



# Signaleren van bovengemiddelde intelligentie bij startende kleuters

Marijke J. van Eijsden en Julia C. Vrolijk

3632466 - 3638448

mei 2014

Academische Lerarenopleiding Primair Onderwijs

Begeleiding van Frans Prins

## Abstract

In dit onderzoek zijn vier signaleringsinstrumenten (Sidi-R, Digitaal Handelingsprotocol Hoogbegaafdheid, Vragenlijst Beginkenmerken, Surplus-model) voor bovengemiddelde intelligentie bij startende kleuters vergeleken op de kwaliteitseisen betrouwbaarheid, validiteit, aanvaardbaarheid en bruikbaarheid voor gebruik op de St. Dominicussschool. De betrouwbaarheid is gemeten door van verschillende proefafnames door ouders en/of leerkrachten het percentage overeenstemming en de Cohen's Kappa te berekenen. De validiteit, aanvaardbaarheid en bruikbaarheid zijn beoordeeld met behulp van een rubric. Daarnaast is de bruikbaarheid ook gemeten met behulp van vragenlijsten voor de deelnemers. Uit dit onderzoek bleek dat alle instrumenten, behalve het DHH, geschikt zijn. Bij elk instrument moet echter een aantal interventies uitgevoerd worden door de school, om het instrument optimaal te laten aansluiten bij de praktijk. Hierbij ligt de Vragenlijst Beginkenmerken het meest voor de hand, gezien de hoge validiteit en de hoge waardering door de deelnemers.

Keywords: Signaleren, intelligentie, Sidi-R, Digitaal Handelingsprotocol Hoogbegaafdheid, Vragenlijst Beginkenmerken, Surplus-model, kleuters

De laatste jaren wordt steeds meer aandacht besteed aan bovengemiddeld intelligente kinderen. Uit onderzoek is gebleken dat intelligentie niet alleen afhankelijk is van erfelijkheid, maar ook van de omgeving (Al-Shabatat, Abbas & Ismail, 2011; Tannenbaum, 2003, in Morrissey, 2011). Wanneer de omgeving op vroege leeftijd niet aansluit bij de capaciteiten van het kind, kan dit onderpresteren als gevolg hebben. Onderpresteren wordt gedefinieerd als het falen in het omzetten van uitzonderlijke capaciteiten in uitzonderlijke prestaties (Gagné, 2004, in Snyder & Linnenbrink-Garcia, 2013). Onderpresteren kan volgens Snyder en Linnenbrink-Garcia (2013) voorkomen wanneer een kind niet voldoende uitdaging krijgt in zijn directe omgeving, met als gevolg dat een kind niet meer gemotiveerd is om zich in te zetten op een later moment waarop hij wel in contact komt met een uitdagende omgeving. Om die reden wordt gesproken over het zo vroeg mogelijk signaleren van bovengemiddelde intelligentie, te beginnen bij de start van het onderwijs. Hiervoor zijn echter nauwelijks bestaande signaleringsmethoden en over het signaleren is nog weinig kennis beschikbaar (Keeman, 2013). In dit onderzoek zal de kennis over het signaleren van bovengemiddelde intelligentie bij kleuters worden uitgebreid. Dit is belangrijk, omdat scholen behoefte hebben aan manieren om onderpresteren te voorkomen. Onder andere op katholieke basisschool St. Dominicus te Utrecht wordt het gebrek aan signaleringsmethoden als een probleem ervaren. In de huidige situatie worden alleen kinderen waarvan hun groepsleerkracht verwacht dat ze bovengemiddeld intelligent zijn in groep 3 of hoger onderworpen aan uitgebreide signaleringsprotocollen voor hoogbegaafdheid. Dat zijn de Sidi-R (De Bruin - De Boer & Kuipers, 2004) en het Digitaal Handelingsprotocol Hoogbegaafdheid (Van Gerven & Drent, 2004). Deze situatie wordt niet als wenselijk ervaren, omdat hierdoor in de eerste schooljaren niet volledig tegemoet kan worden gekomen aan de capaciteiten van de leerlingen. Daardoor bestaat de kans dat een bovengemiddeld intelligente leerling gaat onderpresteren. In dit onderzoek wordt onderzocht in hoeverre beschikbare signaleringsinstrumenten geschikt zijn om bovengemiddelde intelligentie te meten bij de startende kleuters van katholieke basisschool St. Dominicus.

