmerkingen maken over de profielen van hogere politiemedewerkers. Het gaat hierbij om functieniveau 5 en hoger waarbij er sprake is van specifieke werkzaamheden die belangrijk zijn voor de bedrijfsvoering van de politie maar die niet direct betrekking hebben op primair politiewerk op straat (e.g., ordehandhaving, aanhouding arrestanten). Uit de verschillende bronnen van informatie is gebleken dat de politie op termijn steeds meer personeel nodig zal hebben met specifieke kennis. Men kan daarbij bijvoorbeeld denken aan kennis over ICT voor het oplossen van cybercrime of andere zaken die belangrijk zijn voor de bedrijfsvoering, zoals de boekhouding of management.

In de huidige situatie wordt van personeel op hoger niveau verwacht dat zij ook voldoen aan de basiskenmerken van politiepersoneel. Dat wil zeggen dat voor dit hoger personeel dezelfde fysieke en mentale eisen gelden als voor de agent op straat. Uit verschillende interviews is naar voren gekomen dat dit wellicht een onwenselijk situatie is. Door de algemene politie-eisen ook te laten gelden voor specifiek personeel op hoger niveau is de kans groot dat enkele potentieel goede werknemers niet door de selectie heenkomen en men daardoor talent onbenut laat. Een beleidsmaker van de afdeling Werving en Selectie schetste bijvoorbeeld de hypothetische situatie waarbij een computerdeskundige die heel goed is in zijn werk op digitaal vlak, niet bij de politie zou kunnen komen omdat hij de fysieke eis van de sportbaan niet zou halen. Dit valt volgens hem niet onder het optimaal gebruiken van arbeidspotentieel.

Drie van de geïnterviewden op het gebied van werving en selectie pleitten er dan ook voor om meer te kijken naar de specifieke kennis en vaardigheden die gezocht worden en om (gedeeltelijk) de eis te laten varen dat medewerkers met die kennis/vaardigheden aan alle algemene politie-eisen moeten voldoen. De onderzoekers van het huidige project ondersteunen dit idee (zie aanbeveling hieronder). Voor een optimale bedrijfsvoering van de politie is het belangrijk dat medewerkers met name goed zijn in die werkzaamheden waar zij voor zijn aangenomen. Uiteraard geldt dit idee voor sommige politie-eisen sterker dan voor anderen. Sommige fysieke eisen zou men kunnen versoepelen. Andere algemene kenmerken zoals integriteit blijven daarentegen ook voor hoger of specialistischer personeel van belang.

Conclusie 3: Het is te overwegen om bij het aannemen van hoger (niveau \geq 5) politiepersoneel met specifieke kennis/vaardigheden niet langer volledig vast te houden aan alle algemene

aanstellingseisen die gelden voor politiepersoneel. Met sommige van deze eisen (e.g., fysieke conditie, psychologische eigenschappen belangrijk voor straatwerk) zou mogelijkerwijs soepeler omgegaan kunnen worden, terwijl andere algemene kenmerken wel van belang blijven (e.g., integriteit, psychische belastbaarheid). Kortom: De politie zou kunnen verkennen of het werkbaar en wenselijk is om de algemene eisen bij sommige hogere functies te versoepelen of zelfs los te laten.

3.4 De waarde van competenties

Bij het vaststellen van criteria van goed politiewerk wordt tegenwoordig vaak uitgegaan van het begrip competenties. De term 'competenties' is een aanduiding voor de gedragskenmerken waaraan een goede agent zou moeten voldoen. Het concept 'competenties' heeft praktische voordelen. Het omzeilt namelijk ingewikkelde wetenschappelijke discussies over aangeboren versus aangeleerde vaardigheden en stelt selecteurs in staat om simpelweg te beschrijven welke gedragskenmerken iemand moet bezitten. Een voorbeeld daarvan is te vinden in de notitie '*Open advisering en bindend advies*' (Politieacademie, 2002) waarin de competenties die aspirant agenten moeten bezitten worden beschreven. Een daarvan is de competentie 'communicatieve vaardigheden'. Voor assistent politiemedewerkers (niveau 2) en voor politiemedewerkers (niveau 3) is deze gewenste competentie als volgt vastgesteld:

"Spreekt begrijpelijk. Articuleert voldoende en spreekt met voldoende volume voor de boodschap. Maakt weinig grammaticale fouten in uitgesproken zinnen. Geeft informatie aan anderen in voor hen begrijpelijke taal. Verduidelijkt vaktaal meestal. Brengt structuur aan in een betoog." (Politieacademie, 2002, p. 3).

Dit komt overeen met het behalen van een score van 4 op een 7-puntsschaal. Het werken met dit soort competenties is helder en nuttig. De onderzoekers van het huidige project zien de relevantie van dit soort criteria in. Toch vinden wij het nuttig om juist de wetenschappelijke literatuur en discussies te evalueren en nog eens kritisch naar het begrip competenties te kijken. Bovendien is het zinvol om mogelijke aanvullende wetenschappelijke informatie over criteria van politiewerk te onderzoeken.

3.5 Competenties in de wetenschap

Het meten van competenties wordt in toenemende mate toegepast in Human Resource Management, zoals in selectie of bij prestatiebeoordelingen (Cook & Bernthal, 1998). Sinds de jaren 90 is het toepassen van dit begrip in het bedrijfsleven sterk toegenomen maar in de wetenschap werd deze ontwikkeling aanvankelijk vooral gezien als een hype in plaats van een relevant wetenschappelijk onderwerp. Dit is een van de redenen dat het begrip competentie relatief weinig literatuurvorming kent (Roe, 2002). Toch zijn er in de wetenschap verschillende visies op het begrip 'competentie'.

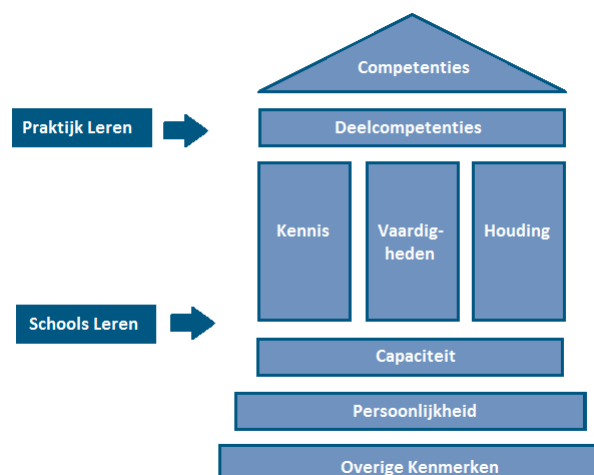
Spencer en Spencer (1993) definieerden competentie als een karaktereigenschap onderliggend aan een effectieve of superieure prestatie. De focus is hierbij met name op de karaktereigenschappen waardoor het begrip competentie sterk overlapt met persoonlijkheid. Green (1999) vatte onder competentie twee factoren, namelijk gedragsgewoontes en persoonlijke vaardigheden om doelen op het werk te behalen. De focus bij Green lag op gedrag,

met andere woorden, in deze definitie is een competentie dat wat iemand laat zien bij het uitvoeren van taken.

Fleisman, Wetrogen, Ulman, en Marshall-Mies (1993) hanteerden een nog bredere definitie en omschreven een competentie als een mix van kennis, vaardigheden, motivatie, overtuigingen, waarden en interesses die leiden tot een prestatie. In tegenstelling tot de eerdere definities wordt door Fleisman et al. (1993) de nadruk gelegd op de verschillende aspecten die ten grondslag liggen aan een bepaalde competentie.

Alhoewel het zeker nuttig is om vereisten voor een specifieke baan te omschrijven in termen van competenties, heeft het gebruik van deze termen ook nadelen. Ten eerste worden competenties over het algemeen gezien als een uiting van veel verschillende aspecten van een persoon. Dit wordt duidelijk in het model van Roe (2002) in Figuur 3.1. Hierin wordt weergegeven dat de mate waarin iemand een bepaalde competentie beheerst afhankelijk is van stabiele kenmerken zoals persoonlijkheid en capaciteiten en van variabele kenmerken zoals wat men heeft geleerd. Het nadeel van zo'n brede conceptualisatie is dat niet direct duidelijk wordt of het ontbreken van een specifieke competentie te wijten valt aan een het ontbreken van stabiele basiskenmerken of dat het simpelweg nog niet is geleerd. Dit heeft dus ook betrekking op de *ontwikkelbaarheid* van een competentie. Indien iemand tijdens de selectie wel de basale eigenschappen heeft om een bepaalde competentie te leren maar deze nog niet in het gedragsrepertoire heeft opgenomen zou men een kandidaat kunnen afwijzen die de competentie nog wel goed zou kunnen leren. In deze situatie laat men potentieel talent dus onbenut.

Er zijn twee verschillende gebieden waarin competenties een rol spelen bij de politie. Enerzijds is er het competentieprofiel van de agent. Het gaat hierbij niet direct om selectie maar juist om de vraag welke kenmerken een goede agent moet bezitten om het werk adequaat uit te voeren. In deze context praat men over de competenties als **criteria**. Anderzijds, zijn er de **voorspellers** van deze criteria, oftewel de competenties zoals deze worden gemeten in de selectie of tijdens de opleiding. De twee gebieden zijn uiteraard sterk met elkaar verbonden. Idealiter zijn de criteria en de competenties die worden gemeten in de selectie identiek. Bijvoorbeeld, als men een goede timmerman zou willen selecteren voor een bouwbedrijf dan zou men een kandidaat kunnen vragen om te laten zien hoe goed zijn timmerwerk is. Bij veel andere beroepen, zoals bij de politie, is dit vaak minder eenduidig. Potentiële aspirant-agenten die zich aanmelden kunnen niet laten zien hoe goed zij politiewerk kunnen uitvoeren omdat ze daar nog geen ervaring mee hebben. Daarom moet men gebruik maken van meer indirecte voorspellers.



Figuur 3.1. Architectuurmodel van competenties volgens Roe (2002)

In onderstaande tabel 3.1 staan de 11 competenties die in het huidige selectiegesprek bij de politie worden gemeten. Deze competenties zijn gebaseerd op eerdere functie-analyses van politiewerk (zie ook: rijksdocument Aanstellingseisen politie, 2002). Dat wil dus zeggen dat de competenties worden gezien als *criteria* en als *voorspellers* van politiewerk.

De competenties beschreven in tabel 3.1 zijn in principe helder en zijn logischerwijs te verbinden met politiewerk. Idealiter zouden goede agenten hoog moeten scoren op *al* deze competenties. Belangrijker is echter dat het ministerie van Veiligheid en Justitie heeft bepaald dat er niet extreem laag wordt gescoord op een of meerdere van de competenties. Hier geldt het principe van de zwakste schakel. Als een agent extreem laag zou scoren op bijvoorbeeld integriteit dan kan dit niet worden gecompenseerd door hoge scores op andere competenties. Met andere woorden, er is vooral aandacht voor de minimale eisen. Hetzelfde geldt voor stressbestendigheid, communicatieve vaardigheden, overzicht etc. Het gaat hier dus voor een groot deel om zogenaamde conjunctieve in plaats van compensatorische selectie.

Tabel 3.1. Competentiemodel Nationale Politie

Competentiemodel Nationale Politie	
1.	Communicatieve vaardigheden
2.	Contactgerichtheid
3.	Inlevingsvermogen
4.	Initiatief nemen
5.	Flexibiliteit
6.	Besluitvaardigheid
7.	Overzicht
8.	Zelfinzicht
9.	Stressbestendigheid
10.	Tolerantie
11.	Integriteit

Het onderscheid tussen competentieniveaus als criteria in politiewerk en als voorspellers in de selectie wordt expliciet erkend in de recente notitie “Ontwerp selectiecommissie Nationale Politie: Eindgesprek” (Kwartiermaker Nationale Politie, 2012). In deze notitie wordt een definitie gegeven van passendheid van een kandidaat voor de politie die ervan uitgaat dat bepaalde kenmerken nog te ontwikkelen zijn tijdens de opleiding.

Definitie: ‘Een kandidaat is passend als hij/zij voldoet aan de basis van de vier kernwaarden van de Nationale Politie (integer, betrouwbaar, moedig, verbindend), waarbij rekening wordt gehouden met een bepaalde mate van ontwikkelbaarheid op deze kernwaarden’. (Kwartiermaker Nationale Politie, 2012, p. 7)

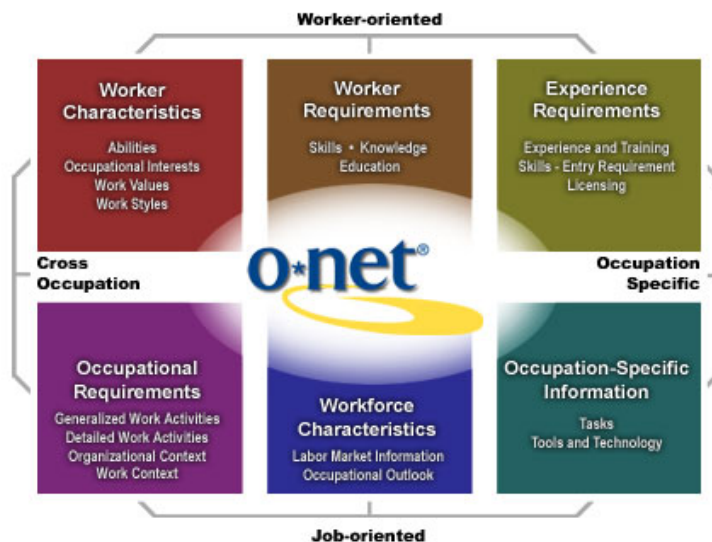
Opvallend is dat in de notitie van 2012 i) wordt gesproken over kernwaarden in plaats van competenties, ii) dat deze kernwaarden worden gezien als drijfveren van de kandidaat die in kaart moeten worden gebracht, en iii) dat er met name nadruk wordt gelegd op vier kernwaarden die tot zekere hoogte overlappen met de competenties zoals beschreven in tabel 3.1. Bijvoorbeeld, één van de kernwaarden die wordt genoemd in de notitie van 2012, is ‘integer’. Deze overlapt met een van de competenties in tabel 3.1, namelijk ‘integriteit’. De volgende kernwaarde zijn beschreven in de notitie van 2012:

- **Integer:** De mate waarin men eerlijk en open is, en zich fatsoenlijk gedraagt naar anderen.
- **Betrouwbaar:** De mate waarin men consequent is in handelen en verantwoordelijkheidsgevoel heeft.
- **Moedig:** De mate waarin men doortastend kan optreden, soms met gevaar voor eigen leven.
- **Verbindend.** De mate waarin men nauw kan samenwerken met burgers en veiligheidspartners.

Het belang van criteria en de mogelijke ontwikkelbaarheid van gedrag waren voor de onderzoekers in het huidige rapport aanleiding om te kijken naar O*Net. O*Net is een grote internationale online database waarin eigenschappen van velen verschillende beroepen zijn beschreven. Een voordeel van het gebruik van deze procedure is dat O*Net specifiek is over de categorieën waaronder een competentie kan worden ingedeeld. Dat wil zeggen; het geeft aanvullend inzicht over welke eigenschappen van een persoon relatief stabiel zullen zijn en welke nog relatief goed te leren zijn.

3.6 Eigenschappen van politiemedewerkers volgens O*Net: Stabiele versus ontwikkelbare kenmerken

O*Net maakt gebruik van een zogeheten ‘content model’, een raamwerk met de belangrijkste informatie over banen, -geïntegreerd in één model. Het model maakt onderscheid tussen specifieke karakteristieken van een baan (‘job-oriented descriptors’) en persoonskenmerken die nodig zijn om een baan goed te kunnen uitvoeren (‘worker-oriented descriptors’) (zie Figuur 3.2).



Figuur 3.2. Content model in O*Net

In feite wordt in O*Net niet gesproken over competenties maar wordt onderscheid gemaakt tussen *domeinen*. Deze zijn: Bekwaamheden, Vaardigheden, Werkstijlen, Werkwaarden, Interesses, Werkcontext, Kennis, Ervaring en Opleiding. In de huidige evaluatie worden Ervaring en Opleiding buiten beschouwing gelaten omdat daar reeds heel specifieke eisen aan worden gesteld bij de politie. De overige domeinen worden hieronder uitgelegd:

1) Bekwaamheden (in het Engels *Abilities*) worden gezien als relatief stabiele attributen van een persoon die bijdragen aan prestatie. In tabel 3.2 is te zien dat voor politiewerk vooral cognitieve bekwaamheden worden vermeld zoals redeneren, aandacht, en reactietijden, maar ook mondelinge uitdrukkingsvaardigheid.

2) Interesses worden in O*Net beschreven als iemands voorkeuren voor een specifieke werkomgeving. O*Net gaat hierbij uit van het Occupational Interest Profile dat vergelijkbaar is met de beroepsinteresseprofielen zoals beschreven in de theorie van Holland (1996). Dit is interessant want voor zover bekend wordt er in de selectieprocedure bij de politie nog geen gebruik gemaakt van formele interessemodellen zoals die van Holland (1996) of andere onderzoekers uit dit veld.

De interesses die volgens O*Net relevant zijn voor politiewerk zijn: Realistisch, Sociaal, Ondernemend en Conventioneel. Interesses kunnen wel enigszins veranderen over tijd maar worden voor een groot deel toch ook gezien als tamelijk stabiele kenmerken.

3) Werkwaarden: In tegenstelling tot de waarden beschreven in de notitie van de kwartiermakers van de politie (2012) worden waarden in O*Net gezien als motivationele aspecten in het werk die belangrijk zijn voor de werktevredenheid. Het verschil met interesses is dat waarden over banen heen relevant zijn, terwijl werkwaarden betrekking hebben op de specifieke aspecten van een baan. Volgens O*Net is een belangrijke waarde voor politiemedewerkers bijvoorbeeld steun, oftewel de behoefte aan management en collega's die ondersteunend zijn. Werkwaarden hebben daarmee ook een sterke relatie met psychische

weerbaarheid (zie ook tabel 3.2 voor een overzicht van een completer overzicht van werkwaarden).

4) Werkstijlen worden in O*Net beschreven als in-principe niet beïnvloedbare persoonlijkheidskenmerken die iemands prestatie kunnen beïnvloeden. Voor politiewerk wordt in O*Net bijvoorbeeld genoemd: Consciëntieusheid.

5) Basisvaardigheden, zijn *ontwikkelbare* vaardigheden die het leren van nieuwe kennis/vaardigheden kunnen faciliteren. Bijvoorbeeld het vermogen om zichzelf of anderen te kunnen evalueren waardoor gedrag verbeterd kan worden (bij de politie wordt dit gemeten als 'zelfinzicht'). Er kan opgemerkt worden dat deze vaardigheden door O*Net expliciet worden omschreven als goed ontwikkelbaar.

6) Crossfunctionele vaardigheden, oftewel die kenmerken die minder sterk gebonden zijn aan prestatie op specifieke taken maar juist een brede invloed op prestatie kunnen hebben. Voor politiewerk noemt O*Net sociale vaardigheden in deze categorie.

In onderstaande tabel 3.2 worden eigenschappen van politiemedewerkers beschreven. Het gaat hier om eigenschappen van agenten die basispolitiewerk uitvoeren zoals surveilleren, hulp bieden, en verdachten aanhouden (police patrol officers). O*Net noemt kenmerken en geeft daarbij tevens informatie over hoe relevant een specifiek kenmerk voor politiewerk is (uitgedrukt in percentages). In de tabel zijn de 10 belangrijkste kenmerken weergegeven inclusief hun percentages. In sommige gevallen zijn er echter minder dan tien kenmerken weergegeven aangezien het belang dat werd genoemd 50% of minder betrof.

Tabel 3.2. Kennis, bekwaamheden, vaardigheden, werkstijlen, werkcontext, werkwaarden en interesses van een politie agent (O*Net)

Politie agent	Onderdelen	Relevantie –percentage volgens experts
Kennis over...	Publieke veiligheid en beveiliging	91
	Wetten en overheid	88
	Nederlandse taal	82
	Psychologie	75
	Klantenservice	71
	Telecommunicatie	60
	Sociologie en antropologie	58
	Administratie	57
Bekwaamheden	Aanvoelen van problemen	78
	Mondelinge uitdrukkingsvaardigheid	75
	Deductief redeneren	72
	Inductief redeneren	69
	Begrijpend luisteren	69
	Selectieve aandacht	69
	Goed zicht (van veraf)	66

	Flexibiliteit in probleemoplossing	66
	Goed zicht (dichtbij)	66
	Reactietijd	66
Vaardigheden	Actief luisteren	75
	Kritisch denken	69
	Monitoren van eigen prestatie	66
	Kunnen inschatten sociale situaties	66
	Effectief spreken	63
	Coördineren	60
	Onderhandelen	60
	Overtuigen	60
	Begrijpend lezen	60
	Service/hulpgerichtheid	60
Werkstijlen	Integriteit	95
	Zelfcontrole	93
	Aandacht voor detail	92
	Stressbestendigheid	92
	Betrouwbaarheid	89
	Samenwerken	86
	Initiatief nemen	83
	Medeleven	82
	Aanpassen/flexibiliteit	81
	Leiderschap	80
Werkcontext	In contact staan met externe partijen	98
	Contact met anderen	97
	Omgaan met boze en agressieve mensen	95
	Werken in een gesloten auto	95
	Face-to-face discussies voeren	93
	Vrijheid om beslissingen te nemen	93
	Gelegenheid om beslissingen te nemen	93
	Belang van exact en accuraat werken	92
	Werken in een groep of team	92
	Buitenwerk	91
Werkwaarden	De behoefte om te willen presteren	78
	De behoefte aan onafhankelijkheid	78
	De behoefte aan relaties met anderen	78
	De behoefte aan steun	78
	De behoeften aan erkenning	67
	De behoefte aan gunstige werkomstandigheden	61
Interesses	Realistisch – problemen oplossen met mensen, ipv gebruiksvorwerpen	78
	Ondernemend – verantwoordelijk voor eigen project	72

Conventioneel – procedures en routines	56
Sociaal – sociaal beroep	50

3.7 Fysieke eisen

Opvallend aan bovenstaande beschrijving van O*Net is dat fysieke eisen/kenmerken nergens in de top tien voorkomen. De duidelijkste verwijzingen naar fysieke eisen worden bij O*Net genoemd onder de noemer 'Bekwaamheden'. De eerste directe verwijzing naar fysieke eisen komt hierbij op de 17^e plaats en betreft de 'coördinatie van ledematen' met een belang van 56%.

Enkele fysieke bekwaamheden die volgen zijn: op de 19^e plaats 'arm-hand vastigheid' (53%), 'fijne motoriek van vingers' (plaats 24, 50%), 'gehoor' (plaats 25, 50%), 'handsnelheid' (plaats 26, 50%), 'snelheid van beweging ledematen' (plaats 27, 50%), 'algehele lichaamscoördinatie' (plaats 30, 47%), 'uithoudingsvermogen' (plaats 31, 47%), 'statische kracht' (tillen, trekken, etc.) (plaats 32, 47%), rompkracht (e.g., buikspieren) (plaats 36, 44%), explosieve kracht (plaats 37, 41%).

Volgens O*Net ligt het belang van fysieke conditie in politiewerk dus rond de 50%. Men zou hierbij kunnen aanmerken dat zodra er sprake is van een voldoende basisniveau, fysieke conditie minder belangrijk is dan andere eigenschappen. Dit idee past goed bij de huidige selectieprocedure van de politie. Daar worden kandidaten pas toegelaten nadat ze aan de minimale fysieke eisen hebben voldaan.

Wel is het zo dat een aantal van de fysieke kenmerken die worden genoemd door O*Net in de huidige selectieprocedure (en opleiding) van de politie ogenschijnlijk niet meer als zodanig erkend worden. Een voorbeeld daarvan komt uit een interview met een instructeur op de Politieacademie. Deze gaf aan dat aspirant-agenten tijdens de opleiding op het fysieke en motorische vlak soms te kort komen. Hij gaf tevens aan dat vroeger *motorische vaardigheden* expliciet getest werden in de selectie maar dat dit tegenwoordig niet meer het geval is. Het gevolg is dat sommige aspirant-agenten moeilijk op te leiden zijn wat betreft hun sportieve vaardigheden en hiermee blijven worstelen tijdens hun loopbaan.

Bovenstaande is een goede illustratie van verandering in inzichten wat betreft de criteria van politiewerk en de invloed daarvan op selectie. Specifiek: het lijkt erop dat men in het verleden op een bepaald moment besloten heeft dat motoriek niet een essentieel criterium is bij politiewerk. Diengevolgen heeft men het testen van motorische capaciteiten uit de selectie gehaald. Volgens minstens een van de geïnterviewde experts heeft dit echter geleid tot de toelating van agenten bij wie motoriek een probleem is.

Over het algemeen kan men concluderen dat lichamelijke conditie een relevant criterium is voor politiewerk. Lichamelijke conditie is voor een groot deel afhankelijk van factoren zoals iemands motivatie om het lichaam in conditie te houden door onder andere sport en goede voeding. Een relevant criterium voor politiewerk zou daarom niet alleen kunnen zijn dat men in relatief goede conditie is –wat een momentopname is- maar ook dat men de motivatie heeft om dat zo te houden. Laatstgenoemde criterium is waarschijnlijk wel meer voorspellend voor lichamelijke conditie op de lange termijn.

Een opvallend punt dat ook te maken heeft met bovenstaande stabiliteitskwestie is dat in O*Net de fysieke kenmerken met name geschaard worden onder 'Bekwaamheden'. Daarmee wordt impliciet aangegeven dat het gaat om relatief stabiele kenmerken die de prestatie beïnvloeden.

De wetenschappelijke discussie over dit onderwerp is nog open. Enerzijds zijn Uithoudingsvermogen en kracht bijvoorbeeld tamelijk goed te verbeteren door voldoende training. Daar komt bij dat uithoudingsvermogen tevens beïnvloed wordt door levensstijl zoals het voedselpatroon. Anderzijds zijn er studies die hebben aangetoond dat fysieke conditie, bijvoorbeeld zoals gemeten met een VO₂-max test en een 12-minuten loopparcours, relatief stabiel is over grote tijdsperiode (e.g. 20 jaar: Kemper, De Vente, Van Mechelen, & Twisk, 2001). Dit onderstreept het belang om niet alleen uit te gaan van huidige lichamelijke conditieniveaus maar ook van stabiele factoren die de conditie over langere periodes beïnvloeden (e.g., motivatie).

Een aspect dat nog niet in deze context genoemd is maar wel mogelijk van belang is voor de uitvoering van politiewerk is de Body Mass Index (BMI). De BMI speelt een rol in de conditie en gezondheid van een agent en zou men dus als criterium voor politiewerk kunnen beschouwen. Recent onderzoek gaf aan dat een te hoog BMI een groeiend probleem is voor aspirant-agenten alsook voor agenten in de praktijk (zie: <http://nos.nl/artikel/253018-zeker-30-%-van-agenten-niet-fit.html>). Aanvullend onderzoek zou moeten aantonen vanaf welk niveau van BMI er (gemiddeld genomen) onaanvaardbare risico's ontstaan wat betreft de functie-uitvoering, zowel direct als op de langere termijn.

Conclusie 4: Concluderend kan worden vermeld dat het criteriumprofiel voor politiewerk (niveau 2 t/m 4) kan worden aangescherpt. Als we ons beperken tot het gebied van de criteria (en dus nog niet de selectie) kunnen voornamelijk de volgende aanbevelingen worden geformuleerd:

1) Het is nuttig om het criteriumprofiel van de fysieke eisen voor politiewerk nog eens kritisch te bekijken en waar mogelijk aan te vullen en/of specificeren.

Daarbij kan men op het volgende letten:

Blijkbaar worden enkele fysieke eisen niet langer expliciet erkend als criteria van politiewerk terwijl deze mogelijkerwijze wel van belang zijn. Denk daarbij onder andere aan motorische vaardigheden zoals ook genoemd in O*Net.

2) Bij het vaststellen van fysieke criteria voor politiewerk zou men niet alleen moeten uitgaan van inclusiecriteria (minimale conditie, minimale kracht etc.) maar ook de exclusiecriteria moeten aanscherpen (bijvoorbeeld problemen met trommelvliezen, knieën etc.).

3) Een interessante mogelijkheid is om bij fysieke eisen niet alleen uit te gaan van directe uitingen van lichamelijke gezondheid en conditie maar ook naar onderliggende (psychologische) factoren die hierop van invloed kunnen zijn. Men kan hierbij bijvoorbeeld denken aan motivatie om gezond/fit te blijven (e.g., Curry & Weiss, 1989; Rejeski & Kennedy, 1988).

3.8 Discussie, conclusies, en verder aanbevelingen betreffende de criteria voor politiewerk (niveau 2 t/m 4)

Aangezien de bespreking van competenties in het huidige hoofdstuk vooral plaats vindt vanuit het criteriumoogpunt, zijn de belangrijkste eindvragen²:

Is het huidige criteriaprofiel van agenten compleet?

Welke van de competenties zijn ontwikkelbaar en welke liggen relatief vast?

3.8.1 is het criteria-profiel compleet?

Wat betreft vraag 1 (compleetheid competentieprofiel) is onze conclusie dat veel van de criteria van politiewerk reeds zijn beschreven in de formele aanstellingseisen zoals die wettelijk zijn vastgelegd door de overheid (BWBR0014138). Deze formele aanstellingseisen overlappen met de criteria zoals die worden genoemd in O*Net.

In het overzicht van de profielen zoals gebruikt bij de Politieacademie valt ons wel op dat er weinig expliciet wordt vermeldt over interesses. Echter, interesses spelen waarschijnlijk een belangrijke rol in de mate waarin een aspirant-agent geneigd is uit te vallen tijdens de studie of de politieorganisatie te verlaten. Bij het invoegen van interesses kan men gebruik maken van classificaties zoals die genoemd worden in O*Net.

Conclusie 5: Interesses zijn een nuttige aanvulling op het competentieprofiel van agenten. Uitgaande van O*Net zijn ‘realistisch’, ‘ondernemend’, ‘conventioneel’, en ‘sociaal’ logische keuzes van interesses.

3.8.2 ontwikkelbaarheid van competenties

De tweede vraag betreft de ontwikkelbaarheid van een competentie. O*Net maakt een aantal aannames hierover. Alle kenmerken genoemd onder bekwaamheden worden bijvoorbeeld gezien als relatief stabiele kenmerken. Hieronder valt cognitieve capaciteit. De wetenschappelijke literatuur is hierover tamelijk eenduidig: cognitieve capaciteit wordt gezien als een zeer stabiel construct met één van de hoogste erfelijkheidscoëfficiënten in de psychologie (Jensen, 1998). Deze stabiliteit van cognitieve capaciteit wordt reeds erkend in de selectieprocedure bij de politie.

Fysieke kenmerken vallen in O*Net ook onder bekwaamheden (zie 3.7 Fysieke eisen). Dit is opmerkelijk omdat zaken zoals conditie en kracht door middel van training makkelijk te

² Hierbij dient opgemerkt te worden dat een aantal vragen over competenties later zullen worden behandeld in een hoofdstuk waarin we de competenties bekijken vanuit het oogpunt van de selectie. Dit zijn vragen zoals: i) hoeveel en welke van de competenties moeten worden gemeten in de selectie, ii) hoe dat moet gebeuren en iii) welk relatief belang ze moeten krijgen in de eindbeslissing.

verbeteren zijn. De stabiliteit van deze fysieke kenmerken is dus van heel andere orde dan die van de cognitieve capaciteiten.

In O*Net worden de eigenschappen genoemd onder werkstijlen ook gezien als relatief stabiel (slecht ontwikkelbaar). De eenduidigheid van deze indeling wordt echter minder breed ondersteund door de literatuur. De stabiliteit van integriteit en zelfcontrole wordt daarentegen wel goed ondersteund door de literatuur (Ones, Viswesvaran, & Schmidt, 1993). Integriteit wordt over het algemeen gezien als een persoonlijkheidseigenschap. Integriteit is essentieel voor politiewerk vanwege de machtspositie en geweldsmonopolie van agenten. Voor de stabiliteit van zelfcontrole is ook veel bewijs. Bijvoorbeeld, verschillende studies hebben aangetoond dat tekenen van zelfcontrole op zeer jonge leeftijd samenhangen met levensuitkomsten op volwassen leeftijd. Men kan hierbij denken aan gezondheid, carrière succes en de mate waarin men in aanraking komt met de politie (e.g., Tangney, Baumeister, & Boone, 2004).

In O*Net wordt stressbestendigheid overigens ook gezien als stabiel kenmerk. Dit komt overeen met het idee dat stressbestendigheid een persoonlijkheidskenmerk is die gedeeltelijk overlapt met neuroticisme (of het omgekeerde hiervan: emotionele stabiliteit).

Conclusie 6: Vragen over de stabiliteit versus beïnvloedbaarheid van een (gedrags)kenmerk zijn op grond van de wetenschappelijke literatuur tamelijk lastig eenduidig te beantwoorden. Voor vele kenmerken is de wetenschappelijk discussie over dit onderwerp nog open. O*Net kan hier wel een richtlijn bieden omdat deze kenmerken indeelt in categorieën die gekoppeld zijn aan een bepaalde, veronderstelde mate van ontwikkelbaarheid.

Deel 3

Evaluatie van de selectieprocedure

Hoofdstuk 4

Methodeverantwoording

4. Methode verantwoording

4.1 Interviews

Ten behoeve van de evaluatie van de selectieprocedure zijn er interviews gehouden om inzicht te krijgen in de besluitvorming omtrent de selectieprocedure en om achtergrondinformatie te verkrijgen over de huidige stand van zaken van de selectie. Het merendeel van de geïnterviewden is nauw betrokken bij de werving en selectie bij de politie, waarvan sommigen uitvoerende taken hebben en anderen zich meer bezighouden met beleid rond selectie. Sommige geïnterviewden doen uitvoerend politiewerk.

We maakten tijdens de interviews gebruik van standaardvragen die aan alle geïnterviewden werden gesteld. Bijvoorbeeld de vraag 'wat maakt volgens u een goede agent?'. Een deel van de vragen was afhankelijk van het gebied waarin de geïnterviewde werkte (e.g., beleid, selectie) en van de manier waarop het gesprek zich ontwikkelde. Er werd met andere woorden gebruik gemaakt van een semi-gestructureerd interviewschema.

Schriftelijke samenvattingen van elk interview zijn opgesteld met daarin een overzicht van alle belangrijke punten die door de geïnterviewden werden vermeld. In totaal zijn er 17 personen geïnterviewd in sessies die ongeveer één á anderhalf uur duurden. De geïnterviewden werden bezocht op locatie. Een lijst met namen van de geïnterviewden en hun functie is te vinden in Bijlage 1.

4.2 Verantwoording data-analyses

4.2.1 Bepaling van de predictoren en de criteria

Een belangrijk onderdeel van het huidige project was de analyse van selectiedata van de Politieacademie. Sinds 2001 worden bijna alle gegevens uit de selectieprocedure bewaard. Het gaat hier om de uitkomsten van de kandidaten in deel A en B van de selectieprocedure. Sinds 2006 worden opleidingsgegevens bewaard in de vorm van de uitslagen op zogeheten proeven van bekwaamheid en de onderdelen daarvan.

Naast informatie over de proeven van bekwaamheid heeft de Politieacademie informatie over de opleidingsstatus van de aspirant-agenten. Van de kandidaten die worden toegelaten tot de Politieacademie wordt ten behoeve van de opleidingsstatus bijgehouden of zij i) de opleiding hebben afgerond met een diploma, ii) vroegtijdig de opleiding hebben verlaten zonder diploma, of iii) dat zij momenteel nog studeren.

Hieronder beschrijven we op welke wijze de predictor en criteriuminformatie is gebruikt in het onderzoek en welke beslissingen zijn genomen. De totale dataset van de Politieacademie bestaat uit ongeveer 30.000 kandidaten die de selectieprocedure hebben doorlopen. Van die kandidaten zijn de demografische kenmerken bekend, zoals geslacht en afkomst, en van de kandidaten die zijn aangenomen zijn de uitkomsten op de proeven van bekwaamheid bekend (criteriuminformatie).

4.2.1.1 Predictorinformatie. Niet alle kandidaten in de selectiedataset zijn gemeten met dezelfde selectie-instrumenten. Ten eerste heeft de Politieacademie in augustus 2006 de cognitieve test vervangen. Voorafgaand aan augustus 2006 was de Politie Intelligentie Test (PIT)

het instrument om cognitieve capaciteit te meten. Na die tijd is overgegaan op *de Cognitieve Adaptieve Intelligentie Test (CAIT)*, geleverd door het bedrijf CEBIR. Ten tweede is sinds maart 2008 de meting van persoonlijkheid veranderd. Voor maart 2008 werd er gebruik gemaakt van de M5Q (van Leeuwen, 2000) om de zogeheten Big Five dimensies te meten (zie hoofdstuk 8 over persoonlijkheid). In maart 2008 is overgegaan op het gebruik van de Nederlandse versie van de *NEO-PI-R* (Hoekstra, Ormel, & de Fruyt, 1993), een van de bekendste persoonlijkheidsmaten in het domein van de persoonlijkheidspsychologie. De competenties die gemeten worden in het selectiegesprek en tijdens de praktijkopdracht zijn sinds 2001 bijvoorbeeld niet substantieel veranderd.

Bovengenoemde veranderingen in de instrumenten impliceerde dat het niet mogelijk was om alle data uit verschillende tijdsperiodes (e.g., voor en na augustus 2006, voor en na maart 2008) bij elkaar te voegen. Bovendien heeft het feit dat sommige instrumenten niet meer gebruikt worden de relevantie van die data verminderd. Ten behoeve van de evaluatie is daarom besloten om *alleen* data mee te nemen die na maart 2008 zijn verzameld.

Een andere reden om alleen data na maart 2008 mee te nemen was dat in haar proefschrift en in een gepubliceerd artikel, De Meijer reeds heeft gekeken naar de voorspellende waarde van de selectie-instrumenten die tot die tijd werden gebruikt (e.g., De Meijer et al., 2008). De meeste informatie hierover is terug te vinden in het proefschrift van De Meijer en het bijbehorend artikel.

In drie gevallen is er gebruik gemaakt van data vóór die periode. Dit geldt ten eerste voor het onderzoeken van de factorstructuur van de competenties. Aangezien de competenties voor 2008 hetzelfde waren als daarna was het mogelijk om bij specifieke analyses *alle* data te analyseren. Hetzelfde geldt, ten tweede, voor de praktijkopdracht en, ten derde, maar in mindere mate voor intelligentie (na 2006 is gekozen voor een andere intelligentietest). In het rapport zullen we telkens aangeven welke steekproefgrootte is gehanteerd.

4.2.2 Vaststellen van de kwaliteit van de criteriuminformatie

De opties om de voorspellende waarde van de selectie-instrumenten te evalueren waren in het huidige project beperkt door de criteria die we tot onze beschikking hadden. Er waren twee bronnen van criteriuminformatie die konden worden aangeleverd door de Politieacademie. De eerste bron betreft informatie over de opleidingsstatus: Van de kandidaten was vastgesteld of zij eind 2012 reeds afgestudeerd waren met diploma, vroegtijdig de opleiding hadden verlaten zonder diploma, of nog studeerden. Met betrekking tot de data na maart 2008 had het gebruik van deze maat wel de beperking dat op het moment van dataverzameling nog maar weinig aspirant-agenten de opleiding al met succes konden hebben afgerond. Wel was in deze data te zien welke personen er reeds vroegtijdig waren uitgevallen.

De tweede bron van criteriuminformatie betreft het wel of niet slagen op de zogenaamde proeven van bekwaamheid tijdens de opleiding. Het gaat hierbij om dichtome data (wel/niet geslaagd) die op twee niveaus zijn vastgesteld, namelijk het aantal losse onderdelen van proeven die is behaald en het totaal aantal proeven dat is behaald.

Ter illustratie: Het eindoordeel (wel/niet gehaald) van de Proeve “Optreden bij delicten tegen de openbare orde” bestaat uit de onderdelen ‘toezicht houden openbare ordegebied’, ‘optreden delicten openbare orde’, ‘rechercheren’, ‘afnemen getuigenverhoor’, ‘afnemen verdachtenverhoor’, ‘voeren eindgesprek naar aanleiding van procesverslag’ en ‘analyseren dilemma’. Men hoeft niet noodzakelijke *alle* onderdelen te halen om een proeve als geheel met een voldoende af te sluiten. Binnen elke proeve zijn er wel een aantal onderdelen die zeker behaald zijn maar er zijn ook enkele onderdelen waar men voor kan zakken en toch een voldoende kan krijgen als eindoordeel.

De data betreffende de proeven die beschikbaar waren bestonden per aspirant-agent uit zo’n 50 tot 70 proeven per instroomniveau (2 t/m 4).

Om de data te reduceren tot werkbare proporties, is besloten twee maten te construeren: i) de proportie behaalde onderdelen van proeven van bekwaamheid in de eerste poging, en ii) de proportie behaalde proeven bekwaamheid in de eerste poging (eindscore proeve van bekwaamheid).

De redenering achter de keuze voor deze maten was dat het succes op de proeven tijdens de eerste poging een indicatie zou moeten zijn voor de kwaliteit van de aspirant-agent.

De genoemde maten worden verondersteld een betere indicatie te zijn dan of een aspirant-agent uiteindelijk, na een aantal herkansingen, een proeve haalt of niet. In dat laatste geval zou bijvoorbeeld het onderscheid wegvallen tussen een aspirant-agent die de meeste proeven direct haalt en een aspirant-agent die bij veel proeven een of meerdere herkansingen nodig heeft.

Om het idee te toetsen dat de bovengenoemde proportiematen inderdaad een indicatie zijn voor de kwaliteit van de aspirant-agent is een aantal analyses uitgevoerd die gedetailleerd zijn beschreven in Bijlage 2. De belangrijkste uitkomsten van deze analyses zijn:

- 1) De meeste aspirant-agenten halen veel van de proeven van bekwaamheid tijdens de eerste poging (gemiddeld tussen de 87 en 95%).
- 2) Er is een significant verschil in de proportie behaalde toetsen tussen degenen die vroegtijdig met de opleiding stoppen (zonder diploma) en degenen die de opleiding afronden met een diploma of in 2012 nog bezig waren. Deze verschillen, hoewel significant, waren klein: De proportie behaalde proeven van bekwaamheid (eerste poging) was niet de belangrijkste factor om te stoppen met de opleiding.

De algemene conclusie die we uit de analyses kunnen trekken is als volgt:

Conclusie 7: Empirisch bewijs ondersteunt het idee dat de proportie behaalde proeven van bekwaamheid (eerste poging) een indicatie is voor de kwaliteit van de aspirant-agent. Specifiek: degenen die vroegtijdig uit zijn gevallen behaalden een significant lagere proportie proeven tijdens de opleiding. De proportie behaalde proeven is overigens niet de belangrijkste factor om vroegtijdig te stoppen. Blijkbaar spelen er andere (vooralsnog) onbekende factoren een grotere rol.

4.3. Belangrijke opmerkingen over het gebruik en de beperkingen van de gehanteerde criteria

In het huidige project is gekozen voor opleidingsstatus en opleidingsprestatie als criteria voor onderzoek naar de voorspellende validiteit van de selectieonderdelen. Alhoewel deze criteria enige validiteit hebben (zie 4.2), moet vermeld worden dat ze ook een aantal belangrijke beperkingen hebben.

Ten eerste is de informatie over proeven van bekwaamheid waarschijnlijk weinig onderscheidend wat betreft de prestatie van de aspirant-agenten. De meeste aspirant-agenten behaalden de proeven tijdens hun eerste poging. Bovendien is de informatie over slagen dichotoom, dat wil zeggen in termen van het wel of niet behalen van een proeve. Tot slot, zal een aspirant-agent pas de proeven van bekwaamheid uitvoeren wanneer hij/zij in overleg met zijn/haar begeleider meent er klaar voor te zijn. Kortom, de gehanteerde maat voor opleidingsprestatie kan maar beperkt informatie opleveren over de kwaliteit van een aspirant-agent.