### **Intelligentietheorieën**

Om intelligentie te kunnen meten wordt vaak gebruik gemaakt van intelligentietesten, welke het intelligentiequotiënt meten (Wicks-Nelson & Israël, 2009). Wanneer een kind een IQ-score hoger dan 120 heeft, wordt het volgens Wechsler (1991) beschouwd als een kind met een bovengemiddelde intelligentie. Aangezien de intelligentie van het kind hoger ligt dan het gemiddelde van 100, kan het kind enkele 'leerstofmaanden' of 'leerstofjaren' hoger functioneren dan zijn leeftijdsgenoten (Mooij, 2011). In de literatuur wordt, wanneer het gaat over een hoge intelligentie, voornamelijk gesproken over hoogbegaafdheid. Hoogbegaafdheid wordt op verschillende manieren gedefinieerd, maar een algemeen aanvaarde definitie is dat sprake is van hoogbegaafdheid bij een IQ hoger dan 130 (Gagné, 1998). Pfeiffer (2012) zegt echter dat het in de praktijk onmogelijk is om zo een scherpe grens te trekken en dat kenmerken van hoogbegaafdheid ook al bij een IQ dat iets lager is dan 130 voor kunnen komen. Daarnaast benadrukt ook Borland (2009) dat de term hoogbegaafdheid slechts een sociaal construct is en dat de term niet gebruikt moet worden om personen te categoriseren. Hierdoor is het mogelijk om de theorie die bekend is over hoogbegaafdheid ook te betrekken op kinderen

waarvan een vermoeden bestaat dat ze bovengemiddeld intelligent zijn, maar waarvan hun IQ nog niet bekend is, mits hier met enige voorzichtigheid mee om wordt gegaan. In het vervolg zal dan ook geen onderscheid meer worden gemaakt tussen hoogbegaafdheid en bovengemiddelde intelligentie en zal enkel worden gesproken over de term bovengemiddelde intelligentie.

Om bovengemiddelde intelligentie te kunnen signaleren, zonder gebruik te maken van een intelligentietest, is kennis nodig van kenmerken die passen bij een bovengemiddeld intelligent kind. In de literatuur zijn in verschillende theorieën en modellen meerdere gebieden beschreven waarop een bovengemiddeld intelligent kind zou moeten uitblinken. Ten eerste bestaat volgens de triarchische theorie van Sternberg (2003) succesvolle intelligentie uit drie gebieden: Analytische intelligentie, oftewel informatieverwerkingsvaardigheden; Creatieve intelligentie, de capaciteit om nieuwe problemen op te lossen; en Praktische intelligentie, het toepassen van intellectuele vaardigheden in alledaagse situaties. Ten tweede gaat het triadisch interdependentiemodel van hoogbegaafdheid van Mönks (Wijnekus & Pluymakers, 2007) uit van de volgende gebieden: Motivatie, Creativiteit en Buitengewone capaciteiten. Tot slot noemt *The Munich Model of Giftedness and Talent* van Heller (2007) zeven gebieden waarop bovengemiddeld intelligente kinderen hoog kunnen presteren: wiskunde, natuurwetenschappen, techniek, computervaardigheden, kunstzinnige vaardigheden, taal, sport en sociale relaties. Over de gebieden waarop een bovengemiddeld intelligent kind zou moeten uitblinken, bestaat dus geen eenduidigheid. Daarnaast onderscheiden Bélanger en Gagné (2006) twee perspectieven waarop tegen bovengemiddelde intelligentie kan worden aangekeken. Het disjunctieve perspectief houdt in dat men bovengemiddeld intelligent is wanneer hij op gebied A, B óf C uitblinkt. Bij het conjunctieve perspectief blinkt men uit in zowel gebied A, B als C. Volgens Bélanger en Gagné zijn wetenschappers het niet eens over de vraag of al over een bovengemiddeld intelligent kind gesproken mag worden als hij op één gebied uitblinkt, of pas wanneer hij dit op meerdere gebieden doet.

De school bekijkt bovengemiddelde intelligentie vanuit het conjunctieve perspectief, omdat zij een kind als bovengemiddeld intelligent beschouwt wanneer het op meerdere gebieden uitblinkt. De gebieden die voor de school belangrijk zijn, zijn de gebieden waarop de school kan inspelen en welke zij optimaal kan en wil ontwikkelen. Dit is voornamelijk het intellectuele vermogen, maar ook de creativiteit om dat toe te passen in allerlei situaties. Deze gebieden komen overeen met de triarchische theorie van succesvolle intelligentie (Sternberg, 2003). Om deze gebieden concreet te kunnen meten, wordt gebruik gemaakt van kwalitatieve gedragsindicaties. De school hanteert de kwalitatieve gedragsindicaties van Drent en Van Gerven (2002). Deze zijn onderverdeeld in leereigenschappen en persoonlijkheidseigenschappen. Onder de leereigenschappen vallen: snel van begrip zijn; grote denk- en leerstappen maken; beschikken over een goed geheugen; een brede algemene interesse en kennis hebben; beschikken over een groot probleemoplossend vermogen; in staat zijn verworven kennis toe te passen; in staat zijn nieuwe kennis met oude kennis te integreren; en het beschikken over een groot analytisch vermogen. De persoonlijkheidseigenschappen die hierbij horen zijn: taalvaardig zijn en kunnen spelen met taal; komen met creatieve en originele oplossingen; geestelijk vroegrijp zijn; houden van uitdagingen; beschikken over een groot doorzettingsvermogen; op een gezonde manier perfectionistisch zijn; een intuïtieve denker zijn; behoefte hebben aan een

hoge mate van autonomie; beschikken over het vermogen tot (zelf)reflectie; en sociaal competent zijn. Om bovengemiddeld intelligente kinderen te kunnen signaleren is het daarom van belang om een instrument in handen te hebben dat deze gedragsindicaties zo optimaal mogelijk meet.

### **Signaleren van bovengemiddelde intelligentie**

Om een geschikt signaleringsinstrument te vinden voor het signaleren van bovengemiddelde intelligentie bij kleuters zal eerst gekeken worden welke instrumenten hiervoor bestaan. In Nederland zijn al veel instrumenten beschikbaar, maar niet alle instrumenten voldoen aan de criteria van de school. De school heeft aangegeven dat het instrument geschikt moet zijn voor de doelgroep, namelijk kinderen van vier jaar. Daarnaast moet het instrument het intellectueel vermogen meten, moet het instrument door leerkrachten en/of ouders gebruikt kunnen worden, mag het instrument niet te arbeidsintensief zijn en daarom slechts vijftien minuten per kind in beslag nemen en tot slot moet het verkrijgbaar zijn in Nederland.

**Psychometrische testen.** Een instrument dat regelmatig wordt ingezet om het intellectueel vermogen te meten is een intelligentietest. In Nederland zijn veelgebruikte intelligentietesten de WISC-III (Wechsler, 1991), de RAKIT-2 (Pieters, Dek & Kooij, 2013), de SON-R (Winkel & Tellegen, 2001) en de K-ABC-II (Lichtenberger & Kaufman, 2010). Deze testen moeten echter altijd door een specialist worden afgenomen waardoor ze niet aan alle criteria voldoen (Berk, 2009).

Een andere manier is gebruik maken van een ontwikkelingsschaal. Veel gebruikte ontwikkelingsschalen in Nederland zijn de Denver Ontwikkeling Screeningstest; de Bayley Ontwikkelingsschalen (BOS 2-30); de Groningse Ontwikkelingsschalen (GOS 2.5-4.5) en de McCarthy Ontwikkelingsschalen (MOS 2.5-8.5) (Winkel & Tellegen, 2001). Deze ontwikkelingsschalen zijn specifiek geschikt voor jonge kinderen. Daarentegen moeten ze net als intelligentietesten door een specialist worden afgenomen en voldoen zij niet aan alle criteria.

Naast meten met intelligentietesten en ontwikkelingsschalen, komt uit het onderzoek van Calero, Belen en Robles (2011) naar voren dat bovengemiddelde intelligentie goed te meten is door *dynamic assessment*. Dit houdt in dat gekeken wordt naar het leerpotentieel van het kind. Aangezien deze manier van testen zeer arbeidsintensief is, voldoet ook dit niet aan alle criteria.

**Protocollen.** Verder blijkt uit het onderzoek van Kuo, Maker, Su en Hu (2010) dat gestructureerde vragenlijsten die leerkrachten of ouders aan de hand van hun observaties invullen, nuttige informatiebronnen kunnen zijn. Naast losstaande gestructureerde vragenlijsten, bestaan ook protocollen. Bij deze protocollen kunnen meerdere stappen worden doorlopen. De eerste stap is vaak een korte gestructureerde vragenlijst, om in vaststaande vervolgstappen de kinderen waarbij een vermoeden van bovengemiddelde intelligentie bestaat verder te onderzoeken. Hierna worden de vier protocollen besproken die in Nederland gebruikt worden om begaafdheid te signaleren.

Het Digitaal Handelingsprotocol Hoogbegaafdheid is een digitaal protocol dat uit meerdere stappen bestaat (Van Gerven & Drent, 2004). Ten eerste wordt voor iedere leerling een quickscan uitgevoerd, waarbij met een korte vragenlijst gekeken wordt welke kinderen in een groep mogelijk bovengemiddeld intelligent zijn. Vervolgens kan voor deze kinderen een uitgebreidere signalering

ingevuld worden om een gedetailleerd beeld van deze kinderen te krijgen. Dit kan vanaf de kleutergroepen worden ingezet. De quickscan voor kleuters bestaat uit vijftien items, waarbij de leerkracht kan aangeven of deze wel of niet van toepassing zijn op het kind. Hierbij wordt automatisch een uitslag gegenereerd, welke adviseert of het gewenst is om verder onderzoek uit te voeren bij het kind.

Het SIDI-R protocol is een protocol dat vanaf de kleutergroep kan worden ingezet bij ieder kind (De Bruin - De Boer & Kuipers, 2004). In de kleutergroep wordt voor iedere leerling de groepssignalerings ingevuld, waarbij aan de hand van zeventien items wordt gekeken naar de onderdelen werkhouding-, omgang-, zelfbeeld- en leeraspecten. De kinderen waarbij een vermoeden is ontstaan van bovengemiddelde intelligentie worden diepgaander onderzocht in de diagnosefase aan de hand van een uitgebreide vragenlijst.