Ten tweede is niet bekend in hoeverre de criteria uit de opleiding in verhouding staan tot het daadwerkelijk functioneren als agent in de praktijk. Dit is een belangrijke beperking. Uit de literatuur is bekend dat er een discrepantie is tussen hoe iemand functioneert tijdens een opleiding en tijdens de daadwerkelijke baanuitvoering (Guion, 2011).

Ondanks bovenstaande beperkingen is er gekozen voor de huidige criteria omdat:

1) Er geen andere criteria beschikbaar zijn. Zoals we in het rapport meerdere malen zullen opmerken is dit een belangrijke observatie die samengaat met de aanbevelingen om in de toekomst meer en betere criteriuminformatie te verzamelen. Dit is noodzakelijk omdat er anders geen empirisch gefundeerde uitspraken gedaan kunnen worden over de voorspellende kracht van de selectieprocedure voor de werkprestatie van agenten.

2) Soortgelijke criteria zijn gebruikt in eerder onderzoek naar de selectieprocedure bij de politie. Zowel door de Politieacademie als door externe onderzoekers is informatie over de opleidingsstatus en over de proeven van bekwaamheid gebruikt om de waarde van selectieonderdelen te toetsen. Men kan hierbij denken aan het onderzoek zoals dat is beschreven in het rapport 'Selectie onder Loep' van de Politieacademie, en de wetenschappelijke gepubliceerde artikelen van De Meijer et al. (2008).

4.4 Correcties voor beperking van variantie in scores (Range Restriction)

Wanneer men in een selectiecontext de correlaties berekent tussen een voorspeller van opleiding- of werkprestatie en een criteriummaat dan is er bijna altijd sprake van een onderschatting van de correlatie. De reden hiervoor is dat in een selectieprocedure kandidaten die laag scoren over het algemeen niet worden aangenomen. Kandidaten die wel worden aangenomen vormen daardoor een selecte groep die de hoogste scores hebben behaald op de tests. Het duidelijkste voorbeeld hierbij is intelligentie. We weten dat in de gehele bevolking, het gemiddelde van intelligentie is vastgesteld op 100 en de standaard deviatie op 15. Stel nu dat voor een bepaalde baan alleen kandidaten worden aangenomen die minstens 120 scoren op een intelligentietest. In dat geval zal het gemiddelde van de groep kandidaten die is aangenomen hoger dan 120 liggen (bijvoorbeeld 125 of 130) en de spreiding in scores tussen de

aangenomen kandidaten zal kleiner zijn dan in de populatie (bijvoorbeeld een standaard deviatie van 10). Door de kleinere spreiding in de groep van aangenomen kandidaten is het verband tussen intelligentie en werkprestatie lager dan in de totale groep van kandidaten die zich heeft aangemeld. Het gevolg daarvan is een onderschatting van het verband tussen werkprestatie en intelligentie en daarmee ook een onderschatting van de rol van de selectie.

In de literatuur is er een mogelijkheid bekend om in de analyses voor bovengenoemde effecten te corrigeren. Dit wordt vaak correctie voor Range restriction genoemd. Range restriction is een Engelse term die aangeeft in hoeverre op een specifieke voorspeller de variantie van de aangenomen kandidaten afwijkt van de variantie van het totaal aantal kandidaten dat zich voor de baan had aangemeld.

Voor zover mogelijk zijn in de analyses ook de relaties tussen voorspellers en criteriummaten aangegeven nadat is gecorrigeerd voor range restriction. De mate van range restriction is berekend met behulp van de formule van Hunter en Schmidt (2003). Deze is gebaseerd op de verhouding tussen de standaard deviatie in de totale groep kandidaten met de standaard deviatie in de groep kandidaten die is aangenomen. Deze formule luidt:

$$\rho = U * r / \sqrt{(U^2 - 1) r^2 + 1}$$

Waarbij in het huidige project:

ρ = de correlatie, gecorrigeerd voor range restriction

r = geobserveerde correlatie tussen voorspeller en criteriummaat

$u = SD_{\text{aangenomen kandidaten}} / SD_{\text{totale groep kandidaten}}$

$U = 1/u$

In de analyses in de volgende hoofdstukken hebben we zowel de geobserveerde relaties tussen voorspellers en criteria weergegeven en, wanneer mogelijk, de relaties gecorrigeerd voor range restriction.

Selectieonderdeel A:
Cognitieve capaciteitentest, fysiek-motorisch
onderzoek en de taaltoets

Hoofdstuk 5

Cognitieve capaciteit

5. Cognitieve capaciteit

5.1 Achtergrond cognitieve capaciteit als voorspeller van werkprestatie

Van alle individuele verschillen die bij personeelsselectie worden gebruikt, is cognitieve capaciteit over het algemeen de beste voorspeller van werkprestatie (Schmidt & Hunter, 2002). Cognitieve capaciteit kan worden gedefinieerd als het vermogen om abstract te denken en het vermogen om nieuwe problemen op te lossen. Cognitieve capaciteit beïnvloedt de mate waarin men nieuwe informatie op kan nemen. Dit vermogen hangt daarom sterk samen met leervermogen (Jensen, 1998). Mensen kunnen verschillen in hun *profiel* van cognitieve capaciteiten. Dat wil zeggen, sommigen zijn beter in het verwerken van verbale informatie en verbaal redeneren, terwijl anderen juist beter zijn in ruimtelijk inzicht of in rekenen. Desalniettemin is er consensus in de wetenschap dat er zoiets bestaat als algemene cognitieve capaciteit ongeacht iemands specifieke profiel. Deze algemene capaciteit wordt ook aangeduid als algemene intelligentie en komt meestal tot uiting in een IQ score.

Meer dan honderd jaar onderzoek en verschillende meta-analyses hebben inmiddels bevestigd dat algemene cognitieve capaciteit een goede voorspeller is voor beroeps- maar ook trainingsprestaties. Standaardwerken op dit gebied zijn de meta-analyses van Schmidt en Hunter (1998; 2004). Deze onderzoekers hebben verschillende beroepsgroepen bekeken en vonden dat cognitieve capaciteit een voorspellende waarde heeft die vaak beduidend hoger is dan andere bekende voorspellers, zoals persoonlijkheid en interesses. In Tabel 5.1 is de voorspellende waarde (predictieve validiteit) van cognitieve capaciteit weergegeven ten opzichte van andere voorspellers. Hieruit blijkt dat cognitieve capaciteit samen met de arbeidsproef (de zogeheten work sample- een test waarin een sollicitant een taak moet uitvoeren die lijkt op een taak in de baan) en het gestructureerde interview de hoogste voorspellende waarden laten zien voor werkprestaties.

Een eerdere Amerikaanse meta-analyse van Hirsch, Nortrup en Schmidt (1986) onder wethandhavers rapporteerde dat cognitieve capaciteit een wat lagere voorspellende waarde heeft op de prestatie van politiewerk dan de waarde van .51 die in Tabel 5.1 wordt gegeven. Hirsch et al. vonden namelijk een waarde van .38. Hierbij moet vermeld worden dat het gaat om een waarde die gecorrigeerd is voor diverse methodologische zaken zoals onbetrouwbaarheid van de meetmethoden.

Relevant voor de huidige evaluatie is een uitvoerige meta-analyse van Salgado et al. (2003) die de voorspellende waarde van cognitieve capaciteit op werkprestatie specifiek *in Europa* onderzochten.

Opvallend in deze studie is dat zij rapporteerden dat de voorspellende waarde van cognitieve capaciteit bij de politie relatief laag bleek te zijn ten opzichte van andere beroepen. Ter illustratie, er werden verschillende beroepsgroepen onderscheiden waaronder managers, verkopers, ingenieurs en administratieve medewerkers, maar ook politiemedewerkers. De meta-analytisch (gecorrigeerde) voorspellende waarde van cognitieve capaciteit bij de politie was .24. Dit was de laagste voorspellende waarde onder de beroepsgroepen. In de andere beroepsgroepen had cognitieve capaciteit voorspellende waarden variërend van .45 tot .67.

Deze resultaten zijn weergegeven in Tabel 5.2. Let bij deze tabel wederom op het verschil tussen de geobserveerde waarden en de waarden nadat gecorrigeerd is voor onbetrouwbaarheid en beperking van range van metingen. De geobserveerde voorspellende waarde van cognitieve capaciteit op prestatie bij de politie was .12; Dit is een relatief lage waarde.

Tabel 5.1. Hunter & Schmidt, 1998

Predictive Validity for Overall Job Performance of General Mental Ability (GMA) Scores Combined With a Second Predictor Using (Standardized) Multiple Regression

Personnel measures	Validity (r)	Multiple R	Gain in validity from adding supplement	% increase in validity	Standardized regression weights	
					GMA	Supplement
GMA tests ^a	.51					
Work sample tests ^b	.54	.63	.12	24%	.36	.41
Integrity tests ^c	.41	.65	.14	27%	.51	.41
Conscientiousness tests ^d	.31	.60	.09	18%	.51	.31
Employment interviews (structured) ^e	.51	.63	.12	24%	.39	.39
Employment interviews (unstructured) ^f	.38	.55	.04	8%	.43	.22
Job knowledge tests ^g	.48	.58	.07	14%	.36	.31
Job tryout procedure ^h	.44	.58	.07	14%	.40	.20
Peer ratings ⁱ	.49	.58	.07	14%	.35	.31
T & E behavioral consistency method ^j	.45	.58	.07	14%	.39	.31
Reference checks ^k	.26	.57	.06	12%	.51	.26
Job experience (years) ^l	.18	.54	.03	6%	.51	.18
Biographical data measures ^m	.35	.52	.01	2%	.45	.13
Assessment centers ⁿ	.37	.53	.02	4%	.43	.15
T & E point method ^o	.11	.52	.01	2%	.39	.29
Years of education ^p	.10	.52	.01	2%	.51	.10
Interests ^q	.10	.52	.01	2%	.51	.10
Graphology ^r	.02	.51	.00	0%	.51	.02
Age ^s	-.01	.51	.00	0%	.51	-.01

Bovenstaande bevindingen en conclusies zijn gebaseerd op de wetenschappelijke literatuur waarin steekproeven zijn gebruikt uit veel landen. Echter, een onderzoek uit het proefschrift van De Meijer (2008) geeft informatie over de voorspellende waarde van cognitieve capaciteit bij de Nederlandse politie in de tijd dat er nog gebruik werd gemaakt van de Politie Intelligentie Test (PIT). In het onderzoek van De Meijer et al. zijn criteria gebruikt die zijn verkregen uit de politieopleiding. Het betreft onder andere het al dan niet halen van proeven van bekwaamheid op de Politieacademie. De Meijer et al. vonden hierbij een voorspellende waarde van cognitieve capaciteit die lager was dan die gerapporteerd in Salgado et al. (2003). Specifiek vonden De Meijer et al. een gemiddelde voorspellende waarde van .04 (variërend van -.04 tot .11) bij autochtone aspirant-agenten op de Politieacademie. Bij allochtonen aspirant-agenten was de gemiddelde voorspellende waarde iets hoger dan bij de autochtone groep, maar nog steeds laag, namelijk .14 (variërend van -.05 < tot .28). Het is belangrijk om te vermelden dat het in bovengenoemde waarden uit het onderzoek van De Meijer et al. (2008) eveneens gaat om waarden die reeds gecorrigeerd zijn voor onbetrouwbaarheid. De geobserveerde waarden zullen dus nog een stuk lager liggen. Dit betekent dat intelligentie een lage directe correlatie zou hebben met prestatie van politieagenten op de opleiding.

De bevindingen van De Meijer et al. (2008) zijn opvallend als men meeneemt dat het hier gaat om criteria uit de *politieopleiding*. Gewoonlijk blijkt cognitieve capaciteit een betere voorspeller van trainingsprestatie dan van de uiteindelijke werkprestatie omdat de uiteindelijke

werkprestatie verder in de toekomst ligt en over het algemeen minder cognitief van aard is dan een opleidingsprestatie (Schmidt, 2004).

In het eerder genoemde onderzoek van Hirsch et al. (1986) werd bijvoorbeeld een voorspellende waarde gevonden van .76 van cognitieve capaciteit op trainingsprestatie bij de politie in de Verenigde Staten. In het Europese onderzoek van Salgado et al. (2003) had cognitieve capaciteit een voorspellende waarde van .25 voor opleidingssucces bij de politie.

Tabel 5.2. Salgado et al., 2003. De voorspellende waarden van cognitieve capaciteit in Europees onderzoek door Salgado et al., 2003 voor verschillende beroepen, waaronder opleidingssucces bij de politie

Meta-Analysis of General Mental Ability–Occupation Combinations for Predicting Job Performance Ratings

Source	<i>K</i>	<i>N</i>	<i>r</i>	S_o^2	SD_o	% VE	S_e^2	ρ	SD_ρ	90% CV	L_ρ	NSD	LCV
Driver	5	394	.22	.025	.159	64	.012	.45	.191	.21	.40	.225	.11
Electrician	3	280	.28	.023	.150	65	.009	.54	.171	.32	.45	.215	.17
Information clerk	5	890	.31	.006	.081	100	.005	.61	.000	.61	.56	.175	.33
Engineer	9	837	.23	.014	.118	100	.010	.63	.000	.63	.57	.181	.34
Manager	6	783	.25	.039	.197	40	.007	.67	.407	.15	.61	.433	.06
Police	5	619	.12	.015	.121	31	.007	.24	.151	.05	.22	.159	.01
Sales	5	394	.34	.018	.133	100	.010	.66	.000	.66	.60	.189	.36
Skilled worker	7	994	.28	.002	.045	100	.006	.55	.000	.55	.51	.161	.30
Typing	23	1,870	.23	.032	.179	47	.011	.45	.256	.12	.41	.276	.05

Note. *K* = number of studies; *r* = observed mean validity; S_o^2 = observed variance; S_e^2 = sampling error variance; SD_o = observed standard deviation; ρ = operational validity; SD_ρ = standard deviation of the operational validity; % VE = variance accounted for by artifactual errors; L_ρ = lowest (hypothetical) rho; NSD = hypothetical standard deviation of L_ρ ; LCV = lowest (hypothetical) CV.

5.2 Voorlopige conclusie gebaseerd op de literatuur

Een tussentijdse conclusie die kan worden getrokken uit de literatuur is dat wanneer het gaat om politiewerk, cognitieve capaciteit net als in andere banen, voorspellende waarde heeft voor latere opleidings- en werkprestaties. Deze waarde is echter lager dan de waarde voor veel andere beroepen. Deze algemene uitkomst bleek bevestigd voor de groep respondenten die ook het onderwerp zijn van het huidige onderzoek, namelijk aspirant-agenten op de Politieacademie in Nederland (De Meijer et al. 2008). Kanttekeningen die hierbij dienen te worden gemaakt zijn dat de criteria in het onderzoek van De Meijer et al. kwamen uit de opleiding. Het is vooralsnog onduidelijk in hoeverre deze opleidingscriteria samenhangen met prestatie in het werkveld. Een tweede kanttekening is dat de criteria die door De Meijer et al. zijn gebruikt dichotoom waren (dat wil zeggen een tweedeling betroffen, namelijk wel/niet gehaald) waardoor de variantie van de scores op het criterium lager is dan wanneer er meer differentiatie in scoring is (e.g., een tienpuntsschaal): Een lagere variantie zorgt ervoor dat de relatie met de voorspeller lager wordt.

Er zijn verschillende verklaringen mogelijk voor de relatief lage voorspellende waarde van cognitieve capaciteit bij de politie. Eén daarvan is het feit dat politiewerk een tamelijk sterke sociale component heeft in verhouding tot de cognitieve component. Het uitvoeren van goed politiewerk vereist vaak het inschatten van een sociale situatie en het adequaat daarop inspringen. Deze mogelijk verklaring wordt elders nog verder uitgewerkt (zie hoofdstuk 14 over emotionele intelligentie).

Ondanks de blijkbaar relatief lage voorspellende waarde van cognitieve capaciteit bij de politie is het belangrijk om te vermelden dat deze voorspeller voldoende ondersteuning in de

wetenschappelijke literatuur heeft om hem te blijven gebruiken in de selectie. Daar komt bij dat de criteria in huidige onderzoeken bij de politie zoveel beperkingen kennen dat er bij gebruik van deze criteria niet eenduidig kan worden geconcludeerd dat cognitieve capaciteit geen goede voorspeller is. Ook is de voorspellende kracht van cognitieve capaciteit vaak nog relatief gunstig ten opzichte van andere voorspellende maten zoals persoonlijkheid (De Meijer et al., 2008). Een ander belangrijk punt is dat de daadwerkelijke voorspellende waarde van cognitieve capaciteit goed moeten worden onderscheiden van effecten die optreden door beperkingen in het meten. Dat wil zeggen, om de waarde van cognitieve capaciteit bij politiewerk goed in te schatten is het essentieel om na te gaan in hoeverre de specifieke test die deze capaciteit meet inderdaad valide is (zie hierover de volgende paragraaf). Verder zouden we ook graag willen weten in hoeverre de gebruikte criteria een valide weergave zijn van de daadwerkelijke kwaliteit van het politiewerk. Onderstaande paragrafen gaan verder op deze punten in door bij de politie gehanteerde meetinstrumenten voor cognitieve capaciteit te bestuderen en eerder onderzoek hiernaar bij de Politieacademie weer te geven.

5.3 Het meten van cognitieve capaciteit bij de politie: De Politie Intelligentie Test (PIT) en de cognitief Adaptieve Intelligentie Test (CAIT)

Voor het meten van cognitieve capaciteit ten behoeve van het selecteren van aspirant-agenten werd tot 2006 gebruik gemaakt van de Politie Intelligentie Test (PIT: Rijkspsychologische dienst, 1975). Deze test bestaat uit 107 items verdeeld over verschillende subtesten zoals verbaal redeneren, ruimtelijk inzicht en rekenvaardigheid. De resultaten uit het onderzoek van De Meijer et al. (2008: zie boven) zijn gebaseerd op de PIT. Het betreffende onderzoek van De Meijer et al. heeft aangetoond dat de betrouwbaarheden van de onderdelen van de PIT voldoende zijn (zie Tabel 5.3). Uit Tabel 5.3. valt tevens af te leiden dat de betrouwbaarheden niet wezenlijk verschillen tussen allochtonen en autochtonen en dus in beide groepen betrouwbaar zijn. Deze achtergrondinformatie is belangrijk om de waarde van de uitkomsten zoals vermeld door De Meijer et al. te interpreteren; namelijk dat de test in ieder geval voldoende betrouwbaar was.

Sinds 2006 werkt de Politieacademie niet meer met de PIT maar met een nieuwe cognitieve capaciteitentest die adaptief meet, namelijk de **CAIT** (CEBIR Adaptieve intelligentietesten; Bakker, 2009).

Adaptief testen betekent dat niet iedereen dezelfde items (i.e., vragen) doorloopt, zoals bij de PIT, maar dat de items die een kandidaat krijgt afhankelijk zijn van zijn of haar prestaties op eerdere items in de test. De CAIT bestaat uit drie verschillende onderdelen, genaamd Abalet, Anabo, en Cijfcrea.

- Abalet meet het inductief redeneervermogen. De Abalet subtest bestaat uit 158 beschikbare items en de tijdslimiet voor de test is 35 minuten. Volgens de handleiding van de testuitgever is de separatie-index (composiete betrouwbaarheid) van deze test .91.

- Anabo meet inductief en deductief redeneren. Hierbij moet de kandidaat de letterlijke betekenis van het betreffende woord kennen, maar ook inzicht hebben in het achterliggende begrip van dat woord. De test meet verbaal redeneervermogen. De Anabo subtest bestaat uit

116 beschikbare items en de tijdslimiet voor de test is 12 minuten. Volgens de testuitgever is de separatie-index (composiete betrouwbaarheid) van deze test .89.

- Cijfercrea meet het inductief redeneren op basis van configuraties van cijfers. Om de ingewikkelde items op te lossen moet een probleem vanuit verschillende perspectieven worden bekeken. Daarom meet deze test volgens de handleiding dan ook zogeheten lateraal ofwel creatief denken. De subtest Cijfercrea bestaat uit 227 beschikbare items en de tijdslimiet voor de test is 30 minuten. De composiete betrouwbaarheid van de test is .93.

Tabel 5.3 Betrouwbaarheden en intercorrelaties van de onderdelen van de Politie IntelligentieTest (PIT) zoals vermeld in De Meijer et al. 2008

	Cronbach's α		Mean (SD)		Correlation coefficients					
	Ethnic majority	Ethnic minority	Ethnic majority	Ethnic minority	1	2	3	4	5	6
Cognitieve capaciteit: PIT										
1. Verbaal begrip	.62	.68	10.25 (2.25)	9.53 (2.45)	-	.49**	.45**	.48**	.39**	.41**
2. Inductief Redeneren	.70	.75	11.03 (2.30)	10.41 (2.45)	.43**	-	.59**	.37**	.58**	.49**
3. Numeriek Redeneren	.82	.80	8.29 (3.28)	7.39 (3.33)	.43**	.52**	-	.39**	.46**	.38**
4. Woordenschat	.81	.79	9.74 (3.54)	8.75 (3.44)	.40**	.34**	.41**	-	.34**	.37**
5. Ruimtelijk inzicht	.81	.83	2.87 (4.08)	19.68 (4.22)	.34**	.53**	.38**	.28**	-	.39**
6. Figuren ordenen	.64	.75	8.32 (1.60)	8.05 (1.74)	.28**	.36**	.25**	.26**	.29**	-

Boven de diagonaal staan de correlaties in de allochtone groep, onder de diagonaal die van de autochtone groep

Op grond van resultaten van een pilot-testafname van de CAIT (zie tabellen B3.1 tot en met B3.3 in Bijlage 3) zijn er door de Politieacademie zelf normgroepen vastgesteld. Er zijn aparte normgroepen voor mannen en vrouwen. Dit impliceert dat men er vanuit gaat dat mannen en vrouwen zullen verschillen in hun prestaties op de testen. Verder wordt er onderscheid gemaakt naar instroomniveau (2 t/m 6) gebaseerd op vooropleiding. Het gebruik van normscores is overigens een onderdeel dat nog open is voor discussie. We komen hier later op terug (zie hoofdstuk 12 over normscores).

Op grond van de achtergrondinformatie die tot onze beschikking stond viel nog geen eenduidige conclusie te trekken over validiteit van de test in de selectieprocedure bij de politie. Een aantal eerdere onderzoeken naar deze test bij de politie lieten tegenstrijdige resultaten zien. Met name de test-hertest betrouwbaarheid blijft nog onduidelijk. Test-hertest betrouwbaarheid is de mate waarin dezelfde kandidaat, indien deze opnieuw zou worden getest, dezelfde score zou halen. Uit een rapport van de Politieacademie (Bakker, 2009) blijkt

dat wanneer de CAIT wordt afgenomen volgens de standaard procedure, de test-hertest betrouwbaarheid vooralsnog als onvoldoende beschouwd moet worden.

Wanneer de test echter wordt afgenomen volgens een procedure waarbij kandidaten aanvullende oefenitems krijgt, zou de test wel acceptabele betrouwbaarheden bereiken. Anderzijds is er in één van de interviews aangegeven dat de CAIT een goede test is in selectie. De tegenstrijdigheid in informatie geeft aanleiding tot het uitvoerig analyseren van de CAIT. We splitsen de bespreking van de CAIT hieronder op in drie psychometrische kenmerken, namelijk betrouwbaarheid, begripsvaliditeit (meet de test dat wat het dient te meten), en predictieve validiteit (voorspellende kracht).

5.4 Betrouwbaarheid CAIT

In verschillende onderzoeken van de Politieacademie is de betrouwbaarheid van de CAIT onderzocht. Een onderzoek van Westerveld uit 2007 ($N = 19$) leek erop te wijzen dat de testscores konden verschillen bij herhaalde metingen. Vanwege de zeer kleine steekproef dient dit resultaat zeer voorzichtig te worden geïnterpreteerd. Een vervolgstudie (Bakker, 2009) met 409 kandidaten liet zien dat de test-hertest betrouwbaarheid van de CAIT uitkwam op een waarde van .71, wat neerkomt op 50% gedeelde variantie tussen testafnames. De gewenste betrouwbaarheid behoort volgens rapporten van de Politieacademie echter minimaal .80 te zijn.

Uiteraard zijn er wel enige verschuivingen in testscores te verwachten. Maar verschuivingen worden als problematisch beschouwd wanneer er een verschil is van minimaal 4 vignitielen tussen de testscores. Een vignitiel is een indeling waarin scores worden ingedeeld in 20 delen. Een score van vignitiel 1 betekent dat men tot de groep laagste scores behoort, een score van vignitiel 20 betekent dat men tot de hoogst-scorende behoort. Een verschuiving van 4 of meer vignitielen naar boven tijdens de tweede meting van cognitieve capaciteit was te zien in 23% van de gevallen. Een verschuiving van 4 of meer vignitielen naar beneden was te zien in 18% van de gevallen. Op grond van deze bevindingen werd door onderzoekers van de politie academie geconcludeerd dat de test onvoldoende betrouwbaar was (Bakker, 2009).

Naar aanleiding van het onderzoek van Westerveld (2007) is een vervolgonderzoek gedaan (Bakker, 2009) naar de test-hertest betrouwbaarheid van de onderdelen van de CAIT. Daarbij zijn twee verschillende groepen onderzocht. Eén groep heeft de test tweemaal gemaakt volgens de oorspronkelijke procedure waarbij de kandidaten per onderdeel twee oefenopgaven kregen. Een andere groep heeft de CAIT ook tweemaal gemaakt maar zij kregen voor elk onderdeel *acht* aanvullende oefenitems. Het aantal oefenitems was dus verviervoudigd in de nieuwe afname procedure.

Bij de kandidaten die getest werden met de standaard afnameprocedure varieerden de test-hertest betrouwbaarheden van net voldoende tot onvoldoende (zie Tabel 5.4). De betrouwbaarheid van de test als geheel (alle drie subtesten samen genomen) was .70. Door het toevoegen van de extra oefenitems stegen de test-hertest betrouwbaarheden, hoewel die van de Cijfecrea subtest nog onder de maat bleef. De betrouwbaarheid van de test in zijn geheel was .79 in de groep met aanvullende oefenitems.

Tabel 5.4. Betrouwbaarheden CAIT in groepen met of zonder oefenitems (naar onderzoek van Bakker, 2009)

	Groep zonder oefenitems	Groep met oefenitems
Gehele test	.70	.79
Abalet	.60	.64
Anabo	.54	.63
Cijfcrea	.36	.52

Wat betreft de betrouwbaarheid van de CAIT kan vooralsnog geen eenduidige conclusie worden getrokken. Enerzijds geven meerdere onderzoeken aan dat er relatief veel grote verschillen kunnen optreden in de scores van herhaalde metingen, hetgeen een negatieve invloed kan hebben op de waarde van de test in de selectie. Anderzijds geeft de informatie uit de interviews aan dat de betrouwbaarheid van de test goed zou zijn. Het werd niet duidelijk waar laatstgenoemde uitspraken tijdens de interviews op gebaseerd waren. Blijkbaar kan de test-hertest betrouwbaarheid worden verhoogd door het aanbieden van meer (8) oefenitems. Deze procedure roept echter ook een aantal vragen op. Een belangrijke vraag is of de voorspellende waarde niet wordt verminderd door het toedienen van aanvullende oefenitems. Indien oefenitems ervoor zorgen dat er een grotere mate van testbekendheid ontstaat (zogenoemde 'test wiseness') dan betekent dat over het algemeen dat de test het construct cognitieve capaciteit minder goed meet. Het is mogelijk dat in zo'n geval de voorspellende waarde van de test minder wordt (Schmidt, 2004). De enige manier om deze mogelijkheid na te gaan is door een onderzoek als dat van Bakker (2009) te combineren met gegevens over de voorspellende waarde van de test.

Conclusie 8: De status van de betrouwbaarheid van de CAIT is nog deels onbekend. Het toevoegen van oefenitems heeft een positief effect op de betrouwbaarheid, maar het gevolg voor de voorspellende waarde van de test is vooralsnog onbekend. Het is mogelijk dat een toename in betrouwbaarheid de voorspellende waarde van de test verhoogd, maar het is ook mogelijk dat de methode van toevoegen van oefenitems testbekendheid beïnvloed waardoor wellicht de voorspellende waarde juist zou afnemen.

5.5 Begripsvaliditeit CAIT

Een volgend belangrijk onderdeel in de evaluatie van de test betreft de mate waarin deze ook echt het construct cognitieve capaciteit meet.

Er kon geen informatie worden gevonden waarin scores op de subtests of de algemene score van de CAIT zijn vergeleken met scores op andere gevalideerde intelligentietesten, zoals de RAVEN, de General Aptitude Test Battery (GATB), of de Weschler Adult Intelligence Scale (WAIS). Direct contact met de uitgever van de CAIT hierover heeft geen aanvullende informatie opgeleverd. Naar onze mening zou een dergelijke vergelijking nuttig zijn teneinde inzicht te verkrijgen in de mate waarin de CAIT inderdaad die constructen meet die het beoogt te meten. De wetenschappelijke literatuur op dit gebied toont aan dat de scores op verschillende cognitieve testbatterijen relatief sterk met elkaar moeten samenhangen. In het handboek van

Jensen (1998) over intelligentie wordt gesproken over gemiddelde correlaties tussen cognitieve testbatterijen van .70.

Het is opmerkelijk dat de politie voor een belangrijk construct als cognitieve capaciteit een test gebruikt waarvan de betrouwbaarheid en validiteit nog niet goed is vastgesteld.

Wel bleek er onderzoek beschikbaar waarin de scores op de CAIT worden vergeleken met de scores op niet-adaptieve varianten van cognitieve testen van hetzelfde bedrijf dat ook de CAIT uitbrengt (CEBIR). De correlaties die hierbij worden vermeld zijn in sommige gevallen acceptabel maar soms ook wat aan de lage kant. Soms zijn de correlaties tussen tests die een soortgelijk begrip zouden meten (bijvoorbeeld numeriek redeneren) lager dan de correlaties tussen verschillende subtests.

Tabel 5.5. Correlaties van Num met andere tests (CEBIR)

Test	Meetpretentie/Aard	N	r_{xy}
Anabo NL IO	Verbale analogieën	24667	0.58
Abalet IO	Letterpatronen	24668	0.66
OT_Num A	Cijferpatronen	19647	0.58
OT_Num B	Cijferpatronen	5021	0.43

Tabel 5.6. Correlatiedata 2005. Vergelijking Adaptief – Klassiek. Adaptieve tests vetgedrukt (N varieert van 789 tot 1046)

	Abalet	Anabo	Num	Analogieën	Fig. Reeks	Cijf. Reeks
Abalet						
Anabo	.47					
Num	.57	.39				
Analogieën	.40	.47	.39			
Fig. Reeks	.50	.40	.47	.46		
Cijf. Reeks	.49	.40	.59	.42	.51	

5.5.1 Meet de CAIT algemene intelligentie?

Een belangrijke toets met betrekking tot de CAIT is in hoeverre deze inderdaad algemene intelligentie meet. Eén van de manieren om deze vraag gedeeltelijk te beantwoorden is om te toetsen of er een sterke algemene factor aanwezig is in de scores op de subtesten. Zo'n algemene factor in cognitieve testen wordt in de literatuur vaak beschouwd als de aanduiding van algemene intelligentie (Jensen, 1998).

Een Factoranalyse op scores van de groep kandidaten die getest zijn in de periode januari 2009 tot eind december 2012 laat zien dat er inderdaad één sterke algemene factor aanwezig is in de subtesten. Deze factor verklaart 70 % van de variantie in de testcores en de drie subtesten laten elk substantiële ladingen zien op deze factor (variërend van .79 [Anabo] tot .87 [Cijfercrea]).

De vraag is of deze factor ook echt intelligentie aanduidt. Een eerste indicatie dat dit inderdaad het geval is komt uit een onderlinge vergelijking van de instroomniveaus bij de Politieacademie. In tabel 5.7 is te zien dat hoe hoger het instroomniveau is, hoe hoger de kandidaat scoort op de algemene intelligentiefactor in de CAIT. Dit is wat men zou verwachten als de CAIT intelligentie meet.

Tabel 5.7. Intelligentiescore gemeten met de CAIT, opgesplitst naar instroomniveau

instroomniveau	Gemiddelde	SD	N
niveau 2	-.57	.76	224
niveau 3	-.05	.74	950
niveau 4	.22	.73	738
niveau 5	.74	.54	2

SD = standaard deviatie

Conclusie 9: Er is relatief weinig directe informatie aanwezig waaruit de begripsvaliditeit van de CAIT kan worden afgeleid (denk hierbij aan relaties met bekende intelligentietesten). Wel zijn er een aantal indirecte aanwijzingen dat de CAIT aspecten van intelligentie meet. Een voorbeeld zijn de verschillende testcores tussen instroomniveaus in de verwachte richting.

5.6 Predictieve validiteit van intelligentie bij de Nederlandse politie

Een relevant onderdeel in het bepalen van de rol cognitieve capaciteit in de selectie bestaat uit het testen van de voorspellende waarde. Zoals reeds aangegeven zijn de criteria die ter beschikking stonden i) de opleidingsgegevens in termen van het al dan niet gehaald hebben van een diploma en ii) de proportie van (onderdelen van) proeven van bekwaamheid die de kandidaten binnen één keer hebben gehaald. De laatstgenoemde criteria overlappen voor een groot deel met het soort criteriumdata dat gerapporteerd is in het onderzoek van De Meijer et al. (2008).

De Meijer et al. hebben de relatie bekeken tussen cognitieve capaciteit zoals destijds gemeten met de PIT enerzijds, en de opleidingscriteria anderzijds. Een reeds eerder genoemde uitkomst uit hun onderzoek was dat voor autochtonen kandidaten, de voorspellende waarde van cognitieve capaciteit vrijwel gelijk was aan nul (.04; gecorrigeerd voor methodologische artefacten). Voor allochtonen vonden zij een zwakke voorspellende waarde van cognitieve capaciteit voor opleidingssucces (.14; gecorrigeerd).

Voor het huidige rapport is de rol van cognitieve capaciteit in de selectie van aspirant-agenten opnieuw onderzocht. Anders dan in het eerdere onderzoek van De Meijer et al. (2008) is dat sinds 2007 bij de politie gebruikte adaptieve-cognitieve capaciteiten test (zie eerdere beschrijving) is onderzocht, terwijl het onderzoek van Meijer et al. (2008) zich richtte op data die voor 2006 waren verzameld.

De voorspellende waarde van cognitieve capaciteit is op verschillende manieren onderzocht.

Ten eerste door aspirant-agenten die voortijdig zijn gestopt met de opleiding te vergelijken met kandidaten die nog bezig waren op de tijd van de meting. Doordat er in 2012 helaas te weinig kandidaten in de dataset aanwezig zijn die de opleiding reeds met succes hadden afgerond konden laatstgenoemde kandidaten niet worden meegenomen in de statistische analyses.

Ten tweede is onderzocht of cognitieve capaciteit een directe verband had met opleidingsprestatie. Prestatie was gemeten door middel van de proportie behaalde proeven van bekwaamheid, of onderdelen daarvan (zie beschrijving methodesectie). De details van alle analyses zijn beschreven in bijlage 3. De huidige paragrafen geven de algemene bevindingen weer.

Bij het vergelijken van de opleidingsstatus (i.e., wel of niet de studie afgerond, of nog studierend) is geen onderscheid gemaakt tussen allochtone en autochtone aspirant-agenten. Er waren hiervoor helaas te weinig kandidaten per groep. Bij het toetsen van de relatie tussen cognitieve capaciteit en opleidingsprestatie kon dit onderscheid wel worden beschreven.

5.6.1 De relatie tussen CAIT-scores en opleidingsstatus

Een vergelijking tussen kandidaten die de opleiding succesvol afgerond hadden en kandidaten die de opleiding voortijdig hadden verlaten, liet zien dat er geen enkele verschillen aanwezig waren op de algemene CAIT scores. Met andere woorden, er waren geen aanwijzingen dat degenen die uitvielen minder intelligent waren dan degene die op het moment van meting nog met de opleiding bezig waren. Deze conclusie was hetzelfde voor de drie subtestscores. Kortom, cognitieve capaciteit lijkt geen belangrijke factor te zijn in de voorspelling van uitval uit de opleiding.

5.6.2 De relatie tussen CAIT-scores en opleidingsprestatie

De correlaties tussen intelligentie (score op de CAIT) en opleidingsprestatie bleken zeer laag te zijn. Over het geheel correleerde intelligentie slechts $r = .05$ met het behalen van onderdelen van proeven en zelfs nog lager $r = .02$ wanneer het ging om het behalen van de proeven van bekwaamheid zelf.

Bij opsplitsing naar functieniveau trad er niet veel verandering op (zie Bijlage 3 voor de details). Het leek er wel op dat hoe hoger het niveau was, en dus hoe complexer het werk, hoe sterker de voorspellende waarde van intelligentie was. Zo'n bevinding komt overeen met de literatuur (Zie Schmidt & Hunter, 1998, 2002). Ondanks deze niveau-afhankelijke verschillen bleef de absolute voorspellende waarde van intelligentie zeer laag.

Uit eerder onderzoek van De Meijer et al. (2008) is gebleken dat er, betreffende de voorspellende waarde van cognitieve capaciteit, relevante verschillen zijn tussen autochtone en allochtone aspirant-agenten. In overeenstemming met dit eerder onderzoek, geeft een opsplitsing naar instroomniveau en afkomst (allochtoon vs. autochtoon) het meest gedetailleerde beeld. Dit is hieronder weergegeven in Tabel 5.8. Opvallend is dat de hoogste validiteitscoëfficiënten die voor de CAIT werden gevonden negatief zijn. Op niveau 2 bij allochtonen werd bijvoorbeeld een waarde gevonden van $r = -.21$, wat betekent dat hoe hoger

men scoort op algemene intelligentie, hoe lager de kans is dat men een toets de eerste keer haalt. Deze bevinding is niet direct goed te verklaren.

5.6.3 Bespreking resultaten

Een lage voorspellende waarde van cognitieve capaciteit zoals die in dit onderzoek naar voren komt wijkt af van de resultaten die gewoonlijk worden gerapporteerd in de wetenschappelijke literatuur (e.g., Jensen, 1998; Hunter & Schmidt, 2004).

Er zijn drie mogelijke algemene redenen die ten grondslag kunnen liggen aan deze afwijkende bevindingen:

- Ten eerste is het mogelijk dat de betrouwbaarheid en/of validiteit van de cognitieve capaciteitentest zelf, de CAIT, onvoldoende is waardoor de relaties met de criteria onderschat worden. Een punt van aandacht hierbij is dat met de gegevens die tot onze beschikking stonden de betrouwbaarheid en validiteit nog niet goed vast konden worden gesteld (zie ook de eerdere conclusie hierover).

Tabel 5.8. Relatie tussen cognitieve capaciteit en opleidingsprestatie, afzonderlijk voor instroomniveau en auto- en allochtonen.

Instroomniveau		Onderdelen proeven	Eindoordeel proeven
Niveau 2	<u>autochtoon</u>	.09 (.12)	.04 (.05)
	<u>allochtoon</u>	.03 (.04)	-.21** (-.27)
Niveau 3	<u>autochtoon</u>	.10** (.13)	.06 (.08)
	<u>allochtoon</u>	.01 (.01)	-.06 (-.08)
Niveau 4	<u>autochtoon</u>	.08* (.10)	.07 (.09)
	<u>allochtoon</u>	.04 (.05)	-.11* (-.14)

** Correlatie is significant op het 0.01 niveau

* Correlatie is significant op het 0.05 niveau

Waarde tussen haakjes zijn correlaties gecorrigeerd voor range restriction

- Ten tweede is er de mogelijkheid dat de gebruikte criteria niet valide of betrouwbaar zijn. De kwaliteit van de criteria is deels afhankelijk van de manier waarop opleidingssucces is geoperationaliseerd in het onderzoek. We hebben als criterium de proportie van (onderdelen van) proeven gebruikt die een kandidaat haalt bij de eerste poging. Deze proportie zou een indicatie moeten zijn van het studiesucces van een kandidaat tijdens de opleiding. Dit idee werd gedeeltelijk bevestigd door de analyses die aantonen dat kandidaten die uitvallen zonder diploma significant minder proeven van bekwaamheid haalden dan degenen die afstuderen met diploma of dan degenen die nog studeren. Desondanks is er altijd nog de mogelijkheid dat alleen de aller-slechtste kandidaten de proeven niet halen en dat dit vooral voorkomt vanwege houdingsproblemen en niet door gebrek aan capaciteit. Tijdens enkele interviews werd namelijk een paar keer geopperd dat het niet moeilijk is om te slagen voor de proeven en dat de meeste aspirant-agenten in staat zouden moeten zijn om deze te halen.

- Een derde mogelijkheid voor de afwijkende bevindingen is dat er daadwerkelijk geen relatie is tussen capaciteit en gebruikte opleidingscriteria. Ook dit kan weer meerdere oorzaken hebben. Eén daarvan is dat de criteria die waren gekozen toch niet indicatief zijn voor het politiewerk in de praktijk. Een andere mogelijke reden is dat cognitieve capaciteit daadwerkelijk geen sterke invloed heeft op de kwaliteit van het politiewerk. Dit is eerder geopperd in voorgaand onderzoek (De Meijer, 2008; Salgado, 2003). Het idee hierbij is dat politiewerk een tamelijk sterke sociale component zou hebben waardoor het vermogen om met mensen om te gaan en overwicht te vertonen belangrijker zou zijn dan cognitieve capaciteit. Anderzijds is bekend dat cognitieve capaciteit een relatief goed voorspeller is voor *alle* beroepen inclusief banen die minder complex zijn of een sterke sociale component bevatten.

5.7 Geen informatie aanwezig?

Eén van de meest opvallende conclusies die getrokken zou kunnen worden uit de interviews en uit eerder onderzoek, is dat er vrijwel geen recente gegevens beschikbaar zijn over de relatie tussen metingen van cognitieve capaciteit in de selectie en het daadwerkelijk functioneren als agent in het korps. Er valt wat betreft criterium-gerelateerd onderzoek nog veel te doen. Er is wel enig onderzoek uitgevoerd naar de relatie tussen cognitieve capaciteit en prestatie tijdens de politieopleiding. Uit onderzoek van Prins (2007: in Prins & Terlouw, 2007) bleek dat cognitieve capaciteit een significante, doch zeer geringe samenhang vertoont met studievoortgang. Een mogelijke verklaring die hiervoor werd gegeven (Prins & Terlouw, 2007) is dat de kandidaten die door de selectie zijn gekomen, in principe voldoende niveau hebben om de opleiding goed af te ronden. Daardoor ontstaat *restriction of range* en wordt cognitieve capaciteit mogelijk minder belangrijk dan andere factoren. Deze verklaring komt echter niet geheel overeen met gebruikelijke bevindingen in de literatuur die aantonen dat cognitieve capaciteit vaak een relevante voorspeller blijft bij prestatie en nog sterker is wanneer het gaat om training/opleiding. Het idee dat cognitieve capaciteit niet goed voorspelt is ook in strijd met bevindingen uit Exit-onderzoek waaruit blijkt dat een belangrijke reden voor uitval is: "Het niet kunnen voldoen aan de opleidingseisen qua intellectueel niveau".

Er moet hier worden vermeld dat bovenstaande onderzoek (Prins & Terlouw, 2007) is uitgevoerd met gegevens verkregen uit de PIT, de cognitieve capaciteitentest die werd gebruikt voor 2006. Het is daarom niet duidelijk in hoeverre deze bevindingen gelden voor de nieuwe test (de CAIT). Anderzijds mag men aannemen dat verschillende maten van cognitieve capaciteit voldoende overlap hebben waardoor ze hetzelfde construct meten (zie bijvoorbeeld Jensen, 1997).

Conclusie 10: Onderzoek ten behoeve van het huidige rapport laat zien dat cognitieve capaciteit een zeer geringe (tot geen) voorspellende waarde heeft voor uitval uit de opleiding en voor opleidingsprestatie. Het dient echter benadrukt te worden dat het vooralsnog moeilijk is om vast te stellen *waardoor* precies de voorspellende waarde zo laag blijkt te zijn. Oorzaken kunnen liggen bij de beperkingen van de gebruikte cognitieve testen, de kwaliteit van de criteria, de relatie tussen opleidingscriteria en prestatie in het werkveld, of een daadwerkelijke lage relatie tussen cognitieve capaciteit en politiewerk.