Het Surplus-model is een signaleringsinstrument dat uit vier onderdelen bestaat (Kort & Van Dijk, 2012). Hierbij is een schets de eerste signalering, waarbij met behulp van een vragenlijst met drieëndertig vragen voor ouders en leerkrachten een beeld wordt geschetst van de intelligentie van iedere leerling. Hierbij wordt gekeken naar de volgende onderdelen: ontwikkelingsvoorsprong, motivatie-werkhouding, sociaal-emotioneel functioneren, talentaccenten en schoolresultaten. Het model adviseert na het invullen van de vragen of een vervolgonderzoek gewenst is. Eventuele vervolgstappen zijn de standaardtekening, de uitgebreide tekening en de foto.

(Hoog)begaafdheid In-Zicht is een uitgebreid signaleringsinstrument waarbij aan de hand van uitgebreide vragenlijsten een profiel wordt gevormd waarin duidelijk wordt hoe groot de kans is dat een kind (hoog)begaafd is en wat voor begaafdheidsprofiel (Betts & Neihart, 1988; Betts & Neihart, 2010) bij dit kind past. Doordat (Hoog)begaafdheid In-Zicht modulair is opgezet, kunnen losse onderdelen worden ingezet zonder gebruik van de overige onderdelen. Alle vragenlijsten worden door de ouder(s) en leerkracht ingevuld, waarna de gegevens met elkaar vergeleken worden. Het is echter een digitale methode, waarbij het gehele instrument aangeschaft dient te worden.

De eerste drie protocollen zijn allemaal geschikt als signaleringsinstrument in de praktijksituatie, aangezien de eerste vragenlijst of quickscan als proefexemplaar beschikbaar is, of al in het bezit van de school. Verder voldoen ze ook aan de overige criteria. Bovendien zijn alle drie de vragenlijsten volgens Van Gerven (2009) gangbaar en veelgebruikt in Nederland. Dit zou kunnen betekenen dat ze ook geschikt zijn voor de St. Dominicusschool. De overige onderdelen van de protocollen zullen in dit onderzoek niet gebruikt worden, aangezien het hier alleen om de eerste signalering gaat van alle kinderen. Vervolgstappen vallen buiten het kader van dit onderzoek. (Hoog)begaafdheid In-Zicht is alleen digitaal in te vullen en hiervoor zou het gehele instrument aangeschaft moeten worden, terwijl slechts de quickscan van belang is. Daarom zal deze niet meegenomen worden in dit onderzoek.

**Vragenlijsten.** Een derde vorm van signaleren kan gerealiseerd worden aan de hand van vragenlijsten die los staan van vaststaande vervolgstappen. De vragenlijsten worden ingevuld door leerkrachten en/of ouders en met behulp van deze gegevens kan een vermoeden worden

uitgesproken over de intelligentie van het kind. Hieronder worden de twee vragenlijsten beschreven die beschikbaar zijn voor kinderen in de kleuterleeftijd.

De eerste vragenlijst is de Vragenlijst Beginkenmerken welke bestaat uit achtentwintig vragen die onderverdeeld zijn in zeven schalen: sociaal-communicatief niveau, algemeen cognitief niveau, taalniveau, (voorbereidend) rekenniveau, senso-motorisch niveau, emotioneel-expressief niveau, onderwijsgedrag/onderwijsmotivatie (Mooij, 2000). Deze vragenlijst wordt ingevuld door de ouders en door de leerkrachten nadat het kind ongeveer vier weken op school zit. Per kind worden de ontwikkelingsniveaus op de verschillende schalen ingeschat om te bepalen welke kinderen mogelijk bovengemiddeld intelligent zijn. Op dit moment loopt een longitudinaal interventieonderzoek op meerdere Nederlandse basisscholen naar verdere uitwerking en ontwikkeling van het instrument (Ministerie OCW, 2014).

Daarnaast is de Gifted Rating Scale for Preschool (GRS-P) een veelgebruikt instrument voor het signaleren van bovengemiddelde intelligentie. Dit is een vragenlijst voor leerkrachten waarbij vijf schalen zijn gevormd van in totaal zestig vragen (Pfeiffer & Petscher, 2008) die de waarschijnlijkheid aangeven of een kind bovengemiddeld intelligent is. De hierbij genoemde schalen zijn *intellectual ability*, *academic ability*, *creativity*, *artistic talent* en *motivation*. Aan de hand van een T score wordt de waarschijnlijkheid dat een kind (hoog)begaafd is aangegeven. Hoe hoger deze T, hoe hoger de waarschijnlijkheid. Dit instrument is echter nog niet in het Nederlands beschikbaar en zou daarom eerst vertaald moeten worden.

De Vragenlijst Beginkenmerken is geschikt voor gebruik in de praktijk, aangezien leerkrachten en ouders deze zelfstandig kunnen invullen en dit niet te veel tijd inneemt. De GRS-P voldoet aan alle criteria, behalve het criterium beschikbaarheid. In Nederland is geen proefexemplaar verkrijgbaar en daarom zal deze niet worden meegenomen in het onderzoek.

**Spelsituaties.** De laatste vorm van signaleren is aan de hand van instrumenten die tijdens het spel van kleuters ingezet kunnen worden. Door een kleuter te observeren in een spelsituatie, wordt een natuurlijke situatie gecreëerd. Dit zorgt dat de resultaten betrouwbaarder zijn (Gravetter & Forzano, 2009).

Een veelgebruikte methode is de menstekening. Hierbij wordt aan een kind gevraagd een tekening van een mens te maken, waarna dit resultaat geïnterpreteerd wordt (Bossink, 2004). Hierbij zijn verschillende stadia in de tekenontwikkeling te onderscheiden die overeenkomen met de cognitieve ontwikkeling. Hierdoor kan een uitspraak worden gedaan over de intellectuele ontwikkeling van een kind. Dit wordt echter afgeraden door Bossink (2004), aangezien deze test slechts een grove schatting van de intellectuele ontwikkeling biedt. Bovendien stellen Thomas & Silk (1993, in Bossink, 2004) dat de test-hertest betrouwbaarheid niet groot is.

Daarnaast is het signaleringsinstrument 'Pientere Peuters' van Grip op Talent een methode die gebruik maakt van een spelsituatie om de ontwikkeling van het kind te toetsen (Keeman, 2013). Dit instrument is bedoeld voor kinderopvangcentra en voorschoolse educatie. De pedagogisch medewerker of leerkracht speelt verschillende spellen met een kind, waarbij geobserveerd wordt wat

het kind al weet en kan. Aan de hand van deze gegevens kan een observatielijst worden ingevuld, waarmee een eventuele ontwikkelingsvoorsprong kan worden gesignaleerd.

Bovengenoemde instrumenten zijn beide ongeschikt. Doordat de menstekening geen geschikt resultaat oplevert volgens Bossink (2004) en Thomas en Silk (1993, in Bossink, 2004), zal dit instrument niet gebruikt worden. Ook Pientere Peuters zal in dit onderzoek geen rol spelen, aangezien het te arbeidsintensief is voor een leerkracht om iedere kleuter met behulp van dit instrument te observeren.

**Selectie.** Aan de hand van bovenstaande beschouwing is een selectie gemaakt van instrumenten. Aangezien ze niet allemaal voldoen aan de door de school opgestelde criteria, zullen alleen de volgende instrumenten onderzocht worden: de quickscan van het Digitaal Handelingsprotocol Hoogbegaafdheid, de groepssignalering van de Sidi-R, het Surplus-model en de Vragenlijst Beginkenmerken.

### **Kwaliteitseisen om signaleringsinstrumenten te beoordelen**

Om de kwaliteit te beoordelen van de geselecteerde instrumenten, is het van belang om ze op de aspecten betrouwbaarheid, validiteit, aanvaardbaarheid en bruikbaarheid te onderzoeken (Stokking & Van der Schaaf, 1999).

Het eerste aspect is betrouwbaarheid. Dit is de mate waarin scores consistent, nauwkeurig en reproduceerbaar zijn en dus vrij van meetfouten (Lubbe & Hollenberg, 2011; Stokking & Van der Schaaf, 1999). Dochy en Janssens (2009) noemen vier onderdelen waarop de betrouwbaarheid beoordeeld kan worden: het instrument, de afnamesituatie, de beoordeling en de beoordeelde. De betrouwbaarheid van het instrument hangt af van de objectiviteit en de omvang. De afnamesituatie kan de betrouwbaarheid beïnvloeden door de manier van afname en de omstandigheden waaronder de toets wordt afgenomen. Van der Vleuten et al. (2012) stellen daarnaast dat assessoren een belangrijke rol spelen in het beoordelingsproces. Hun expertise, de mate waarin zij de beoordeling serieus nemen en de tijd die zij hieraan willen besteden bepalen of het beoordelingsproces betrouwbaar verloopt. De betrouwbaarheid van de beoordeling hangt af van de eenduidigheid van de scoringsregels. Tot slot kunnen ook factoren bij het kind meespelen. Door meerdere participanten eenzelfde kind te laten beoordelen, kan gekeken worden of dit instrument bij meerdere afnames een eenduidig en daarmee betrouwbaar resultaat biedt.

Het tweede aspect is validiteit. Dit houdt in dat met de beoordeling moet worden gemeten wat men wil meten en niet (ook) iets anders (Stokking & Van der Schaaf, 1999). Volgens Crooks, Kane en Cohen (1996) is de validiteit van een toetsinstrument het belangrijkste kenmerk om de kwaliteit te meten. Zij hebben een acht-stappen-model ontwikkeld waarbij gestructureerd wordt gekeken naar verschillende aspecten om een zo compleet mogelijk beeld te vormen van de validiteit van een instrument (tabel 1). De stappen *decision* en *impact* zullen in het huidige onderzoek buiten beschouwing gelaten worden, aangezien deze stappen gericht zijn op behandeling van bovengemiddeld intelligente kinderen. Dit onderdeel zal pas in vervolgonderzoek aan bod komen.