Meer concluderende opmerkingen over cognitieve capaciteit in de selectieprocedure worden beschreven in hoofdstuk 11, waarin de onderdelen van de selectie en hun plaats in de gehele procedure nog eens stapsgewijs worden besproken.

Hoofdstuk 6

Fysieke tests

6. Het fysiek-motorisch onderzoek

6.1 Achtergrond fysiek-motorische tests en baanprestatie

Het werk van politiefunctionarissen op niveaus 2 tot en met 4 bevat fysieke componenten. Hierbij valt te denken aan de lichamelijke eisen die worden gesteld aan het aanhouden van verdachten, zoals kunnen hardlopen. Om die reden spelen fysieke en motorische factoren een relevante rol in de selectie van aspirant-agenten. Er zijn in de wetenschappelijke literatuur enkele onderzoeken te vinden over de rol van fysieke-motorische tests in de personeelsselectie (Arvey, Landon, 1992; Campion, 1983). Hoewel dit aspect ten opzichte van andere factoren zoals intelligentie en persoonlijkheid tamelijk onderbelicht blijft.

Wetenschappelijke onderzoek naar de kwaliteit van fysieke en motorische tests in personeelsselectie richt zich met name op de vraag welke grote factoren van fysieke capaciteiten men kan onderscheiden (Hogan, 1991), de vraag hoe groot de voorspellende kracht van fysiek-motorische tests is (Arvey et al., 1992; Campion, 1983), en de vraag hoe men om moet gaan met verschillen tussen groepen, zoals tussen mannen en vrouwen. Bij dit laatste worden vaak ook juridische aspecten betrokken. Als vrouwen bijvoorbeeld minder hoog zouden scoren op fysieke tests dan heeft dit, bij gelijke normering, tot gevolg dat zij minder kans maken om aangenomen te worden. Zulke effecten worden in de internationale literatuur geschaard onder de noemer 'adverse impact' en kunnen leiden tot rechtszaken. Daarom wordt in de literatuur het belang van de validiteit van fysieke tests benadrukt en of de vereisten aan de fysieke kenmerken van de kandidaat noodzakelijke eisen zijn die direct herleidbaar zijn uit de baan (Courtright, 2013; Hogan, 1991).

Een andere relevante kwestie is in hoeverre iemands fysieke conditie stabiel is over de tijd. Met name in een selectiecontext is dit belangrijk omdat men graag zou willen dat aspirant-agenten die relatief fit zijn tijdens hun aanname, dat ook blijven tijdens hun carrière als uitvoerend politiefunctionaris.

6.1.1 Predictieve validiteit van fysieke en motorische tests

Het meeste onderzoek naar de rol van fysieke capaciteiten in personeelsselectie is uitgevoerd in de jaren tachtig en negentig. Een belangwekkend artikel is dat van Hogan (1991). Zij beschrijft hierin een (ongepubliceerde) meta-analyse van Lewis (1989) die de relaties tussen fysieke tests en trainings- en werkprestatie heeft onderzocht. Verschillende testen van uithoudingsvermogen en kracht vertoonden in de meta-analyse validiteiten van .23 tot .30 met betrekking tot trainingscriteria, en .37 tot .82 voor werkprestatie-indicatoren. Deze resultaten laten zien dat fysieke tests relevant zijn voor de baanuitvoering (i.e., de predictieve validiteit). Campion (1983) gaf in dit kader aan dat de relatie tussen fysieke capaciteit en werkprestatie-indicatoren vooral sterk is als de fysieke capaciteiten tijdens de selectie zijn gemeten met behulp van 'job-samples' of assessment oefeningen waarin de fysieke aspecten van het werk goed naar voren komen. In een politiecontext betekent dit dat fysieke test vooral de prestatie zouden voorspellen op taken waarbij agenten zich moeten inspannen (e.g., worsteling met een verdachte of een achtervolging).

Verschillende onderzoekers benadrukken het belang van goede functieanalyses om de fysieke aspecten van een baan expliciet te maken (Campion, 1983; Robertson & Schmidt, 2001). Voor de personeelsselectie bij de Nederlandse politie is er op dit gebied reeds veel werk verricht: Veel van de fysieke eisen van politiewerk zijn in het recente verleden expliciet geanalyseerd en vertaald in een reeks van fysieke vaardigheden die een agent moet bezitten. Deze vaardigheden zijn uitvoerig beschreven in de wettelijke aanstellingseisen van politiepersoneel (Rijksdocument aanstellingseisen politie, 2002) en worden in het huidige rapport besproken in hoofdstuk 3 (criteria van politiewerk).

Wat betreft de factorstructuur van fysieke tests, concludeert Hogan (1991) dat er drie grote onderliggende categorieën te onderscheiden zijn, namelijk: i) spierkracht, ii) (cardiovasculair) uithoudingsvermogen, en iii) kwaliteit van beweging. Deze indeling wordt in de huidige literatuur nog steeds veel gebruikt (e.g., Courtright et al., 2013). De drie grote factoren in fysieke tests worden in sommige gevallen verder uitgewerkt in subcategorieën:

- 1) Spierkracht:** 1) spierspanning, nodig om objecten te verplaatsen (e.g., trekken, duwen, optillen), 2) spierkracht, en 3) spieruithoudingsvermogen.
- 2) (Cardiovasculair) Uithoudingsvermogen:** het leveren van inspanning die de hartslag verhoogd.
- 3) Bewegingskwaliteit:** 1) flexibiliteit, 2) balans (het bewaren van het evenwicht), 3) coördinatie (het uitvoeren van verschillende handelingen op een zinvolle taakgerichte manier).

De fysieke vaardigheden zoals beschreven in de aanstellingseisen voor politiepersoneel zijn goed onder te brengen onder de genoemde categorieën. Het fysiek-motorisch onderzoek (FMO) in de selectieprocedure bevat tests die een beroep doen op verschillende aspecten van kracht (categorie 1 hierboven), uithoudingsvermogen (categorie 2), en coördinatie (categorie 3, subcategorie 3).

De procedures die de politie heeft gevolgd bij het ontwikkelen van de fysieke tests zijn in overeenstemming met aanbevelingen in de literatuur. De praktijkonderdelen waarop de fysieke tests bij de politie zijn gebaseerd (achtervolging te voet; het onder controle brengen van de verdachte; het handmatig verplaatsen van zware objecten) lijken veel op de centrale situaties van politiewerk zoals geschetst in het kernartikel van Arvey et al. (1992). Wat betreft de fysiek-motorische eisen bij de politie kan het volgende worden geconcludeerd:

Conclusie 11: De hoofdcategorieën van fysieke vaardigheden zoals deze worden vermeld in de literatuur zijn goed gerepresenteerd in de huidige selectieprocedure van aspirant-agenten.

Op bovenstaande conclusie volgt de vraag of de huidige fysieke eisen in de selectieprocedure bij de politie *afdoende* goed gepresenteerd zijn? Dat wil zeggen zouden er nog aanvullende eisen, en daaruit volgende tests, moeten worden toegevoegd waarvan de verhouding van selectiekosten versus opbrengsten positief is?

Het beantwoorden van deze vraag zou een aanvullend onderzoek vereisen dat buiten het huidige project valt. Er zijn wel enkele signalen dat de fysieke eisen wellicht aangescherpt

kunnen worden. Zo zijn er bijvoorbeeld signalen dat er onder het politiepersoneel relatief veel individuen zouden zijn met overgewicht (persoonlijke communicatie).

Daarnaast heeft tenminste één van de geïnterviewde experts aangegeven dat er één en ander kan verbeteren in de fysieke eisen in de selectie. De specifieke mening van de geïnterviewde is hier relevant omdat het de persoon betrof die expert is op het gebied van fysieke eisen en opleiding van politiepersoneel. Uit het interview kwam naar voren dat een deel van de huidige aspirant-agenten bijvoorbeeld problemen zou hebben met motoriek en coördinatie of een aantal fysieke kenmerken bezitten die afronding van de Politieacademie bemoeilijken (zie ook beschrijving fysieke kenmerken in hoofdstuk 3). In hoeverre er hier sprake is van incidenten of van een structureel probleem kon niet goed worden vastgesteld. Bovendien kan moeilijk worden bepaald in hoeverre de kosten van het selecteren op aanvullende fysieke eisen opwegen tegen de voordelen van de selectie. Om deze verhouding vast te stellen moet men immers beschikken over duidelijke uitkomstmaten van de selectieprocedure. Een belangrijke conclusie uit het huidige project is echter juist dat zulke uitkomstmaten grotendeels ontbreken. Daarom moeten we dan ook concluderen:

Conclusie 12: Er kan niet goed worden vastgesteld of de kosten van het selecteren op aanvullende fysieke eisen opwegen tegen de voordelen die voortkomen uit zulke veranderingen (e.g., verbeterde werkprestatie van agenten, minder uitval van [aspirant-]agenten om fysieke redenen). Meer onderzoek op dit gebied is noodzakelijk.

Een terugkerend thema in de wetenschappelijke literatuur over fysieke tests in personeelsselectie betreft verschillen tussen groepen, met name geslachtsverschillen. Een recente meta-analyse van Courtright et al. (2013) laat zien dat er grote verschillen zijn tussen mannen en vrouwen in de categorieën spierkracht en cardiovasculair uithoudingsvermogen, namelijk verschillen van ongeveer twee standaard deviaties. Dit zijn aanzienlijke verschillen die zouden betekenen dat slechts 2.5 procent van de vrouwen in staat is om de gemiddelde spierkracht van mannen (gemiddeld wil hier zeggen dat 50% van de mannen minimaal scoort op dit niveau) te evenaren.

In de categorie bewegingskwaliteit (flexibiliteit, balans, coördinatie) werden geen geslachtsverschillen gevonden.

De auteurs stellen overigens ook dat er noch in de wetenschappelijke literatuur noch in de praktijk definitieve oplossingen zijn gevonden voor het omgaan met deze verschillen (Courtright et al., 2013). Zij wijzen op overeenkomsten tussen enerzijds de geslachtsverschillen in fysieke capaciteiten en anderzijds verschillen in de score op cognitieve testen tussen andere groepen (e.g., etnisch verschillende groepen). Op beide gebieden spelen dezelfde problemen en vragen, zoals de mate waarin men moet werken met absolute versus relatieve normen, of de wijze waarop men de adverse impact zo klein mogelijk kan houden. Zoals gezegd, de literatuur geeft nog geen afdoende antwoorden op deze vragen en zodoende blijven de voor- en nadelen van absolute versus relatieve normen vooralsnog onduidelijk.

6.1.2 Stabiliteit fysieke conditie

De stabiliteit van de fysieke conditie van kandidaten is belangrijk omdat men personen zou willen aannemen die fit zijn tijdens selectie maar die ook in een later stadium fit blijven. O*Net

schaart fysieke capaciteiten vaak onder bekwaamheden, daarmee implicerend dat het stabiele kenmerken zijn (Zie hoofdstuk 3). Men kan zich daarbij iets voorstellen: Atletisch potentieel is immers deels genetisch bepaald (Ostrande, Huson, Ostrande, 2009). Anderzijds is voor het uitvoeren van politiewerk atletisch potentieel geen kernfactor. Belangrijker is dat de politiefunctionaris een basisconditie heeft die afdoende is om taken die fysieke inspanning vereisen adequaat uit te voeren. Een voldoende basisconditie is goed door training te bereiken voor de meeste aspirant-agenten, maar blijven deze agenten ook fit tijdens hun carrière als uitvoerend politiefunctionaris?.

In een grootschalig longitudinaal onderzoek dat personen volgde over een periode van 20 jaar (van 13 tot 33 jaar), toonden Kemper, De Vente, Van Mechelen en Twisk (2001) aan dat hun lichamelijke conditie een bepaalde mate van stabiliteit vertoonde. Fysieke tests vertoonden stabiliteitscoëfficiënten variërend van .35 tot .82. De 12-minuten hardlooptest bleek de meeste stabiele test te zijn (.82). Maten van frequentie van fysieke activiteit (e.g., sporten) vertoonden lagere stabiliteitscoëfficiënten namelijk tussen de .14 en .43. Een factor die in zulke uitkomsten waarschijnlijk een rol zal spelen is de motivatie om fit en/of actief te blijven. Zoals in hoofdstuk 3 werd aangegeven zou het nuttig kunnen zijn om in de selectie van aspirant-agenten rekening te houden met dit soort motivaties. Een beperking hierbij is dat de literatuur over dit onderwerp momenteel nog beperkt van omvang is en er, zover bij ons bekend, nog geen eenduidige en algemeen geaccepteerde maten zijn voor het meten van de motivatie om fit te blijven.

Ten behoeve van de selectie van aspirant-agenten zou men daarom eigen onderzoek kunnen doen naar de factoren die een rol spelen bij het op peil houden van fysieke conditie in het verloop van de carrière. Metingen van fysieke kenmerken (e.g., BMI, uitslagen fysieke tests) van huidige agenten in functie zouden gekoppeld kunnen worden aan uitkomsten van hun selectieprocedure. Aangezien de Politieacademie alle selectiegegevens sinds 2006 bewaard moet dit praktisch haalbaar zijn (e.g., op grond van BSN-nummer). Of deze aanpak ook juridisch mogelijk zou zijn zou onderzocht moeten worden. Er zou daarbij ook een test ontwikkeld kunnen worden om 'de motivatie om fit te blijven' te meten. Dit is echter wetenschappelijk braakliggend terrein.

Conclusie 13: Het zou nuttig zijn om naast het meten van de huidige fysieke conditie in de selectieprocedure, ook gebruik te maken van voorspellers van latere fysieke conditie (tijdens het werk als uitvoerend agent). De motivatie om fit/actief te blijven zou een mogelijke voorspeller kunnen zijn. Verder zou het wenselijk zijn om het verband te bepalen tussen de scores op het fysiek-motorisch onderzoek in de selectieprocedure en de fysieke conditie van agenten tijdens hun loopbaan.

6.2 Het huidige fysiek-motorisch onderzoek in de selectie van politiepersoneel

In het Fysiek-Motorisch Onderzoek (FMO) wordt de fysieke toestand van een kandidaat gemeten. Er wordt onder meer gemeten in hoeveel tijd men een standaardparcours kan afleggen. Dit parcours is ontworpen door experts en wordt verondersteld een goede afspiegeling te zijn van praktijksituaties die een beroep doen op de conditie van de agent (zie ook bespreking bij 6.1.1). Daarnaast wordt er onder andere gevraagd naar sportbeoefening en blessures en wordt getest of men hoogtevrees heeft. De specifieke richtlijnen voor de

lichamelijke conditie van agenten zijn beschreven in de wettelijke toelatingseisen voor agenten. Deze richtlijnen zijn vastgesteld door een panel van experts.

In de selectie maakt het FMO deel uit van onderdeel A samen met de capaciteitentesten en de taaltoets (zie Hoofdstuk 7). Kandidaten mogen alleen verder naar de volgende stap in de selectieprocedure als zij de toetsen in onderdeel A, inclusief het FMO, hebben doorstaan.

Op grond van de wettelijke toelatingseisen zijn er norm- en cutoff scores vastgesteld voor het FMO in de selectie. Hierbij houdt men rekening met leeftijd en geslacht. Het uitgangspunt bij de scores is dat ongeveer 80% van de kandidaten de verschillende FMO-tests moet halen. In september 2011 zijn enkele normscores binnen de FMO aangepast. Een voorbeeld daarvan is de sporttest waarbij kandidaten moeten proberen zo snel mogelijk een parcours af te leggen. De normscores vanaf die periode zijn als volgt:

Tabel 6.1. Normscores sporttest vanaf September 2011

Leeftijdsgroep	P80-normtijd mannen	P80-normtijd vrouwen
< 19 jr.	3 min. 07 sec.	3 min. 43 sec.
19-24 jr.	3 min. 08 sec.	3 min. 45 sec.
25-29 jr.	3 min. 10 sec.	3 min. 46 sec.
30-34 jr.	3 min. 13 sec.	3 min. 52 sec.
35-39 jr.	3 min. 19 sec.	4 min. 00 sec.
40-44 jr.	3 min. 24 sec.	4 min. 07 sec.
45-49 jr.	3 min. 30 sec.	4 min. 13 sec.
50-54 jr.	3 min. 38 sec.	4 min. 25 sec.
55-59 jr.	3 min. 53 sec.	4 min. 40 sec.
60-64 jr.	4 min. 00 sec.	4 min. 50 sec.

Aangezien het uitgangspunt bij het opstellen van de normen was dat gemiddeld 80% van de kandidaten zou moeten kunnen slagen, is er in 2012 een evaluatieonderzoek uitgevoerd naar de slagingspercentages. Uit dit onderzoek kwam naar voren dat gemiddeld over alle groepen, 22.4% van de kandidaten de sporttoets niet haalt. Dit komt dus tamelijk goed overeen met de beoogde 80% slagingspercentage. Wel waren er verschillen tussen mannen en vrouwen. Gemiddeld viel 19.7% van de mannen uit op de sporttoets, en 25.9% van de vrouwen (zie ook tabel 8.2). Een opmerkelijke bevinding was dat het uitvalpercentage verschilde per testlocatie. In Amsterdam was het uitvalpercentage veel hoger dan in Apeldoorn, namelijk respectievelijk 26.8% en 15.6% van de mannen, en respectievelijk 36.1% en 20.3% van de vrouwen. Verder onderzoek suggereerde dat dit verschil in uitvalpercentage deels te maken had met verschil in demografische variabelen van de kandidaten in de verschillende locaties. Wanneer men echter corrigeerde voor demografische variabelen bleef er nog steeds een significant verschil in uitvalpercentage tussen Amsterdam en Apeldoorn.

Met als uitgangspunt een 80% slagingspercentage, is het niet verwonderlijk dat men op grond van het onderzoek uit 2012 tot de conclusie kwam dat de normscores voor vrouwen zouden

moeten worden aangepast. Uit onderstaande tabel 6.2 wordt namelijk duidelijk dat in alle leeftijdscategorieën het percentage vrouwen dat uitvalt op de sporttoets groter is dan beoogd.

Tabel 6.2. Slagings- en uitvalspercentages bij sporttoets volgens normen van september 2011

Leeftijdsgroep	Mannen				Vrouwen			
	Af		Door		Af		Door	
	Aantal	%	Aantal	%	Aantal	%	Aantal	%
< 19 jr.	127	19.8%	515	80.2%	77	27.6%	202	72.4%
19-24 jr.	463	18.8%	2000	81.2%	261	24.9%	789	75.1%
25-29 jr.	214	19.9%	859	80.1%	104	22.8%	352	77.2%
30-34 jr.	82	21.0%	308	79.0%	37	28.5%	93	71.5%
35-39 jr.	52	25.4%	153	74.6%	30	34.5%	57	65.5%
40-44 jr.	25	19.1%	106	80.9%	19	36.5%	33	63.5%
45-49 jr.	16	23.2%	53	76.8%	8	32.0%	17	68.0%
50-54 jr.	1	7.7%	12	92.3%	4	57.1%	3	42.9%
Totaal	980	19.7%	4006	80.3%	540	25.9%	1546	74.1%

Het hanteren van een relatieve norm is een manier om er zeker van te zijn dat een specifiek aantal kandidaten aan een bepaalde eis kan voldoen. Een voor de hand liggend gevaar van het hanteren van zulke normen is dat de absolute eis omlaag zou kunnen gaan als de algehele conditie in de populatie achteruit gaat. Stel dat bijvoorbeeld op termijn de bevolking door een veranderde levenswijze gemiddeld minder fit zou worden, dan zou dat betekenen dat de absolute eis voor de fysieke test omlaag gaat (bijvoorbeeld van 3.07 sec voor personen < 19 jaar naar 3.30 voor diezelfde groep). Anderzijds kan men zich afvragen of dit als een probleem kan worden beschouwd aangezien in dat geval ook degenen die bijvoorbeeld achtervolgd moeten worden wellicht ook gemiddeld minder fit zal zijn. Het hanteren van absolute versus relatieve normscores blijft ook in de wetenschappelijke literatuur een voortdurend discussiepunt. Een ander belangrijke punt bij het hanteren van normscores is de verantwoording ervan. Deze kwestie is gerelateerd aan de vraag waar de absolute grens ligt van de capaciteiten die men nodig heeft om goed te kunnen functioneren. Stel bijvoorbeeld dat het afleggen van een parcours in 4 minuten (zie norm vrouwen in de leeftijd 35-39 jaar) een minimumeis is om politiewerk adequaat uit te voeren. In hoeverre is het dan gerechtvaardigd om een mannelijke kandidaat van 23 jaar af te wijzen die aan alle psychologische eisen voldoet maar die niet de gestelde fysieke norm haalt van 3.07 minuten volgens zijn normgroep? Het gaat hier om het spanningsveld tussen enerzijds de belangen van de politieorganisatie (minimum normscores) en anderzijds de belangen van de verschillende kandidaatsgroepen. Deze wegingskwestie heeft eerder met belangen te maken dan met wetenschappelijke argumenten maar kan wel implicaties hebben voor de selectieprocedure (juridisch).

6.2.1 Predictieve validiteit van de uitkomsten van het fysiek-motorisch onderzoek

Veel FMO-onderdelen hebben rechtstreeks betrekking op de lichamelijke conditie van de kandidaten en zijn daarom van een andere aard dan de persoonlijkheids- en capaciteitentesten, het interview en de praktijkopdrachten die psychologische constructen en gedrag meten. Toch is het zinvol om, net als de bij de andere selectieonderdelen, na te gaan wat de predictieve

validiteit is van de FMO-onderdelen. Politiewerk is voor een deel fysiek en indien men niet in staat is om aan de fysieke eisen van politiewerk te voldoen zou men de opleiding niet halen. Vanuit dit oogpunt is getoetst wat de voorspellende waarde is van de onderdelen van het FMO voor de opleidingscriteria. Met name de uitvalpercentages zijn hier relevant. Ook zijn de studieresultaten betrokken in de voorspelling -in termen van de proportie behaalde proeven tijdens de opleiding. Niet alle proeven van bekwaamheid bevatten fysieke aspecten, desalniettemin bevat een proportie van de proeven van bekwaamheid componenten die een beroep doen op fysieke conditie.

De FMO meet verschillende aspecten van fysieke conditie. De volgende onderdelen zijn meegenomen in het onderzoek naar de voorspellende waarde:

- **Parcoursprestatie:** De snelheid (in seconden) waarmee men een standaard parcours kan afleggen.
- **Intensiteit sportbeoefening:** In aantal keren per week
- **Blessures:** Tijd dat men geblesseerd was in categorieën variërend van niet geblesseerd tot meer dan drie weken achtereenvolgens geblesseerd.
- **Conditie:** VO2 max (i.e., de maximale longinhoud van de kandidaten)
- **Vetpercentage**

De details van de analyses met betrekking tot FMO-uitkomsten zijn beschreven in Bijlage 4. Een belangrijke bevinding is dat kandidaten die sneller waren op het parcours tijdens de selectie, significant minder vaak vroegtijdig de opleiding verlaten. De uitslag op het parcours bleek dus een relevante voorspeller voor het afmaken van de opleiding.

De mate van fysieke activiteit (i.e. sporten) bleek geen voorspeller van het al dan niet afmaken van de opleiding. De uitkomsten met betrekking tot blessures waren complex. Aspirant-agenten die de opleiding met succes afronden hadden de minste blessuretijd. Aspirant-agenten die waren uitgevallen hadden meer blessuretijd dan degenen die een diploma hadden behaald, maar ze hadden echter minder blessuretijd dan degenen die op moment van de analyses nog in opleiding zaten.

Conditie en vetpercentages lieten enigszins onverwachte resultaten zien. Aspirant-agenten die vroegtijdig (zonder diploma) de opleiding hadden verlaten hadden namelijk een *lager* vetpercentage dan degenen met een diploma of die nog studeerden, en ze hadden ook een *hogere* VO2 max.

De correlatieve analyses lieten zien dat de onderdelen van het FMO geen sterke relatie vertonen met het de totale studieprestatie. Een uitzondering is dat voor zowel mannen als vrouwen met instroomniveau 3, degenen met een hoger vetpercentage vaker proeven behaalden tijdens hun eerste poging. Vrouwen met een betere conditie behaalden ook significant meer proeven. Mannen die sneller waren op het parcours haalden iets meer proeven. De effectgroottes waren echter klein.

Tabel 6.3. Correlaties tussen FMO onderdelen en (onderdelen van) de proeven

	Onderdelen		Proeven	
	man	vrouw	man	vrouw
parcoursprestatie	-.06*	-.03	-.05*	-.07
Intensiteit sportbeoefening	-.01	-.03	-.04	.03
Blessures	.02	.05	.01	-.03
Conditie	.01	.01	.01	.08*
Vetpercentage	.08**	.11**	.04	.08*

Conclusie 14: Enkelen FMO-onderdelen hebben voorspellende waarde op wie al dan niet succesvol de opleiding zal afronden. Met name Parcourstijd voorspelt tot op zekere hoogte iemands studie-uitval. In sommige gevallen is de relatie contra-intuïtief; kandidaten met een hogere VO2 max en een lager vetpercentage hebben minder kans om de opleiding succesvol af te ronden.

In overeenstemming met een recent onderzoek van de Politieacademie (2009) is gevonden dat FMO-uitkomsten de criteria van opleidingsprestatie nauwelijks voorspellen met uitzondering van parcourstijd. De criteria waren echter tamelijk breed gedefinieerd; ze omvatten bijvoorbeeld ook alle cognitief en sociaal georiënteerde taken in de opleiding. Het is dus niet verwonderlijk dat lichamelijk conditie niet sterk samenhangt met deze criteria. Daarnaast is een fundamentele punt dat vooralsnog onduidelijk is in hoeverre de gekozen criteria goede maten zijn voor de prestatie van de aspirant-agent tijdens de opleiding en in hoeverre de criteria het latere werk als agent goed voorspellen.

Hoofdstuk 7

De taaltoets

7. De taaltoets

7.1 Taaltoetsen in de selectieprocedure: Achtergrond

Politiewerk vereist een adequate taalvaardigheid. Een deel van het politiewerk bestaat immers uit schrijftaken zoals het opmaken van een proces-verbaal. Een proces-verbaal is een beschrijving van een delict en de bijbehorende aanklacht die beiden een rol spelen in het rechtsproces. Het is daarom van groot belang dat het proces-verbaal op een correcte en begrijpelijke manier wordt opgemaakt.

Eén van de eerste stappen in de selectieprocedure bij de politie is dan ook een taaltoets. Deze toets speelt een rol in het besluit of kandidaten al dan niet kunnen toetreden tot de volgende stap in de selectieprocedure. In de wetenschappelijke literatuur over personeelsselectie wordt weinig geschreven over de rol van taalvaardigheidstoetsen als selectie-instrument. Wanneer er in de literatuur wordt verwezen naar taalvaardigheid in personeelsselectie is dat vooral in de context van minderheden (e.g., Hough & Oswald, 2000). Het idee hierbij is dat kandidaten uit sommige minderheidsgroepen minder zijn blootgesteld aan de Nederlandse taal, bijvoorbeeld omdat hun ouders thuis geen Nederlands spraken of omdat zij zelf op latere leeftijd naar Nederland zijn gekomen, en dat deze taalachterstand een mogelijk nadeel oplevert tijdens selectieprocedures.

Bij de selectieprocedure van de politie is de taalvaardigheid van minderheden ook een relevant onderwerp. Gerelateerd aan het diversiteitsbeleid bij de politie gaat het om de vraag in hoeverre taal een barrière is voor minderheden. In de wetenschap is de discussie over de rol van taalvaardigheid bij minderheden nog open en de visies van experts lopen sterk uiteen. Een voor de hand liggende visie is dat een verminderde taalvaardigheid de selectie-uitkomsten van minderheden ten onrechte benadeelt. Dat wil zeggen, minderheden zouden minder goed scoren op cognitieve en persoonlijkheidstesten, en het minder goed doen tijdens interviews en assessments *omdat* ze de taal niet goed beheersen en niet omdat ze noodzakelijk lager op de gemeten constructen zouden scoren (De Meijer et al., 2008). Men spreekt in dit geval van een vertekening of bias in de toetsing ten nadele van minderheden.

Er zijn daarentegen ook artikelen waarin wordt gesuggereerd dat in veel gevallen taalvaardigheid niet een beperkende factor is (Te Nijenhuis & Van der Flier, 1999). In laatstgenoemde visie wordt er vanuit gegaan dat de testresultaten een correcte weergaven zijn van verschillen tussen groepen in plaats van dat ze een vertekend beeld geven.

De argumenten en onderzoeksmethoden in de discussie over dit onderwerp zijn tamelijk complex. Zo hoeft een verminderde taalvaardigheid niet alleen een *oorzaak* te zijn van slechtere scores op een test maar kan het tevens een *indicator* zijn van capaciteiten. Het is bijvoorbeeld bekend dat intelligentie en woordenschat relatief sterk met elkaar samenhangen (Jensen, 1998).

Een ander vaak gebruikt argument in deze wetenschappelijke discussie is dat de taalvaardigheid zoals gemeten tijdens selectieprocedures, daadwerkelijk een voorspellende waarde heeft voor werkprestatie.

De kern van bovenstaande is dat men er niet zonder meer van uit kan gaan dat taalvaardigheid een onterechte oorzaak is van verschillen in alle testcores tussen groepen. Dit was echter wel de boodschap die, bijvoorbeeld de media en wellicht daarmee ook de politiek, oppikte uit eerder onderzoek van De Meijer et al.(2008) naar selectie bij de Nederlandse politie. Zo beginnen verschillende nieuwsberichten uit die periode met uitspraken als ‘taalvaardigheid beïnvloed personeelsselectie’ (<http://www.managersonline.nl/nieuws/7564/taalvaardigheid-benvloedt-personeelsselectie.html>), ‘selectieprocedures politie verkleinen kansen allochtone kandidaten’ (<http://www.profnews.nl/896069/selectieprocedures-politie-verkleinen-kans-allochtone-kandidaten>), en ‘taal is handicap tijdens sollicitatie’ (<http://www.kennislink.nl/publicaties/taal-is-handicap-tijdens-sollicitatie>). We zijn van mening dat zulke uitspraken te voorbarig zijn en in sommige gevallen zelfs kunnen leiden tot maatregelen die de kwaliteit van de selectieprocedure kunnen verminderen (bijvoorbeeld wanneer men normscores voor cognitieve capaciteitentests zou corrigeren voor taalvaardigheidsscores).

Los van het ingewikkelde vraagstuk omtrent minderheden en de rol van taalvaardigheid in de selectie is de vraag van belang aan welke taalvaardigheidseisen aspirant-agenten zouden moeten voldoen en of de gebruikte toetsen een betrouwbaar en valide beeld geven van hun taalvaardigheid. Daarnaast is het de vraag of de procedure voor het vaststellen van taalvaardigheid in de selectie geoptimaliseerd zou kunnen worden. Deze vragen komen hieronder aan bod.

7.2 Taaltoetsen in de selectieprocedure van de Nederlandse politie

De taaltoets in de selectieprocedure bij de Nederlandse politie meet zowel passieve (lezen) als actieve (schrijven) taalvaardigheid. Tot 2010 werd er bij de selectie van nieuwe agenten gebruik gemaakt van de Digibo, een taaltoets die zowel schrijf- als luistervaardigheden en woordenschat meet. Vanaf 2010 is de Digibo echter vervangen door de TOA. Eén van de voordelen van de nieuwe taaltoets is dat verschillende onderdelen expliciet en afzonderlijk worden getoetst. Op die manier kan er een gedifferentieerder beeld worden gevormd van de taalvaardigheden van een kandidaat. Een ander groot voordeel van de TOA is dat de uitslagen van de toets kunnen worden geïnterpreteerd aan de hand van algemene taalvaardigheidsniveaus zoals geformuleerd in het Common European Framework for the References of Languages (CEF). De categorieën die in de CEF worden onderscheiden zijn hieronder weergegeven in tabel 7.1 en gaan uit van basisniveau, professioneel niveau, en specialistisch niveau.

Met taalkennis op het basisniveau (A1 en A2) wordt iemand geacht zich goed te kunnen redden in alledaagse (sociale) situaties, zoals het lezen van een krant. Het volgende niveau betreft taalkennis waarmee iemand goed kan functioneren in een beroepssituatie of in een onderwijssituatie (e.g., het volgen van een cursus). Dit betekent dat de persoon in staat moet zijn om complexere teksten te kunnen begrijpen zoals boeken die worden gebruikt in het onderwijs. Op het specialistische niveau moet iemand in staat zijn om teksten te kunnen begrijpen en gesprekken te kunnen voeren binnen een specialistisch vakgebied (e.g., op universiteiten).

Tabel 7.1. Taal categorieën zoals geformuleerd in het Common European Framework for the references of languages (CEF)

Sociaal redzaam		Educatief/Professioneel		Specialistisch	
Basic		Independent		Proficient	
A1	A2	B1	B2	C1	C2

In de selectieprocedure bij de politie wordt de taaltoets relatief vroeg in de procedure afgenomen als component van onderdeel A, naast het capaciteitenonderzoek en het fysiek-motorisch onderzoek. Het huidige project had geen beschikking over de uitkomsten van de taaltoetsen in de selectie. Bij de bespreking van de waarde van taaltoetsen kon echter wel gebruik worden gemaakt van onderzoeksrapporten van de Politieacademie zelf. Deze rapporten vermelden onder andere de uitvalspercentages afhankelijk van scores op de taaltoetsen.

7.2.1 Norm- en cutoffscores

Om de taaltoets adequaat te kunnen gebruiken in de selectieprocedure is het essentieel om te weten welke taalniveaus nodig zijn om goed te functioneren als agent. Hierbij kan men onderscheid maken tussen de agenten op niveau 2, 3 en 4. Een beeld van de benodigde taalvaardigheid van agenten kan men afleiden uit algemene functiebeschrijvingen van politiewerk zoals op O*Net. Wat betreft de Nederlandse situatie is er een specifieke evaluatie uitgevoerd van de benodigde taalkennis voor agenten. Deze evaluatie is beschreven in het rapport 'Lezen en schrijven bij de politie' van het adviesbureau ICE (Pit, 2010). Dit rapport beschrijft een systematische analyse van de rol van taalvaardigheid in politiewerk. Het betreffende rapport heeft geleid tot algemene richtlijnen over de benodigde taalvaardigheid van agenten volgens de indeling van de CEF. Deze richtlijnen luiden:

Tabel 7.2. Aanbeveling van de benodigde taalkennis van agenten door ICE

Niveau	Lezen	Schrijven
Niveau 2 Assistent politiemedewerker	B1	B1
Niveau 3 Politiemedewerker	B2	B2
Niveau 4 Allround politiemedewerker	C1	B2

Een aanname van ICE is dat wanneer een kandidaat het betreffende taalniveau bezit, deze in staat zou moeten zijn om de politieopleiding met succes te kunnen afronden. In het huidige project is dit rapport kritisch bekeken. Op grond hiervan kan worden geconcludeerd dat de analyse van taalkennis grondig is uitgevoerd. Bovendien lijken de aanbevelingen die worden gedaan redelijk en in overeenstemming met de beoogde taken van politiewerk.

7.2.2 Beoordeling en uitval op taaltoets

De nieuw ingevoerde taaltoets bestaat uit drie delen, namelijk een voorschatter, een leestoets en een schrijftoets. De voorschatter is bedoeld om te bepalen op welk niveau de lees- en

schrijftoetsen moeten worden aangeboden. De lees- en schrijftoetsen kunnen worden aangeboden op drie niveaus, namelijk A2-B1, B1-B2, en B2-C1. Het niveau van de schrijftoets dat wordt aangeboden aan een kandidaat is altijd afhankelijk van het behaalde resultaat op de leestoets.

Op de voorschatter wordt slechts een zeer klein deel van de kandidaten die zich hebben aangemeld afgewezen (0.4%). Het grootste deel van de kandidaten krijgt naar aanleiding van de resultaten op de voorschatter de leestoets aangeboden op het B1-B2 niveau (62%). Van de kandidaten die door zijn na de voorschatter blijkt 8 tot 10% achteraf het vereiste taalniveau alsnog niet te halen. Deze kandidaten vallen daarom af in selectieonderdeel A. Van degenen die de leestoets op niveau B1-B2 krijgen en de schrijftoets op niveau A2-B1 valt zo'n 8% uit. Van degenen die zowel de lees- als schrijftoets op niveau A2-B1 hebben gemaakt valt ongeveer 50% uit. Het rapport van de Politieacademie geeft aan dat wanneer lezen en schrijven beiden op B1-B2 niveau zijn gemaakt, het onmogelijk is om in deel A afgewezen te worden op de taaltoets. Deze procedure lijkt enigszins ongebruikelijk: als kandidaten die deze test maken niet meer kunnen worden afgewezen lijkt het overbodig om ze nog een taaltoets te laten maken in deel A van de selectieprocedure.

Conclusie 15: Voor zover met de beschikbare informatie kan worden ingeschat, lijken de gevolgde procedures en normen met betrekking tot de taaltoets recent grondig onderbouwd in de selectieprocedure bij de politie. De vervanging van de oude taaltoets (Digibo) door de nieuwe TOA lijkt een goede keuze omdat de TOA duidelijk aansluit bij Europese normen.

Toch zijn er wel een aantal aspecten die ons opvielen en waar, wat ons betreft, vragen over blijven bestaan. Deze bespreken we hieronder.

7.2.3 Scoring taaltoets

Eén van die aspecten die bij ons vragen oproep zijn de resultaten van een onderzoek naar de beoordeling van taaltoetsen (Politieacademie, 2011). Dit onderzoek liet zien dat er relevante verschillen kunnen bestaan tussen beoordelaars. Zo werden taaltoetsbeoordelingen door medewerkers van de afdeling Werving en Selectie bij de politie vergeleken met beoordelingen door externen (Medewerkers van adviesbureau ICE). Hieruit bleek dat er tussen de groepen belangrijke verschillen waren in taaltoetsbeoordelingen. Deze verschillen hadden effect op het wel of niet aannemen van kandidaten. De uitkomsten van dit onderzoek benadrukken de subjectiviteit van de beoordelingen van de taaltoetsen. Ons voorstel is dan ook om aandacht te besteden aan het zoveel mogelijk standaardiseren van de beoordeling op de taaltoets.

Conclusie 16: De beoordeling van de taaltoets is ten dele subjectief. Hierdoor kunnen er grote verschillen in scoring ontstaan tussen (groepen van) beoordelaars. Besteedt daarom aanvullende aandacht aan het standaardiseren van de beoordeling op de taaltoetsen. Te denken valt aan verbeterde nakijksleutels, bijvoorbeeld volgens de methode van Angoff. Deze laatstgenoemde methode komt oorspronkelijk uit de onderwijspsychologie maar wordt ook vaak gebruikt bij het bepalen van normscores in selectieprocedures. De methode is gebaseerd op de meningen van verschillende experts (Subject Matter Experts) die inschatten welk

percentage van minimaal gekwalificeerde kandidaten een bepaald item of bepaalde taak goed kunnen volbrengen.

7.4 Verantwoordelijkheid taaltoetsing

Bij de taaltoetsing kan de vraag gesteld worden bij wie de primaire verantwoordelijkheid ligt om aan te tonen dat een kandidaat over voldoende taalbeheersing beschikt. In de huidige procedure lijkt deze verantwoordelijkheid vooral bij de politieorganisatie te liggen omdat de toets een onderdeel vormt van de selectieprocedure. Het valt te overwegen om deze verantwoordelijkheid te verschuiven omdat verwacht mag worden dat taalkennis voldoende aanwezig is als een kandidaat een bepaalde vorm van middelbaar onderwijs succesvol heeft afgerond. In deze lijn past het wellicht om de 'bewijslast' bij de kandidaat te leggen. Een mogelijkheid hiertoe is het uitbesteden van de taaltoetsen aan een externe organisatie die voor een relatief klein bedrag een korte taaltoets afneemt zodat een goede en snelle indicatie verkregen wordt van iemands taalniveau.

Als een kandidaat zou willen solliciteren bij de politie zou deze zich eerst moeten melden bij de betreffende organisatie om tegen betaling een taalschatter te maken. Mits de kandidaat hiervoor slaagt, kan hij of zij zich met bewijs daarvan verder aanmelden bij de politie en toetreden tot deel A, dat in dit geval zou bestaan uit de capaciteitentoets en het FMO. De norm van de taalschatter zou dusdanig ingesteld moeten zijn dat men er bij het slagen op de toets tamelijk zeker van kan zijn dat de kandidaat over het vereiste taalniveau beschikt. Het verschil met de huidige procedure is dat de kandidaat geen taaltoets meer hoeft te maken nadat deze de taalschatter met goed resultaat heeft gemaakt.

Het zou verkieselijk zijn om de grenswaarde van de 'voorschatter nieuwe stijl' hoger af te stellen dan de norm van de huidige voorschatter. In de huidige procedure valt nog zo'n 8% van de kandidaten die de voorschatter hebben behaald, af op taalvaardigheid. Als men de norm van de beoogde externe taalschatter enigszins hoger zou instellen dan kan men dit wellicht reduceren tot zo'n 4%.

Indien men zakt voor de initiële taalschatter zou men de mogelijkheid moeten hebben om aan te kunnen tonen dat men wel degelijk over het gewenste taalniveau beschikt door aanvullende en meer uitgebreide taaltoetsen te maken. Deze taaltoetsen worden dan wel tegen een meerprijs aangeboden.

In bovenstaande opzet is de verantwoordelijkheid voor het aantonen van het eigen taalniveau verschoven naar de kandidaat. Een nadeel is dat deze procedure voor sommige kandidaten een extra drempel kan zijn om zich aan te melden bij de politie. Anderzijds, is te verwachten dat de kandidaten die zich dan aanmelden gemotiveerd zijn omdat ze deze extra stap hebben doorlopen en zelf een bedrag hebben geïnvesteerd. Een ander voordeel is dat de kosten voor de voorschatter en de taaltoets daarmee verdwijnen voor de politie.

Conclusie 17: Het zou nuttig zijn om te overwegen een deel van de verantwoordelijkheid voor het aantonen van taalvaardigheid bij de kandidaat te leggen. Het model zoals hierboven geschetst vormt een mogelijkheid dit te implementeren.

Selectieonderdeel B:
De persoonlijkheidstest,
het selectiegesprek en de praktijkopdracht,

Hoofdstuk 8

Persoonlijkheid

8. Persoonlijkheid

8.1 Achtergrond persoonlijkheid en prestatie

In de wetenschappelijke literatuur is de relatie tussen persoonlijkheid en prestatie altijd een onderwerp van discussie geweest (Ghiselli, 1973; Guion & Gottier, 1965). Enkele decennia geleden verliep het onderzoek op dit gebied moeizaam omdat er geen consensus bestond over de basiskenmerken van persoonlijkheid. Verschillende onderzoekers gebruikten verschillende persoonlijkheidsmodellen en uiteenlopende maten. Onderzoeken waren dus nauwelijks vergelijkbaar. Daar kwam verandering in toen er zogeheten structurele modellen van persoonlijkheid werden ontwikkeld. Alhoewel er tot op heden nog verschillende modellen van persoonlijkheid bestaan is de algemene wildgroei aan manieren om naar persoonlijkheid te kijken teruggebracht tot een aantal modellen waar enige consensus bestaat bij wetenschappers. Voorbeelden van zulke modellen zijn Cattells 16-factor model, Eysencks drie-factor model of het HEXACO model dat zes persoonlijkheidsfactoren onderscheidt. Momenteel is het bekendste en wereldwijd meest gebruikte model het zogeheten vijf factoren model, of kortweg de Big Five. Dit model veronderstelt dat er vijf basisdimensies zijn die het grootste deel van individuele verschillen in persoonlijkheid kunnen beschrijven. De vijf dimensies in dit model zijn:

- 1) Openheid voor nieuwe ervaringen: De mate waarin men op zoek gaat naar nieuwe ideeën en ervaringen en openstaat voor andere meningen en of levenswijzen.
- 2) Consciëntieusheid: De mate waarin men betrouwbaar en zorgvuldig is en geneigd is regels te volgen.
- 3) Extraversie: Het gemak waarmee men met andere mensen kan omgaan, nieuwe vrienden maakt en geneigd is om de leiding te nemen.
- 4) Altruïsme /Vriendelijkheid: Een dimensie die representeert in hoeverre men zich vriendelijk opstelt tegenover anderen.
- 5) Neuroticisme: De neiging om negatieve gevoelens te ervaren en heftig te reageren op stress. De tegenhanger van neuroticisme is emotionele stabiliteit.