Tabel 1

*Acht-stappen-model van Crooks et al. (1996)*

1. <i>Administration</i>	Hierbij gaat het om de condities waaronder het instrument wordt gebruikt.
2. <i>Scoring</i>	Hierbij wordt gekeken hoe de verschillende onderdelen van het instrument worden meegenomen in de beoordeling.
3. <i>Aggregation</i>	De verschillende onderdelen die een gezamenlijk resultaat tot stand brengen, moeten coherent zijn.
4. <i>Generalization</i>	De scores die behaald worden moeten representatief zijn voor de daadwerkelijke capaciteiten van het kind.
5. <i>Extrapolation</i>	Hierbij wordt gekeken of de verschillende vaardigheden die getest moeten worden ook daadwerkelijk met dit instrument getest worden.
6. <i>Evaluation</i>	Wordt op de juiste manier een oordeel verbonden aan de resultaten?
7. <i>Decision</i>	Wat wordt vervolgens met de resultaten gedaan, welke consequenties hebben ze?
8. <i>Impact</i>	Hierbij staat centraal of de vervolgaanpak naar aanleiding van de resultaten een positief effect heeft op het kind in vergelijking met de eerdere aanpak.

Het derde aspect is aanvaardbaarheid. De procedure en criteria moeten duidelijk en inzichtelijk zijn, kunnen worden uitgelegd aan anderen en objectief en dus verdedigbaar zijn (Stokking & Van der Schaaf, 1999). Ook Admiraal, Hoeksma, Van de Kamp en Van Duin (2011) vinden duidelijke communicatie over de procedure en de criteria van belang en spreken in dit verband over transparantie. Volgens Stokking (2010) bestaat aanvaardbaarheid uit drie criteria. Ten eerste de objectiviteit waarbij de metingen voldoende gestandaardiseerd zijn en de toets weinig interpretatie behoeft. Ten tweede noemt hij vraagzuiverheid, dit betreft de afwezigheid van bias. Het laatste criterium is normhandhaving, waarbij alle items door verschillende deelnemers op gelijke wijze gehanteerd moeten worden.

Het laatste kwaliteitsaspect is bruikbaarheid. Het instrument moet in de praktijk toepasbaar zijn en betekenis hebben (Stokking & Van der Schaaf, 1999). Stokking (2010) noemt vier aspecten die gerelateerd zijn aan de bruikbaarheid. Ten eerste de moeilijkheid, wat in dit verband geïnterpreteerd kan worden als de voorwaarde dat het instrument moet aansluiten bij de capaciteiten van de doelgroep. Ten tweede het onderscheidend vermogen, om verschillen tussen leerlingen aan te tonen. Ten derde moet het doel van het instrument aansluiten bij het gebruik in de praktijk, dus het signaleren van bovengemiddelde intelligentie van leerlingen. Tot slot mag de hoeveelheid tijd en inspanning niet te groot zijn voor de uitvoerende.

**Huidig onderzoek**

In dit onderzoek zullen de vier geselecteerde instrumenten beoordeeld worden op de vier genoemde kwaliteitseisen. De onderzoeksvraag is: In hoeverre zijn de quickscan van het Digitaal Handelingsprotocol Hoogbegaafdheid, de groepssignalering van de SIDI-R, de schets van het



Surplus-model en de Vragenlijst Beginkenmerken geschikt voor het signaleren van bovengemiddelde intelligentie bij nieuwe kleuters op de St. Dominicuschool, wanneer gekeken wordt naar de betrouwbaarheid, validiteit, aanvaardbaarheid en bruikbaarheid? De verwachting is dat de quickscan van het DHH, de groepssignalering van de Sidi-R en de schets van het Surplus-model geschikt zullen zijn, aangezien deze drie modellen gangbaar en veelgebruikt zijn in Nederland (Van Gerven, 2009). Daarentegen wordt de Vragenlijst Beginkenmerken nog onderzocht, dus op dit moment bestaat de kans dat de kwaliteit niet optimaal is.

## **Methode**

### **Design**

Het onderzoek bestond uit vier onderdelen, te weten het meten van de vier kwaliteitsaspecten per instrument. Het eerste onderdeel, betrouwbaarheid, is kwantitatief geanalyseerd door de interbeoordelaarsbetrouwbaarheid te meten met behulp van proefafnames. De onderdelen validiteit, aanvaardbaarheid en bruikbaarheid zijn kwalitatief geanalyseerd. Hiervoor is een rubric ontwikkeld. Daarnaast is de bruikbaarheid ook geanalyseerd met behulp van vragenlijsten die door de deelnemers zijn ingevuld.

### **Deelnemers**

Voor de kwaliteitsaspecten betrouwbaarheid en bruikbaarheid zijn twaalf deelnemers benaderd voor het invullen van de vragenlijsten. Deze steekproef bestond uit acht leerkrachten en vier ouders. De leerkrachten zijn allemaal vrouw en kleuterleerkrachten op de St. Dominicuschool. De leeftijd varieert tussen de 21 en 60 jaar, met een gemiddelde van 38 jaar. De ouders zijn de biologische ouders van twee leerlingen waarover de instrumenten zijn ingevuld. Deze leeftijd varieert tussen de 33 en 38 jaar, met een gemiddelde van 36 jaar. Alle deelnemers hebben minimaal een HBO opleiding afgerond.