De dimensies in het Big Five model kan men gezamenlijk afkorten met het acroniem OCEAN. Belangrijke aannames in het Big Five model zijn dat de OCEAN dimensies de breedste, meest algemeen zinvolle beschrijvingen zijn van persoonlijkheid en dat ze conceptueel onderscheidbaar en onderling onafhankelijke dimensies vormen (Goldberg, 1983).

De introductie van het Big Five-model heeft ertoe geleid dat het mogelijk werd om op een meer systematische manier onderzoek te doen naar de relatie tussen persoonlijkheid en werkprestatie. De in dit kader relevante publicaties van Barrick en Mount (1991), Barrick, Mount en Judge (2002), en Salgado et al. (2003) zijn meta-analyses over de relatie tussen de Big Five en werkprestatie. Het artikel van Salgado et al. (2003) beschrijft de relatie tussen persoonlijkheid en werkprestatie specifiek in Europa. Deze drie publicaties laten zien dat de meeste Big Five dimensies zwakke tot matige verbanden hebben met werk- en trainingsprestatie (zie ook Tabellen 8.1 en 8.2). Consciëntieusheid vertoont de sterkste en meest consistente positieve relatie met werkprestatie. De niet voor methodologische factoren gecorrigeerde waarde, dat wil zeggen de geobserveerde waarde daarvan ligt rond de .17, de

gecorrigeerde waarde rond de .28 (Barrick & Mount, 1991). Emotionele stabiliteit is de andere dimensie die met werkprestatie lijkt samen te hangen, alhoewel de relatie met werkprestatie lager is dan die van consciëntieusheid. De resterende Big Five dimensies vertonen lagere relaties met werkprestatie en vaak is hun voorspellende waarde afhankelijk van het type baan. Extraversie blijkt bijvoorbeeld relatief belangrijk te zijn bij verkoopfuncties en voor managers waarin men veel met anderen moet praten en hen moet overtuigen (zie Tabel 8.2).

Tabel 8.1. Meta-analytische bevindingen van de relatie tussen persoonlijkheid en beoordelingen van werkprestatie (Barrick, Mount, & Judge, 2002)

Personality dimension	K	N	r	ρ	SDr	SD ρ	VE	CV
Conscientieusheid FFM	90	19460	0.17	0.28	0.13	0.16	40	0.07
Conscientieusheid NFFM	36	5874	0.11	0.18	0.13	0.17	40	-0.04
Emotionale stabiliteit FFM	72	10786	0.09	0.16	0.10	0.10	70	0.04
Emotionale stabiliteit NFFM	25	4541	0.03	0.05	0.09	0.07	74	-0.05
Openheid FFM	48	7562	0.05	0.08	0.10	0.10	63	-0.05
Openheid NFFM	29	4364	0.05	0.08	0.09	0.04	93	0.03
Vriendelijkheid FFM	68	10716	0.08	0.13	0.12	0.15	48	-0.05
Vriendelijkheid NFFM	31	4573	0.08	0.13	0.11	0.12	60	-0.02
Extraversie FFM	75	11940	0.04	0.07	0.12	0.14	47	-0.10
Extraversie NFFM	26	4338	0.05	0.08	0.10	0.10	61	-0.06

Noot. K=aantal onderzoeken; N=totale steekproef; r= gemiddelde waargenomen validiteit; SDr=standaard deviatie van r; ρ =validiteit gecorrigeerd voor onbetrouwbaarheid en range restriction; SD ρ = ρ standaard deviatie; VE= percentage variantie dat wordt verklaard door meetfouten en artefacten; NFFM= niet Five Factor Model; CV=90% betrouwbaarheidswaarde.

Met name relevant voor de politie zijn de gegevens uit het artikel van Barrick en Mount (1991) omdat zij daarin specifiek ingaan op de relatie tussen persoonlijkheid en werkprestatie van *politied medewerkers* (Zie Tabel 8.2). Daaruit komt naar voren dat wat betreft politiewerk, de Big Five relaties met werkprestatie vertonen van .00, .22, .09, .10 en -.10 voor respectievelijk O, C, E, A, en N. Ook hier is te zien dat de hoogste waarde wordt bereikt voor consciëntieusheid. Net als bij verschillende andere beroepen worden deze waarden gezien als laag tot matig. Het gaat om overigens om gecorrigeerde waarden. De geobserveerde waarden liggen nog lager en zijn 0, .13, .05, .06, en -.06 voor respectievelijk O, C, E, A, en N.

8.2 Persoonlijkheidsmaten in de selectie van Nederlandse politie

Bij de selectie van de Nederlandse politie worden Big Five maten gebruikt om aspirant-agenten te selecteren. Voor maart 2008 gebruikte de politie de M5Q (Klinkenberg & Van Leeuwen, 2003; Van Leeuwen, 2000). Daarna is de politie overgestapt op de NEO PI-R, een van de bekendste en wereldwijd meest gevalideerde Big Five persoonlijkheidsvragenlijsten in de persoonlijkheidspsychologie. De NEO-PI-R bestaat uit 200 vragen die samen 30 *persoonlijkheidsfacetten* vormen. Persoonlijkheidsfacetten zijn afgebakende clusters van gevoelens en gedragingen die ten grondslag zouden liggen aan de Big Five dimensies. Er worden 6 facetten per Big Five dimensie onderscheiden. De factorstructuur en betrouwbaarheid van de NEO-PI-R is reeds goed onderzocht door de afdeling werving en

selectie van de Politieacademie en bevestigt dat de facetten volgens verwachting onder te verdelen zijn in de betreffende Big Five dimensies. Bovendien zijn de scores op de Big Five dimensies betrouwbaar. Uit de analyses ten behoeve van het huidige project blijkt dat de factorstructuur van de Big Five dimensies overeenstemt met de bevindingen van de Politieacademie en dus de factorstructuur van de Big Five ondersteunen.

Tabel 8.2. (Engelstalige) Tabel met relaties tussen persoonlijkheid en werkprestatie in verschillende beroepen uit Barrick en Mount (1991)

TABLE 2
*Meta-Analysis Results for Personality Dimension-Occupation
Combinations (all Criterion Types Included)*

Occupational group	Total <i>N</i>	Number of <i>r</i> 's	Obs \bar{r}	$\hat{\rho}$	SD_{ρ}	90% C.V.	% Variance accounted
Extraversion							
Professionals	476	4	-.05	-.09	.05	-.03	92
Police	1,496	16	.05	.09	.00	.09	127
Managers	11,335	59	.11	.18	.13	.01	48
Sales	2,316	22	.09	.15	.16	-.05	54
Skilled/Semi-Skilled	3,888	23	.01	.01	.08	-.10	72
Mean (across occupations)			.08	.13	.11	-.01	69 ^a
Emotional stability							
Professionals	518	5	-.07	-.13	.04	-.07	92
Police	1,697	18	.06	.10	.00	.10	138
Managers	10,324	55	.05	.08	.09	-.04	65
Sales	2,486	19	.04	.07	.19	-.18	38
Skilled/ Semi-Skilled	3,694	26	.05	.12	.10	-.06	50
Mean (across occupations)			.05	.08	.10	-.05	63 ^a
Agreeableness							
Professionals	557	7	.01	.02	0	.02	158
Police	1,437	14	.06	.10	0	.10	121
Managers	8,597	47	.05	.10	.03	.06	94
Sales	2,344	16	.00	.00	.24	-.31	25
Skilled/Semi-Skilled	4,585	28	.04	.06	.17	-.16	37
Mean (across occupations)			.04	.07	.09	-.05	54 ^a
Conscientiousness							
Professionals	767	6	.11	.20	.00	.20	106
Police	2,045	19	.13	.22	.20	-.03	40
Managers	10,058	52	.13	.22	.10	.09	64
Sales	2,263	21	.09	.23	.00	.23	150
Skilled/Semi-Skilled	4,588	25	.12	.21	.09	.09	67
Mean (across occupations)			.13	.22	.09	.10	70 ^a
Openness to experience							
Professionals	476	4	-.05	-.08	.04	-.03	94
Police	1,364	13	.00	.00	.00	.00	181
Managers	7,611	37	.05	.08	.16	-.12	37
Sales	1,566	12	-.01	-.02	.16	.18	46
Skilled/Semi-Skilled	3,219	16	.01	.01	.12	-.15	49
Mean (across occupations)			.03	.04	.13	.13	59 ^a

^a An unbiased estimate of mean percentage of variance accounted for across meta-analyses, calculated by taking the reciprocal of the average of reciprocals of individual predicted to observed variance ratios (Hunter & Schmidt, 1990).

8.3 Voorspellende waarde (predictieve validiteit) van persoonlijkheid

In deze sectie staat de voorspellende waarde van persoonlijkheid voor uitval tijdens de opleiding en opleidingsprestatie centraal. Daarbij is dezelfde aanpak gehanteerd als in Hoofdstuk 5 en 6. De details van de methode en de analyses zijn weergegeven in Bijlage 4. De nu volgende paragrafen beschrijven de algemene bevindingen en conclusies.

8.3.1 Persoonlijkheid en opleidingsstatus

Met opleidingsstatus bleek er geen enkel significant verschil te zijn in de scores op de Big Five schalen tussen aspirant-agenten die vroegtijdig met de opleiding waren gestopt en diegenen die nog aan het studeren waren. Persoonlijkheid kon dus op geen enkele manier voorspellen wie er tijdens de opleiding zouden uitvallen. Wel scoorden aspirant-agenten met een hoger instroomniveau gemiddeld ook hoger op Openheid voor nieuwe ervaringen. Op de andere Big Five dimensies waren er geen verschillen tussen de instroomniveaus.

8.3.2 Persoonlijkheid en opleidingsprestatie

De relaties tussen persoonlijkheid en opleidingsprestatie (Proeven van bekwaamheid) waren over het algemeen zwak en niet significant (zie Tabel 8.3). Persoonlijkheid kon over het algemeen niet voorspellen hoe goed een aspirant-kandidaat het zou doen op de proeven van bekwaamheid tijdens de opleiding.

Tabel 8.3. Correlaties tussen de Big Five dimensies en proportie geslaagde proeven (alle instroomniveaus)

	Onderdelen	Proeven
O	.00	-.00
C	.02	.02
E	.01	.02
A	.02	.03
N	.00	.02

* $p < .05$.

Bij een onderscheid naar instroomniveau blijft het beeld in grote lijnen hetzelfde. Dat wil zeggen, ongeacht het instroomniveau waren er over het algemeen geen betekenisvolle relaties tussen persoonlijkheid en het behalen van proeven van bekwaamheid. Wel bleek dat, op instroomniveau 2, aspirant-agenten die enigszins consciëntieuzer en aardiger waren gemiddeld meer proeven van bekwaamheid haalden tijdens hun eerste poging (zie Tabel 8.4). Deze twee kenmerken zijn blijkbaar de belangrijkste persoonlijkheidsvoorspellers wanneer de (opleidingsprestatie) van aspirant-agenten op niveau 2 zou moeten worden voorspeld.

8.3.3 Bespreking resultaten voorspellende waarde van persoonlijkheid

Concluderend lijkt dat persoonlijkheid nauwelijks kan voorspellen wie zal uitvallen tijdens de opleiding en hoe goed de aspirant-agenten het zullen doen op de proeven van bekwaamheid. Uitzonderingen hierop zijn de dimensies consciëntieusheid en vriendelijkheid voor aspirant-agenten op instroomniveau 2. Deze specifieke dimensies vertonen voorspellende waarden die

in overeenstemming zijn met de literatuur (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2002).

Even als dat het geval was voor intelligentie is het ook weer belangrijk aan te geven dat er verschillende oorzaken kunnen zijn voor de relatief lage validiteiten, te weten: i) beperkingen in de operationalisatie van de criteria, ii) suboptimale keuze van de criteria, en iii) een daadwerkelijk afwezige relatie tussen persoonlijkheid en prestatie.

Anders dan bij de intelligentiematen is het overigens onwaarschijnlijk dat de problemen liggen bij *de meting* van de voorspellers. Zoals eerder aangegeven is de NEO-PI-R namelijk één van de meest gebruikte en erkende instrumenten om de Big Five te meten.

Tabel 8.4. Correlaties tussen Big Five en proportie geslaagde proeven (opgesplitst naar instroomniveau)

	Niveau 2 (n= 239)	Niveau 3 (n = 1027)	Niveau 4 (n = 802)
O	.01	.03	-.01
C	.13*	.03	.01
E	.11	.02	.06
A	.15*	.05	-.01
N	-.01	-.01	.03

* $p < .05$.

Hoewel de geobserveerde correlaties laag zijn, zijn ze gedeeltelijk in overeenstemming met de waarden uit eerdere meta-analyses (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2002; Salgado, 1997). Tot slot, valt op dat de huidige resultaten met betrekking tot persoonlijkheid overeenkomen met bevindingen uit eerder onderzoek bij de politie (De Meijer et al., 2008). De Meijer et al. rapporteerden validiteiten van de Big Five variërend van .03 tot .08, waarvan geen enkele waarde significant was. De Meijer et al. (2008) gebruikten opleidingsprestatie-indicatoren die lijken op de indicatoren die in het huidige onderzoek zijn gebruikt. Ten tijde van het onderzoek van De Meijer et al. (2008) werd nog gebruik maakte van de M5Q om persoonlijkheid te meten in plaats van de NEO-PI-R. Met enige voorzichtigheid valt te concluderen dat het overstappen van de M5Q naar de NEO-PI-R in 2008 geen duidelijke positieve effecten heeft gehad op de voorspellende waarde van persoonlijkheid voor de opleidingscriteria.

Conclusie 18: Persoonlijkheid (Big Five) heeft geen significante relaties met uitval tijdens de opleiding. Wat betreft opleidingsprestatie zijn er slechts twee dimensies die voorspellende waarden hebben (Consciëntieusheid en vriendelijkheid). Deze dimensies hebben echter alleen voorspellende waarde bij aspirant-agenten op instroomniveau 2. Het is onwaarschijnlijk dat het groot aantal nulbevindingen te wijten is aan het instrument waarmee persoonlijkheid wordt gemeten.

Hoofdstuk 9

Competenties (in de selectie):

Het selectiegesprek en de praktijkopdracht

9 .Competentiemetingen in de selectie

In hoofdstuk 3 kwamen competenties al deels aan de orde met betrekking tot het vaststellen van de criteria van goed politiewerk. Het huidige hoofdstuk gaat specifiek in op het meten van competenties *in de selectieprocedure*. Het gaat hierbij om competenties die worden gemeten in het selectiegesprek en tijdens de praktijkopdracht. Deze metingen behoren tot deel B van het geschiktheidsonderzoek. Over deze competenties, gemeten in deel B, kunnen de volgende vragen worden gesteld:

1) Sluiten de gemeten competenties aan bij de criteria van politiewerk? Hierbij aansluitend, hoeveel criteria en competenties moeten worden gemeten?

2) Meten de competenties ook datgene wat ze beogen te meten (begripsvaliditeit)? Hiermee samenhangend, wat is de relatie tussen de competenties en andere meer fundamentele kenmerken van de kandidaten zoals persoonlijkheid en cognitieve capaciteit?

3) Hebben de gemeten competenties voorspellende waarde voor de prestatie van (aspirant)-agenten (de predictieve validiteit)?

Deze vragen zullen hieronder worden behandeld.

9.1 Sluiten de gemeten competenties aan bij de criteria van politiewerk?

9.1.1 Competenties in het selectiegesprek

De politieorganisatie heeft de belangrijkste criteria voor politiewerk vastgesteld (wettelijke aanstellingseisen 2002). Deze zijn nagenoeg één-op-één vertaald naar de competenties die worden gemeten in de selectie. Een eerste belangrijke vraag die men zich hierbij kan stellen is of, uitgaande van O*Net, de competenties die gemeten worden in de selectie compleet zijn?

In de selectie van aspirant-agenten worden competenties gemeten in het selectiegesprek en in de praktijkopdracht. In het selectiegesprek komen 11 competenties aan bod die zijn beschreven in Tabel 9.1 hieronder. In Tabel 9.1 valt te zien dat de 11 competenties kunnen worden herkend in O*Net. Alleen tolerantie wordt niet in O*Net genoemd onder de kenmerken benodigd voor een politiefunctie.

Wat betreft de meting van competenties zijn door de afdeling werving en selectie van de Politieacademie goede beoordelingsschalen ontwikkeld. Een beschrijving van deze schalen is te vinden in het rapport 'Open advisering en bindend advies' uit oktober 2002 (in het vervolg wordt de afkorting OABA gebruikt om dit rapport aan te duiden). Daarin wordt beschreven dat er sinds 2002 gebruik wordt gemaakt van een 7-puntsschaal om de kandidaten te scoren omdat deze schaal beter kan differentiëren tussen het niveau van de kandidaten op de betreffende competenties. Vóór 2002 werden 5-puntsschalen gebruikt.

De beschrijvingen in het OABD rapport dienen als richtlijnen voor de psychologen die de selectiegesprekken houden. Een sterk punt van deze richtlijnen is dat er gebruik wordt gemaakt van zogenaamde 'Behavioral Anchored Rating Scales' oftewel gedragsverankerde schalen. Dat wil zeggen dat er bij elk punt op de 7-puntsschaal nauwkeurig beschreven is wat een score inhoudt in termen van gedragingen. De onderzoekers van het huidige project ondersteunen deze aanpak aangezien bekend is dat het gebruik van schaalankers de kwaliteit van beoordelingen verbetert en er meer overeenstemming is tussen verschillende beoordelaars (Guion, 2011). De beschrijvingen van de schaalankers in het OABD rapport zijn van goede kwaliteit. Elk punt op de 7-puntschalen is in concreet gedrag en in duidelijke oplopende stappen weergegeven.

Een voorbeeld hiervan is contactgerichtheid (zie tabel 9.1). Bij de laagste score (1) wordt hier het volgende schaalanker bij gegeven:

Is gesloten, zwijgzaam en stug. Voelt zich slecht op gemak in sociale situaties. Voelt zich geremd contact te leggen met anderen. Het onderhouden van werkcontacten kost veel energie. Reageert niet adequaat op de ander. Toont nauwelijks interesse in anderen. Toont weinig respect. Laat anderen regelmatig niet uitspreken. Reageert nauwelijks op hetgeen de ander naar voren brengt.

Bij de hoogste score 7 wordt beschreven:

Is spontaan, open en extravert. Communiceert gemakkelijk met anderen, ongeacht afkomst en achtergrond. Legt graag nieuwe contacten en voelt zich prettig en ontspannen in uiteenlopende sociale situaties. Heeft durf in de contacten met anderen. Treedt anderen op ontspannen wijze tegemoet en heeft er geen moeite mee om op de voorgrond te treden. Kan over een groot aantal onderwerpen meepraten. Weet door een grote interesse in de belevingswereld van anderen, mensen aan zich te binden. Staat ook in spanningsvolle situaties open voor nieuwe informatie en reageert dan nog adequaat op wat de ander zegt. Let ook op subtiele signalen en hints.

Zoals uit bovenstaande voorbeelden blijkt, zijn de beschrijvingen concreet en bieden ze goede handvatten voor de gesprekspsychologen om de kandidaten te scoren. Daarnaast zijn bij de afdeling werving en selectie van de academie aanvullende documenten beschikbaar die de selectiepsychologen kunnen ondersteunen bij het bepalen en interpreteren van de competenties.

In Tabel 9.1 komt naar voren dat er soms inhoudelijke overlap lijkt te zijn tussen de competenties. Zodoende is het niet altijd even duidelijk hoe competenties het best gekoppeld kunnen worden aan kenmerken in O*Net. Het criterium 'begrijpend luisteren' kan bijvoorbeeld opgevat worden als een onderdeel van communicatieve vaardigheden, maar ook als een onderdeel van inlevingsvermogen.

Tabel 9.1. Competenties zoals gemeten in het selectiegesprek en de vergelijking met enkele criteria zoals genoemd in O*Net

Competentiemodel Nationale Politie	Kenmerken zoals genoemd op O*Net
1. Communicatieve vaardigheden	Mondelinge uitdrukingsvaardigheid Effectief spreken Face-to-face gesprekken voeren Begrijpend luisteren
2. Contactgerichtheid	Serviceoriëntatie Behoefte aan relaties met anderen Sociale oriëntatie
3. Inlevingsvermogen	Sociaal inschattingsvermogen Medeleven
4. Initiatief nemen	Initiatief
5. Flexibiliteit	Aanpasbaarheid/flexibiliteit
6. Besluitvaardigheid	Reactietijd
7. Overzicht	Leiderschap
8. Zelfinzicht	Monitoren van de eigen prestatie
9. Stressbestendigheid	Stress tolerantie
10. Tolerantie	- *
11. Integriteit	Integriteit

*Voor tolerantie is geen vergelijkbaar kenmerk aanwezig

9.1.2 Competenties in de praktijkopdracht

In de praktijkopdracht worden drie hoofdcategorieën van gedrag beoordeeld, te weten: Denken, Voelen, en Kracht. Bovendien krijgen kandidaten na afloop nog een score op algemene effectiviteit, die is gebaseerd op de algehele prestatie tijdens de proef. Bij de Politieacademie zijn de drie hoofdcategorieën als volgt gedefinieerd:

I Denken: Deze competentie bestaat uit de onderdelen Analyseren, Plannen, Oordeelsvorming, Begrip, Interventie/Visie. Over het algemeen gaat het hier om hoe de kandidaat problemen aanpakt en of deze systematisch te werk gaat, inzicht heeft in de problemen en oplossingen kan genereren.

II Voelen: Deze competentie bestaat uit de onderdelen Inlevingsvermogen, Sociabiliteit/Flexibiliteit, Samenwerking, en Relatiebeheer. Het gaat hier dus duidelijk om hoe

men met andere omgaat en of men in staat is zich in de anderen te verplaatsten en samen te werken.

III Kracht: Deze competentie bestaat uit de onderdelen Initiatief, Aansturen, Overtuigingskracht, Verantwoordelijkheid, Besluitvaardigheid, Stressbestendigheid, en Resultaatgerichtheid. Deze competentie uit de praktijkopdracht wordt dus gekenmerkt door overwicht en daadkracht.

Bovenstaande competenties hebben enige overlap met competenties die in het selectiegesprek aan bod komen. Het idee hierachter is de zogeheten *multi-trait multi-method* aanpak. Met andere woorden, als kandidaten een competentie bezitten zou deze zichtbaar moeten in zijn het gesprek en ook in de praktijkopdracht. Dit impliceert dat sommige competenties in het selectiegesprek relatief sterk dienen samen te hangen met competenties in de praktijkopdracht. Een voorbeeld hiervan is de competentie 'Inlevingsvermogen' in het gesprek, waarvan te verwachten is dat deze zou samenhangen met de competentie 'Voelen' in de praktijkopdracht. Later in dit hoofdstuk komen de resultaten aan bod waarin dit idee expliciet is getest.

Conclusie 19: Uit de kwalitatieve evaluatie van het selectiegesprek en de praktijkopdracht valt te concluderen dat de competenties die gemeten worden in de selectieprocedure logisch lijken (face validity hebben) en conceptueel duidelijk zijn afgebakend. Bovendien zijn de kenmerken van politiewerk die worden genoemd in O*Net goed te herkennen in de competenties gemeten door de Politieacademie. De afdeling werving en selectie doet in dit opzicht veel om de kwaliteit van de metingen van competenties te blijven garanderen en te optimaliseren.

9.2 Hoeveel competenties moeten er worden gemeten?

Uit de beschrijving van O*Net wordt duidelijk dat er veel criteria van politiewerk zijn te onderscheiden. Het gaat om een lange lijst van criteria, waar Tabel 9.1 slechts een aantal van weergeeft (zie ook Tabel 3.2 in Hoofdstuk 3). Het zou teveel tijd kosten en te duur worden om de kandidaten op *al* deze criteria te beoordelen tijdens de selectie. Als twee of meer competenties elkaar voor een groot deel overlappen kan men zich bovendien afvragen of het überhaupt zinvol is om kandidaten op elk te beoordelen in de selectie (Lievens, 1998).

Los van mogelijk conceptuele overlap tussen competenties is er ook de vraag in hoeverre de selecteurs zelf in staat zijn om een groot aantal verschillende competenties te beoordelen. Het cognitieve verwerkingsvermogen van mensen beperkt vaak het aantal kenmerken waar men tegelijk op kan letten. Als er teveel kenmerken zijn waarop een psycholoog moeten letten wordt de cognitieve verwerkingslast erg hoog waardoor de kwaliteit van de beoordeling omlaag gaat. Dit is een bekend verschijnsel in de literatuur (bijv. Lievens, 1998).

Samenvattend: Er zijn twee manieren waarop competenties met elkaar kunnen overlappen. Eén daarvan is de *daadwerkelijke* inhoudelijke overlap, bijvoorbeeld als waarin deelvaardigheden van een competentie ook van belang zijn in een andere competentie (bijv. communicatieve vaardigheden en sociale vaardigheden).

De tweede manier is het *niet in staat zijn door de beoordelaars* om de competenties te onderscheiden vanwege een te grote aanspraak op cognitieve verwerkingssystemen.

Het definitieve antwoord op de vraag hoeveel competenties gemeten moeten worden in de selectie is niet te geven. Er kan wel worden getoetst in hoeverre competenties met elkaar overlappen (zie hieronder). Er kan echter niet worden getoetst of clusters van competenties de prestatie van de aspirant-agenten kunnen voorspellen omdat informatie over prestatie niet beschikbaar is.

Verschillende interviews met selectiepsychologen bevestigden het beeld dat het maar de vraag is of de testpsychologen goed in staat zijn om de kandidaten op de veelheid van verschillende competenties goed te beoordelen.

Eén van de geïnterviewden was bijvoorbeeld van mening dat het beoordelen van kandidaten op 11 competenties teveel is. Hij vermeldde dat het een illusie is om te denken dat psychologen zo gedifferentieerd naar een kandidaat kunnen kijken dat ze de 11 competenties goed kunnen onderscheiden. Hij pleitte daarentegen voor *vier competentieclusters* die belangrijke gebieden van politiewerk zouden omvatten, te weten:

Houding/gedrag, sociaal gedrag, stressbestendigheid, en kracht. De argumenten hiervoor waren dat met deze vier brede competentieclusters de belangrijkste criteria voor politiewerk afgedekt zouden zijn terwijl de mentale werklast van de psychologen minder is zodat de competentieclusters beter kunnen worden onderscheiden. Ook in een ander interview werd gepleit voor minder competenties, namelijk 6 á 7.

De literatuur is echter niet eenduidig over het optimaal aantal te beoordelen competenties. Onderzoek van Gaugler en Thornton (1989) toonde wel aan dat het beoordelen op een kleiner aantal competenties de kwaliteit van de beoordelingen verbetert. Deze onderzoekers trainden studenten als assessoren die vervolgens kandidaten in een assessment moesten beoordelen. Een groep studenten moest drie competenties beoordelen, een andere groep zes, en weer een andere groep negen. Uit de resultaten bleek dat het beoordelen van minder competenties samenhangt met meer accurate beoordelingen.

Arthur, Day, McNelly en Edens (2003) voerden een grote meta-analyse uit waarin ze kandidaat-beoordelingen op 168 verschillende competenties bestudeerden. Uiteindelijk bleken de 168 competenties terug te brengen te zijn tot zes brede categorieën, te weten: i) aandacht/bewustwording van anderen, ii) communicatie, iii) drive, iv) beïnvloeden van anderen, v) organiseren en plannen en, v) het oplossen van problemen. Het lijkt erop dat er maximaal zes competenties zouden moeten zijn waarop kandidaten worden beoordeeld in een selectiecontext.

Conclusie 20: Er zijn inhoudelijke gronden om het aantal gemeten competenties in het selectiegesprek terug te brengen tot een kleiner aantal. De indeling van Arthur, Day, McNelly en Edens (2003) kan daarbij als richtlijn dienen. Bovendien is er onder geïnterviewden consensus over het idee om minder competenties te hanteren. Welke competenties dit precies zouden moeten zijn is met name een empirische vraag.

9.2.1 Empirisch onderzoek naar clusters binnen de competenties

Om het onderscheid, tussen de competenties te onderzoeken zijn verschillende analyses uitgevoerd. Bijlage 6 bevat de details van de methode en de analyses. In het huidige hoofdstuk komen de algemene uitkomsten en conclusies aan de orde.

9.2.1.1 Selectiegesprek.

In het selectiegesprek lijkt het aantal competenties samen te vallen in één á twee grote clusters. Blijkbaar vinden beoordelaars het dus moeilijk om de competenties afzonderlijk te beoordelen.

Het eerste grote cluster wordt vooral gekenmerkt door contactgerichtheid en flexibiliteit. Emotionele weerbaarheid, drukbestendigheid en besluitvaardigheid vallen ook onder dit cluster. Kandidaten die hoog scoren op dit cluster zijn stevige en communicatieve personen. Een mogelijk label voor dit cluster zou kunnen zijn 'Kracht'.

Het tweede cluster wordt gekenmerkt door een hoge mate van zelfinzicht, tolerantie en integriteit. Er is ook enige overlap met communicatievaardigheden. Globaal lijkt deze cluster een indicatie te zijn van een soort algemene 'Wijsheid en Eerlijkheid/Mildheid'.

Er zijn verschillen tussen mannen en vrouwen op de clusters. Mannen scoren hoger dan vrouwen op het cluster Kracht, terwijl vrouwen hoger scoren op het cluster Wijsheid/Mildheid.

Naast de twee clusters, bleek er ook een algemene score op het selectiegesprek aanwezig. Dat wil zeggen dat de beoordelaars blijkbaar (onbewust?) een totaalindruk van de kandidaat vormen die dwars door alle afzonderlijke competenties heenloopt.

De vraag is of deze algemene indruk moet worden gezien als een ongewenste vertekening in de oordeelsvorming van de psychologen of dat het een relevant kenmerk van de kandidaat weergeeft. De discussie hierover is complex en wordt reeds decennia gevoerd in de wetenschappelijke literatuur (onder andere, Anusic et al. 2009; Van der Linden, Te Nijenhuis, & Bakker, 2010). Er valt hierover vooralsnog geen eenduidig antwoord te geven.

Conclusie 21: De competenties gemeten in het selectiegesprek lijken te clusteren in twee overkoepelende factoren: I) Kracht en II) Wijsheid/Mildheid. Bovendien is er ook een sterke algemene cluster waarneembaar die een globale beoordeling van een kandidaat lijkt weer te geven. Deze informatie geeft aan dat beoordelaars geen duidelijk onderscheid maken tussen de afzonderlijke competenties in het selectiegesprek.

In toekomstige beslissingen over het aantal en welke competenties moeten worden gemeten in het selectiegesprek kunnen deze resultaten een rol spelen. Hierbij moet wel worden aangemerkt dat veranderingen in de procedure van het selectiegesprek goed dienen worden ondersteund met onderzoek: Er kan bijvoorbeeld niet zonder meer worden aangenomen dat

brede categorieën van competenties werkprestaties ook beter kunnen voorspellen. Dit laatste blijft een empirische kwestie.

9.2.1.2 De praktijkopdracht

Bij de analyse van de praktijkopdracht is dezelfde aanpak gehanteerd als bij het selectiegesprek (Zie Bijlage 5 voor de details). Ten eerste moet worden vermeld dat de competenties in de praktijkopdracht al bij voorbaat veel breder gedefinieerd zijn dan in het selectiegesprek. Uitgaande van het idee dat het beoordelen van een kandidaat op een klein aantal competenties minder cognitief belastend zou zijn voor de beoordelaars (Lievens, 1998) zou de voorspelling zijn dat de (brede) competenties in de praktijkopdracht beter te onderscheiden zijn van elkaar. Dit idee kwam echter niet overeen met de resultaten.

De belangrijkste bevinding bij de praktijkopdracht was namelijk dat de drie competenties (Denken, Voelen, Kracht) nauwelijks worden onderscheiden door de beoordelaars. Dat wil zeggen, zowel de acteur als de beoordelaar lijken eerder een algemene indruk te hebben van de wijze waarop een kandidaat heeft gefunctioneerd tijdens de praktijkopdracht dan dat ze de kandidaat verschillende scores op afzonderlijke competenties geven.

Conclusie 22: Alhoewel kandidaten in de praktijkopdracht formeel op slechts drie brede competenties moeten worden beoordeeld, blijkt dat kandidaten op een algemene wijze worden beoordeeld in termen van hoe goed ze de opdracht hebben uitgevoerd.

9.3 Meten de competenties ook datgene wat ze beogen te meten (begripsvaliditeit)?

De begripsvaliditeit van de competenties is op twee manieren getoetst. Ten eerste door na te gaan of competenties uit het gesprek en de praktijkopdracht die met elkaar zouden moeten samenhangen dat ook daadwerkelijk doen. Ten tweede door na te gaan in hoeverre competenties zinvol overlappen met meer fundamentele kenmerken van de kandidaat zoals persoonlijkheid en cognitieve capaciteit.

Tabel 9.2 Correlaties tussen competenties uit het selectie gesprek en de praktijkopdracht

	Denken	Voelen	Kracht
Zelfinzicht	.32**	.26**	.24**
Contactgerichtheid	.35**	.34**	.29**
Flexibiliteit	.28**	.23**	.20**
Emotionele weerbaarheid	.23**	.15**	.24**
Drukbestendigheid	.30**	.21**	.32**
Tolerantie	.14**	.17**	.06**
Integriteit	.14**	.11**	.11**
Communicatieve vaardigheden	.38**	.26**	.32**

** Correlatie is significant op $p < 0.01$ level.

Tabel 9.2 laat zien dat er significante correlaties zijn tussen de competenties uit het gesprek en de praktijkopdracht. De competenties uit de praktijkopdracht hangen opvallend genoeg met bijna alle competenties uit het selectiegesprek samen. Dit suggereert een zekere consistentie in het beoordelen van een kandidaat. Met andere woorden, een kandidaat die het redelijk goed

doet in het selectiegesprek heeft een grotere kans het ook goed te doen in de praktijkopdracht. Meenemend dat er in het selectiegesprek en de praktijkopdracht verschillende beoordelaars zijn kan deze consistentie als positief worden beschouwd: De consistentie ondersteunt namelijk de (convergente) validiteit van totaaloordelen van beide opdrachten.

9.3.1 De verbanden tussen de competenties en persoonlijkheid en intelligentie

In het bepalen van de validiteit van competenties is het zinvol om te toetsen in hoeverre deze samenhangen met meer fundamentele kenmerken zoals persoonlijkheid en intelligentie. De details van deze toetsen zijn weergegeven in Bijlage 5.

De competenties uit het selectiegesprek en de praktijkopdracht hangen in sommige gevallen wel significant samen met persoonlijkheid en intelligentie, maar de effectgroottes zijn over het algemeen klein. Ondanks de relatief zwakke relaties worden er wel logische patronen gevonden. Zo blijkt bijvoorbeeld dat kandidaten die hoog scoren op neuroticisme, laag scoren op emotionele weerbaarheid en drukbestendigheid. Soortgelijke logische verbanden vind men tevens tussen contactgerichtheid (competentie) en extraversie (persoonlijkheid), en tussen enerzijds tolerantie (competentie) en anderzijds openheid voor nieuwe ervaringen en vriendelijkheid (persoonlijkheid). Kortom, de inhoud van de competenties in het gesprek en de praktijkopdracht worden ondersteund door de relaties met persoonlijkheid. Een andere opvallend patroon is dat intelligentie significant samenhangt met bijna alle competenties (behalve zelfinzicht).

Conclusie 23: Er zijn aanwijzingen dat er onder verschillende beoordelaars in de selectie en tijdens verschillende onderdelen (gesprek, praktijkopdracht) consensus is over wie al dan niet een goede kandidaat is. Dit ondersteunt de validiteit van de beoordelingen.

De relatie tussen enerzijds de competenties en anderzijds persoonlijkheid en intelligentie zijn relatief zwak. Dit suggereert dat de competenties toegevoegde waarde kunnen hebben in de selectie en zeker niet hetzelfde meten als persoonlijkheid en intelligentie.

9.4 Hebben de gemeten competenties voorspellende waarde voor de prestatie van (aspirant)-agenten (predictieve validiteit)?

De details van het onderzoek naar de voorspellende waarde van de competenties zijn beschreven in Bijlage 5. De criteria die in dit onderzoek zijn gebruikt betreffen wederom opleidingsstatus (vroegtijdige uitval, afgerond, nog studierend) en opleidingsprestatie (proeven van bekwaamheid).

9.4.1 Competenties en opleidingsstatus

Met betrekking tot de opleidingsstatus bleek dat, vergeleken met aspirant-agenten die nog studeerden of vroegtijdig waren uitgevallen, aspirant-agenten die de opleiding reeds hadden afgerond hoger scoorden op de volgende competenties: emotionele weerbaarheid, drukbestendigheid, besluitvaardigheid (gemeten in het selectiegesprek) en de competenties Denken en Kracht (gemeten in de praktijkopdracht).

Conclusie 24: Emotionele weerbaarheid, drukbestendigheid, en besluitvaardigheid zijn de competenties uit het selectiegesprek die significant voorspellende waarden hebben voor vroegtijdig stoppen met de opleiding. Voor competenties gemeten in de praktijkopdracht geldt dit voor Denken en Kracht.

9.4.2 Competenties en opleidingsprestatie (proeven van bekwaamheid)

De correlaties tussen de competenties en de proeven van bekwaamheid zijn weergegeven in tabel 9.3 hieronder.

Tabel 9.3 Correlaties tussen competentie scores (gesprek en praktijkopdracht) en de prestatie op toetsen (N ≈ 3000)

	Onderdelen proeven	Eindoordeel proeven
Zelfinzicht	.03 (.03)	.01 (.01)
Contactgerichtheid	.01 (.01)	.00 (.00)
Flexibiliteit	-.01 (-.01)	-.02 (-.02)
Emotionele Weerbaarheid	.02 (.02)	-.02 (-.02)
Drukbestendigheid	.01 (.01)	.00 (.00)
Tolerantie	.01 (.01)	-.02 (-.02)
Integriteit	.05* (.06)	.03 (.03)
Communicatieve vaardigheden	.05* (.06)	.02 (.02)
Besluitvaardigheid	.03 (.03)	.01 (.01)
Denken	.04 (.04)	-.01 (-.01)
Voelen	.00 (.00)	-.04* (-.04)
Kracht	.05* (.06)	.02 (.02)
Effectiviteit	.04 (.04)	-.01 (-.01)

* Correlatie is significant op het 0.05 niveau

Tussen haakjes staan de waarden gecorrigeerd voor range restriction

Uit deze tabel blijkt duidelijk dat de competenties vrijwel geen significante verbanden vertonen met de gehanteerde criteria, zelfs niet bij de grote hoeveelheden proefpersonen als in het huidige onderzoek. Dit patroon van bevindingen veranderde niet noemenswaardig wanneer onderscheid werd gemaakt naar instroomniveau (zie tabel B5.8 in Bijlage 5). Deze bevindingen leiden tot de volgende conclusie:

Conclusie 25: Over het algemeen kunnen de competenties niet sterk voorspellen hoe goed aspirant-agenten zullen presteren op de proeven van bekwaamheid tijdens de opleiding. Bij de interpretatie van deze uitkomsten moeten wederom de inherente beperkingen van de gehanteerde criteria in acht worden meegenomen (zie ook hoofdstuk 5, 6 en 8).

Hoofdstuk 10

Het eindgesprek (voorheen korpsgesprek)

10 Het eindgesprek (voorheen korpsgesprek)

10.1 Het korpsgesprek (tot 2013)

Nadat een kandidaat alle stappen in de centrale selectie bij de politie met succes heeft doorlopen volgde tot 2013 een eindgesprek bij het korps, waarin de passendheid bij het betreffende korps werd bepaald. Voordat de Nationale Politie in 2013 een feit was, vormde het korpsgesprek een op de centrale selectie aanvullend sollicitatiegesprek waarin functionarissen uit de praktijk van het specifieke korps de kandidaat beoordeelden. Bij de politie bestond hiervoor de uitdrukking ‘de kandidaat beoordelen met blauwe ogen’. De insteek van het korpsgesprek was dat de agenten die in een korps werkzaam zijn (de zogenaamde executieve functies) aanvullende informatie kunnen geven over de geschiktheid van de kandidaat.

Het korpsgesprek werd ten dele beschouwd als problematisch en oncontroleerbaar (Eindgesprek, 2012). Dat wil zeggen, vaak werd niet inzichtelijk op welke gronden kandidaten een positief of negatief advies kregen na het gesprek. Bovendien werden bij sommige korpsen tijdens de eindgesprekken de kandidaten opnieuw beoordeeld op competenties die ook al in het centrale deel van de selectie aan bod waren gekomen. In dit geval waren de normen echter minder expliciet dan in het centrale selectiedeel.

Een ander probleem van de korpsgesprekken was dat er weinig uniformiteit bestond in de toelatingspercentages tussen korpsen. Bij sommige korpsen werd bijvoorbeeld één op de 58 kandidaten afgewezen en bij andere was dit één op de drie (Rapport ‘Eindgesprek’, 2012).

10.2 Het eindgesprek: Structuur

In het licht van bovenstaande heeft de politie het korpsgesprek aangepast. Deze aanpassingen zijn ingevoerd aan het begin van 2013, tegelijk met de invoering van de Nationale Politie. De specifieke veranderingen en de argumentatie hiervoor zijn uitvoerig beschreven in het rapport ‘Ontwerp selectiecommissie Nationale Politie (SNP): Eindgesprek’ (Kwartiermaker Nationale Politie, 2012). De benaming ‘korpsgesprek’ is veranderd in ‘eindgesprek’ omdat die term de strekking van het gesprek beter zou weergeven.

Eén van de belangrijkste inhoudelijke veranderingen is dat een kandidaat naar aanleiding van het eindgesprek niet meer kan worden afgewezen op geschiktheid maar alleen op passendheid. Of een kandidaat geschikt is zou immers reeds zorgvuldig zijn vastgesteld in het centrale deel van de selectie.

Voorts is de structuur van het eindgesprek geformaliseerd. Het eindgesprek bevat drie vaste onderdelen namelijk, 1) een thuiswerkopdracht, 2) een vragenlijst over drijfveren die de kandidaten reeds in deel B van de centrale selectie hebben ingevuld, en 3) een semi-gestructureerd gesprek met leden van de selectiecommissie.

In de thuiswerkopdracht wordt de kandidaat gevraagd een korte presentatie voor te bereiden over waarom hij of zij meent geschikt te zijn voor het politiewerk. Het doel van de presentatie is dat de commissieleden aanvullende informatie krijgen over de presentatievaardigheden en

drijfveren van de kandidaat. Er wordt vooraf vermeld dat de presentatie meetelt in de besluitvorming over het eindoordeel van de kandidaat.

Na de presentatie gaat de selectiecommissie verder in gesprek met de kandidaat en stelt deze een aantal vragen volgens een semi-gestructureerd schema. De aard van de vragen is deels afhankelijk van de uitkomsten van de drijfverenvragenlijst die is ingevuld in deel B van de centrale selectieprocedure. De selectiecommissie heeft inzicht in de uitkomsten van de drijfverenvragenlijst. De selectiecommissie heeft overigens ook inzicht in het sollicitatieformulier maar *niet* in de psychologische rapportage over geschiktheid voorafgaand aan het gesprek.

10.3 Wat wordt bedoeld met passendheid?

De passendheid van de kandidaat zoals die dient te worden vastgesteld in het eindgesprek is gebaseerd op vier zogenoemde 'waarden'. Deze zijn: Integer, Betrouwbaar, Moedig, en Verbindend. In het rapport uit 2012 (Kwartiermakers) worden deze waarden als volgt omschreven:

- Integer: Onafhankelijk in hun optreden
- Betrouwbaar: Consequent en rechtvaardig in hun optreden
- Moedig: Doortastend in hun optreden, soms met gevaar voor eigen leven.
- Verbindend: Nauw samenwerkend met burgers en veiligheidspartners, vanuit gedeelde verantwoordelijkheid.