### **Onderzoeksinstrumenten**

Om de kwaliteitseisen validiteit, aanvaardbaarheid en bruikbaarheid te analyseren is een rubric ontwikkeld, op basis van de theorieën van Crooks et al. (1996) en Stokking (2010). Het gebruik van een rubric is een betrouwbare methode volgens Jonsson en Svingby (2007), omdat in een rubric duidelijk wordt weergegeven wat en hoe beoordeeld moet worden en hierdoor consistente scores verkregen worden. In tabel 2 is een gedeelte van deze rubric getoond. De volledige rubric is weergegeven in bijlage 1. Voor de bruikbaarheid zijn daarnaast twee vragenlijsten ontwikkeld met een aantal vragen voor respectievelijk leerkrachten en ouders. Elke vraag kan worden beantwoord aan de hand van een Likert-schaal met een bereik van één tot tien. Dit brede bereik zorgt ervoor dat deelnemers de mogelijkheid hebben een klein nuanceverschil zichtbaar te maken. Bovendien heeft dit bereik geen midden, waardoor de deelnemers gedwongen worden een positieve of negatieve keuze te maken per vraag. De vragen gaan inhoudelijk in op de bruikbaarheid van het instrument. Voorbeelden van vragen zijn: 'Had u het idee dat u na drie weken voldoende informatie had over de leerling om de vragenlijst in te vullen?' en 'In hoeverre vond u alle vragen duidelijk en eenduidig gesteld?' (zie bijlage 2 en 3).

Tabel 2

*Rubric voor de validiteit, aanvaardbaarheid en bruikbaarheid van de instrumenten*

	Onvoldoende (0 pt.)	Voldoende (2 pt.)	Goed (4 pt.)
	<b>Validiteit</b>		
Extrapolation	Minder dan 75% van de kwalitatieve gedragsindicaties uit het theoretisch kader komen terug in de items van het instrument.	Tenminste 75% van de kwalitatieve gedragsindicaties uit het theoretisch kader komen terug in de items van het instrument.	Alle kwalitatieve gedragsindicaties uit het theoretisch kader komen terug in de items van het instrument.
...	...	...	...
...	...	...	...
	<b>Aanvaardbaarheid</b>		
Objectiviteit	De resultaten kunnen op meerdere manieren geïnterpreteerd worden.	Over de resultaten kan enige twijfel bestaan.	De resultaten kunnen slechts op één manier geïnterpreteerd worden.
...	...	...	...
...	...	...	...
	<b>Bruikbaarheid</b>		
Moeilijkheid	Minder dan 75% van alle items van het instrument sluiten aan bij de capaciteiten van een bovengemiddeld intelligent kind in vergelijking met leeftijdsgenoten.	Tenminste 75% van alle items van het instrument sluiten aan bij de capaciteiten van een bovengemiddeld intelligent kind in vergelijking met leeftijdsgenoten.	Alle items van het instrument sluiten aan bij de capaciteiten van een bovengemiddeld intelligent kind in vergelijking met leeftijdsgenoten.
...	...	...	...
...	...	...	...

**Procedure**

Om de kwaliteitseisen in kaart te brengen, is ten eerste een mail gestuurd aan alle kleuterleerkrachten waarin het doel van het onderzoek werd beschreven en de vraag of zij wilden deelnemen aan het onderzoek. De leerkrachten die hiermee instemden en die een nieuwe leerling kregen in de maanden februari of maart 2014 zijn vervolgens benaderd. Uit de beschikbare leerlingen is een random selectie gemaakt van vier leerlingen. Wanneer de ouders van de leerling instemden met dit onderzoek, kregen de leerkrachten zowel mondeling als schriftelijk uitleg over een van de instrumenten en de bijbehorende procedure. Hierbij werd duidelijk gemaakt dat de resultaten alleen voor het onderzoek zijn en verder geen consequenties hebben voor het kind. De leerkrachtprocedure bestond uit het doornemen van het signaleringsinstrument, het observeren van het betreffende kind gedurende drie weken en vervolgens het onafhankelijk invullen van het signaleringsinstrument in een rustige

omgeving. Elk van de vier kinderen werd afzonderlijk beoordeeld door de twee leerkrachten van wie het les krijgt. De signaleringsinstrumenten Vragenlijst Beginkenmerken en het Surplus-model moesten ook door ouders ingevuld worden. De ouderprocedure bestond voor zowel de vader als moeder uit het individueel doornemen en invullen van het instrument in een rustige omgeving binnen vijf dagen nadat hun kind op school was begonnen. Dit werd via een brief aan de ouders toegelicht, waarbij ook benadrukt werd dat de resultaten alleen voor het onderzoek gebruikt werden en verder geen consequenties hadden voor het kind. Alle participanten vulden na het voltooien van het instrument tevens een vragenlijst met betrekking tot de bruikbaarheid in.