Iedere waarde wordt verder uitgewerkt in een aantal bijbehorende drijfveren. Enkele drijfveren die horen bij Moedig zijn bijvoorbeeld: 'het streven naar spanning en avontuur en het verleggen van grenzen' en 'het streven naar succes, status en controle'. Voorbeelden van drijfveren bij Verbindend zijn, 'de behoefte aan sociale contacten en interactie' en 'de wil om andere te helpen en ondersteunen'. Bij de waarde Betrouwbaar horen onder andere 'streven naar hoge kwaliteit en foutloosheid', en 'de wil om eigen talenten en kennis verder te ontwikkelen'. Een aantal drijfveren behorend bij Integer zijn 'de behoefte om te voldoen aan verwachtingen' en 'het belangrijke vinden te kunnen doen wat je zelf wilt doen'.

10.4 Het eindgesprek: Beoogde afwijzingspercentage en procedure

Aangezien het eindgesprek een onderdeel is van de selectieprocedure zullen er een aantal kandidaten afgewezen kunnen worden na dit gesprek. In de eerdere korpsgesprekken was het uitvalspercentage relatief hoog en vertoonde dit percentage bovendien een grote variatie tussen de verschillende korpsen. In het nieuwe eindgesprek wil de politie de afwijzing echter meer standaardiseren. Er wordt in het eindgesprek nieuwe vorm een uitvalspercentage van rond de 12% verwacht.

De selectiecommissie die het eindgesprek voert voor aspirant-agenten (tot en met niveau 4) bestaat uit één persoon 'uit het blauw' en één persoon van de afdeling Werving en Selectie van de betreffende eenheid. De persoon 'uit het blauw' is een executieve functionaris die in het betreffende veld werkzaam is. Executieve functionarissen voor de selectiecommissies zullen worden geworven door middel van vacatures. Functionarissen die solliciteren voor een plek in

een selectiecommissie zullen worden geselecteerd op een aantal competenties, te weten: open-mindedness, contactvaardigheid, oordeelsvorming, besluitvaardigheid, en zelfreflectie. De functionarissen die worden aangenomen voor de commissies zullen een training krijgen waarin een uniforme werkwijze van selectiemethode zal worden aangeleerd.

In totaal zal een selectiecommissie ongeveer 90 minuten per kandidaat bezig zijn met de voorbereiding, het luisteren naar de presentatie van de kandidaat, het semi-gestructureerd interview, de afronding, en het uitreiken van de bescheiden van intentie tot aanstelling.

10.5 Discussie eindgesprek

De onderzoekers van het huidige project zijn van mening dat veel van de (beoogde) veranderingen in het eindgesprek nuttig zijn en goed onderbouwd met argumenten en kennis over selectie. Desalniettemin zijn er enkele punten ter reflectie aan te geven die hieronder stapsgewijs worden beschreven.

10.5.1 Uitvalspercentages

Momenteel is het verwachte uitvalspercentage gesteld op 12%. Gegeven de beschikbare informatie lijkt dit getal tamelijk willekeurig te zijn vastgesteld. Het is voorstelbaar dat dit uitvalspercentage deels is gebaseerd op de beschikbare plekken die moeten worden opgevuld in combinatie met de kosten van de selectieprocedure. Desalniettemin is het de vraag of er ook inhoudelijke argumenten voor zijn.

Tabel 2.1 in hoofdstuk 2 geeft aan dat er gemiddeld 27% van de kandidaten werd afgewezen in de korpsgesprekken. In het nieuwe eindgesprek beoogt men dus een afwijzingspercentage dat 66% lager ligt. Nu zijn er in de selectie meerdere procedures veranderd. Mogelijk is de 12% dus een juiste inschatting maar het werd in het project niet duidelijk waar deze op gebaseerd is. Dit zou expliciet gemaakt moeten worden.

Een alternatieve manier om naar dit onderwerp te kijken is als volgt: Stel dat in het merendeel van de gevallen de inschattingen van de functionarissen in het korpsgesprek correct waren (zeg, 70% correct, dus een valide afwijzingspercentage van $.70 * 27 \approx 19\%$). Dit zou betekenen dat door de grens op 12% te zetten er een discrepantie ontstaat van 7%. Met andere woorden, er zullen dan 7% aspirant-agenten aangenomen worden die later daadwerkelijk suboptimaal zullen functioneren. Op de lange termijn kan dat aanzienlijke kosten met zich meebrengen en bovendien de kwaliteit van het korps verlagen.

Conclusie 26: Het wordt niet duidelijk waar het verwachte toelatingspercentage in de korpsgesprekken op is gebaseerd. Daarom valt te overwegen aannames en keuzes hierbij explicieter te maken zodat mogelijke terechte en onterechte veronderstellingen in de achterliggende redenering eerder en gemakkelijker opgemerkt kunnen worden.

10.5.2 Selectiecommissie: Samenstelling en training

In de recente opzet van het eindgesprek zal de selectiecommissie bestaan uit één executief functionaris en één medewerker van werving en selectie. Dit lijkt in beginsel een goede opzet. Mogelijk heeft deze samenstelling echter ook enkele nadelen. Eén daarvan is dat de eigen inbreng van de executieve functionaris en de unieke visie typerend voor 'kijken met blauwe ogen' gemakkelijk ondergesneeuwd wordt in deze opzet. De executieve functionaris is namelijk de enige persoon in de selectieprocedure die de praktijkvisie representeert. Bovendien krijgt deze executieve functionaris een training waarin de uniformiteit van de selectieprocedure benadrukt wordt. De methode en uniformiteit in de training zal waarschijnlijk gebaseerd zijn op wetenschappelijke inzichten en zal daardoor mogelijk meer gaan lijken op de procedure zoals gehanteerd in het centrale deel van de selectie bij de Politieacademie, terwijl juist het intuïtieve oordeel van de persoon in het blauw onderscheidend is in het eindgesprek ten behoeve van het meten van de passendheid.

In de wetenschappelijke literatuur heerst nog veel onduidelijkheid over de rol van dit soort processen in personeelsselectie. Enerzijds is bekend dat het ongestructureerde interview, waarbij men mag aannemen dat de intuïtie een grotere rol speelt, een lage voorspellende waarde heeft (Hunter & Schmidt, 1998). Anderzijds zijn er in de recente literatuur steeds meer aanwijzingen dat globale, intuïtieve inschattingen van persoonlijkheid vaak een kern van waarheid bevatten (e.g., Beer & Watson, 2008) en daardoor informatief en nuttig zijn. Zo is bijvoorbeeld aangetoond dat personen in staat zijn om de persoonlijkheid in te schatten van anderen waar zij slechts zeven minuten mee in contact waren (Rauthmann, 2012)

Conclusie 27: De rol van 'kijken met blauwe ogen' in de selectieprocedure blijft nog enigszins onduidelijk. Heeft de Politieacademie expliciete ideeën over wat voor aanvullende informatie deze stap in de selectieprocedure oplevert? En zo ja, hoe schat men de waarde van deze informatie in, in termen van betrouwbaarheid, construct validiteit, en voorspellende waarde? Hierop volgend stellen we voor: Wees expliciet over hoe men de rol van kennis van executieve functionarissen in de selectieprocedure beschouwd (e.g., in welke mate gewenst, valide, etc.) en wees bewust van de mogelijke voor- en/of nadelen van het minimaliseren van deze kennis.

10.5.3 Waarden versus persoonlijkheid/competenties

In het rapport over de opzet van het eindgesprek wordt vermeld dat kandidaten niet meer worden getoetst op geschiktheid maar alleen op passendheid. In overeenstemming hiermee wordt gesteld dat men niet langer zal toetsen op competenties maar op *vier waarden* (zie eerdere beschrijving hierboven). Als men echter de definities van de waarden en hun drijfveren bestudeert dan wordt het onderscheid tussen competenties, persoonlijkheid en waarden niet direct duidelijk. De waarde Betrouwbaarheid bijvoorbeeld heeft als een van de drijfveren 'streven naar hoge kwaliteit en foutloosheid'. Dit komt over als een belangrijk aspect van de persoonlijkheidstrek consciëntieusheid. De waarde Verbindend heeft als drijfveer 'behoefte aan sociale contacten en interactie', hetgeen veel lijkt op een aspect van extraversie. Als de selectiecommissie op grond van de uitslagen van dergelijke drijfveren besluit een negatief advies te geven dan zou dit mogelijk betekenen dat de kandidaat in feite als niet voldoende wordt beschouwd op deze persoonlijkheidsaspecten die eigenlijk ook al in de centrale selectie zijn gemeten.

Conclusie 28: Het onderscheid tussen enerzijds de beoordeling van waarden en anderzijds de meting van competenties en persoonlijkheid is niet volledig duidelijk. Daarom zou het nuttig zijn om empirisch na te gaan wat het eindgesprek onderscheidt van de karakteristieken die in de centrale selectie worden gemeten. Hiertoe zouden, zodra er voldoende gegevens beschikbaar zijn, de statistisch verbanden tussen de waarden, competenties en persoonlijkheid worden bepaald.

Mogelijkerwijs betreft het onderscheid het verschil tussen typische en maximale prestatie (in het Engels wordt dit aangeduid als *typical versus maximum performance*: Klehe & Latham, 2006). Tijdens het selectiegesprek op de academie en de praktijkopdracht laat een kandidaat zich wellicht van zijn beste kant (maximale prestatie) zien terwijl de waarden meer betrekking hebben op hoe een kandidaat zich typisch gedraagt dan wel hoe de kandidaat zich idealiter zou willen gedragen.

Hoofdstuk 11

Discussie gehele selectieprocedure

11. Discussie gehele selectieprocedure

De voorgaande hoofdstukken beschreven de afzonderlijke selectie-instrumenten die onderdeel vormen van de selectieprocedure (testen, gesprek, proeven, etc.). Per selectie-instrument werd een aantal conclusies getrokken. Hoofdstuk 11 bevat een algemene discussie en conclusies over de selectie-instrumenten in de context van de gehele selectieprocedure.

De vragen die in dit hoofdstuk besproken worden komen overeen met de vragen die zijn gesteld in de startnotitie van het WODC (2012) en het projectvoorstel (Van der Linden, & Born, 2012). Het gaat om de volgende vragen: Wat valt er te zeggen over de voorspellende waarde van de selectie-instrumenten? Welke selectie-instrumenten werken goed en welke zouden kunnen worden vervangen of herzien?

11.1 Voorspellende waarde selectieonderdelen

11.1.1 De effectgrootten van de selectie. De selectieprocedure voor politiepersoneel wordt naar onze inschatting grondig uitgevoerd. Over de verschillende onderdelen wordt voortdurend goed nagedacht door de afdeling werving en selectie. Bovendien blijkt onder de medewerkers bij de afdeling werving en selectie naar onze indruk een sterke motivatie aanwezig te zijn om de kwaliteit continue te verbeteren. Veel van de gehanteerde instrumenten of procedures zijn gebaseerd op breed geaccepteerde en up-to-date inzichten in de literatuur over personeelsselectie. Binnen de afdeling werving en selectie blijkt ook een goede wetenschappelijke cultuur te heersen die samengaat met de wens om bij voortduring nieuwe inzichten te verwerven over de kwaliteit van de selectieprocedures.

Met de beschikbare gegevens over opleidingsstatus blijkt geen van de selectieonderdelen een sterke voorspellende waarde te hebben. Op sommige onderdelen scoren aspirant-agenten die de opleiding vroegtijdig verlaten statistisch significant lager dan degenen die de opleiding met succes hebben afgerond of nog bezig zijn met studeren. Deze statistisch significante verschillen zijn over het algemeen overigens klein. De verklaarde variantie betrof slechts enkele procenten (tussen de 0.5% en 3%). Dit wil zeggen dat deze selectie-instrumenten voor een klein deel kunnen voorspellen wie zal gaan uitvallen maar dat er andere, vooralsnog onbekende factoren zijn die een veel grotere rol blijken te spelen.

De relatie tussen de selectie-instrumenten en de opleidingscriteria (behaalde proeven van bekwaamheid) waren eveneens laag en veelal niet significant.

Wat valt uit deze bevindingen af te leiden? Gezien de gebruikte opleidingscriteria (opleidingsstatus en proeven van bekwaamheid) dient onze conclusie omtrent de voorspellende waarde van selectie-instrumenten uitermate voorzichtig te worden geformuleerd:

Conclusie 29: Op grond van de analyses is de conclusie dat de selectieonderdelen slechts beperkt de opleidingscriteria voorspellen. Er dienen echter twee belangrijke kanttekeningen bij deze conclusie geplaatst te worden.

Ten eerste blijft vooral nog onbekend in hoeverre de criteria indicatief zijn voor daadwerkelijk opleidingssucces en, nog belangrijker, voor latere werkprestatie. De criteria hebben een aantal beperkingen die de interpretatie van de in dit rapport beschreven uitkomsten bemoeilijken (zie beschrijving van de criteria in Hoofdstuk 3).

Ten tweede zijn de relatief lage relaties tussen selectieonderdelen en uitkomstmaten (criteria) niet specifiek voor selectie bij de politie. Een recent artikel over opleidingsuitval bij de Nederlandse landmacht laat bijvoorbeeld zien dat persoonlijkheidsmaten (De Big Five) nauwelijks kunnen voorspellen wie er tijdens de opleiding zal uitvallen (Van der Linden, Te Nijenhuis, Cremer, & Van der Ven, 2011). Wat betreft de Big Five persoonlijkheidsdimensies is bovendien bekend dat de validiteiten van persoonlijkheidskenmerken laag zijn ongeacht welk criterium men wilt voorspellen (Barrick & Mount, 1991). Er is minder wetenschappelijke literatuur beschikbaar over competenties dan over persoonlijkheid. Desalniettemin lijken voorgaande studies ook aan te geven dat de voorspellende waarde van competenties vaak relatief laag is.

Betekent dit dat sommige onderdelen van de selectie dan overbodig zouden zijn? Deze vraag kan niet goed worden beantwoord op grond van de beschikbare data. Het is waarschijnlijk dat veel kandidaten met de sterkste risicofactoren op disfunctioneren niet door de selectie heenkomen. Zonder selectie zou het waarschijnlijker worden dat kandidaten met een laag cognitief vermogen of met persoonlijkheidsstoornissen aangenomen zouden worden. Selectie blijft in dit opzicht dus essentieel. Wat uit de resultaten echter naar voren komt is dat het moeilijk is om de bijdragen van de selectie direct te kwantificeren. Elke schatting gebaseerd op de uitkomsten van de selectieonderdelen zal waarschijnlijk een sterke onderschatting geven van de daadwerkelijke bijdragen van de selectie. Bovendien is de kwaliteit en relevantie van de beschikbare opleidingscriteria nog onduidelijk.

11.1.2 Lage voorspellende waarde: Een criteriumprobleem?

In de voorgaande paragraaf, zowel als op andere plaatsen in het huidige rapport, wordt de waarde van de criteria steeds ter discussie gesteld. In Hoofdstuk 4 staan enkele resultaten die de operationalisatie van de criteria ondersteunen. Desalniettemin blijft het om verschillende redenen mogelijk dat deze operationalisatie suboptimaal was.

Zo waren de scores op de proeven van bekwaamheid dichotoom, hetgeen betekent dat iemand een proef wel of niet heeft gehaald. Er wordt niet geregistreerd *hoe goed* iemand aan de proef heeft voldaan. Er is dus weinig differentiatie tussen aspirant-agenten die de proeven hebben behaald. Daar komt bij dat de aspirant-agenten een proeve van bekwaamheid pas uitvoerden als zij, in overleg met hun begeleider, het idee hadden daar 'klaar voor te zijn'. Gemiddeld 90 procent van de aspirant-agenten haalt een proeve van bekwaamheid tijdens de eerste poging. Er kan ook ten dele sprake zijn van een plafondeffect: De proeven zijn te makkelijk om onderscheidend te zijn waardoor de kans dat de selectie-instrumenten een goede voorspellende waarde kunnen laten zien niet groot is.

Conclusie 30: De huidige opleidingsmaten (proeven van bekwaamheid en opleidingsstatus) hebben een aantal kenmerken die het moeilijk maken om goed onderzoek te doen naar de relatie tussen de selectie-instrumenten en opleidingssucces. Men zou kunnen overwegen om bij de beoordeling van aspirant-agenten in de toekomst, scores te gebruiken die beter kunnen differentiëren tussen aspirant-agenten dan nu het geval is. Dit levert aanvullende informatie op voor de opleiding over de aspirant-agenten en kan tevens leiden tot betere inzichten in de efficiency van werving en selectie. Men kan bijvoorbeeld denken aan het invoeren van standaard schoolcijfers (1 t/m 10) bij de proeven van bekwaamheid.

Aanvullende informatie over aspirant-agenten tijdens de opleiding zou niet alleen nuttig zijn om de wervings- en selectie-inspanningen te valideren maar geeft tevens meer mogelijkheden om de kenmerken van de opleiding *zelf* te onderzoeken. Idealiter zou de prestatie op de proeven van bekwaamheid tijdens de opleiding moeten kunnen voorspellen hoe goed een agent later zal functioneren op straat. De opleidingsgegevens kunnen dan in toekomstig onderzoek gebruikt worden als *voorspellers van daadwerkelijk politiewerk*.

11.1.3 Criteriumonderzoek

Een terugkerend thema is dat er slechts beperkte opleidingscriteria beschikbaar zijn en dat prestatie-indicatoren van agenten in uitvoering onbeschikbaar zijn. Opvallend is dat een soortgelijke conclusie reeds werd getrokken in een rapport over de evaluatie van de selectieprocedure bij de politie uit 1990 (Wilbrink, Hoorn, Van der Kamp, & Algera, 1990, "*Selectie voor politie officier: Onderzoek naar de selectieprocedure voor de toelating tot de Nederlandse politie academie*"). Blijkbaar is er sinds 1990 nog niet veel veranderd.

Conclusie 31: Voor een adequate evaluatie van zowel de activiteiten van werving en selectie en van de politieopleiding is het essentieel dat er onderzoek wordt uitgevoerd met zoveel mogelijk directe prestatie-indicatoren tijdens het politiewerk. Men kan daarbij denken aan beoordelingen van leidinggevendenden of anderen directe prestatie-indicatoren.

Van de verschillende aanbevelingen die worden gedaan moet worden benadrukt dat de laatstgenoemde aanbeveling over criteriumonderzoek de hoogste prioriteit zou moeten krijgen. Indien een criteriumonderzoek goed wordt uitgevoerd kan het één van de pijlers vormen waarop veel beslissingen over werving, selectie en opleidingen worden gebaseerd.

11.1.3.1 Welke criteriummaten? Indien de politie zou besluiten een criteriumonderzoek te starten dient nagegaan te worden welke werkprestatiematen gebruik moeten worden. Er is uitvoerige wetenschappelijke literatuur over dit onderwerp (e.g., Cook, 2009; Guion, 2011). Veel gebruikte prestatie-indicatoren in de literatuur betreffen de beoordelingen van direct leidinggevendenden. Zulke beoordelingen hebben een aantal inherente beperkingen. Er kunnen bijvoorbeeld vertekeningen in de beoordeling optreden. Toch zijn er ook duidelijke aanwijzingen dat ondanks deze beperkingen, leidinggevendenden kunnen inschatten hoe goed hun ondergeschikten presteren (e.g., Ones, Viswesvaran, Schmidt, 1997).

In sommige beroepen heeft men naast de subjectieve beoordelingen tevens toegang tot meer objectieve prestatie-indicatoren. Als het gaat om verkoopfuncties of financiële functies is een voor de hand liggende maat het aantal verkochten producten of de omzet. Bij politiewerk zal het waarschijnlijk lastiger zijn om objectieve prestatie-indicatoren te identificeren.

11. 2 De status van de selectie-instrumenten/onderdelen bij de politie

Hierboven kwam een aantal overkoepelende aspecten van de selectieprocedure aan de orde. In de volgende paragrafen hebben de conclusies betrekking op afzonderlijke instrumenten en selectie-onderdelen en hun status in de selectieprocedure.

11.2.1 De cognitieve test

Uit het onderzoek kwam naar voren dat de cognitieve test een lage voorspellende waarde heeft met betrekking tot de opleidingscriteria. De hoogste waargenomen predictieve validiteiten van cognitieve capaciteit werden gevonden onder allochtonen maar deze validiteiten waren *negatief*. Dat wil zeggen; hoe intelligenter de aspirant-agent hoe lager de kans dat hij tijdens de eerste poging een proeve zou behalen. Deze bevinding is opmerkelijk. Voor de groep autochtone aspirant-agenten waren de voorspellende waarden laag. De lage voorspellende waarde van cognitieve capaciteit voor beide etnische groepen komt overeen met de bevindingen van De Meijer et al. (2008) die eerder onderzoek deden naar criterium-gerelateerde validiteit in de selectie bij de Nederlandse politie. Dit onderzoek van De Meijer et al. heeft data gebruikt van de Politie Intelligentie Test (PIT). Het huidige project waarin is gekeken naar de gegevens van de recent ingevoerde Cognitieve Adaptieve Intelligentie Test (CAIT) laat zien dat er geen grote verbetering is opgetreden in de voorspellende waarde van cognitieve capaciteit. Zoals reeds is opgemerkt is het natuurlijk mogelijk dat cognitieve capaciteit daadwerkelijk slechts zwak samenhangt met prestatie van politiewerk. Een illustratie hiervan is de bevinding van Salgado et al. (2003) die rapporteerden dat cognitieve capaciteit een waargenomen predictieve validiteit vertoonde van $r = .12$ die na correcte voor methodologische factoren steeg tot $p = .25$. Deze waarden zijn hoger dan de waarden die zijn gevonden in het huidige project, maar nog steeds laag ten opzichte van veel andere beroepen.

In de beoordeling van de rol van cognitieve capaciteit in selectie bij de politie was de conclusie dat er nog weinig informatie beschikbaar was over de validiteit/kwaliteit van de test die momenteel gebruikt wordt. Voor zover bekend was er geen informatie beschikbaar waarin scores op de CAIT duidelijk gevalideerd waren met behulp van andere bekende cognitieve capaciteitstesten. Daar komt bij dat in de wetenschappelijke literatuur de huidige status van adaptief testen nog niet onbetwist is. Er is nog geen consensus of adaptief testen even goed of beter is dan traditioneel meten. De politie heeft mogelijk een risico genomen door in de selectie te vertrouwen op een cognitieve test die een nog niet breed gesteunde methode hanteert (adaptief testen) en waarvan belangrijke validiteitsinformatie ontbreekt.

Conclusie 32: Het is belangrijk dat er meer inzicht komt in de psychometrische kenmerken alsook de voorspellende waarde van de gebruikte cognitieve tests. Indien deze niet beschikbaar is valt een vervanging van de test te overwegen.

11.2.2 Persoonlijkheid.

Ondanks de rijke traditie van onderzoek op dit gebied wordt de rol van persoonlijkheid in selectie tot op heden door sommigen nog in twijfel getrokken. De reden daarvoor is dat persoonlijkheid over het algemeen lage voorspellende waarden voor werkprestatie laat zien. Over verschillende beroepen heen rapporteerden Barrick, Mount en Judge (2002) meta-analytische waarden van .05, .17, .04, .08 en -.09 bij het voorspellen van werkprestatie voor respectievelijk openheid voor nieuwe ervaringen, consciëntieusheid, extraversie, vriendelijkheid en neuroticisme. Voor opleidingscriteria van politiewerk worden soms nog lagere validiteiten gerapporteerd (De Meijer et al., 2008). De bevindingen uit het huidige project zijn wat dit betreft niet veel anders: Er werden lage en vaak niet significante waarden gevonden. Wat kan er geconcludeerd worden over de zinvolheid van het meten van persoonlijkheid in de selectie bij de politie?

Ten eerste kan uiteraard hetzelfde worden geconcludeerd als bij cognitieve capaciteit, namelijk dat de inschatting van de waarde van persoonlijkheid beperkt wordt door de kwaliteit van de criteria.

Ten tweede blijkt persoonlijkheid in bijna elke selectieprocedure een relatief lage voorspellende waarde te hebben voor werkprestatie. Het lijkt erop dat dit voor de politie niet anders is.

Ten derde kost het relatief weinig moeite om de persoonlijkheidstest af te nemen en om de scores te berekenen en interpreteren, hoewel persoonlijkheid lage validiteiten vertoont. Daarom doen we in dit opzicht de volgende aanbeveling:

Conclusie 33: Ondanks de lage validiteiten, stellen we voor de persoonlijkheidsmaten te handhaven in de selectieprocedure van de politie. De status van persoonlijkheid is in de literatuur dusdanig dat het een te groot risico zou zijn om van het gebruik van deze maten af te zien. Het instrument dat hier in de huidige situatie voor wordt gebruikt, namelijk de NEO-PI-R is één van de bekendste en meest gevalideerde in zijn gebied en zou dus volstaan.

Dit neemt niet weg dat het in de toekomst zinvol kan zijn om mogelijkheden te exploreren om de validiteit van persoonlijkheidsmetingen te verbeteren. Bijvoorbeeld door na te gaan of gecontextualiseerde persoonlijkheidsvragenlijsten (beter toegespitst op politieorganisatie) of ander-beoordelingen (in plaats van traditionele zelfbeoordeling) de predictieve validiteit kan verhogen.

Toelichting: Gecontextualiseerde persoonlijkheidsvragenlijsten zijn metingen waarin vragen over persoonlijkheid worden gesteld die duidelijk verbonden worden met een specifieke werksituatie. Als bijvoorbeeld de beoogde context 'werk' is dan gebruikt men bijvoorbeeld niet de stelling: 'ik ben zorgvuldig' maar vervangt men deze door de stelling: 'ik ben zorgvuldig op mijn werk'.

11.2.3 Selectiegesprek

Het selectiegesprek wordt bij de politie gestructureerd en grondig uitgevoerd. De afdeling werving en selectie doet er alles aan tot onderbouwde kandidaat-scores te komen op een reeks competenties.

Over het aantal en de aard van de competenties die in het gesprek worden gemeten kan het volgende worden opgemerkt. Lievens (1998) komt tot de conclusie dat het beoordelen van meer dan zes tot negen competenties tijdens een selectie-instrument (een gesprek, een opdracht, een rollenspel etc.), problematisch is. Het aantal competenties dat momenteel in het gesprek wordt gemeten zou gereduceerd kunnen worden. Deze suggestie wordt ondersteund door de kwantitatieve analyses, die lieten zien dat er in het selectiegesprek een substantiële algemene factor te identificeren valt die bijna 40 % van de variantie in de scores verklaart. In hoeverre het ontstaan van deze algemene factor het gevolg is van een vertekening in de beoordeling tijdens het gesprek of dat de factor het ware gedrag van de kandidaat weergeeft valt niet definitief te beantwoorden.

Er zijn aanwijzingen dat de algemene indruk van een kandidaat, die tot uitdrukking komt in een algemene factor, voor een belangrijk deel gebaseerd is op waar gedrag van de kandidaat (Ones, Viswesvaran, & Schmidt, 2006). Bovendien bleek de algemene indruk uit het selectiegesprek tussen de $r = .45$ en $r = .50$ te correleren met de algemene indruk die afkomstig is van andere beoordelaars dan uit het selectiegesprek.

Conclusie 34: Op grond van de literatuur en de uitkomsten van de analyse van de selectiedata valt te overwegen om minder competenties te beoordelen in het selectiegesprek. Het zou wellicht beter zijn om kandidaten in te schatten op bredere gedragscategorieën die relevant zijn voor politiewerk. Eén van de mogelijkheden hiervoor is de kandidaten te beoordelen op vier categorieën, te weten: Daadkracht, Sociale effectiviteit, Weerbaarheid, en Houding/Integriteit. Het voordeel van zo'n aanpak is dat de cognitieve last van de psycholoog wat betreft de beoordeling wordt verminderd en de beoordelingen daardoor betrouwbaarder en robuuster worden. Indien er een verandering in de te beoordelen competenties plaats zou vinden dienen de uitkomsten daarvan uiteraard geëvalueerd te worden door de voorspellende waarde voor opleidingssucces en, het liefst, voor criteria op het werk (prestaties, verloop, tevredenheid) te bepalen.

11.2.4 Praktijkopdracht

De punten die aan bod zijn gekomen bij het selectiegesprek zijn deels ook van toepassing op de praktijkopdracht. Wat betreft het aantal competenties is er bij de praktijkopdracht echter iets anders aan de hand. In de praktijkproeven worden slechts drie grote competentie-categorieën beoordeeld namelijk Denken, Voelen en Kracht. Het beoordelen van slechts drie categorieën zou het onderscheidend vermogen en de accurateid van de beoordeling ten goede moeten komen. De analyses laten echter een ander beeld zien aangezien de algemene indruk over een kandidaat 67 % van de variantie op de afzonderlijke categorieën verklaart. Dit houdt in dat de drie categorieën niet goed te onderscheiden zijn.

In het kader van het efficiënter maken van de selectieprocedure is door één van de geïnterviewden de vraag gesteld of het nodig zou zijn om de praktijkopdracht te handhaven. Op grond van de beschikbare gegevens is deze vraag moeilijk te beantwoorden. De predictieve validiteit van de praktijkopdracht voor de opleidingscriteria was niet hoog. De voorspellende kracht van de praktijkopdracht was echter niet slechter dan de andere onderdelen van de selectie zoals het selectiegesprek, de persoonlijkheidstesten en de cognitieve testen. De resultaten van het empirisch onderzoek lieten zien dat bij opsplitsing naar instroomniveau, de

algemene score op de praktijkopdracht zelfs de beste voorspeller voor opleidingsprestatie op niveau 2 was. Dit was bovendien de hoogste positieve voorspellende waarde die is gevonden in het huidige project³.

Uitval van de opleiding werd door twee van de drie competentiegebieden die in de praktijkopdracht worden gemeten op significante wijze voorspeld (de derde vertoonde een trend met $p = .09$). In dit opzicht deed de praktijkopdracht het dus iets beter dan het selectiegesprek. De veiligste conclusie die op dit moment te trekken valt is dan ook de volgende:

Conclusie 35: De gegevens laten *niet* duidelijk zien dat de praktijkopdracht als selectieonderdeel slechter zou werken dan de andere selectieonderdelen. In sommige opzichten doet de proef het zelfs iets beter. Onze aanbeveling is dan ook om de praktijkopdracht in de selectieprocedure te behouden. In elk geval tot meer informatie beschikbaar is.

11.3 Utiliteitsanalyses van de selectieprocedure

Bij de evaluatie van de selectieprocedure zou het nuttig zijn om gebruik te maken van zogeheten *utiliteitsanalyses*. Hierbij wordt er een inschatting gemaakt van de bijdrage van de selectieprocedure aan de kwaliteit van het personeel (bijvoorbeeld in termen van percentage geselecteerde dat succesvol blijkt te zijn), en ook van de financiële gevolgen van de selectieprocedure.

Wij stellen dan ook voor om in de toekomst gebruik te maken van dit soort technieken. Oorspronkelijk was het plan om in het huidige project met behulp van de selectiegegevens een indicatie te geven van de utiliteit. Echter, in de loop van het project werd duidelijk dat de mogelijkheden hiertoe beperkt waren. Er zijn namelijk geen goede criteriummaten beschikbaar. In onderstaande paragrafen zal één van de bekendste utiliteitsmodellen worden beschreven en zal worden aangegeven hoe deze kan worden toegepast bij de selectie van aspirant-agenten.

11.3.1 Taylor-Russell

Het meest eenvoudige utiliteitsmodel is dat van Taylor-Russell (1939). In dit model gaat men uit van een tweedeling van kandidaten, namelijk zij die geschikt zijn voor een specifieke functie (succesvol zouden functioneren) en zij die ongeschikt zijn voor die functie (geen goede agenten zouden zijn). Deze indeling is enigszins kunstmatig, want in de praktijk is het beter om te spreken van een continuüm van geschiktheid. Als het gaat om politiewerk zou men bijvoorbeeld kunnen zeggen dat er een proportie kandidaten is die zeker geschikt zou zijn voor politiewerk, maar dat er ook een proportie mensen is die zeker niet geschikt is voor politiewerk. De meeste mensen zullen echter ergens in het middengebied zitten. Dat wil zeggen dat zij tot op zekere hoogte succesvol zouden kunnen functioneren als agent.

³ Strikt genomen werd de hoogste voorspellende waarde gevonden bij de cognitieve test bij allochtone aspirant-agenten, namelijk $r = -.21$ (ongecorrigeerd), maar deze waarde was negatief hetgeen impliceert dat hoe slimmer men is hoe minder kans men heeft om te slagen op de tests.

Voor de berekeningen van utiliteit met behulp van de Taylor-Russell methode is het nodig om de dichotomie van geschiktheid versus ongeschiktheid aan te houden. Verder zijn er in dit model enkele sleutelvariabelen die een rol spelen bij de bepaling van de utiliteit van de selectie. De eerste is de proportie van kandidaten die in principe geschikt zou zijn voor het werk. Deze proportie wordt ook wel de toevalsratio of *base rate* genoemd. Deze proportie van potentieel geschikte kandidaten moet men inschatten, bijvoorbeeld op grond van eerder onderzoek, door redenering, of op grond van overleg tussen experts.

De tweede sleutelvariabele is de selectieratio. Met andere woorden, welk percentage aangemelde kandidaten wordt daadwerkelijk aangenomen. Over het algemeen is de selectieratio bekend. Bij de politie is bijvoorbeeld vastgesteld dat in de huidige selectieprocedure ongeveer 1 van de 10 á 11 kandidaten door de selectie heen komt. In dit geval is de selectieratio dus ongeveer .10.

De laatste sleutelvariabele betreft het verband tussen de instrumenten die gebruikt worden in de selectie en de uitkomstmaten oftewel de correlatie tussen de voorspellers en de criteria. De selectieprocedure als geheel kan worden geëvalueerd door de totale verklaarde variantie van de criteria door de voorspellers te bepalen. Deze laatste methode is toegepast in het huidige project.

Indien de waarde van de drie sleutelvariabelen zijn vastgesteld kan met behulp van Taylor-Russell tabellen worden nagaan wat de utiliteit is van een selectieprocedure. Deze wordt weergegeven aan de hand van een vergelijking van de proportie aangenomen geschikte personen versus de proportie geschikte personen die zouden zijn aangenomen als kandidaten aselect toegelaten zouden worden.

11.3.1.1 Voorbeelden uitwerking Taylor-Russell utiliteitsberekeningen

Hieronder staan enkele uitgewerkte voorbeelden waarin de utiliteit is berekend. De voorbeelden geven niet de directe utiliteit van de selectieprocedure bij de politie weer. Ze zijn eerder ter illustratie van hoe het model gebruikt kan worden.

- Uit statistische regressie-analyse blijkt dat de (multiple) correlatie (voorspellende waarde) tussen enerzijds de voorspellers en anderzijds de criteria (in dit geval: proeven van bekwaamheid) uitkomt op ongeveer $r = .10$.

- De proportie kandidaten in de selectieprocedure dat potentieel geschikt zou zijn voor de functie is moeilijk exact in te schatten. Daarom werken we voorbeelden uit waarin we een inschatting maken van .30, waarbij we ervan uitgaan dat 30% van de kandidaten die zich melden voor de selectie potentieel geschikt is.

- De selectieratio is .10 (10% van de kandidaten wordt tot de opleiding toegelaten).

Bij een selectieratio van .10, een toevalsratio (base rate) van .30 en een voorspellende waarde van .10, kan de utiliteit worden berekend met behulp van de Taylor-Russell tabellen: de utiliteit

is gelijk aan .36. Dat wil zeggen, bij aanname van 100 kandidaten blijken er hieronder 36 geschikt te zijn. Omdat er bij aselecte aanname van 100 kandidaten (zonder selectieprocedure) 30 geschikt zouden zijn voor het werk (dit is gelijk aan de toevalratio), blijkt dat door de invoering van de selectie er meer geschikten zijn geselecteerd (36 in plaats van 30).

In bovenstaand voorbeeld is uitgegaan van een multiple correlatie van .10, gebaseerd op een regressieanalyse. Stel nu echter dat de onderdelen van de selectie gezamenlijk .50 gecorreleerd waren met de criteria terwijl de selectieratio nog steeds gelijk is aan .10 en de toevalratio nog steeds gelijk is aan .30, dan zou dat betekenen dat door de selectieprocedures, het aantal geschikten onder de 100 toegelaten kandidaten stijgt van 30 (toevalratio) naar 52 (Dit kan afgeleid worden uit de Taylor-Russel tabellen). In dit geval gaat het dus om een winst van 22 geschikte kandidaten door het toepassen van de selectieprocedure.

Zoals vermeld zijn bovenstaande voorbeelden slechts illustraties van de wijze waarop men de Taylor-Russell methode zou kunnen toepassen bij de selectie van de politie. De uitkomsten kunnen niet opgevat worden als de daadwerkelijke utiliteit van de selectieprocedure omdat de kwaliteit van de gebruikte criteriummaten onduidelijk is.

Utiliteitsberekeningen, mits deze worden uitgevoerd met de juiste en betrouwbare gegevens, kunnen zinvol zijn om een inschatting te maken van de financiële gevolgen van de selectieprocedure. Dit is echter alleen mogelijk als er een reële inschatting te maken valt van de diverse soorten kosten die het aannemen van ongeschikte aspiranten voor politiewerk met zich meebrengt (e.g., kosten uitval opleiding, vervanging personeel etc.) en de financiële opbrengsten van succesvol functionerend politiepersoneel.

11.4 Keuze selectie-instrumenten en kosten

Eén van de beoogde doelstellingen van het huidige project betrof het doen van aanbevelingen voor het gebruik van tests in de selectieprocedure en het eventueel inzicht te geven in de kosten van tests. Met het voortschrijdend inzicht in de loop van het project werd echter duidelijk dat met de informatie die tot onze beschikking stond, er slechts beperkte uitspraken mogelijk zijn. Er zijn verschillende redenen die leiden tot deze conclusie:

De eerste, en voornaamste reden is dat er beperkte criteriuminformatie aanwezig was. In een selectieonderzoek zijn goede criteria een essentiële voorwaarde. Zonder goede criteriuminformatie kan er weinig geconcludeerd worden over de voorspellende waarde van specifieke selectie-instrumenten en/of onderdelen.

Een tweede reden dat we geen specifieke voorstellen hebben gedaan voor vernieuwde of efficiëntere tests is dat naar onze mening de meeste tests die worden gebruikt in de huidige selectie voldoende psychometrische eigenschappen bevatten, voor zover dit is vast te stellen met de ons beschikbare informatie. Voor het meten van persoonlijkheid wordt bijvoorbeeld gebruik gemaakt van de NEO-PI-R, één van de meeste bekende, gebruikte, en gevalideerde tests die op dit gebied beschikbaar is. Ook ander meetmethoden (selectiegesprek, praktijkopdracht) lijken zorgvuldig opgezet. Zo zijn de maten in deel A en B ontworpen volgens

recente wetenschappelijke kennis op dit gebied en zijn ze in veel gevallen gestandaardiseerd (met mogelijke uitzondering de cognitieve test).

Een derde reden is dat duidelijk is geworden dat in de totale kosten van selectie, de kosten van de tests slechts een klein deel beslaan. Ter illustratie, schattingen betreffende de kosten van selectie voor de politie liggen ergens rond de 1500 euro per kandidaat. Van dit bedrag bestaat ongeveer 30 euro uit de kosten voor de tests. Dit is slechts 2% van de 1500 euro. Zelfs wanneer een aanzienlijke reductie in testkosten mogelijk zou zijn, bijvoorbeeld wanneer de kosten met een derde verminderd zouden kunnen worden, zou dit slechts een besparing van 0.7% van de totale kosten betekenen.

Gezien bovenstaande zou er meer kostenefficiënt gewerkt kunnen worden wanneer i) bepaalde onderdelen van de selectie verwijderd zouden worden uit de procedure of ii) dat sommige onderdelen zouden worden ingekort. Gezien de huidige resultaten van de analyses en andere beschikbare informatie kan er ons inziens geen onderbouwde argumentatie worden gegeven waarom sommige onderdelen ingekort of verwijderd zouden dienen te worden. Een uitzondering hierop is wellicht de taaltoets (zie onze aanbeveling hierover in hoofdstuk 7).

Hoofdstuk 12

Norm en cut-off scores

12. Norm- en cut-off scores

De psychometrische eigenschappen van de instrumenten zijn essentieel voor de kwaliteit van de selectieprocedure. Maar de norm en cut-off scores van de selectieonderdelen zijn evenzo belangrijk. In theorie zou zelfs wanneer men een perfect betrouwbaar en valide instrument heeft om prestatie te voorspellen, de selectie nog suboptimaal kunnen zijn als men de norm verkeerd legt. Het bepalen van goede normscores is over het algemeen moeilijk. De reden daarvoor is dat er vaak geen objectieve norm valt vast te stellen en er veel factoren betrokken zijn in besluiten over normscores. Financiële, politieke, morele en empirische kwesties spelen allen een rol.

Daarnaast is er ook in de wetenschappelijke selectieliteratuur nog discussie over het gebruik van normscores (Guion, 2011). Eén van de vragen hieromtrent is onder welke omstandigheden men überhaupt gebruik moet maken van relatieve normscores en wanneer men beter met absolute scores kan werken. Een argument tegen het gebruik van relatieve normscores is dat er vaak een minimale competentie nodig is om bepaalde taken uit te voeren. Als een baan bijvoorbeeld vereist dat objecten van rond de 80 kg handmatig verplaatst moeten worden dan heeft het wat dit betreft geen zin om relatieve normen te stellen (e.g., afhankelijk van leeftijd, geslacht). Men zou een soortgelijk voorbeeld kunnen verzinnen maar dan met cognitieve capaciteit of sociale vaardigheden in plaats van het verplaatsen van gewicht.

Twee argumenten *voor* het gebruik van relatieve normen is dat in veel beroepen, zoals bij politiewerk, de eisen niet zo duidelijk afgebakend zijn als in bovenstaand voorbeeld en dat het hanteren van een absolute norm kan leiden tot een sterke adverse impact. Dat wil zeggen, dat individuen uit bepaalde groepen minder kans hebben om aangenomen te worden.

Daar komt bij dat er door de veelzijdigheid van het politiewerk het waarschijnlijk mogelijk is om te compenseren voor minder ontwikkelde competenties. Alhoewel iedere agent in staat moet zijn om alle politietaken naar behoren uit te voeren blijkt er in de praktijk wel differentiatie te ontstaan. Zo werd uit één van de interviews duidelijk dat zich in de praktijk de volgende situatie voor kan doen:

Wanneer er iemand achtervolgd moet worden of wanneer de kans groot is dat geweld gebruikt moet worden stuurt men de grootste of meeste daadkrachtige agent als eerste op pad. In situaties die sociale hulp of de-escalatievaardigheden vereisen stuurt men daarentegen eerder de agent op pad die wat meer sociaal intelligent is.

Dit voorbeeld laat zien dat er, zoals bij veel andere banen, dus een zekere mate van *compensatie in competenties* mogelijk is bij de politie.

In de huidige selectieprocedure bij de politie maakt men reeds gebruik van de mogelijkheid tot compensatie. Sommige aspirant-agenten zullen relatief daadkrachtig zijn maar iets minder goed zijn in het inleven in anderen. Andere aspirant-agenten zullen wat beter scoren op sociale vaardigheden terwijl ze qua krachtig optreden minder goed scoren. Zolang er maar aan de minimale eisen is voldaan, kunnen dit soort kandidaten echter toch aangenomen worden.

Een belangrijke vraag is of de ondergrenzen en normscores die men hanteert in de huidige selectieprocedures afdoende zijn.

Wat betreft aandachtsgebieden bij de politie zoals psychische weerbaarheid en daadkrachtig optreden zou men zich bijvoorbeeld kunnen voorstellen dat simpelweg de normscores of ondergrenzen worden verhoogd. Wat men hierbij moet beseffen is dat dit negatieve gevolgen kan hebben voor de bedrijfsvoering. De huidige selectieratio bij de politie ligt bijvoorbeeld ergens rond de 10 %. Dit betekent dat voor elke 100 aspirant-agenten die men wilt aannemen er 1000 kandidaten moeten worden getest. Dit brengt vaste financiële kosten met zich mee. Wanneer men de ondergrenzen relatief laag houdt zullen er meer kandidaten doorstromen en de selectieratio hoger zijn waardoor de selectiekosten laag blijven.

Daar staat tegenover dat het hanteren van soepele normen kan betekenen dat sommige kandidaten die doorstromen achteraf toch problemen zullen opleveren. Hier zitten ook kosten aan vast zoals psychische begeleiding, extra training, verloren opleidingskosten of ziektekosten.

In theorie zou het verhogen van de ondergrenzen op de relevante competenties leiden tot een lager percentage van aspirant-agenten die problemen zullen ondervinden. Men kan bijvoorbeeld besluiten om alleen de meest emotioneel stabiele en daadkrachtige kandidaten aan te nemen. Het gevolg daarvan is echter dat de selectieratio daalt. Om 100 aspirant-agenten te vinden zou men nu bijvoorbeeld structureel 2000 kandidaten moeten testen. Dit betekent tevens een structurele stijging in de kosten van de selectie. Of men als organisatie al dan niet bereid is om deze kosten te maken gaat verder dan wetenschappelijke selectie-aspecten en heeft er onder andere mee te maken in hoeverre men bepaalde uitkomsten al dan niet als probleemgebieden in de organisatie ziet. Een specifiek voorbeeld daarvan is stressbestendigheid. Dit voorbeeld wordt hieronder uitgewerkt.

12.1 Stressbestendigheid en daadkracht: Een concreet voorbeeld.

In de startnotitie van het WODC (2012), die aanleiding was voor het huidige project, werd aangegeven dat verminderde psychische belastbaarheid een relevant probleem zou kunnen zijn in de politieorganisatie. Volgens de uitkomsten van een rapport uit 2010 zou 20 tot 30% van de politiemedewerkers te maken hebben met verminderde psychische belastbaarheid. Op grond van deze uitkomsten heeft het ministerie besloten aanvullend onderzoek op dit gebied te financieren en te laten onderzoeken welke mogelijkheden er zijn om de belastbaarheid van agenten te verhogen. Verminderde psychische belastbaarheid werd in deze context dus gezien als een relevant probleem in de bedrijfsvoering van de politie dat grote kosten met zich meebrengt.

In de looptijd van het huidige project is echter ook duidelijk geworden dat het probleem omtrent verminderde psychische belastbaarheid wellicht minder groot is dan men aanvankelijk had gedacht. Nieuwe gegevens suggereren dat de psychische belastbaarheid van politiepersoneel niet slechter is dan die van personeel in veel andere bedrijfstakken en dat indien er sprake is van verminderde belastbaarheid dit vaak te maken heeft met organisatorische stressoren (Van der Velden, 2012).

Bovenstaande maakt duidelijk dat er binnen de politieorganisatie momenteel onduidelijkheid is over de status van verminderde psychische belastbaarheid. Dientengevolge zal ook onduidelijk

zijn hoeveel financiële middelen men beschikbaar zou willen stellen in de selectie om problemen met verminderde belastbaarheid te voorkomen. Over de ondergrens of normscore van competenties kunnen we de volgende resultaten presenteren met betrekking tot psychische belastbaarheid:

De competentie uit de selectieprocedure die direct te maken heeft met psychische belastbaarheid is 'Stressbestendigheid' zoals gemeten in het selectiegesprek. In de notitie van de Politieacademie uit 2002 is aangegeven dat kandidaten minimaal een score van 2 moeten halen om doorgelaten te worden. Analyse op de data van ongeveer 30.000 kandidaten die de afgelopen jaren zijn getest bij de Politieacademie laat zien dat met gebruikmaking van die eis 1.4 % van de kandidaten zonder meer afgewezen is. Herziening van dit criterium heeft plaatsgevonden waarbij men minstens een 3 moest scoren (score 1 en 2 leiden direct tot afwijzing). In deze situatie wordt 8.4 % van de kandidaten direct afgewezen op dit criterium. Het aantal kandidaten dat door deze maatregel wordt afgewezen is dan 6 maal zo groot!

Stel dat we in het huidige rapport zouden voorstellen om de ondergrens voor stressbestendigheid *nog* hoger te leggen en alleen de meest stressbestendigste kandidaten door te laten die 4 of hoger scoren. In dat geval zou maar liefst *34% van alle kandidaten uitvallen* op deze competentie.

Nu is stressbestendigheid uiteraard niet de enige voorspeller van belang in de selectie van politieagenten. Een ander mogelijk problematisch gebied is daadkracht (zie ook hoofdstuk 1). Stel nu voor dat naast stressbestendigheid ook de ondergrens van Kracht met een punt zou worden opgehoogd. Als de ondergrens van Kracht 2 is, wordt slechts een tiende procent (= 0.1%) van de kandidaten afgewezen. Wanneer de ondergrens wordt verhoogd naar 3 zal 4.3 % worden afgewezen.

Het belangrijkste punt hierbij is dat een verhoging van de ondergrens van stressbestendigheid naar 3 in combinatie met een verhoging van Kracht naar 3, zou betekenen dat 65 % van alle kandidaten zal worden afgewezen op deze twee competenties. Dit is ongeveer zoveel als er normaal gesproken totaal in Deel B wordt afgewezen (dus inclusief alle competenties uit het gesprek en de praktijkopdracht). Met de huidige ondergrens van 2 op beide criteria is dit 5.5 %.

Conclusie 36: Het verhogen van ondergrenzen van competenties zorgt er waarschijnlijk voor dat kandidaten die worden toegelaten minder kans hebben op een aantal problemen zoals verminderde psychische weerbaarheid, conflict etc. Men moet echter beseffen dat er bij verandering van ondergrenzen sprake is van een trade-off. De negatieve kanten van de trade-off zijn een hoger afwijzingspercentage en hogere selectiekosten.

12.2 Normscores en uitkomsten predictieve validiteitsstudie.

Naast de complexiteit van beslissingen omtrent ondergrenzen en normscores, is er nog een andere reden dat het niet direct zinvol is om in dit stadium concrete voorstellen te doen voor veranderingen in normscores en/of ondergrenzen. Deze reden is de relatief lage predictieve validiteit van de selectieonderdelen zoals in dit project naar voren komt. Geen van de

onderzochte selectieonderdelen verklaarden meer dan enkele procenten van de variantie in uitval of prestatie tijdens de opleiding. De mogelijke verklaringen hiervoor zijn in voorgaande hoofdstukken reeds behandeld.

Gezien deze lage validiteiten kan men verwachten dat een verandering in ondergrenzen en/of normscores slechts marginale effecten zal hebben op de daadwerkelijke uitval en prestatie terwijl de selectieratio wel sterk zal stijgen. We kunnen dit relateren aan bovenstaand voorbeeld waarin we spraken over stressbestendigheid en daadkracht. De correlaties tussen die twee competenties en opleidingsprestatie waren zeer zwak. Deze zijn hieronder nogmaals weergegeven (Tabel 12.1):

Tabel 12.1 Correlaties tussen competenties (gesprek en praktijkopdracht) en prestatie op toetsen

	Onderdelen proeven	Eindoordeel proeven
Emotionele Weerbaarheid	.02 (.02)	-.02 (-.02)
Kracht	.05* (.06)	.02 (.02)

** Correlatie is significant op het 0.01 niveau

* Correlatie is significant op het 0.05 niveau

De waarden tussen haakjes zijn de correlaties, gecorrigeerd voor range restriction.

Indien men op grond van deze resultaten zou besluiten om aanpassingen te doen in de ondergrenzen dan valt te zien in paragraaf 12.1 dat dit een relatief sterk effect heeft de selectieratio's, bijvoorbeeld stijgingen van 4 tot 6 maal zoveel kandidaten die worden afgewezen op deze criteria. Gezien de zwakke verbanden met de criteria zullen deze veranderingen echter een minimaal effect hebben op opleidingsprestatie omdat bijvoorbeeld de competentie kracht slechts voor een kwart procent ($0.05 * 0.05 = 0.0025$) verantwoordelijk is voor de variantie in prestatie op deze criteria.

Conclusie 37: De beschikbare gegevens over de selectieonderdelen en de criteria kunnen nauwelijks dienen als uitgangspunt voor beslissingen omtrent ondergrenzen of normscores.

Deel 4

Proces- en beleidsmatige aspecten van de selectie

Hoofdstuk 13
**Proces- en beleidsmatig aspecten van de
selectieprocedure**

13. Proces- en beleidsmatige aspecten van de selectieprocedure

Voorafgaande hoofdstukken waren vooral gericht op de psychometrische kwaliteiten van de selectieonderdelen. Daarnaast zijn uiteraard nog andere proces- en beleidsmatige aspecten die een rol spelen. Enkele daarvan zijn relevant voor het huidige project en bieden mogelijkwerwijs aanknopingspunten voor aanpassingen in de selectieprocedure. Deze worden hieronder puntsgewijs besproken.

13.1 Opleiding als verlenging van selectiemechanisme

Nadat kandidaten zich aangemeld hebben voor een baan bij de politie moeten zij een aantal selectiestappen doorlopen. In elk van die stappen wordt bepaald of de kandidaat doorgaat. Er is hier dus sprake van het zogenaamde multiple-hurde model.

Kandidaten kunnen uitvallen omdat ze de 'verkeerde' motivatie hebben, omdat hun intelligentie of schrijfvaardigheid, of lichamelijke conditie niet voldoende is voor een baan bij de politie, of omdat ze enkele persoonskenmerken bezitten waarvan men veronderstelt dat deze de adequate uitvoering van politiewerk in de weg staan. Tot slot kunnen kandidaten uitvallen bij de medische keuring of door het antecedentenonderzoek.

Over het algemeen kan men dus aannemen dat na het doorlopen van al deze stappen in de selectieprocedure, degenen met de grootste risico's voor disfunctioneren als agent reeds zijn afgewezen. Voor een deel wordt dit idee ondersteund door de cijfers. In vergelijking met het reguliere onderwijs is de uitval bij de Politieacademie laag, zo'n 15 tot 20 %. Daarbij moet overigens wel worden aangemerkt dat een rapport over de selectieprocedure bij de politie uit 1990 (Wilbrink et al.) aangeeft dat de uitval in die periode rond de 10% lag, wat zou betekenen dat het uitvalspercentage sinds die periode bijna is verdubbeld. Waarschijnlijk is dit echter meer een algemeen maatschappelijke trend in onderwijs dan één die specifiek is voor de politieopleiding. Een andere mogelijke verklaring voor het verschil met de uitvalspercentages in 1990 is dat de politieopleiding tegenwoordig dual is. In de nieuwe vorm van de opleiding wisselen onderwijs en de praktijk elkaar af en duurt de opleiding 3 a 4 jaar. In 1990 bestond de opleiding uit een onderwijsperiode van rond de 18 maanden. De langere duur van de opleiding en het feit dat aspirant-agenten tevens 2 a 3 jaar werken als agent kan dus ook te maken hebben met de verhoogde uitval.

Wanneer kandidaten alle selectiestappen met succes hebben doorlopen (inclusief het eindgesprek) krijgen zij een aanstelling bij de politie. Vroeger kregen de kandidaten direct een aanstelling bij één van de korpsen. In de huidige situatie is er maar één Nationaal politiekorps en krijgen kandidaten een aanstelling bij één van de operationele eenheden. Aanvankelijk omvatte dit contract in principe een vaste aanstelling. Zeer recent is de politie echter overgegaan op het aanbieden van tijdelijke contracten tijdens de opleiding.

Alhoewel kandidaten die de selectie hebben doorlopen zich zeker hebben gekwalificeerd om opgeleid te worden tot agent is het belangrijk te benadrukken dat geen enkele selectieprocedure geheel 'waterdicht' is. De selectieprocedure biedt een momentopname. In de korte interactie tussen kandidaten en selecteurs worden niet altijd *alle* stabiele risicofactoren waargenomen. Het gevolg hiervan is dat sommige kandidaten die worden

toegelaten tot de politieopleiding achteraf niet afdoende zullen functioneren tijdens de opleiding. Deze kandidaten zullen mogelijk laat in de opleiding alsnog uitvallen. Wellicht maken zij de opleiding wel af maar zullen later in de beroepsuitvoering tegen problemen aanlopen.

Eén van de mogelijkheden om dit soort situaties te voorkomen zou kunnen liggen in het gebruik van de opleiding als gedeeltelijke verlenging van de selectie. Het idee is dat mogelijke problemen die niet zijn opgemerkt tijdens de selectie wel tot uiting zullen komen tijdens de opleiding. Kandidaten kunnen bepaalde negatieve kanten wellicht verborgen houden tijdens het selectiegesprek of de praktijkopdracht. Zulke negatieve kanten zouden echter moeilijker te verbergen zijn tijdens de opleiding die zich uitstrekt over maanden en waarin de aspirant-agenten dagelijks wordt blootgesteld aan uitdagende situaties (in opleiding of praktijk).

De politieacademie heeft sinds kort enkele stappen ondernomen om het principe van verlengde selectie tijdens de opleiding toe te passen. Eén maatregel betreft het invoeren van tijdelijke contracten voor aspirant-agenten. Een andere maatregel is het bindend studieadvies, waarbij aspirant-agenten uit de opleiding geweerd kunnen worden wanneer enkele tussentermen niet worden behaald. Daarnaast kunnen aspirant-agenten van de opleiding geweerd worden als er sprake is van duidelijk disfunctioneren of in geval van integriteitskwesaties. Uiteraard zijn er bij deze maatregel wel enkele waarborgen die tot doel hebben de aspirant-agenten te beschermen tegen ontslag om onterechte redenen.

Hoewel er dus formeel al sprake is van de invoering van verlengde selectie, werd in enkele interviews wel de suggestie gewekt dat het de praktijk moeilijk blijft om disfunctionerende aspirant-agenten van de opleiding te weren. Binnen de looptijd van het huidige project kon niet met zekerheid worden vastgesteld in welke mate het hier gaat om een algemeen aandachtspunt of juist om idiosyncratische indrukken van de geïnterviewden. Desalniettemin kan het nuttig zijn om het functioneren van de maatregelen op dit gebied goed te monitoren. Het is bijvoorbeeld relevant om na te gaan hoe vaak de hierboven genoemde (nieuwe) maatregelen daadwerkelijk in de praktijk worden toegepast. Verder zou kunnen onderzoeken hoe de betrokkenen (e.g., docenten op de academie, selecteurs) de effectiviteit van de maatregelen beoordelen.

Conclusie 38: Alhoewel de selectieprocedure bij de politie in grote lijnen goed lijkt te functioneren zijn er altijd kandidaten die door zullen stromen voor wie een politiefunctie achteraf niet past of die te grote risicofactoren vertonen wat betreft latere functie-uitvoering. De politieacademie heeft zeer recent enkele stappen ondernomen om flexibel te reageren op dit soort situaties, bijvoorbeeld door het aanbieden van tijdelijke contracten en het invoeren van een bindend studieadvies. Desalniettemin is het zinvol om goed te blijven monitoren in welke mate deze maatregelen in de praktijk op een adequate manier worden toegepast en of ze de gewenste uitkomsten hebben. Vooralsnog kon in het huidige project namelijk onvolledig worden vastgesteld in hoeverre de genoemde maatregelen goed werken.

13.2 Continuïteit en kosten van de bedrijfsvoering bij de afdeling werving en selectie

Het aantal kandidaten dat zich aanmeldt bij de Politieacademie fluctueert afhankelijk van een aantal factoren. Zo zijn er seizoenseffecten en effecten van het al dan niet actief werven voor een baan bij de politie. Zulke fluctuaties zorgen voor variatie in de drukte en werkzaamheden van de afdeling werving en selectie. De meeste factoren die fluctuatie in het aanbod van kandidaten veroorzaken zijn te voorspellen en/of te beïnvloeden. Blijkens de informatie die ons te beschikking stond en enkele interviews is echter duidelijk geworden dat er ook factoren zijn die op een minder voorspelbare wijze het aanbod van kandidaten beïnvloed. Een voorbeeld daarvan zijn de richtlijnen met betrekking tot het aannemen van aspirant-agenten in voorgaande jaren. Vanuit de politiek was er in 2011/2012 de eis om een groot aantal aspirant-agenten aan te nemen. De redenen hiervoor waren i) dat de met de Kamer afgesproken operationele sterkte van 49.500 fte moest worden gewaarborgd en ii) dat er afspraken uit het arbeidsvoorwaardenakkoord waren die stelden dat er in 2011 en 2012 een minimaal aantal aspiranten moest instromen (instroom van 3 x 1600 in de periode 2011-2012). De processen om dit te realiseren kwamen pas laat op gang. Hierdoor moesten in 2011 aanvullende maatregelen worden getroffen. Om aan de hogere instroomeis te voldoen moest er veel aanvullend personeel worden aangetrokken bij de afdeling werving en selectie. Er moesten bijvoorbeeld nieuwe selectiepsychologen worden aangenomen om de selectiegesprekken en praktijkopdrachten uit te voeren. Deze psychologen moesten ook opgeleid worden hetgeen tot een aanzienlijke aanvullende kostenpost heeft geleid.

Na de aanname van veel aspirant-agenten in 2011 en 2012, heeft de situatie zich voorgedaan dat er in 2013 blijkbaar minder behoefte aan nieuwe instroom van aspiranten is. Hierdoor worden er in 2013 veel minder aspirant-agenten geworven dan in de jaren daarvoor. Bovendien moeten kandidaten die reeds (deels) zijn geselecteerd, vaak enige maanden wachten voordat zij daadwerkelijk een aanstelling krijgen en kunnen toetreden tot de opleiding. De exacte reden voor het onverwachte overschot is niet geheel duidelijk geworden in het project. Er zijn aanwijzingen dat dit deels te maken heeft met het uitstellen van pensioen, een trend die blijkbaar in heel Nederland gaande is. Maar of dit de enige reden is kan niet worden bevestigd.

Uiteraard is het onvermijdelijk dat er soms onverwachte externe factoren zijn die fluctuatie in aanbod of werkzaamheden veroorzaken. Toch moet men zich goed bewust blijven van de mogelijke gevolgen hiervan.

Ten eerste zorgt een verhoogde aanname-eis dat er zeer grote verschillen ontstaan in de werkdruk en in de werkwijze. Een onverwachte verhoging van de aanname-eisen kan bijvoorbeeld met zich meebrengen dat men de aanstellingseisen versoepelt. Wanneer meer aspirant-agenten moeten worden aangenomen kan dat alleen bewerkstelligd worden wanneer:

- i) Het aanbod van kandidaten wordt verhoogd. Met de huidige selectieratio van ongeveer 1 op 10 zou dat betekenen dat voor elke extra kandidaat die wordt aangenomen er zich 10 extra kandidaten zouden moeten aanmelden.
- ii) De aanstellingseisen worden (formeel of informeel) verlaagd. In dit geval zou dat dus betekenen dat de kwaliteit van de kandidaten omlaag zou gaan.
- iii) De aanstellingseisen worden aangescherpt. Dat wil zeggen dat men er vanuit gaat dat in een eerdere situatie, er streng geselecteerd werd en er dus veel kandidaten zijn afgewezen die in feite goed hadden kunnen functioneren.

Welke van de bovenstaande strategieën is toepast is moeilijk exact vast te stellen. Wel is het goed zich ervan bewust te zijn dat maatregelen i en ii nadelen met zich meebrengen. In maatregel i gaat het om hogere selectiekosten, en in maatregel ii betreft het een verhoging van de zogeheten false-negatives, dat wil zeggen hogere instroom van kandidaten die in principe ongeschikt zouden zijn voor het betreffende werk.

Bovenstaand voorbeeld dient ter ondersteuning voor de volgende conclusie:

Conclusie 39: Het is wenselijk dat wanneer het gaat om het invoeren van nieuw of tijdelijk beleid men rekening houdt met de regelmaat in de bedrijfsvoering en mogelijke financiële gevolgen.

Deel 5

Nieuwe constructen in de selectieprocedure?

Hoofdstuk 14

Nieuwe constructen in de selectieprocedure?

14 Nieuwe constructen in de selectieprocedure?

In het huidige hoofdstuk wordt een aantal constructen besproken die mogelijk relevant zijn voor personeelsselectie en daarmee wellicht praktische implicaties hebben. We bespreken achtereenvolgens, de algemene factor van persoonlijkheid, emotionele intelligentie, en de duistere zijde van persoonlijkheid.

14.1 De algemene factor in persoonlijkheidsdata

Gezien de, over het algemeen, relatief lage validiteitscoëfficiënten van de Big Five bij het voorspellen van politiewerk (zie Hoofdstuk 8) vonden we het zinvol om te kijken naar mogelijk nieuwe ontwikkelingen op het gebied van persoonlijkheid. Eén van die nieuwe ontwikkelingen is de zogeheten algemene factor van persoonlijkheid, in het Engels ook wel aangeduid als de *General Factor of Personality* (Figueredo et al., 2006; Musek, 2007; Rushton & Irwing, 2011; Van der Linden, Te Nijenhuis, Bakker, 2010). De afkorting daarvan is de *GFP* die ook in het huidige rapport zal worden gebruikt. Het idee van de GFP is onder andere gebaseerd op de empirische observaties dat de Big Five vaak onderlinge correlaties vertonen. Dit in tegenstelling tot het idee dat deze dimensies conceptueel onafhankelijk van elkaar zijn.

Correlaties tussen persoonlijkheidsdimensies blijven overigens niet beperkt tot de Big Five dimensies maar zijn inmiddels ook aangetoond in zo goed als elk ander persoonlijkheidsmodel of persoonlijkheidsmeting (zie Rushton & Irwing, 2011 voor een overzicht). De correlaties tussen persoonlijkheidsfactoren zorgen ervoor dat er een gemeenschappelijk deel is in persoonlijkheidsmaten dat overlapt met de meeste persoonlijkheidsdimensies. Dit valt te vergelijken met de algemene factor die wordt gevonden in cognitieve capaciteit. In het cognitieve domein kan men capaciteit meten met verschillende soorten tests, zoals ruimtelijk inzicht, verbaal begrip, redeneervermogen, etc. De onderlinge correlaties tussen de cognitieve tests zorgen er echter voor dat er een globale factor te onderscheiden is die algemene intelligentie weergeeft.

Er dient te worden opgemerkt dat in de huidige wetenschappelijke literatuur de status van de GFP nog onderhevig is aan discussie. Grofweg zijn er twee interpretaties te onderscheiden. De eerste is de zogenaamde substantieve verklaring van de GFP. Hierin wordt aangenomen dat de GFP een mix van sociaal wenselijke eigenschappen is. Personen die hoog scoren op de GFP zouden over het algemeen makkelijk in de omgang zijn en omschreven kunnen worden als personen die een open houding hebben, betrouwbaar, productief, sociaal, aardig en emotioneel stabiel. In de literatuur zijn er aanwijzingen dat een hoge score op de GFP samenhangt met een hoge emotionele intelligentie (Van der Linden, Tsaousis, Petrides, 2012). De redenering hierachter is dat wanneer iemand emotioneel intelligent is dit een zeer brede invloed zou hebben op gedrag. Zodoende zou het tevens de scores op alle dimensies in het Big Five model op een positieve manier beïnvloeden wat leidt tot een GFP. Loehlin (2012) omschreef de GFP als een factor die indicatief is voor de mate van sociale participatie. Dit is mogelijk relevant voor politiemedewerkers aangezien van hen verwacht wordt dat zij midden in de samenleving staan en dienstbaar zijn voor de gemeenschap.

De alternatieve verklaring voor de GFP is dat deze factor niet veel meer is dan een weergave van vertekeningen in het meten van persoonlijkheid (e.g., Revelle & Wilt, 2013). Persoonlijkheid wordt vaak gemeten door vragenlijsten en doordat sommigen zich op zo'n vragenlijst beter voor willen doen dan ze daadwerkelijk zijn geven ze zichzelf betere scores op alle dimensies. In deze zogenaamde meetfoutverklaring wordt de GFP gezien als een onwenselijk bijproduct van het meten van persoonlijkheid en heeft als zodanig weinig theoretische of praktische waarde.

De wetenschappelijke literatuur is nog niet eenduidig over de verschillende verklaringen. Enkele onderzoeken lijken de meetfoutverklaring van de GFP te ondersteunen terwijl andere onderzoeken juist meer richting de substantieve verklaring gaan (Voor een overzicht zie bijvoorbeeld: Just, 2010; Rushton & Irwing, 2011). Wat betreft dat laatste is er bijvoorbeeld gevonden dat de GFP een betrekkelijke goede voorspeller is voor werkprestatie (Sitser, Van der Linden, Born, 2013; Van der Linden et al., 2010). Prestatie was daarbij gemeten door de mening van de directe leidinggevende te vragen of door objectieve uitkomsten.

De relatie tussen de GFP en prestatiebeoordelingen door de leidinggevende suggereren dat het hier niet alleen zou gaan om meetfouten in persoonlijkheid maar dat personen die hoog op de GFP scoren daadwerkelijk gedrag vertonen dat door andere observeerbaar is.

Om bovengenoemde reden is het nuttig om na te gaan of de GFP mogelijk een rol zou kunnen spelen in de selectie van politiepersoneel. Een eerste voorwaarde daarvoor is dat de GFP terug te vinden moet zijn in de persoonlijkheidsdata van de politie. Dit hebben we getoetst.

Als initiële toets voor de GFP zijn de onderlinge correlaties tussen de Big Five dimensies, zoals deze zijn gemeten bij de politie, geanalyseerd. Dit laat zien dat er substantiële correlaties zijn tussen de dimensies (zie tabel 14.1). Vervolgens is een factoranalyse uitgevoerd op de scores van de vijf dimensies. De resultaten van deze factoranalyse zijn te zien in Tabel 14.2 en ondersteunen dat er één grote factor te identificeren valt die bijna 50 % van de variantie in de Big Five factoren verklaart.

De factorladingen van de Big Five dimensies op de algemene factor zijn weergegeven in tabel 14.3. Deze factorladingen geven aan in hoeverre de dimensies bijdragen aan de algemene factor. De tabel laat duidelijk zien dat alle Big Five dimensies substantieel bijdragen aan de algemene factor en dat de richting van de ladingen in overeenstemming is met de theorie van de GFP. Specifiek, de algemene factor lijkt een reflectie te zijn van sociaal wenselijke eigenschappen gekenmerkt door een hoge mate van consciëntieusheid, extraversie, aardigheid en emotionele stabiliteit en in iets mindere mate openheid.

Tabel 14.1. Correlatieanalyse op de Big Five Dimensies

	Openheid	Consciëntieusheid	Extraversie	Altruïsme	Neuroticisme
Openheid					
Consciëntieusheid	.15**				
Extraversie	.92**	.37**			
Altruïsme	.14**	.32**	.21**		
Neuroticisme	.88**	-.60**	.88**	-.36**	

Tabel 14.2. Factoranalyse op de Big Five Dimensies van de NEO-PI-R

Component	Eigenwaarde	Verklaarde variantie (Cumulatief)
1	2.35	47.04
2	.96	66.26
3	.73	80.94
4	.57	92.35
5	.38	100.00

Tabel 14.3. Factorladingen van de Big Five dimensies op de algemene factor

	GFP
Openheid	.46
Conscientieusheid	.78
Extraversie	.71
Altruïsme	.62
Neuroticisme	-.79

Uit bovenstaande analyses wordt duidelijk dat er ook in de persoonlijkheidsdata van de politie een overtuigende GFP aanwezig is. Gezien de verschillende verklaringen voor de GFP in de literatuur is het relevant om te toetsen in hoeverre deze factor samenhangt met kandidaatsbeoordelingen uit de andere onderdelen van de selectie, te weten het selectiegesprek en de praktijkopdracht. Uit tabel 14.4 blijkt dat de score op de GFP $r = .28$ samenhangt met de algemene indruk uit het selectie-interview. Een correlatie van $.28$ is van gemiddelde grootte. Men moet hierbij echter beseffen dat het gaat om verschillende methoden en dat de kenmerken die worden gemeten in het interview niet hetzelfde zijn als in de persoonlijkheidsvragenlijst. Onderlinge correlaties tussen beoordelaars op dezelfde aspecten (bijvoorbeeld extraversie) hebben vaak slechts een waarde van rond de $r = .40$. In dit licht laat de correlatie van $.28$ zien dat er redelijke overeenstemming is en dat de GFP niet alleen terug te vinden is in de persoonlijkheidsvragenlijst maar ook gedeeltelijk terugkomt in de beoordeling uit het selectiegesprek. De relatie met de algemene indruk uit de praktijkopdracht was $.13$. Dit is lager maar zoals te zien is in hoofdstuk 7 werd de algemene score op de praktijkopdracht vooral bepaald door intelligentie en probleemaanpak; de relatie daarvan met persoonlijkheid is nog onduidelijk.

Tabel 14.4. Intercorrelaties tussen algemene indrukken uit persoonlijkheidsvragenlijst, selectiegesprek en praktijkopdracht

	1.	2.	3.
1. Persoonlijkheidsfactor	-		
2. Interviewfactor	.28	-	
3. Praktijkfactor	.13	.50	-

Er kan geconcludeerd worden dat de GFP aanwezig is in de persoonlijkheidsdata van de politie en dat de relatie met de beoordelingen uit de andere selectiemethoden (i.e., interview, praktijkopdracht) hiermee samenhangen. Dit laatstgenoemde resultaat levert enige ondersteuning voor de substantieve verklaring van de GFP omdat deze factor blijkbaar niet

alleen terug te vinden is in persoonlijkheidsmetingen (e.g., door meetfouten) maar ook gedeeltelijk tot uiting komt tijdens het gesprek en de praktijkopdracht.

Een univariate analyse is uitgevoerd met als onafhankelijke variabele opleidingsstatus (nog studierend, uitval). De GFP was de afhankelijke variabele. Geen van de effecten was significant ($F_{\text{opleidingsstatus}(1, 1986)} = .31, ns.$; $F_{\text{instroomniveau}(1, 1986)} = .72, ns.$; $F_{\text{interactie}(2,186)} = 2.53, ns.$). Er waren dus geen verschillen in GFP tussen de instroomniveaus en tussen degenen die gestopt waren versus degenen die nog studeerden.

Ook de correlaties tussen de GFP en de (onderdelen van) proeven van bekwaamheid zijn getoetst. De uitkomsten daarvan zijn hieronder weergegeven (zie Tabellen 14.5 t/m 14.7). Ook hier kwamen geen significante resultaten uit. Dit resultaat was dus niet veel anders dan de resultaten die werden gevonden voor de afzonderlijke Big Five factoren.

Tabel 14.5. Correlaties tussen enerzijds GFP en Big Five en anderzijds proportie geslaagde tests (alle instroomniveaus)

	Onderdelen	Proeven
GFP	.01	.01

* $p < .05$.

Tabel 14.6. Correlaties tussen enerzijds GFP en Big Five en anderzijds proportie geslaagde onderdelen van de proeven (opgesplitst op instroomniveaus)

	Niveau 2 (n = 239)	Niveau 3 (n = 1027)	Niveau 4 (n = 802)
GFP	.06	.02	.01

* $p < .05$.

Tabel 14.7. Correlaties tussen enerzijds GFP en Big Five en anderzijds proportie geslaagde proeven (opgesplitst op instroomniveaus)

	Niveau 2 (n = 239)	Niveau 3 (n = 1027)	Niveau 4 (n = 802)
GFP	.10	.03	-.00

* $p < .05$.

Conclusie 40: De GFP was duidelijk aanwezig in de persoonlijkheidsdata van de selectie bij de politie. Deze GFP vertoonde enige overlap met de beoordelingen uit het selectiegesprek en de praktijkopdracht. Dit suggereert een relatief consistente factor. De GFP was echter niet voorspellend voor uitval of opleidingsprestatie.

Gezien de huidige status van persoonlijkheid in de selectie bij de politie is het (nog) niet zinvol om gebruik te maken van nieuwe constructen zoals de GFP en eerst af te wachten op de uitkomsten van onderzoek waarin valide criteria zijn gebruikt.

14.2 Emotionele Intelligentie (EI)

De literatuur over de politie (e.g., Salgado et al., 2003; Steinberg & Figart, 1999; Zeidner, Matthews, & Roberts, 2004) en de interviews in het huidige project wijzen er consistent op dat politiewerk een grote sociale component bevat. Van een agent wordt sociale flexibiliteit verwacht in het omgaan met de verschillende doelgroepen. Zo moet tegen een arrestant

krachtig worden opgetreden maar moet inlevend worden gereageerd op een slachtoffer van een misdrijf. Tevens is van essentieel belang dat een agent sociale situaties goed kan inschatten en aan de hand daarvan adequaat reageert. Uit de interviews komt tevens naar voren dat dit sociale inschattingsvermogen agenten in staat stelt om risico's in te schatten en te bepalen wanneer er moet worden ingegrepen.

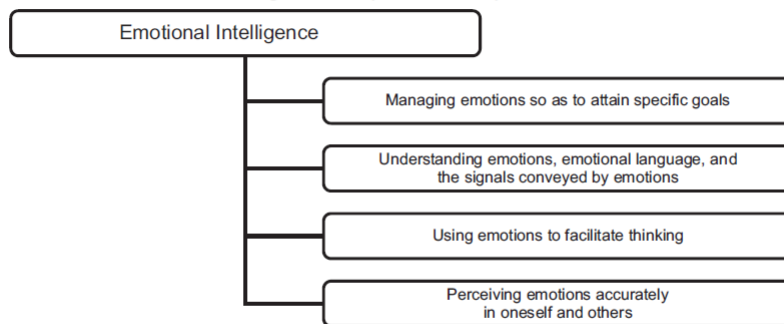
Het inschatten van sociale situaties en het vertonen van adequaat sociaal gedrag is niet alleen van belang voor het werk op straat maar ook voor het functioneren van een agent in teams. In één van de interviews werd bijvoorbeeld geopperd dat sommige agenten heel goed functioneren op straat maar wel moeite hebben met de omgang met collega's en zich minder houden aan sociaal acceptabele omgangsvormen. Hierdoor ontstonden conflicten tussen collega's en problemen met het functioneren in teams.

Wetenschappelijke literatuur toont aan dat sociale vaardigheid inhoudt dat men i) eigen emoties en die van anderen kan herkennen, en ii) in staat is om deze emoties te reguleren. Zulke vaardigheden blijken niet alleen invloed te hebben op hoe iemand reageert in sociale situaties maar hebben ook op het in stand houden van een goede psychische weerbaarheid (Charroci et al., 2002). Het aspect weerbaarheid is reeds besproken in hoofdstuk 12, waarin gesuggereerd werd dat verminderde psychische weerbaarheid een relevant probleem is bij de politie. Uit het bovenstaande worden de potentiële voordelen van kandidaten met goede sociale vaardigheden duidelijk.

In dit kader wordt de mogelijke rol van emotionele intelligentie besproken. Emotionele intelligentie (EI) verwijst naar een begrip dat sinds een jaar of twintig meer aandacht heeft gekregen in de literatuur (Mayer, Salovey, Caruso, 2004; Petrides, 2011). In het begin werd het begrip vooral als problematisch gezien omdat de definities niet eenduidig waren. In meer recente literatuur wordt echter steeds duidelijker dat EI een relevant concept kan zijn in het voorspellen van prestatie en welzijn (Petrides, 2011; Ziedner, Matthews, & Roberts, 2009). In feite lijkt EI tamelijk veel op het construct sociale intelligentie zoals reeds geïdentificeerd door Thorndike in 1920. Thorndike vermeldde dat er naast cognitieve intelligentie nog zoiets bestaat als een "...ability to understand and manage people" (Thorndike & Stein, 1937, p. 275).

Een veel gebruikte beschrijving van EI in de huidige literatuur is die van Mayer en Salovey (1997). Deze luidt als volgt: "The ability to monitor one's own and others' feelings and emotions, to discriminate among them and to use this information to guide one's thinking and actions.". Mayer en Salovey onderscheidden vier dimensies die iemands EI bepalen, te weten i) het vermogen om emoties te herkennen, ii) het gebruiken van emoties, iii) het begrijpen van emoties en iv) het reguleren van emoties.

The Four-Branch Model of Emotional Intelligence (Mayer & Salovey, 1997)



Note. Each branch describes a set of skills that make up overall emotional intelligence. Each branch has its own developmental trajectory, proceeding from relatively easy skills to more sophisticated ones. For example, Perceiving Emotions typically begins with the ability to perceive basic emotions in faces and voice tones and may progress to the accurate perception of emotional blends and to the detection of emotional microexpressions in the face.

Figuur 14.1 Viertakken model van EI

Verder maakt men in de literatuur over het algemeen onderscheid tussen enerzijds EI zoals gemeten met vragenlijsten (vaak zelfrapportage) en anderzijds EI gemeten met behulp van kennis of capaciteitentests. Dit soort tests bevatten onderdelen waarin men emoties moet herkennen of optimale gedragingen moet inschatten. Er is hierbij enige overlap met de zogenaamde Situational Judgement Tests (SJTs) waarin men ook aan de hand van situaties moet inschatten wat adequaat gedrag zou zijn.

EI gemeten met vragenlijsten en EI gemeten met kennis – of capaciteitentests overlappen relatief weinig. Vaak worden correlaties vermeld in de grootte van .20 tot .30 (Van Rooy & Viswesvaran, 2004). Sommige onderzoekers menen dan ook dat de twee verschillende meetmethoden verschillende constructen meten (Cote & Miners, 2006).

14.2.1 De relatie EI en prestatie in de literatuur

Ondanks het feit dat de wetenschappelijke discussie over het construct EI op sommige punten nog open is, is het nuttig om na te gaan of dit construct van dienst kan zijn in de selectie van personeel bij de politie. Er zijn hiervoor meerdere redenen aan te voeren. Eén daarvan is dat in de huidige selectieprocedure, de niet-cognitieve maten die worden gemeten met vragenlijsten relatief lage voorspellende waarden hebben.

Een tweede reden is dat, ondanks enige wetenschappelijke onenigheid over definities, EI een duidelijke relatie lijkt te hebben met aspecten van gedrag die voor politiewerk belangrijk zijn. Een meta-analyse van Van Rooy en Viswesvaran (2004) toonde aan dat EI voorspellende waarde kan hebben voor werkprestatie. De gemiddelde predictieve validiteit van EI was .20. Deze waarde was gebaseerd op 59 verschillende studies met een totaal aantal proefpersonen van $N = 9522$. In de meta-analyse werden ook verschillende domeinen bekeken, namelijk werkprestatie, academische prestatie (e.g., studiesucces), en succes in een reeks van andere levensdomeinen (e.g., relaties). De voorspellende waarden waren .24, .10 en .24 respectievelijk. Dit geeft dus aan dat met uitzondering van academische prestatie EI een relevante voorspeller kan zijn.

14.2.2 Eerder onderzoek EI en politiewerk

Recent zijn er enkele onderzoeken uitgevoerd naar de relatie tussen EI en politiewerk. Ali, Garner en Magadley (2011) onderzochten de relatie tussen EI en werkprestatie van

politieagenten in Abu Dhabi. Dit onderzoek verdient speciale aandacht aangezien prestatie is gemeten door middel van leidinggevende-beoordelingen van agenten in functie. Beoordelingen door leidinggevende worden in de literatuur gezien als goede criteriummaten. In het huidige selectieonderzoek bij de politie in Nederland ontbreken dit soort uitkomstmaten vaak nog. EI werd gemeten met een EI test ontworpen door Schutte et al. (1998), de zogenaamde Self-report Emotional Intelligence Test (SREIT). In het onderzoek werd een relatie gevonden van $r = .43$ tussen EI en prestatiebeoordelingen. Dit lijkt hoger dan de relaties die vaak worden gevonden met de klassieke persoonlijkheidskenmerken (e.g., De Big Five). Een andere belangrijke bevinding was dat EI toegevoegde voorspellende waarde had bovenop cognitieve capaciteit en persoonlijkheid. Dit suggereert dat EI metingen de selectie van politiepersoneel zou kunnen verbeteren.

EI zou niet alleen werkprestatie kunnen voorspellen maar ook psychische weerbaarheid. Vuzzo (2009) onderzocht de relatie tussen EI en stress bij politieagenten uit New Jersey (VS). Stress kan worden gezien als een belangrijke indicator voor verminderde psychische weerbaarheid. Stress werd gemeten met de Police Stress Survey. EI werd gemeten met de Bar-On vragenlijst van EI. De auteur rapporteerde een negatieve relatie tussen totale EI score en stress van $-.46$. Dat wil zeggen dat agenten die hoog scoorden op EI minder stress ervaarden. Tot slot heeft onderzoek uit het proefschrift van Smith (2009) laten zien dat EI een relatie vertoonde met integriteit bij politieagenten. De bevindingen uit dit onderzoek suggereren dat agenten met hogere niveaus van EI hogere niveaus van morele ontwikkeling hebben.

14.2.3 Gebruik EI maten in de selectieprocedure van aspirant agenten?

Wanneer de politieorganisatie zou overwegen om EI maten te gebruiken in de selectie zijn er enkele zaken waar rekening mee moet worden gehouden.

Ten eerste moet worden vermeld dat ondanks bemoedigende resultaten uit eerder onderzoek enige voorzichtigheid moet worden betracht wanneer men EI maten zou willen implementeren in de selectieprocedure. Dit geldt natuurlijk voor het merendeel van de vernieuwingen in selectieprocedures maar moet extra benadrukt worden wanneer het gaat om EI. De status van het begrip wordt nog bediscussieerd in de wetenschappelijke literatuur. Daardoor is het begrip minder goed uitgekristalliseerd in vergelijking met bijvoorbeeld cognitieve capaciteit of persoonlijkheid. Het is daarom niet wenselijk om simpelweg een EI maat in de selectieprocedure te voegen en er, gebaseerd op de literatuur, vanuit gaan dat deze goede voorspellende waarde heeft bij de politie. Wij stellen voor dat, indien de politieorganisatie dit soort maten wilt introduceren in de selectie, dit stapsgewijs gebeurt. Een eerste stap hierbij kan zijn het invoegen van een goede EI maat (hier komen we nog op terug) naast de huidige selectie-instrumenten bij een kleinere steekproef van kandidaten. Belangrijk is dat in deze eerste stap, EI nog niet wordt meegenomen in beslissingen over geschiktheid van de kandidaat maar dat de maat slechts parallel loopt aan de gebruikelijke procedure. Na enige tijd kan dan empirisch worden vastgesteld of de EI maat inderdaad de gewenste voorspellende waarde heeft op prestatie. Indien dat het geval is kan de EI maat op grotere schaal ingevoerd kunnen worden en kan men de EI informatie meenemen in beslissingen betreffende de geschiktheid van kandidaten.

Het tweede punt van aandacht is welke EI maat geschikt zou zijn om mee te nemen in de selectieprocedure. Zoals eerder aangegeven wordt in de literatuur onderscheid gemaakt tussen verschillende EI modellen en maten. EI als capaciteit wordt typisch gemeten met capaciteitstesten zoals de Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). Dit is wereldwijd de meest frequent gebruikte EI-capaciteitentest. Het voordeel van deze test is dat hij relatief ongevoelig is voor sociale wenselijkheid. Sommige onderzoekers menen zelfs dat dit soort capaciteitentests de enige juiste manier is om EI te meten. Nadelen van tests zoals de MSCEIT is dat de afname relatief lang duurt (30 tot 40 minuten) en dat de test niet vrij beschikbaar is. Testuitgevers zullen waarschijnlijk 5 tot 50 euro per afname rekenen. Gezien de grote aantallen kandidaten die zich aanmelden voor de politie kan dit problematisch zijn.

Alternatieven zijn zelfrapportagematen van EI. Een punt van kritiek hierbij is de mogelijke invloed van sociale wenselijkheid als vertekening van de uitkomsten. Daar staat tegenover dat onderzoek heeft aangetoond dat sociale wenselijkheid een minder groot probleem is voor de predictieve validiteit dan aanvankelijk werd aangenomen in de literatuur.

Een ander argument dat pleit voor het gebruik van zelfrapportage maten is dat veel van de onderzoeken naar de voorspellende waarde van EI zijn uitgevoerd met dit soort maten (zie bijvoorbeeld Ali et al., 2011; Smith, 2009; Van Rooy & Viswesvaran, 2004; Vuzzo, 2009).

Indien gebruik gemaakt zou worden van zelfrapportagematen van EI in de selectieprocedure dan is een vervolgvraag welke maten hiervoor in aanmerking kunnen komen. In de literatuur zijn vaak genoemde maten:

- De Wong and Law Emotional Intelligence Scale (WLEIS): Dit is een EI vragenlijst bestaande uit 16 items, ontworpen door Wong en Law (2002). De vragenlijst is gebaseerd op het capaciteitenmodel van EI en bevat dezelfde vier subschalen als in de MSCEIT capaciteitentest. Men kan naar de afzonderlijke schalen kijken maar kan ook een algemene EI-score berekenen als de gemiddelde score op de vier schalen of door middel van het extraheren van de algemene EI-factor uit de subschalen.

- Triat Emotional Intelligence Questionnaire (TEIQue): Deze schaal is ontworpen door Furnham en Petrides (2001) en is gebaseerd op het idee van trait EI. Dat wil zeggen, EI als een algemeen kenmerk dat tot uiting komt in verschillende persoonlijkheidsdimensies. Van de TEIQue bestaat een lange (xx items) en een korte versie (32 items). De vragenlijst meet de volgende vier subschalen van EI, maar in onderzoek wordt vaak gebruik gemaakt van de algemene EI-score.

- De Emotion Quotient Inventory (EQ-I): Ontworpen door Ba-On (2006) en bestaat uit 133 items die samen 15 subschalen van EI vormen. Men kan gebruik maken van de algemene score of van de scores van de subschalen.

- Self Report Emotional Intelligence Test (SREIT): Een EI-vragenlijst ontworpen door Schutte et al. (1998) bestaande uit 33 items die drie aspecten van EI meten zoals genoemd worden in het model van Mayer en Salovey (waarnemen van emoties [perceiving emotions] wordt niet gemeten in deze test).

Conclusie 41:

- Emotionele intelligentie (EI) wordt in de literatuur beschreven als een begrip bestaande uit kennis over sociale situaties en emoties en het vermogen om deze kennis te gebruiken om doelen te bereiken.

- EI blijkt samen te hangen met baanprestatie, psychische weerbaarheid, en integriteit. De mogelijk toegevoegde waarde van het meten van EI is dat men een concept gebruikt dat een brede invloed heeft op gedrag en daardoor mogelijk veel van de belangrijke aspecten van politiewerk kan voorspellen. De literatuur onderscheidt EI als capaciteit en EI als persoonlijkheidstrek die andere persoonlijkheidsdimensies beïnvloed. Bovendien wordt onderscheid gemaakt in het meten van EI met capaciteitentest in tegenstelling tot meten met behulp van zelfrapportage.

Wij stellen voor de mogelijkheid te verkennen om EI metingen te introduceren in de selectieprocedure bij de politie.

14.3 De duistere kanten van persoonlijkheid (Dark side of personality)

De meeste modellen in persoonlijkheidsonderzoek hebben tot doel een taxonomie te geven van persoonlijkheid in het algemeen. De Big Five en Eysenck's Giant Three zijn daar voorbeelden van. Recent is er in dit onderzoeksgebied echter steeds meer aandacht voor de negatieve kanten van persoonlijkheid (e.g., Paulhus & Williams, 2002) Het gaat hierbij om die stabiele kenmerken in individuen die ofwel sociale interactie bemoeilijken of het risico verhogen op immoreel gedrag.

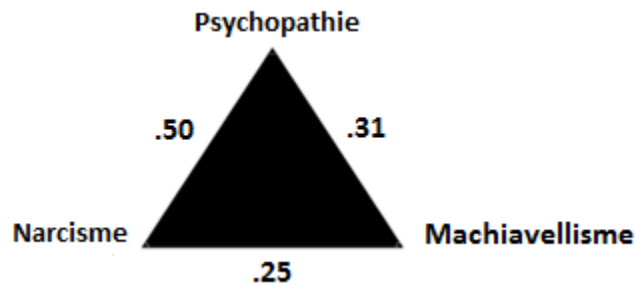
Voor selectie bij de politie zijn deze ontwikkelingen mogelijk relevant omdat de medewerkers van de politie bij uitstek een voorbeeldfunctie hebben in de maatschappij. In de gevallen dat er agenten 'de fout in gaan' en immoreel gedrag vertonen is er dan vaak ook grote verontwaardiging onder burgers. Er zijn verschillende voorbeelden te noemen op nationaal of internationaal gebied van agenten die gedrag vertoonden dat volgens velen immoreel was en dat heeft geleid tot grote maatschappelijke onrust. In de VS is de situatie rondom Rodney King een saillant voorbeeld waarin agenten volgens velen excessief geweld gebruikten. Soms komen er ook gevallen van corruptie naar buiten waaruit blijkt dat politiemedewerkers hun macht hebben misbruikt.

Bij de Nederlandse politie zijn gevallen van grof of immoreel gedrag door politieagenten relatief schaars. Gezien de ernst van de gevolgen wanneer dit echter wel gebeurt kan het nuttig zijn om na te gaan of er in de selectie constructen zouden kunnen worden gemeten die beter kunnen voorspellen wie kans heeft om immoreel gedrag te vertonen. Metingen van de zogenaamde 'Dark side of personality' zouden hierbij kunnen helpen.

Zoals vaak in de psychologie, zijn er verschillende manieren om naar de donkere kanten van persoonlijkheid te kijken. Een indeling waarover een belangrijke mate van consensus bestaat onder onderzoekers is die van Paulhus en Williams (2002). Zij leggen de nadruk op drie brede negatieve constructen in persoonlijkheid, namelijk psychopathie, narcisme, en machiavellisme. Deze drie constructen vormen volgens Paulhus en Williams (2002) de basis voor wat zij 'the dark triad of personality' oftewel de Duistere drie van persoonlijkheid noemden.

In één van hun eerste onderzoeken op dit gebied lieten Paulhus en Williams (2002) zien dat de drie constructen van de Dark Triad wel aan elkaar gerelateerd zijn maar toch te onderscheiden constructen zijn. In een onderzoek onder 245 studenten vonden zij bijvoorbeeld relaties tussen de constructen zoals weergegeven in Figuur 14.2.

De drie constructen vertoonden ook differentiële relaties met andere bekende psychologische constructen zoals de Big Five en intelligentie zoals hieronder weergegeven in Tabel 14.8.



Figuur 14.2. De Dark Triad en hun onderlinge correlaties

Correlations of the Dark Triad with the Big Five, cognitive ability, and self-enhancement

	Narcissism	Machiavellianism	Psychopathy
<i>Big Five Inventory</i>			
Extraversion	.42	-.05	.34
Agreeableness	-.36	-.47	-.25
Conscientiousness	-.06	-.34	-.24
Neuroticism	.02	.12	-.34
Openness	.38	-.03	.24
<i>Cognitive ability</i>			
OCQ accuracy index	.09	.04	.09
IQ test	.15	.04	.05
Verbal-nonverbal discrepancy	.05	.20	.13
<i>Self-enhancement</i>			
Discrepancy of self-rating vs. IQ test	.24	-.02	.14
Over-claiming bias index	.17	.08	.09

Note. $N = 245$. All correlations in bold are significant at $p < .05$, two-tailed.

Tabel 14.8. Correlaties tussen enerzijds de Dark Triad en anderzijds persoonlijkheids- en capaciteitsmaten

Verschillende onderzoeken hebben gekeken naar de relatie tussen de Dark triad en werkprestatie. Een bekende studie op dit gebied is van Moscoso en Salgado (2004) hierbij maakten ze onderscheid tussen taakprestatie, contextuele prestatie, en een globale aanduiding van prestatie. Taakprestatie had betrekking op het uitvoeren van de kerntaken van een baan. Contextuele prestatie is de mate waarin werknemers positief werkgedrag vertonen dat buiten hun formele werkbeschrijving omgaat, zoals het helpen van collega's of het aannemen van aanvullende taken die ten goede komen van de organisatie. De globale prestatie-indicatie was een algemene index die bestond uit een combinatie van taak- en contextuele prestatie.

Moscoso en Salgado (2004) keken naar veertien facetten van disfunctionele persoonlijkheid die ten grondslag liggen aan de Dark Triad. Zij vonden dat bijna alle disfunctionele facetten van persoonlijkheid een negatieve relatie vertoonden met de prestatie-indicatoren. Een overzicht van hun resultaten is weergegeven in Tabel 14.9 hieronder.

Wat voor de selectie van agenten een belangrijke bevinding kan zijn, is dat scores op de Dark Triad positief samen lijken te hangen met negatief werkgedrag, in het Engels vaak aangeduid met de term counter productive behavior. Dit slaat op grove onethische handelingen zoals diefstal op het werk en corruptie, maar ook op lichtere vormen van negatief gedrag zoals het besteden van werktijd voor niet-werk gerelateerde taken. In een meta-analyse van O'Boyle, Forsyth, Banks & McDaniel (2012) kwam naar voren dat alle drie de aspecten van de Dark Triad samenhangen met counter productive behavior, met waardes van $\rho = .25$ voor machiavellisme, $\rho = .43$ voor narcisme, en $\rho = .07$ voor psychopathie.

Interessant was overigens dat er ook modererende factoren zijn gevonden in deze relaties. In dit geval waren de moderatoren de mate van autoriteit en de cultuur die heerste met betrekking tot autoriteitsrollen. Hoe hoger de positie in de organisatie, hoe sterker de relatie was tussen de Dark Triad en negatieve prestatie.

Figuur 14.9. Relaties tussen Dark Triad facetten en prestatie

Personality disorder	Task performance	Contextual performance	Overall performance
Narcissistic	-.06	-.22*	-.18
Paranoid	-.25*	-.32**	-.33**
Sadistic	-.09	-.11	-.12
Avoidant	-.31**	-.27*	-.31**
Depressive	-.30**	-.27*	-.31**
Passive-aggressive	-.32**	-.34**	-.37**
Self-defeated	-.34**	-.27*	-.33**
Schizotypal	-.35**	-.43**	-.45**
Borderline	-.15	-.15	-.17
Schizoid	-.19	-.17	-.20
Compulsive	.10	.05	.07
Histrionic	-.04	.07	.03
Antisocial	-.31**	-.21*	-.27*
Dependent	-.24*	-.12	-.18

N = 85, * $p < .05$, ** $p < .01$.

Er zijn in de literatuur verschillende maten bekend die de Dark triad meten. Een bekend meetinstrument is bijvoorbeeld de 'Dirty Dozen' van Jonason en Webster. Dit instrument bevat 12 items die een korte maar betrouwbare en valide meting zouden geven van de Dark Triad. De psychometrische eigenschappen van deze vragenlijst zijn uitvoerig beschreven in verschillende artikelen.

Het zou in de selectieprocedure bij de politie het overwegen waard zijn om aandacht te schenken aan de duistere kanten van persoonlijkheid en hun voorspellende waarde op onethisch of negatief werkgedrag.

Conclusie 42: In de selectie van aspirant-agenten zou men kunnen overwegen om aanvullende maten mee te nemen die onethisch en/of negatief werkgedrag kunnen voorspellen. Daarbij kan men bijvoorbeeld denken aan maten die de duistere kanten van persoonlijkheid beogen te meten zoals de Dark Triad: psychopathie, narcisme, en machiavellisme.

Hoofdstuk 15

Overzicht aanbevelingen

15 Overzicht Conclusies en aanbevelingen

15.1 Concluderende opmerkingen

Dit hoofdstuk geeft een overzicht van de conclusies en bijbehorende aanbevelingen in het huidige rapport. De aanbevelingen betreffen in sommige gevallen mogelijkheden om de selectieprocedure te verbeteren, efficiënter, of goedkoper te maken. In sommige gevallen betreft het echter suggesties voor vervolgonderzoek omdat er niet genoeg informatie beschikbaar is om stevige uitspraken te doen.

In de inleiding van het rapport zijn enkele algemene vragen geformuleerd. Het is zinvol om na te gaan in hoeverre deze vragen konden worden beantwoord. De vragen uit de inleiding worden daarom hieronder nogmaals genoemd

A) Wat kenmerkt een goed agent?

Dit betrof de volgende vragen:

- Welke eigenschappen en competenties moet een agent bezitten (of ontwikkelen) om politiewerk adequaat uit te voeren? Het huidige rapport richt zich bij de beantwoording van deze vraag specifiek op executief politiewerk op de niveaus twee tot en met vier.
- Welke van deze eigenschappen/competenties zijn stabiel en welke kunnen nog goed worden ontwikkeld tijdens de opleiding?

Over het algemeen kan worden vermeld dat veel van de criteria van politiewerk bekend zijn bij de politieacademie en reeds zijn verwerkt in de selectieprocedure. In Hoofdstuk 3 zijn deze criteria beschreven. Tevens worden in Hoofdstuk 3 enkele suggesties gedaan om de lijst van criteria uit te breiden of aan te scherpen. Hoofdstuk 3 gaat ook in op ontwikkelbaarheid van competenties en eigenschappen. Sommige hiervan blijken tamelijk stabiel te zijn (cognitieve capaciteit, persoonlijkheid), anderen zullen meer ontwikkelbaar zijn maar in veel gevallen geeft de wetenschappelijke literatuur geen uitsluitsel over de mate van ontwikkelbaarheid. Een voorbeeld hiervan is lichamelijke conditie die tot zekere hoogte stabiel lijkt te zijn maar de mate waarin is vooralsnog onduidelijk.

B) Wat is de waarde van de verschillende selectieonderdelen en de meetinstrumenten die worden gebruikt?

Deze vraag kon worden onderverdeeld in de volgende deelvragen:

- Hoe valide en betrouwbaar zijn de meetinstrumenten in de selectieprocedure? Daarbij moet worden gedacht aan de cognitieve capaciteitentesten, het fysiek-motorisch onderzoek, de taaltoets, persoonlijkheidstesten, het selectiegesprek en de praktijkopdrachten.
- Wat is de voorspellende waarde van meetinstrumenten in de selectieprocedure?
- Zijn de stappen in de selectieprocedure in overeenstemming met huidige wetenschappelijke inzichten in de personeelsselectie?

Het beantwoorden van deze vragen over de selectie-onderdelen omvatte een substantieel deel van het huidige rapport (Hoofdstukken 5 t/m 11). Over het algemeen zijn de selectie-instrumenten in overeenstemming met de huidige wetenschappelijk inzichten in de

personeelspsychologie. Bij sommige selectie-instrumenten kan men wel enkele fundamentele vragen stellen. Zoals de vraag hoeveel en welke competenties men idealiter zou willen meten in het selectiegesprek en de praktijkopdracht. De wetenschappelijke literatuur is nog niet dusdanig ontwikkeld dat hier definitieve antwoorden op kan worden gegeven. Bovendien was er in het huidige project onvoldoende criteriuminformatie beschikbaar om deze vragen op een empirische manier te beantwoorden. Wel worden in het rapport enkele conclusies getrokken over de betrouwbaarheid en validiteit van de selectie-instrumenten.

Een van de conclusies die als een rode draad door het rapport heen liep is dat er momenteel onvoldoende criteriuminformatie aanwezig is om te toetsen in hoeverre de gehanteerde selectiemethoden daadwerkelijk samenhangen met prestatie van agenten in de praktijk. Dit is belangrijk omdat zonder zulke informatie de fundering ontbreekt om sterke empirische uitspraken te doen over de waarde van (onderdelen van) de selectie. Uitgaande van de literatuur en specifieke informatie die beschikbaar is over de procedures kan men uiteraard enkele suggesties formuleren, maar hun daadwerkelijke effecten op kwaliteit van agenten zal vooralsnog niet direct duidelijk worden. Dit is met name een risico omdat de werving en selectie niet alleen afhankelijk is van wetenschappelijke feiten maar ook onderhevig is aan geldende denkbeelden (e.g., over wat een goede agent is) en politieke keuzes. Het is belangrijk om regelmatig te monitoren wat de effecten zijn van beslissingen in de selectie op de aard van de agenten die worden aangenomen en de kwaliteit van hun politiewerk. Kortom, criteriuminformatie is in dit opzicht cruciaal.

C) Afhankelijk van de antwoorden op bovenstaande vragen, worden de volgende vervolgvragen gesteld:

- Zijn er op grond van de evaluatie van de selectieprocedure aanpassingen nodig in het gebruik van, of de normering van de selectie-instrumenten? Zo ja, welke en wat zijn de verwachte financiële gevolgen van de veranderingen?
- Is het op grond van de uitkomsten van de evaluatie mogelijk om aanbevelingen te doen voor het verminderen van de uitval uit de opleiding? Zo ja, welke aanbevelingen en wat zijn de mogelijke financiële consequenties?
- Is het op grond van de evaluatie mogelijk om aanbevelingen te doen voor een kostenreductie van de selectieprocedure? Zo ja, welke?

Gezien de conclusies betreffende de afwezigheid van goede criteriuminformatie, zou het onverantwoord zijn om sterke aanbevelingen te doen over het vervangen of weglaten van onderdelen in de selectie. Tevens zou er voor concrete financiële uitspraken een empirische basis ontbreken. Desalniettemin, bevat het huidige rapport enkele conclusies en aanbevelingen die kunnen dienen als uitgangspunt voor het optimaliseren van de selectieprocedure bij de politie. Zo is er gewezen op enkele mogelijke tekortkomingen van de cognitieve test (Hoofdstuk 5, Conclusies 8 en 9). Bovendien zijn er een aantal voorstellen gedaan die mogelijk kostenbesparend zou kunnen zijn. Men kan daarbij bijvoorbeeld denken aan de conclusies/aanbevelingen met betrekking tot taalvaardigheid (Hoofdstuk 7, Conclusie 17) of aanbevelingen over contracttype bij aanstelling (Hoofdstuk 13, Conclusie 38).

Los van de specifieke vragen die zijn behandeld in het huidige rapport is een conclusie die hier genoemd moet worden dat er, naar onze inschatting, onder de betrokken medewerkers bij de politie een wetenschappelijke instelling heerst en een sterke behoefte om constant de kwaliteit van de werving en selectie te monitoren en waar mogelijk te verbeteren. Dit komt onder andere tot uiting in de reeks van onderzoeken die de Politieacademie jaarlijks uitvoert. Veel van de huidige procedures in de werving en selectie komen voort uit grondige kennis over de wetenschappelijke literatuur in combinatie met de praktische overwegingen die spelen bij de selectie van aspirant-agenten (e.g., wat is de gewenste selectie-ratio, hoeveel geld is er beschikbaar, wat zijn de diversiteitseisen).

In het algemeen hopen we dat de conclusies en aanbevelingen in het huidige rapport aanleiding kunnen zijn om de selectieprocedure van agenten te optimaliseren en zullen leiden tot een Nationaal politiekorps dat van Nederlanders het aanzien en respect krijgt dat het verdient.

15.2 Overzicht conclusies

Conclusie 1: Door beleidsveranderingen in 2011 is de uitval per selectieonderdeel veranderd. In elk onderdeel is het uitvalspercentage verminderd. Er kon geen definitieve verklaring worden gegeven voor de toegenomen selectieratio, ook niet door de betrokken partijen. Bovendien is vooralsnog onduidelijk wat voor effecten de veranderde selectieratio heeft op de kwaliteit van de aspirant-agenten die worden aangenomen.

Maatregelen die zichtbaar effect hebben op uitvalspercentages en/of de selectieratio's hebben bij een gelijkblijvende pool van kandidaten per definitie effect op het type personeel dat wordt aangenomen. In het kader van geïnformeerde beleidsvoering, stellen we voor dat er in dit soort gevallen wordt onderzocht wat de mogelijke lange –termijneffecten in de praktijk zijn van zulke selectieveranderingen. Goed opgezette cohortstudies waarin aspirant-agenten worden gevolgd vanaf de selectie tot en met het afronden van de opleiding op de politieacademie (en mogelijk nog daarna) kunnen hiertoe een goed middel zijn.

Ter illustratie van bovenstaande aanbeveling: We refereerden al eerder naar geluiden dat een te grote nadruk op de sociale aspecten van politiewerk in de jaren negentig mogelijk hebben geleid tot agenten die minder goed in staat zouden zijn om geweld te gebruiken (zie de startnotitie WODC, 2012). Dit is een voorbeeld van hoe veranderingen in de selectie van personeel mogelijke lange-termijn negatieve gevolgen kunnen hebben op de kwaliteit van politiewerk.

Conclusie 2: Ondanks organisatorische (e.g., Nationale Politie) en technologische ontwikkelingen, zullen de eisen die aan uitvoerend politiepersoneel worden gesteld niet fundamenteel veranderen.

Conclusie 3 : Het is te overwegen om bij het aannemen van hoger (niveau ≥ 5) politiepersoneel met specifieke kennis/vaardigheden niet langer volledig vast te houden aan alle algemene aanstellingseisen die gelden voor politiepersoneel. Met sommige van deze eisen (e.g., fysieke conditie, psychologische eigenschappen belangrijk voor straatwerk) zou

mogelijkerwijs soepeler omgegaan kunnen worden, terwijl andere algemene kenmerken wel van belang blijven (e.g., integriteit, psychische belastbaarheid). Kortom: De politie zou kunnen verkennen of het werkbaar en wenselijk is om de algemene eisen bij sommige hogere functies te versoepelen of zelfs los te laten.

Conclusie 4: Concluderend kan worden vermeld dat het criteriumprofiel voor politiewerk (niveau 2 t/m 4) kan worden aangescherpt. Als we ons beperken tot het gebied van de criteria (en dus nog niet de selectie) kunnen voorsnog de volgende aanbevelingen worden geformuleerd:

1) Het is nuttig om het criteriumprofiel van de fysieke eisen voor politiewerk nog eens kritisch te bekijken en waar mogelijk aan te vullen en/of specificeren.

Daarbij kan men op het volgende letten:

Blijkbaar worden enkele fysieke eisen niet langer expliciet erkend als criteria van politiewerk terwijl deze mogelijkerwijze wel van belang zijn. Denk daarbij onder andere aan motorische vaardigheden zoals ook genoemd in O*Net.

2) Bij het vaststellen van fysieke criteria voor politiewerk zou men niet alleen moeten uitgaan van inclusiecriteria (minimale conditie, minimale kracht etc.) maar ook de exclusiecriteria moeten aanscherpen (bijvoorbeeld problemen met trommelvliezen, knieën etc.).

3) Een interessante mogelijkheid is om bij fysieke eisen niet alleen uit te gaan van directe uitingen van lichamelijke gezondheid en conditie maar ook naar onderliggende (psychologische) factoren die hierop van invloed kunnen zijn. Men kan hierbij bijvoorbeeld denken aan motivatie om gezond/fit te blijven (e.g., Curry & Weiss, 1989; Rejeski & Kennedy, 1988).

Conclusie 5: Interesses zijn een nuttige aanvulling op het competentieprofiel van agenten. Uitgaande van O*Net zijn 'realistisch', 'ondernemend', 'conventioneel', en 'sociaal' logische keuzes van interesses.

Conclusie 6: Vragen over de stabiliteit versus beïnvloedbaarheid van een (gedrags)kenmerk zijn op grond van de wetenschappelijke literatuur tamelijk lastig eenduidig te beantwoorden. Voor vele kenmerken is de wetenschappelijk discussie over dit onderwerp nog open. O*Net kan hier wel een richtlijn bieden omdat deze kenmerken indeelt in categorieën die gekoppeld zijn aan een bepaalde, veronderstelde mate van ontwikkelbaarheid.

Conclusie 7: Empirisch bewijs ondersteunt het idee dat de proportie behaalde proeven van bekwaamheid (eerste poging) een indicatie is voor de kwaliteit van de aspirant-agent. Specifiek: degenen die vroegtijdig uit zijn gevallen behaalden een significant lagere proportie proeven tijdens de opleiding. De proportie behaalde proeven is overigens niet de belangrijkste factor om vroegtijdig te stoppen. Blijkbaar spelen er andere (voorsnog) onbekende factoren een grotere rol.

Conclusie 8: De status van de betrouwbaarheid van de CAIT is nog deels onbekend. Het toevoegen van oefenitems heeft een positief effect op de betrouwbaarheid maar het gevolg voor de voorspellende waarde van de test is vooralsnog onbekend. Het is mogelijk dat een toename in betrouwbaarheid de voorspellende waarde van de test verhoogd, maar het is ook mogelijk dat de methode van toevoegen van oefenitems testbekendheid beïnvloed waardoor wellicht de voorspellende waarde juist zou afnemen.

Conclusie 9: Er is relatief weinig directe informatie aanwezig waaruit de begripsvaliditeit van de CAIT kan worden afgeleid (denk hierbij aan relaties met bekende intelligentietesten). Wel zijn er een aantal indirecte aanwijzingen dat de CAIT aspecten van intelligentie meet. Een voorbeeld zijn de verschillende testscores tussen instroomniveaus in de verwachte richting.

Conclusie 10: Onderzoek ten behoeve van het huidige rapport laat zien dat cognitieve capaciteit een zeer geringe (tot geen) voorspellende waarde heeft voor uitval uit de opleiding en voor opleidingsprestatie. Het dient echter benadrukt te worden dat het vooralsnog moeilijk is om vast te stellen *waardoor* precies de voorspellende waarde zo laag blijkt te zijn. Oorzaken kunnen liggen bij de beperkingen van de gebruikte cognitieve testen, de kwaliteit van de criteria, de relatie tussen opleidingscriteria en prestatie in het werkveld, of een daadwerkelijke lage relatie tussen cognitieve capaciteit en politiewerk.

Conclusie 11: De hoofdcategorieën van fysieke vaardigheden zoals deze worden vermeld in de literatuur zijn goed gerepresenteerd in de huidige selectieprocedure van aspirant-agenten.

Conclusie 12: Er kan niet goed worden vastgesteld of de kosten van het selecteren op aanvullende fysieke eisen opwegen tegen de voordelen die voortkomen uit zulke veranderingen (e.g., verbeterde werkprestatie van agenten, minder uitval van [aspirant-]agenten om fysieke redenen). Meer onderzoek op dit gebied is noodzakelijk.

Conclusie 13: Het zou nuttig zijn om naast het meten van de huidige fysieke conditie in de selectieprocedure, ook gebruik te maken van voorspellers van latere fysieke conditie (tijdens het werk als uitvoerend agent). De motivatie om fit/actief te blijven zou een mogelijke voorspeller kunnen zijn. Verder zou het wenselijk zijn om het verband te bepalen tussen de scores op het fysiek-motorisch onderzoek in de selectieprocedure en de fysieke conditie van agenten tijdens hun loopbaan.

Conclusie 14: Enkelen FMO-onderdelen hebben voorspellende waarde op wie al dan niet succesvol de opleiding zal afronden. Met name Parcourstijd voorspelt tot op zekere hoogte iemands studie-uitval. In sommige gevallen is de relatie contra-intuïtief; kandidaten met een hogere VO₂ max en een lager vetpercentage hebben minder kans om de opleiding succesvol af te ronden.

In overeenstemming met een recent onderzoek van de Politieacademie (2009) is gevonden dat FMO-uitkomsten de criteria van opleidingsprestatie nauwelijks voorspellen met uitzondering van parcourstijd. De criteria waren echter tamelijk breed gedefinieerd; ze omvatten bijvoorbeeld ook alle cognitief en sociaal georiënteerde taken in de opleiding. Het is dus niet

verwonderlijk dat lichamelijke conditie niet sterk samenhangt met deze criteria. Daarnaast is een fundamenteeler punt dat vooralsnog onduidelijk is in hoeverre de gekozen criteria goede maten zijn voor de prestatie van de aspirant-agent tijdens de opleiding en in hoeverre de criteria het latere werk als agent goed voorspellen.

Conclusie 15: Voor zover met de beschikbare informatie kan worden ingeschat, lijken de gevolgde procedures en normen met betrekking tot de taaltoets recent grondig onderbouwd in de selectieprocedure bij de politie. De vervanging van de oude taaltoets (Digibo) door de nieuwe TOA lijkt een goede keuze omdat de TOA duidelijk aansluit bij Europese normen.

Conclusie 16: De beoordeling van de taaltoets is ten dele subjectief. Hierdoor kunnen er grote verschillen in scoring ontstaan tussen (groepen van) beoordelaars. Besteedt daarom aanvullende aandacht aan het standaardiseren van de beoordeling op de taaltoetsen. Te denken valt aan verbeterde nakijksleutels, bijvoorbeeld volgens de methode van Angoff. Deze laatstgenoemde methode komt oorspronkelijk uit de onderwijspsychologie maar wordt ook vaak gebruikt bij het bepalen van normscores in selectieprocedures. De methode is gebaseerd op de meningen van verschillende experts (Subject Matter Experts) die inschatten welk percentage van minimaal gekwalificeerde kandidaten een bepaald item of bepaalde taak goed kunnen volbrengen.

Conclusie 17: Het zou nuttig zijn om te overwegen een deel van de verantwoordelijkheid voor het aantonen van taalvaardigheid bij de kandidaat te leggen. Het model zoals hierboven geschetst vormt een mogelijkheid dit te implementeren.

Conclusie 18: Persoonlijkheid (Big Five) heeft geen significante relaties met uitval tijdens de opleiding. Wat betreft opleidingsprestatie zijn er slechts twee dimensies die voorspellende waarden hebben (consciëntieusheid en vriendelijkheid). Deze dimensies hebben echter alleen voorspellende waarde bij aspirant-agenten op instroomniveau 2. Het is onwaarschijnlijk dat het groot aantal nulbevindingen te wijten is aan het instrument waarmee persoonlijkheid wordt gemeten.

Conclusie 19: Uit de kwalitatieve evaluatie van het selectiegesprek en de praktijkopdracht valt te concluderen dat de competenties die gemeten worden in de selectieprocedure logisch lijken (face validity hebben) en conceptueel duidelijk zijn afgebakend. Bovendien zijn de kenmerken van politiewerk die worden genoemd in O*Net goed te herkennen in de competenties gemeten door de Politieacademie. De afdeling werving en selectie doet in dit opzicht veel om de kwaliteit van de metingen van competenties te blijven garanderen en te optimaliseren.

Conclusie 20: Er zijn inhoudelijke gronden om het aantal gemeten competenties in het selectiegesprek terug te brengen tot een kleiner aantal. De indeling van Arthur, Day, McNelly en Edens (2003) kan daarbij als richtlijn dienen. Bovendien is er onder geïnterviewden consensus over het idee om minder competenties te hanteren. Welke competenties dit precies zouden moeten zijn is met name een empirische vraag.

Conclusie 21: De competenties gemeten in het selectiegesprek lijken te clusteren in twee overkoepelende factoren: I) Kracht en II) Wijsheid/Mildheid. Bovendien is er ook een sterke algemene cluster waarneembaar die een globale beoordeling van een kandidaat lijkt weer te geven. Deze informatie geeft aan dat beoordelaars geen duidelijk onderscheid maken tussen de afzonderlijke competenties in het selectiegesprek.

In toekomstige beslissingen over het aantal en welke competenties moeten worden gemeten in het selectiegesprek kunnen deze resultaten een rol spelen. Hierbij moet wel worden aangemerkt dat veranderingen in de procedure van het selectiegesprek goed dienen worden ondersteund met onderzoek: Er kan bijvoorbeeld niet zonder meer worden aangenomen dat brede categorieën van competenties werkprestaties ook beter kunnen voorspellen. Dit laatste blijft een empirische kwestie.

Conclusie 22: Alhoewel kandidaten in de praktijkopdracht formeel op slechts drie brede competenties moeten worden beoordeeld, blijkt dat kandidaten op een algemene wijze worden beoordeeld in termen van hoe goed ze de opdracht hebben uitgevoerd.

Conclusie 23: Er zijn aanwijzingen dat er onder verschillende beoordelaars in de selectie en tijdens verschillende onderdelen (gesprek, praktijkopdracht) consensus is over wie al dan niet een goede kandidaat is. Dit ondersteunt de validiteit van de beoordelingen.

De relatie tussen enerzijds de competenties en anderzijds persoonlijkheid en intelligentie zijn relatief zwak. Dit suggereert dat de competenties toegevoegde waarde kunnen hebben in de selectie en zeker niet hetzelfde meten als persoonlijkheid en intelligentie.

Conclusie 24: Emotionele weerbaarheid, drukbestendigheid, en besluitvaardigheid zijn de competenties uit het selectiegesprek die significant voorspellende waarden hebben voor vroegtijdig stoppen met de opleiding. Voor competenties gemeten in de praktijkopdracht geldt dit voor Denken en Kracht.

Conclusie 25: Over het algemeen kunnen de competenties niet sterk voorspellen hoe goed aspirant-agenten zullen presteren op de proeven van bekwaamheid tijdens de opleiding. Bij de interpretatie van deze uitkomsten moeten wederom de inherente beperkingen van de gehanteerde criteria in acht worden meegenomen (zie ook hoofdstuk 5, 6 en 8).

Conclusie 26: Het wordt niet duidelijk waar het beoogde toelatingspercentage in de korpsgesprekken op is gebaseerd. Daarom valt te overwegen aannames en keuzes hierbij explicieter te maken zodat mogelijke terechte en onterechte veronderstellingen in de achterliggende redenering eerder en gemakkelijker opgemerkt kunnen worden.

Conclusie 27: De rol van 'kijken met blauwe ogen' in de selectieprocedure blijft nog enigszins onduidelijk. Heeft de Politieacademie expliciete ideeën over wat voor aanvullende informatie deze stap in de selectieprocedure oplevert? En zo ja, hoe schat men de waarde van deze informatie in, in termen van betrouwbaarheid, construct validiteit, en voorspellende waarde?

Hierop volgend stellen we voor: Wees expliciet over hoe men de rol van kennis van executieve functionarissen in de selectieprocedure beschouwd (e.g., in welke mate gewenst, valide, etc.) en wees bewust van de mogelijke voor- en/of nadelen van het minimaliseren van deze kennis.

Conclusie 28: Het onderscheid tussen enerzijds de beoordeling van waarden en anderzijds de meting van competenties en persoonlijkheid is niet volledig duidelijk. Daarom zou het nuttig zijn om empirisch na te gaan wat het eindgesprek onderscheidt van de karakteristieken die in de centrale selectie worden gemeten. Hiertoe zouden, zodra er voldoende gegevens beschikbaar zijn, de statistisch verbanden tussen de waarden, competenties en persoonlijkheid worden bepaald.

Mogelijkerwijs betreft het onderscheid het verschil tussen typische en maximale prestatie (in het Engels wordt dit aangeduid als *typical versus maximum performance*: Klehe & Latham, 2006). Tijdens het selectiegesprek op de academie en de praktijkopdracht laat een kandidaat zich wellicht van zijn beste kant (maximale prestatie) zien terwijl de waarden meer betrekking hebben op hoe een kandidaat zich typisch gedraagt dan wel hoe de kandidaat zich idealiter zou willen gedragen.

Conclusie 29: Op grond van de analyses is de conclusie dat de selectieonderdelen slechts beperkt de opleidingscriteria voorspellen. Er dienen echter twee belangrijke kanttekeningen bij deze conclusie geplaatst te worden.

Ten eerste blijft vooralsnog onbekend in hoeverre de criteria indicatief zijn voor daadwerkelijk opleidingssucces en, nog belangrijker, voor latere werkprestatie. De criteria hebben een aantal beperkingen die de interpretatie van de in dit rapport beschreven uitkomsten bemoeilijken (zie beschrijving van de criteria in Hoofdstuk 3).

Ten tweede zijn de relatief lage relaties tussen selectieonderdelen en uitkomstmaten (criteria) niet specifiek voor selectie bij de politie. Een recent artikel over opleidingsuitval bij de Nederlandse landmacht laat bijvoorbeeld zien dat persoonlijkheidsmaten (De Big Five) nauwelijks kunnen voorspellen wie er tijdens de opleiding zal uitvallen (Van der Linden, Te Nijenhuis, Cremer, & Van der Ven, 2011). Wat betreft de Big Five persoonlijkheidsdimensies is bovendien bekend dat de validiteiten van persoonlijkheidskenmerken laag zijn ongeacht welk criterium men wilt voorspellen (Barrick & Mount, 1991). Er is minder wetenschappelijke literatuur beschikbaar over competenties dan over persoonlijkheid. Desalniettemin lijken voorgaande studies ook aan te geven dat de voorspellende waarde van competenties vaak relatief laag is.

Conclusie 30: De huidige opleidingsmaten (proeven van bekwaamheid en opleidingsstatus) hebben een aantal kenmerken die het moeilijk maken om goed onderzoek te doen naar de relatie tussen de selectie-instrumenten en opleidingssucces. Men zou kunnen overwegen om bij de beoordeling van aspirant-agenten in de toekomst, scores te gebruiken die beter kunnen differentiëren tussen aspirant-agenten dan nu het geval is. Dit levert aanvullende informatie op voor de opleiding over de aspirant-agenten en kan tevens leiden tot betere inzichten in de efficiency van werving en selectie. Men kan bijvoorbeeld denken aan het invoeren van standaard schoolcijfers (1 t/m 10) bij de proeven van bekwaamheid.

Conclusie 31: Voor een adequate evaluatie van zowel de activiteiten van werving en selectie en van de politieopleiding is het essentieel dat er onderzoek wordt uitgevoerd met zoveel mogelijk directe prestatie-indicatoren tijdens het politiewerk. Men kan daarbij denken aan beoordelingen van leidinggevenden of anderen directe prestatie-indicatoren.

Conclusie 32: Het is belangrijk dat er meer inzicht komt in de psychometrische kenmerken alsook de voorspellende waarde van de gebruikte cognitieve tests. Indien deze niet beschikbaar is valt een vervanging van de test te overwegen.

Conclusie 33: Ondanks de lage validiteiten, stellen we voor de persoonlijkheidsmaten te handhaven in de selectieprocedure van de politie. De status van persoonlijkheid is in de literatuur dusdanig dat het een te groot risico zou zijn om van het gebruik van deze maten af te zien. Het instrument dat hier in de huidige situatie voor wordt gebruikt, namelijk de NEO-PI-R is één van de bekendste en meest gevalideerde in zijn gebied en zou dus volstaan. Dit neemt niet weg dat het in de toekomst zinvol kan zijn om mogelijkheden te exploreren om de validiteit van persoonlijkheidsmetingen te verbeteren. Bijvoorbeeld door na te gaan of gecontextualiseerde persoonlijkheidsvragenlijsten (beter toegespitst op politieorganisatie) of ander-beoordelingen (in plaats van traditionele zelfbeoordeling) de predictieve validiteit kan verhogen.

Conclusie 34: Op grond van de literatuur en de uitkomsten van de analyse van de selectiedata valt te overwegen om minder competenties te beoordelen in het selectiegesprek. Het zou wellicht beter zijn om kandidaten in te schatten op bredere gedragscategorieën die relevant zijn voor politiewerk. Eén van de mogelijkheden hiervoor is de kandidaten te beoordelen op vier categorieën, te weten: Daadkracht, Sociale effectiviteit, Weerbaarheid, en Houding/Integriteit. Het voordeel van zo'n aanpak is dat de cognitieve last van de psycholoog wat betreft de beoordeling wordt verminderd en de beoordelingen daardoor betrouwbaarder en robuuster worden. Indien er een verandering in de te beoordelen competenties plaats zou vinden dienen de uitkomsten daarvan uiteraard geëvalueerd te worden door de voorspellende waarde voor opleidingssucces en, het liefst, voor criteria op het werk (prestaties, verloop, tevredenheid) te bepalen.

Conclusie 35: De gegevens laten *niet* duidelijk zien dat de praktijkopdracht als selectieonderdeel slechter zou werken dan de andere selectieonderdelen. In sommige opzichten doet de proef het zelfs iets beter. Onze aanbeveling is dan ook om de praktijkopdracht in de selectieprocedure te behouden. In elk geval tot meer informatie beschikbaar is.

Conclusie 36: Het verhogen van ondergrenzen van competenties zorgt er waarschijnlijk voor dat kandidaten die worden toegelaten minder kans hebben op een aantal problemen zoals verminderde psychische weerbaarheid, conflict etc. Men moet echter beseffen dat er bij verandering van ondergrenzen sprake is van een trade-off. De negatieve kanten van de trade-off zijn een hoger afwijzingspercentage en hogere selectiekosten.

Conclusie 37: De beschikbare gegevens over de selectieonderdelen en de criteria kunnen nauwelijks dienen als uitgangspunt voor beslissingen omtrent ondergrenzen of normscores.

Conclusie 38: Alhoewel de selectieprocedure bij de politie in grote lijnen goed lijkt te functioneren zijn er altijd kandidaten die door zullen stromen voor wie een politiefunctie achteraf niet past of die te grote risicofactoren vertonen wat betreft latere functie-uitvoering. De politieacademie heeft zeer recent enkele stappen ondernomen om flexibel te reageren op dit soort situaties, bijvoorbeeld door het aanbieden van tijdelijke contracten en het invoeren van een bindend studieadvies. Desalniettemin is het zinvol om goed te blijven monitoren in welke mate deze maatregelen in de praktijk op een adequate manier worden toegepast en of ze de gewenste uitkomsten hebben. Vooralsnog kon in het huidige project namelijk onvolledig worden vastgesteld in hoeverre de genoemde maatregelen goed werken.

Conclusie 39: Het is wenselijk dat wanneer het gaat om het invoeren van nieuw of tijdelijk beleid men rekening houdt met de regelmaat in de bedrijfsvoering en mogelijke financiële gevolgen.

Conclusie 40: De GFP was duidelijk aanwezig in de persoonlijkheidsdata van de selectie bij de politie. Deze GFP vertoonde enige overlap met de beoordelingen uit het selectiegesprek en de praktijkopdracht. Dit suggereert een relatief consistente factor. De GFP was echter niet sterk voorspellend voor uitval of opleidingsprestatie.

Gezien de huidige status van persoonlijkheid in de selectie bij de politie is het (nog) niet zinvol om gebruik te maken van nieuwe constructen zoals de GFP en eerst af te wachten op de uitkomsten van meer onderzoek waarin valide criteria zijn gebruikt.

Conclusie 41:

- Emotionele intelligentie (EI) wordt in de literatuur beschreven als een begrip bestaande uit kennis over sociale situaties en emoties en het vermogen om deze kennis te gebruiken om doelen te bereiken.

- EI blijkt samen te hangen met baanprestatie, psychische weerbaarheid, en integriteit. De mogelijk toegevoegde waarde van het meten van EI is dat men een concept gebruikt dat een brede invloed heeft op gedrag en daardoor mogelijk veel van de belangrijke aspecten van politiewerk kan voorspellen. De literatuur onderscheidt EI als capaciteit van EI als persoonlijkheidstrekk die andere persoonlijkheidsdimensies beïnvloed. Bovendien wordt onderscheid gemaakt in het meten van EI met een capaciteitentest in tegenstelling tot meten met behulp van zelfrapportage.

Wij stellen voor de mogelijkheid te verkennen om EI metingen te introduceren in de selectieprocedure bij de politie.

Conclusie 42: In de selectie van aspirant-agenten zou men kunnen overwegen om aanvullende maten mee te nemen die onethisch en/of negatief werkgedrag kunnen voorspellen. Daarbij kan men bijvoorbeeld denken aan maten die de duistere kanten van persoonlijkheid beogen te meten zoals de Dark Triad: psychopathie, narcisme, en machiavellisme.

Referenties

Referenties

- Aanstellingseisen politie (2002). Ministerie van Veiligheid en Justitie. Verkregen van <http://wetten.overheid.nl/BWBR0014138/>
- Ali, A. O. E., Garner, I., & Magadley, W. (2011). An Exploration of the Relationship Between Emotional Intelligence and Job Performance in Police Organizations. *Journal of Police and Criminal Psychology, 27*, 1-8.
- Andersson Elffers Felix (AEF). (2010). *De prijs die je betaalt; Politie: de kosten achter een hoog risicoberoep*, Netherlands
- Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among big five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology, 97*, 1142-1156.
- Avrey, R. D., Landon, T.E., Nutting, S.M., & Maxwell, S.E., (1992). Development of physical tests for police officers: A construct validation approach. *Journal of Applied Psychology, 77*, 996-1009.
- Bakker, S. (2009). Een vervolgonderzoek naar de betrouwbaarheid van de cognitieve capaciteiten test van de Politieacademie.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). The FFM personality dimensions and job performance: Meta-analysis of meta-analyses. Invited submission to a special "selection" issue of *International Journal of Selection and Assessment, 9*, 9–30.
- Campion, M. A. (1983) Personnel selection for physical demanding jobs: Review and recommendations, *Personnel Psychology, 36*, 527-550.
- Cook, K. W., & Bernthal, P. (1998). Job/role competency practices survey report. Bridgeville, PA.
- Cote, S., & Miners, C. T. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly, 51*, 1-28.
- Courtrigh, S. H., McCormick, B. W., Postlethwaite, B. E., Keeves, C. J., Mount, M.K., (2013). A meta-analysis on sex differences in physical ability: Revised estimates and strategies for reducing differences in selection contexts. *Journal of Applied Psychology, 98*, 623-641.

- De Meijer, L. A. (2008). Ethnicity effects in police officer selection: Applicant, assessor, and selection method factors. Proefschrift. Erasmus Universiteit Rotterdam
- De Meijer, L. A. L., Born, M. Ph., Terlouw, G., & Van der Molen, H. T. (2008). Criterion-Related Validity of Dutch Police-Selection Measures and Differences Between Ethnic Groups. *International Journal of Selection and Assessment*, 16, 321-332.
- Directie HRM Nationale Politie (2012). De toekomst begint vandaag.
- Directie Strategie (2012).
- Figueredo, A. J., Vásquez, G., Brumbach, B. H., Schneider, S. M. R., Sefcek, J. A., Tal, I. R., Hill, D., Wenner, C., & Jacobs, W. J. (2006). Consilience and life history theory: From genes to brain to reproductive strategy. *Developmental Review*, 26, 243-275.
- Fleischman, E.A., Wetrogan, L.I., Uhlman, C.E., & Marshall-Miles J.C. (1995) In Petertson N.G., Mumford M.D., Borman, W.C, Jeanneret P.R., Fleishman, E.A (Eds.). Development of prototype occupational information network content model (Vol. 1, pp. 10.1-10.39). Utah: Utah Department of Employment Security.
- Gaugler, B. B., & Thornton, G. C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, 74, 611-618.
- Green, P. C. (1999). *Building robust competencies: Linking human resource systems to organizational strategies*. San Francisco, EUA: Jossey-Bass.
- Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions*. Taylor & Francis.
- Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology*, 39, 399-420.
- Hogan, J. (1991). Structure of physical performance in occupational tasks. *Journal of Applied Psychology*, 76, 495-507.
- Holland, J. L. (1996). Exploring careers with a typology: What we have learned and some new directions. *American Psychologist*, 51,397–406.
- Hough, L.M. & Oswald, F. L., (2000). Personnel selection: Looking toward the future- Remembering the past. *Annual Review of Psychology*, 51, 631-664.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. SAGE Publications, Incorporated.

- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kemper, H. C. G., De Vente, W., Van Mechelen, W. & Twisk, W.J.R., (2001). Adolescent motor skills and performance: Is physical activity at adolescence related adult physical fitness? *American Journal of Human Biology*, *13*, 180-189.
- Klinkenberg, E. L., en Van Leeuwen, A. E. (2003). *Voortgangverslag ontwikkeling M5Q- IWSP*. Culemborg, Nederland: Meurs Personeelsadvies.
- Kwartiermaker Nationale Politie (2012). *Ontwerp selectiecommissie Nationale Politie: Eindgesprek*.
- Kwartiermaker Nationale Politie (2002).
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment*, *6*, 141–152.
- Loehlin, J. C. (2012). The general factor of personality: What lies beyond? *Personality and Individual Differences*, *53*, 463-467.
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–31). New York: Basic Books.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, *15*, 197-215.
- Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, *41*, 1213-1233.
- Onderzoekscluster Afdeling Business Support, Werving & Selectie. (2009). *Politie selectie onder de loep*.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556-563.
- Petrides, K. V. (2011). An application of belief-importance theory with reference to trait emotional intelligence, mood, and somatic complaints. *Scandinavian Journal of Psychology*, *52*, 161-167.
- Pit, M. (2010). *Een onderzoek naar lees- en schrijftaken in de politieopleidingen en de korpsen om de vereiste instroomniveaus voor lees- en schrijfvaardigheid vast te stellen*.
- Politieacademie (2002). *Open advisering en bindend advies*.

- Prins, J., & Terlouw, G. (2007). Instroom, studievoortgang en studieuitval initiële politieopleidingen 2002 t/m 2006.
- Rijksdocument Aanstellingseisen politie (2002).
- RijksPsychologische Dienst. (1975) Politie Intelligentie Test [Police Intelligence Test]. The Hague, the Netherlands: Rijks Psychologische Dienst (RPD).
- Roe, R.A. (2002). Competenties - Een sleutel tot integratie in theorie en praktijk van de A&O psychologie. *Gedrag & organisatie*, 15, 203-224.
- Rushton, J. P., & Irwing, P. (2011). The general factor of personality: Normal and abnormal. In Chamorro-Premuzic, T, von Stumm, S., & Furnham, A (Eds.). *The Wiley- Blackwell handbook of individual differences*. Wiley-Blackwell: London (2011), pp. 132–161.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J. F., Moscoso, S., de Fruyt, F., Anderson, N., Bertua, C., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for occupations in the European community. *Journal of Applied Psychology*, 88, 1068-1081.
- 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.
- Schmidt, F. L., & Hunter. J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162-173.
- Schutte, N.S., Malouff, J.M., Hall, L.E., Haggerty, D.J., Cooper, J.T., Golden, C.J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence, *Personality and Individual Differences*, 25,167-177.
- Smith, M. V. (2009). Police perceptions of integrity: The relationship between emotional intelligence and moral development among police officers. (Doctoral dissertation).
- Sorensen, L., Smolander, J. Louhevaara, V., Korhone, O., Oja, P. (2000) Physical activity, fitness, and body composition of Finish police officers: A 15-year follow up study, *Occupational Medicine*, 50, 3-10.
- Spencer, jr., L. M., & Spencer, S. M. (1993). *Competence at work. Models for superior performance*. New York: John Wiley & Sons.

- Startnotitie WODC-onderzoek (2012). Aanstellingseisen aspiranten.
- Steinberg, R. J., & Figart, D. M. (1999). Emotional demands at work: A job content analysis. *The Annals of the American Academy of Political and Social Science*, 561, 177-191.
- Taris, R. (2012). Uitvalpercentages en selectieratio's in verschillende stappen van het werving- en selectieproces Nederlandse Politie.
- Taylor, H.C., & Russel, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Te Nijenhuis, J. & Van der Flier, (1999). Bias research in the Netherlands: Review and implications, *European Journal of Psychological Assessment*, 15, 165-175.
- Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine*, 140, 227-235.
- Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin*, 34, 275-385.
- Van Leeuwen, A. E. (2000). *Constructie van de M5Q voor IWSP*. Culemborg, Nederland: Meurs Personeelsadvies.
- Van der Linden, D. & Born, M. Ph. (2012). Projectvoorstel 'evaluatie aanstellingseisen aspirant-agenten.' Erasmus Universiteit Rotterdam.
- Van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, 44, 315-327.
- Van der Linden, D., te Nijenhuis, J., Cremers, M., & Van de Ven, C. (2011). General Factor of Personality in Six Datasets and a Criterion-Related Validity Study at the Netherlands Armed Forces. *International Journal of Selection and Assessment*, 19, 157-169.
- Van der Linden, D., Tsaousis, I., & Petrides, K.V. (2012). Overlap between General Factors of Personality in the Big Five, Giant Three, and trait emotional intelligence. *Personality and Individual Differences*, 53, 175-179.
- Velden, P.G. van der, Bosmans, M.W.G., & Brekveld, S. (2012). Politiestress: feiten en mythes. *Het tijdschrift voor de politie*, 74, 6-9
- Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*, 65, 71-95.

- Vuzzo, D. V. (2009). Emotional intelligence and stress on today's law enforcement sergeant. (Doctoral dissertation).
- Wong, C. S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The Leadership Quarterly*, *13*, 243–274.
- Zeidner, M., Matthews, G., & Roberts, R. D. (2004). Emotional intelligence in the workplace: A critical review. *Applied psychology: an international review*, *53*, 371-399.
- Zeidner, M., Matthews, G., & Roberts, R. D. (2009). What we know about emotional intelligence: How it affects learning, work, relationships, and our mental health. London: England, Cambridge, Massachusetts, Institute of Technology.

Bijlagen

Bijlage 1 Geïnterviewden

Naam:	Functie
Bea Marsman	Programmamanager HRM en onderwijs
Conchita Alvarez	Hoofd school hogere politiekunde
Dineke Oldenhof	Directeur HRM Nationale Politie
Ellis Beemsterboer	Medewerker Werving & Selectie
Geoffry Gill	Docent politieacademie
Geraldine van den Brink	Medewerker Werving & Selectie
Gerard Visser	Medewerker Werving & Selectie
Gert Terlouw	Medewerker Werving & Selectie
Jennie Pierik	Docente politieacademie
Jitske Tensen-Schellingerhout	Medewerker Werving & Selectie
Jose Nelis	Beleidsmedewerkster kwartiermaker
Lena Westerveld	Medewerker Werving & Selectie
Mark Halfmouw	Sportdocent politieacademie
Martin Staats	Directeur Werving & Selectie
Maykel Van Rijen	Strategische personeelsplanning
Een executieve politiefunctionaris (anoniem)	

Bijlage 2

METHODEVERANTWOORDING

Om de gebruikte criteriummaten te valideren zijn de proportiematen van proeven van bekwaamheid (zie Hoofdstuk 4 voor beschrijving) vergeleken met de opleidingsstatus van de aspirant-agenten. De resultaten van deze analyse zijn hieronder weergegeven (Tabel B2.1) en laten zien dat er een significant hoofdeffect was van opleidingsstatus (Figuur B2.1). Degenen die vroegtijdig zijn gestopt met de opleiding hebben gemiddeld 90% van de onderdelen van de proeven behaald in hun eerste poging. Dit zijn er echter significant minder dan het aantal dat met diploma is afgestudeerd of nog bezig zijn. Zij behaalden namelijk 93% van de onderdelen in de eerste poging.

Tabel B2.1. Variantie-analyse per Opleidingsniveau (diploma, vroegtijdig stoppen, studerend) en Instroomniveau (2, 3, 4), op proportie behaalde onderdelen van de proeven in eerste poging.

Source	df	Mean Square	F	Sig.
Opleidingsstatus (Opl)	2	.03	12.09	.01
Instroomniveau (Instr)	2	.00	1.79	.16
Opl * Instr	4	.00	2.96	.01
Error	2059	.00		
Total	2068			

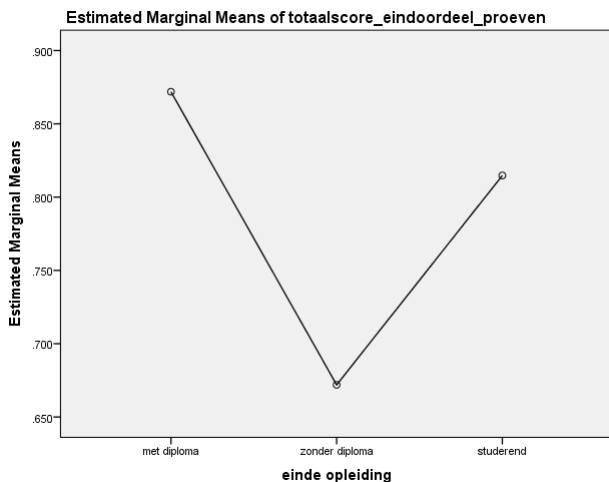
Er was ook een significant interactie-effect van opleidingsstatus en instroomniveau. Er wordt duidelijk dat met name op niveau 4 er een groot verschil is tussen aspirant-agenten met diploma (of studerend) en aspirant-agenten die vroegtijdig zijn uitgevallen. De laatstgenoemden behalen relatief weinig onderdelen in de eerste poging (e.g., .87 voor de uitvallers ten opzichte van .94 voor de gediplomeerden). Op niveau 2 zien we een soortgelijk patroon alhoewel het verschil tussen aspiranten met diploma en degenen die gestopt zijn minder groot is en niet significant. Opvallend is dat er op niveau 3 geen verschil lijkt te zijn in proportie behaalde onderdelen tussen degenen die zijn gestopt en degenen die met diploma de studie hebben afgerond of nog studerend zijn.

Dezelfde analyse maar dan met proportie behaalde proeven van bekwaamheid gaf een consistent beeld (zie ook Figuren B2.1 en B2.2). In deze analyse was er een significant hoofdeffect van opleidingsstatus. Degenen die met diploma afstudeerden of nog studerend waren behaalden meer proeven in de eerste poging dan degenen die vroegtijdig met de opleiding waren gestopt. De proporties behaalde proeven van degenen met diploma en zonder diploma waren respectievelijk .87 en .66. Ook was dit verschil consistent over de verschillende niveaus alhoewel het hier het grootst was voor degenen op niveau 3.

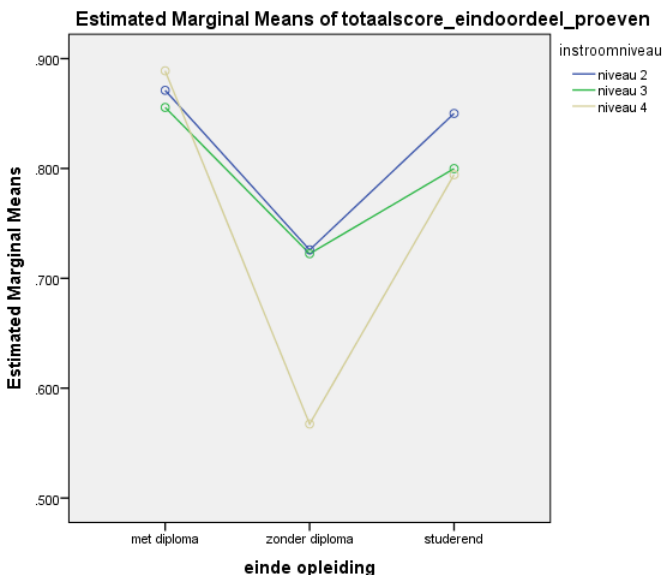
Over het algemeen zijn de relaties tussen de opleidingsmaten en opleidingsstatus zwak. De effecten variëren ergens tussen een half en twee procent van de verklaarde variantie. Er is dus wel een verband tussen het aantal proeven en opleidingsstatus wat suggereert dat de gebruikte criteriummaat enigszins inzicht geeft in de prestatie tijdens de opleiding. De uitkomsten betreffende deze maten moet wel met uiterste voorzichtigheid worden geïnterpreteerd

Tabel B2.2. Variantie-analyse gehaalde proeven per Opleidingsniveau (diploma, vroegtijdig stoppen, studerend) en Instroomniveau (2, 3, 4)

Source	df	Mean Square	F	Sig.
Opleidingstatus (opl)	2	.83	20.48	.00
Instroomniveau (Instr)	2	.06	1.70	.18
Opl * Instr	4	.09	2.36	.05
Error	2059	.04		
Total	2068			



Figuur B2.1, proportie behaalde overall toetsen in eerste poging per opleidingsstatus



Figuur B2.2, proportie behaalde overall toetsen in eerste poging per opleidingsstatus en instroomniveau

Bijlage 3

ANALYSES COGNITIEVE CAPACITEIT

In de huidige bijlage worden de details besproken van de methode en analyses van cognitieve capaciteit. Inhoudelijke besprekingen en samenvatting van de resultaten zijn te vinden in Hoofdstuk 5 van het rapport.

B3.1 Statistische gegevens en analyses cognitieve capaciteit

In de huidige selectieprocedure van de politieacademie maakt men gebruik van een Cognitieve Adaptieve Intelligentie Test (De CAIT). Deze bestaat uit drie subtesten die in Hoofdstuk 5 zijn beschreven.

Tabel B3.1 tot en met B3.3 hieronder geven de beschrijvende statistieken van de drie subtests van de cognitieve testbatterij.

Tabel B3.1. Beschrijvende statistieken van Abalet test

Groep	N	Gem	SD
Niv2 Man	580	-109	106
Niv2 Vrouw	194	-113	113
Niv3 Man	2437	-67	115
Niv3 Vrouw	680	-72	112
Niv4 Man	3161	-13	119
Niv4 Vrouw	1006	-15	119
Niv5 Man	4430	39	129
Niv5 Vrouw	1678	48	131
Niv6 Man	701	74	153
Niv6 Vrouw	446	110	144

Range van scores op test: -510 tot 510

Tabel B3.2. Beschrijvende statistiek van Anabo test

Groep	N	Gem	SD
Niv2 Man	553	-185	86
Niv2 Vrouw	182	-204	72
Niv3 Man	2430	-155	85
Niv3 Vrouw	677	-178	78
Niv4 Man	3277	-110	86
Niv4 Vrouw	1034	-134	78
Niv5 Man	4600	-62	95
Niv5 Vrouw	1733	-90	85
Niv6 Man	1342	17	108
Niv6 Vrouw	1029	-4	99

Range van scores op test: -510 tot 510

Tabel B3.3. Beschrijvende statistiek, Cijfcrea-test

Groep	N	Gem	SD
Niv2 Man	536	-198	169
Niv2 Vrouw	175	-258	140
Niv3 Man	2384	-132	168
Niv3 Vrouw	664	-203	155
Niv4 Man	3232	-41	170
Niv4 Vrouw	1018	-106	165
Niv5 Man	4503	42	174
Niv5 Vrouw	1695	-22	175
Niv6 Man	1307	138	183
Niv6 Vrouw	997	116	174

Range van scores op test: -510 tot 510

B3.2 Resultaten van de factor analyse op de subtests van de CAIT

Om te toetsen of er in de CAIT inderdaad een algemene factor te vinden is die een aanduiding zou kunnen zijn voor algemene intelligentie, werd een factoranalyse uitgevoerd. Daarbij is gebruikt gemaakt van exploratieve factor analyse met als criterium voor factor-extractie een Eigenwaarde van minimaal 1.

Onderstaande tabel B3.4 laat de Eigenwaarden zien van de factoren. Hieruit valt af te leiden dat er slechts één grote en duidelijk factor te onderscheiden valt die maar liefst 70% van de variantie verklaarde. Tabel B3.5 laat de ladingen van de subtests op de algemene factor zien.

Tabel B3.4 Factor analyse uitkomst van de drie CAIT onderdelen

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.1	70.40	70.41
2	.52	17.59	88.00
3	.36	12.00	100.00

Tabel B3.5 Factor ladingen CAIT onderdelen op de algemene factor

	Algemene factor
Abalet	.85
Anabo	.79
Cijfcrea	.87

B3.3 Analyses van de voorspellende waarde van de CAIT op opleidingsprestatie

Zoals beschreven in hoofdstuk 5 van het rapport, is er getoetst of de score op de CAIT verschilt in kandidaten die de opleiding hebben afgemaakt versus kandidaten die vroegtijdig zijn gestopt. Verder zijn de correlaties berekend tussen CAIT scores en opleidingsprestatie. De procedures en uitkomsten zijn hieronder in detail besproken.

B3.3.1 Intelligentie en opleidingsstatus

De eerste analyses bestonden uit multivariate en univariate analyses met algemene intelligentie (*g*-factor of de drie subtest scores) als afhankelijke variabele en opleidingsstatus als onafhankelijke variabele. Bij de onafhankelijke variabele werd in eerste instantie onderscheid gemaakt tussen kandidaten die hun studie inmiddels met diploma hadden afgerond, kandidaten die de opleiding zijn gestopt zonder diploma en kandidaten die op het moment van meten nog aan het studeren waren (zie ook Hoofdstuk 4 methodeverantwoording). Aangezien de steekproef in dit geval betrekking had op de gegevens vanaf 2008 waren er echter relatief weinig kandidaten die de opleiding in 2012 reeds met diploma konden hebben afgerond. Op de niveaus 3 en 4 ging het bijvoorbeeld vaak om minder dan 10 kandidaten (zie tabel B3.6). Dit levert geen betrouwbare resultaten op en om die reden hebben is besloten die kandidaten niet mee te nemen. De categorie van kandidaten die inmiddels waren uitgevallen uit de studie was groter. Uit tabel B3.7 valt te zien dat het hier ging om 22, 65 en 22 kandidaten op respectievelijk niveau 2, 3 en 4. Alhoewel dit analyses toelaat moet ook hier de uitkomsten met enige voorzichtigheid worden geïnterpreteerd omdat de groepen nog steeds klein zijn. De analyses zijn uitgevoerd per instroomniveau. De resultaten met de algemene intelligentiescore laten zien dat er op geen enkel niveau een significant verschil was tussen kandidaten die nog studeren en kandidaten die vroegtijdig met de opleiding zijn gestopt. Soortgelijke met de scores op de subtests leverde geen aanvullende informatie op aangezien de patronen van resultaten over het algemeen hetzelfde zijn als die van de algemene intelligentiemaat. Alhoewel de De Meijer et al. (2008) in eerder onderzoek onderscheid maakte tussen allochtone en autochtone kandidaten was dat bij de huidige analyses niet zinvol omdat daardoor te kleine groepen van analyses ontstonden. In de correlatieve analyses hieronder beschreven nemen we het onderscheid tussen allochtoon en autochtoon echter wel weer wel expliciet mee.

Tabel B3.6. Aantal personen in de cellen Opleidingsstatus * instroomniveau

	instroomniveau			Total
	niveau 2	niveau 3	niveau 4	
einde opleiding met diploma	67	3	6	76
zonder diploma	22	65	22	109
studerend	150	959	774	1883
Totaal	239	1027	802	2068

Tabel B3.7. Cognitieve capaciteit (g-factor) per opleidingsstatus en instroomniveau

Source	df	Mean Square	F	Sig.
Opleiding (Opl)	1	.01	.02	.87
Instroomniveau (instr)	2	9.97	18.39	.000
Opl * instr	2	.48	.88	.41
Error	1839	.54		
Total	1845			

B3.3.2 Intelligentie en opleidingsprestatie

De tweede reeks analyses die zijn uitgevoerd betroffen de correlaties tussen intelligentiescores en de opleidingsgegevens. De algemene correlaties tussen opleidingsprestatie en intelligentie (niet uitsplitst voor instroomniveau) waren $r = .054$ wanneer het gaat om de scores op de onderdelen van de proeven en $r = .023$ wanneer het gaat om de scores op de proeven zelf. Dit is dus zeer laag ten opzicht van de waarden die typisch in de literatuur worden vermeld. Opgesplitst naar instroomniveau waren de correlaties $.051$, $.060$ en $.092$ voor respectievelijk niveau 2, 3 en 4 wanneer het ging om onderdelen van de proeven van bekwaamheid. Voor de scores op de proeven zelf was dit $-.023$, $.047$, en $.070$ voor respectievelijk niveau 2, 3, en 4. Ook op deze manier berekend zijn de associaties laag. Uitgaande van de absolute waarde lijken de relaties wel iets sterker te worden naarmate de complexiteit van de functie stijgt. Dit laatste is in overeenstemming met de literatuur. Desalniettemin zijn ook deze verschillen klein en niet significant.

De meest gedetailleerde analyses betroffen de opsplitsing naar functieniveau en naar allochtoon of autochtoon. De gegevens hiervan zijn gerapporteerd in het rapport in hoofdstuk 5 (Tabel 5.8) en zijn daarom niet in deze bijlage weergegeven. De meest opvallende uitkomst van deze analyses is overigens dat de hoogste voorspellende waarden die zijn gevonden, negatief waren. Deze negatieve voorspellende waarden werden gevonden in de groep van allochtone aspirant-agenten.

Bijlage 4

ANALYSES HET FYSIEK-MOTORISCH ONDERZOEK (FMO)

Voor de analyses van FMO zijn opleidingsstatus (afgerond met diploma, Vroegtijdig gestopt, nog studierend) en instroomniveau (2, 3, 4) als onafhankelijke factoren meegenomen. Gezien de duidelijke geslachtsverschillen hebben we geslacht toegevoegd als extra onafhankelijke variabele in de analyses. Aangezien een significant hoofdeffect van geslacht te verwachten valt werken we in de beschrijving van de analyses hieronder eventuele effecten van geslacht in de meeste gevallen niet verder uit. De onderdelen van het FMO zijn gebruikt als afhankelijke variabele in afzonderlijke analyses.

Uit de analyses kwam naar voren dat er wat betref de parcoursprestatie er significante hoofdeffecten waren van opleidingsstatus ($F(2, 6082) = 25.48, p < .01$), instroomniveau ($F(2, 6082) = 28.26, p < .01$) en geslacht ($F(1, 6082) = 5318.55, p < .01$). Van de interactie-effecten, was alleen het interactie-effect tussen instroomniveau en geslacht significant ($F(2, 6082) = 1711.00, p < .01$).

Post-hoc toetsen lieten zien dat kandidaten die zonder diploma de opleiding hebben verlaten significant het langzaamst waren op het parcours en dat degenen die nog aan het studeren waren significant langzamer waren dan degenen die reeds een diploma hadden behaald, maar wel significant sneller waren dan de kandidaten die zonder diploma de opleiding verlieten. Het absolute verschil tussen de groepen was echter klein, namelijk tussen rond de 5 seconden.

De post-hoc toetsen lieten verder zien dat alle drie de instroomniveaus (niveau 2 tot en met 4) significant van elkaar verschilden. Kandidaten op niveau 4 zijn gemiddeld significant het snelst, gevolgd door de kandidaten op niveau 3 en de kandidaten op niveau 2 zijn het langzaamst.

Wat betreft wekelijkse sportbeoefening was er een significant hoofdeffect van geslacht ($F(1, 6014) = 18.21, p < .01$). Verder waren geen significante hoofdeffecten of interactie-effecten zichtbaar (alle p s $> .05$). Dus het aantal keren sportbeoefening per week had geen relatie met uitval uit de opleiding.

Wat betreft blessures waren er significante hoofdeffecten van instroomniveau ($F(2, 6015) = 9.46, p < .01$) en opleidingsstatus ($F(2, 6015) = 7.92, p < .01$), dat verder werd gekwantificeerd door een significante interactie tussen opleidingsstatus en instroomniveau ($F(4, 6015) = 4.12, p < .01$). De andere hoofd- en interactie-effecten waren niet significant.

De post-hoc toetsen lieten zien dat wat betreft opleidingsstatus alle verschillen significant zijn. De kandidaten met diploma zijn gemiddeld genomen het minst geblesseerd, gevolgd door kandidaten die zonder diploma de opleiding hebben verlaten en de studerende kandidaten hebben gemiddeld de langst durende blessures. Voor instroomniveau is het verschil tussen niveau 2 en 3 significant, maar tussen 3 en 4 niet. Kandidaten op niveau 2 zijn het minst lang geblesseerd, gevolgd door niveau 4 en kandidaten op niveau 3 zijn het langst geblesseerd.

Voor conditie waren de hoofdeffecten van instroomniveau ($F(2, 6011) = 4.44, p = .01$) en opleidingsstatus ($F(2, 6011) = 14.29, p < .01$) significant en was het interactie-effect tussen opleidingsstatus en instroomniveau eveneens significant ($F(4, 6011) = 6.91, p < .01$).

De post-hoc toetsen lieten zien dat er significante verschillen waren tussen aan de ene kant de kandidaten die de opleiding met een diploma hebben verlaten en aan de andere kant de kandidaten die de opleiding zonder diploma hebben verlaten en de kandidaten die nog studeren. De kandidaten die de opleiding met een diploma verlaten hebben gemiddeld namelijk een slechtere conditie dan de andere twee groepen. Verder lieten de post-hoc toetsen zien dat kandidaten op niveau 3 wel significant verschilden van kandidaten op niveau 2, maar niet van kandidaten op niveau 4. De kandidaten op niveau 2 hebben namelijk een slechtere conditie dan de kandidaten op niveau 3.

Vetpercentage heeft een significant hoofdeffect voor opleidingsstatus ($F(2, 6053) = 5.00, p < .01$) en een significant hoofdeffect voor geslacht ($F(1, 6053) = 1626.01, p < .01$). De andere hoofd- of interactie-effecten waren niet significant.

De post hoc-toetsen laten zien dat de nog studerende kandidaten gemiddeld significant een hoger vetpercentage hebben dan de kandidaten die de opleiding zonder diploma hebben verlaten en dat vrouwen gemiddeld genomen een hoger vetpercentage hebben dan mannen.

Bijlage 5

ANALYSES PERSOONLIJKHEID

Bij het analyseren van de voorspellende waarde van persoonlijkheid is dezelfde aanpak gehanteerd als bij intelligentie (in Hoofdstuk 5 en 6 en Bijlage 3 en 4). Bij de analyse van opleidingsstatus worden de scores van degenen die vroegtijdig zijn gestopt zonder diploma vergeleken met degenen die nog studeren. Verder relateren we de persoonlijkheidsscores aan de proportie proeven die de kandidaten hebben gehaald in de eerste poging tijdens de opleiding.

B5.1 Persoonlijkheid en opleidingsstatus

Een multivariate analyse met opleidingsstatus (gestopt, studerend) en instroomniveau (2, 3, 4) als onafhankelijke variabelen en de Big Five dimensies als afhankelijke variabelen liet zien dat er een significant hoofdeffect was van Instroomniveau ($F(10, 3966) = 2.43, p < .01$) dat werd veroorzaakt door significante verschillen tussen de instroomniveaus op openheid voor nieuwe ervaringen ($F(2, 1986) = 25.58, p < .01$). Hoe hoger het instroomniveau hoe hoger de score op openheid. Er was echter geen significant hoofdeffect van opleidingsstatus ($F(5, 1982) = 1.25, ns.$) en geen significante interactie tussen opleidingsstatus en instroomniveau ($F(10, 3966) = .92, ns.$). Dit laat zien dat diegenen die voortijdig zijn gestopt met de opleiding niet significant anders scoorden op de Big Five dimensies dan diegenen die nog studeren.

B5.2. Persoonlijkheid en opleidingsprestatie

Vervolgens is de relatie getoetst tussen persoonlijkheid en opleidingsprestatie. De algemene correlaties waren zwak en niet significant (zie Tabel 8.3 in Hoofdstuk 8). Het opsplitsen naar instroomniveau leverde geen nieuwe inzichten op omdat het patroon van bevindingen in grote lijnen gelijk bleef. Dat wil zeggen, bijna geen van de correlaties waren significant op het .05 niveau. De uitzonderingen waren consciëntieusheid en aardigheid op instroomniveau 2. Consciëntieuzere en aardigere aspirant-agenten op dit niveau behaalde meer proeven van bekwaamheid tijdens hun eerste poging. In Hoofdstuk 6 in tabel 6.4 staan de afzonderlijke correlaties weergegeven tussen de Big Five en proportie behaalde proeven, opgesplitst naar instroomniveau. Hieronder in tabel B4.1 staan ook nog de correlaties tussen persoonlijkheid en behaalde onderdelen van proeven van bekwaamheid.

Tabel B5.1 Correlaties tussen Big Five en proportie geslaagde onderdelen van de proeven (opgesplitst op instroomniveau)

	Niveau 2 (n= 239)	Niveau 3 (n = 1027)	Niveau 4 (n = 802)
O	-.05	.04	-.01
C	.07	.01	.02
E	.04	.00	.06
A	.13*	.03	-.01
N	.02	-.01	.01

* $p < .05$.

Bijlage 6

ANALYSES COMPETENTIES

B6.1 Onderzoek naar het aantal clusters binnen de competenties

Om de samenhang (dan wel het onderscheid) tussen de verschillende competenties te onderzoeken zijn verschillende analyses uitgevoerd. Er is gebruik gemaakt van correlatieve analyses en factoranalyse. De competenties zijn sinds 2002 min of meer gelijk gebleven. Daarom zijn we in staat geweest om nagenoeg de meeste selectiedata van de academie te gebruiken sinds 2002. De *N* daarvan verschilde per competentie maar schommelde rond de 32500.

B6.1.1 Het selectiegesprek

De uitkomsten van de initiële factoranalyses op negen competenties zoals die zijn aangeleverd door de academie, is gerapporteerd in tabel B6.1. Bij deze factoranalyse hebben we gebruik gemaakt van principaal componenten analyse. Voor de extractie van het aantal factoren zijn we in eerste instantie uitgegaan van een Eigenwaarde groter dan één. Op grond van dit criterium zijn er twee factoren te onderscheiden die samen ongeveer 52.7 % van de variantie in de competenties verklaren. De screeplot ondersteunt dat er twee belangrijke factoren zijn.

In Tabel B6.2 is weergegeven hoe de verschillende competenties laden op de twee factoren, dat wil zeggen, in hoeverre de competenties verdeeld worden over de twee factoren. Merk op dat in de tabel ladingen kleiner dan .25 niet zijn weergegeven. De ladingen van de competenties laten zien dat kandidaten die hoog scoren op de eerste factor worden gekenmerkt door een relatief sterke contactgerichtheid en flexibiliteit. Bovendien scoren zij erg hoog op emotionele weerbaarheid, drukbestendigheid en zijn ze zeer besluitvaardig. In globale termen zou deze factor dus een indicatie zijn voor stevige en communicatieve personen. Een mogelijk label voor deze factor zou kunnen zijn 'Kracht'.

De tweede factor wordt gekenmerkt door een hoge mate van zelfinzicht, tolerantie en integriteit en in minder mate laadt communicatievaardigheden hierop. Globaal lijkt deze factor dus een indicatie te zijn van een soort algemene Wijsheid en Eerlijkheid/Mildheid.

Er zijn verschillen tussen mannen en vrouwen op de factoren. Mannen scoren hoger dan vrouwen op de factor kracht, terwijl vrouwen hoger dan mannen scoren op de factor wijsheid/mildheid. Daar moet echter wel bij vermeld worden dat de factorstructuur binnen elk geslacht bijna identiek is. Dat wil zeggen dat de onderverdeling van competenties in de twee factoren niet wordt veroorzaakt door verschillen tussen mannen en vrouwen.

Tabel B6.1. Factor analyse op de competenties uit het selectiegesprek
Component Eigenwaarde Verklaarde variantie

Component	Eigenwaarde	Verklaarde variantie
1	3.43	38.11
2	1.30	14.52
3	.92	10.26
4	.78	8.73
5	.65	7.28
6	.55	6.15
7	.54	6.03
8	.47	5.28
9	.32	3.59

Tabel B6.2. ladingen van de competenties op de twee hogere-order factoren

	Component	
	1	2
Zelfinzicht	.33	.55
Contactgerichtheid	.57	.29
Flexibiliteit	.55	
Emotionele weerbaarheid	.80	
Drukbestendigheid	.86	
Tolerantie		.79
Integriteit		.69
Communicatieve vaardigheden	.43	.42
Besluitvaardigheid	.74	

Extraction Methode: Principal Component Analysis.
 Rotatie Methode: Oblimin met Kaiser Normalization.
 Ladingen < .25 niet weergegeven.

De tabel met de verklaarde variantie (Tabel B6.1) en de zogenaamde screeplot (niet weergegeven in dit rapport) lieten zien dat alhoewel er volgens het Eigenwaarde > 1 criterium twee factoren geïdentificeerd kunnen worden, er wel een relatief sterke algemene factor aanwezig is die bijna veertig procent van de variantie kon verklaren. Op grond hiervan is het interessant om te bekijken wat de aard is van deze eerste ongeroteerde factor. De resultaten daarvan zijn hieronder weergegeven in tabel B6.3.

Tabel B6.3 laat zien dat de meeste competenties substantieel laden op de eerste algemene factor. Wat dit aangeeft is dat de psychologen tijdens de gesprekvoering een algemeen oordeel vormen over de kandidaat en dat dit algemene oordeel terugkomt in de scores op bijna alle competenties.

Tabel B6.3. Ladingen van de competenties op de Eerste ongeroteerde factor

	Eerste factor
Zelfinzicht	.65
Contactgerichtheid	.72
Flexibiliteit	.59
Emotionele weerbaarheid	.67
Drukbestendigheid	.71
Tolerantie	.39
Integriteit	.39
Communicatieve vaardigheden	.66
Besluitvaardigheid	.67

Extractie Methode: Principal Component analysis.

B6.1.2 de Praktijkopdracht

Dezelfde aanpak die is gehanteerd bij de competenties gemeten in het selectiegesprek, kan ook worden toegepast op de competenties die worden gemeten met de praktijkopdracht.

In tabel B6.4 staan ten eerste de relaties tussen de beoordelingen van de acteur en de externe assessor tijdens de praktijkopdracht. Een positief punt dat hieruit naar voren komt is dat de overeenstemming tussen de acteur en de assessor hoog is, variërend van .82 tot .90. Dit impliceert een hoge intersubjectiviteit die een ondersteuning biedt voor de validiteit van de gemeten competenties.

We zien echter ook in tabel B6.4 dat de correlaties tussen de verschillende competenties relatief hoog zijn. Blijkbaar maken de acteurs en assessoren dus relatief moeilijk onderscheid tussen de verschillende competenties in de praktijkopdracht. Dit wordt bevestigd door de factoranalyse. Wanneer we dezelfde statistische criteria voor de factoranalyse hanteren als bij het selectiegesprek is direct te zien dat er bij de praktijkopdracht slechts één factor te onderscheiden valt (Tabel B6.5). Deze factor verklaart maar liefst 67% van de variantie in de competenties en alle afzonderlijke competenties laden hoog op deze factor. Dit wil dus zeggen dat een algemene indruk die over de kandidaat wordt gevormd tijdens de praktijkopdracht sterk is en dat daarbij geen sterk onderscheid wordt gemaakt tussen de afzonderlijke competenties.

Tabel B6.4 Intercorrelaties competenties in praktijkopdracht en overeenstemming beoordelaars (N varieert tussen 16808 tot 16820)

	Denken	Voelen	Effectiviteit	Kracht
Denken	.86**	.61**	.72**	.52**
Voelen	.63**	.85**	.66**	.23**
Effectiviteit	.79**	.63**	.90**	.61**
Kracht	.53**	.25**	.63**	.81**

** p < .01

Boven de diagonaal zijn de beoordelingen van de acteur en onder de diagonaal van de assessor

Tabel B6.5. Factor analyse op competenties uit praktijkopdracht

Component	Initial Eigenvalues	
	Total	% of Variance
1	2.00	66.85
2	.73	24.40
3	.26	8.73

Tabel B6.6. Ladingen competenties praktijkopdracht op algemene factor

	Eerste factor
Denken	.92
Voelen	.79
Kracht	.72

Extractie Methode: Principal Component Analysis.

B6.2 Meten de competenties ook datgene wat ze beogen te meten (begripsvaliditeit)?

Om de begripsvaliditeit te meten kan men ten eerste kijken naar de onderlinge correlaties tussen de competenties uit het gesprek en de praktijkopdracht. Deze correlaties zijn in het rapport weergegeven in Tabel 9.2 in Hoofdstuk 9.

De algemene conclusie die men kan trekken uit Tabel 9.2 is beschreven in paragraaf 9.3 van Hoofdstuk 9 in het rapport. Hieronder wordt een meer gedetailleerde bespreking van de correlaties beschreven:

De correlaties vallen in absolute zin vaak in de range van gemiddelde effectgroottes (ergens tussen de .20 en .40). Als men in ogenschouw neemt dat wanneer het gaat om personeuseigenschappen de correlaties tussen verschillende beoordelaars überhaupt niet erg hoog zijn dan vallen deze correlaties in de verwachte range. Toch zijn er een aantal zaken die opvallen. Ten eerste vertoont het patroon van correlaties weinig differentiatie. In bijna alle gevallen correleert de competentie 'Denken' uit de praktijkopdracht het sterkst (of nagenoeg

het sterkst) met alle competenties uit het gesprek. Dit lijkt erop te wijzen dat de intelligentie van een kandidaat of de algemene (cognitieve) manier waarop hij/zij problemen aanpakt relatief sterk de score op alle competenties in het gesprek bepaalt. Verder valt op dat met uitzondering van tolerantie en integriteit, de correlaties tamelijk dicht bij elkaar in de buurt zitten (.23 tot .39). Dit komt overeen met eerdere bevindingen over de factor structuur van de competenties, waarin tolerantie en integriteit ook een aparte plaats innamen.

Correlationele analyses laten verder zien dat de algemene factor uit de praktijkopdracht $r = .45$ correleert met de algemene factor uit het selectiegesprek. Als we deze correlatie vergelijken met de correlaties die men normaal gesproken rapporteert in de persoonlijkheidsliteratuur dan zien we dat deze overeenkomt. Dat wil zeggen, de gevonden correlatiewaarde ($r = .45$) lijkt sterk op de correlatie die typerend is wanneer verschillende personen iemand beoordelen op persoonlijkheidstrekken. Dus we kunnen concluderen dat op de algemene evaluatie van een kandidaat, de overeenstemming tussen beoordelaars goed is. Dit impliceert dat dit een goede weergave is van hoe een kandidaat zich over het algemeen gedraagt tijdens de selectieprocedure (i.e., tijdens het gesprek en de praktijkopdracht).

B6.3 De verbanden tussen de competenties en persoonlijkheid en intelligentie

In tabel B6.7 staan de correlaties weergegeven tussen enerzijds de competenties zoals die gemeten worden in het selectiegesprek en de praktijkopdracht en anderzijds de metingen van persoonlijkheid en intelligentie. Voor persoonlijkheid is daarbij de Big Five meegenomen. Wat ten eerste opvalt aan de tabel is dat de onderlinge correlaties relatief laag zijn. De hoogste correlatie die wordt gevonden is .25 maar de meesten zijn aanzienlijk lager. Ondanks de lage correlaties zijn er wel enkele consistent patronen zichtbaar. Zo vertoont Extraversie de hoogste correlatie met contactgerichtheid. Neuroticisme toont zoals te verwachten omgekeerde correlaties met emotionele weerbaarheid en drukbestendigheid. De competentie tolerantie lijkt ten gedeeltelijk een mix te zijn van Openheid voor nieuwe ervaringen en vriendelijkheid. Deze patronen komen overeen met wat men op conceptuele gronden zou kunnen verwachten. Een ander patroon dat opvalt is dat de algemene intelligentiescore correlaties vertoont met bijna alle (op zelfinzicht na) competenties.

Tabel B6.7 Correlaties tussen competenties versus persoonlijkheid en intelligentie

	Openheid	Conscientieus	Extraversie	Vriendelijkheid	Neuroticisme	IQ
Zelfinzicht	.11**	.01	.02	.06**	.000	.14**
Contactgerichtheid	.12**	.08**	.25**	.07**	-.11**	.04
Flexibiliteit	.05 ⁺	-.01	.10**	.03	-.06**	.10**
Emotionele weerbaarheid	.00	.10**	.12**	.04 ⁺	-.22**	.06**
Drukbestendigheid	.02	.11**	.08**	.02	-.21**	.09**
Tolerantie	.14**	.06**	.01	.16 ⁺	-.05 ⁺	.03
Integriteit	.07**	.13**	.00	.14**	-.10**	.07**
Communicatieve vaardigheden	.11**	.01	.04 ⁺	.04 ⁺	-.05 ⁺	.15**
Praktijkopdracht : Denken	.08**	-.03	.01	.03	-.01	.19**
Praktijkopdracht: Voelen	.07**	-.01	.01	.03	.00	.10**
Praktijkopdracht: Kracht	.05 ⁺	.01	.02	-.00	-.03	.08**

BIJLAGE 7
BEGELEIDINGSCOMMISSIE

Prof. dr. R.R. Meijer (Voorzitter)	Rijksuniversiteit Groningen, Gedrags- en Maatschappijwetenschappen
Dr. W.H.M. Emons	Universiteit Tilburg, Tilburg School of Social and Behavioral Sciences
Drs. O.L Hendriks	Ministerie van Veiligheid en Justitie, Wetenschappelijk Onderzoek en Documentatiecentrum (WODC)
Drs. P. Jansen	Kwartiersmakersorganisatie Nationale Politie
Mr. M. J. Laan	Ministerie van Veiligheid en Justitie, Directoraat-generaal Politie
Drs. J. van Zielst	Politieacademie