### **Analyse**

De informatie die is verkregen aan de hand van de gegevens uit de proefafnames is geanalyseerd door per instrument de interbeoordelaarsbetrouwbaarheid te berekenen met behulp van het percentage overeenstemming en de Cohen's Kappa. Bij instrumenten met een ordinale schaal is een gewogen Kappa berekend. Hierbij was een minimaal percentage overeenstemming van 60% nodig voor een voldoende en werd voor de Cohen's Kappa een grenswaarde van  $K = .60$  aangehouden. Dit werd gedaan bij de leerkrachten onderling, de ouders onderling en de vier deelnemers per instrument gezamenlijk. Deze gezamenlijke score werd berekend door het gemiddelde te nemen van de scores van alle mogelijke combinaties van deelnemers per instrument. Met behulp van de rubric is aan elk instrument een validiteitsscore, een aanvaardbaarheidsscore en een bruikbaarheidsscore toegekend. Dit werd gedaan door aan de verschillende onderdelen (bijvoorbeeld *aggregation*) nul tot vier punten toe te kennen en deze op te tellen per kwaliteitseis. Voor de validiteit was de maximumscore vierentwintig punten en voor de kwaliteitseisen aanvaardbaarheid en bruikbaarheid was dit twaalf punten. De antwoorden van de vragenlijst om de bruikbaarheid te beoordelen zijn per instrument geanalyseerd. Per deelnemer is gekeken op welke punten hij of zij het instrument hoog of laag beoordeelde en wat aan deze beoordelingen opviel. Per instrument zijn de scores van de vier kwaliteitseisen vergeleken waardoor uiteindelijk een totaalbeeld werd verkregen van de beoordeling. Vervolgens is per kwaliteitseis een vergelijkende analyse gemaakt tussen de verschillende instrumenten.

### **Resultaten**

Om van elk signaleringsinstrument een beeld te krijgen, zijn de vier kwaliteitseisen vergeleken. Hieronder wordt per instrument uiteengezet hoe het scoort op de kwaliteitseisen betrouwbaarheid, validiteit, aanvaardbaarheid en bruikbaarheid. In tabel 3 is overzichtelijk weergegeven welke waarden de instrumenten op de verschillende kwaliteitseisen hebben behaald. Deze tabel wordt in de lopende tekst toegelicht, waarna een vergelijkende analyse tussen de instrumenten wordt gemaakt.

Tabel 3

Overzichtelijke weergave van de beoordeling van de signaleringsinstrumenten per kwaliteitseis

	Sidi-R	DHH	Vragenlijst Beginkenmerken	Surplus- model
<b>Betrouwbaarheid</b>				
<i>Percentage overeenstemming</i>				
Leerkrachten	82.4	86.7	32.1	45.5
Ouders	n.v.t.	n.v.t.	82.1	75.8
Ouders en leerkrachten	n.v.t.	n.v.t.	45.2	32.8
<i>Cohen's Kappa</i>				
Leerkrachten	<.00	.00	.09 <sup>1</sup>	.45 <sup>1</sup>
Ouders	n.v.t.	n.v.t.	.63 <sup>1</sup>	.53 <sup>1</sup>
Ouders en leerkrachten	n.v.t.	n.v.t.	.15	.18
<b>Validiteit</b>				
Totaalscore rubric	50%	33.3%	45.8%	41.7%
<b>Aanvaardbaarheid</b>				
Totaalscore rubric	66.7%	50%	50%	66.7%
<b>Bruikbaarheid</b>				
Totaalscore rubric	100%	66.7%	100%	50%
<i>Eindcijfer</i>				
Leerkracht 1	5.0	2.0	7.0	6.5
Leerkracht 2	6.0	3.0	6.0	5.0
Ouder 1	n.v.t.	n.v.t.	7.0	5.0
Ouder 2	n.v.t.	n.v.t.	6.0	6.0

1. Gewogen kappa in verband met ordinaal niveau.

### Sidi-R

Het eerste instrument dat onderzocht is, is de groepssignalering van de Sidi-R. Dit instrument is door twee leerkrachten ingevuld. Het percentage overeenstemming is 82,4% en de Cohen's Kappa had een waarde van  $K < .00$ . De gevonden Kappa komt onder de grenswaarde uit en is onvoldoende. Daarentegen is het percentage overeenstemming ruim voldoende.

De validiteit is gemeten aan de hand van een rubric. De Sidi-R behaalde hierop acht van de zestien punten. Opvallend hierbij was dat bij de Sidi-R geen sprake is van verschillende onderdelen, omdat de zelfbeeld-, omgang- en werkhoudingaspecten niet worden meegenomen in de beoordeling. Hierdoor hadden de rubric-onderdelen *scoring* en *aggregation* geen waarde. Op de onderdelen *generalization* en *extrapolation* scoorde het instrument geen punten, aangezien slechts vier van de achttien kwalitatieve gedragsindicaties overeenkomen met de items van het instrument. Daarnaast is het opvallend dat in de groepssignalering alleen naar een aantal leeraspecten wordt gekeken, terwijl in de handleiding een meer omvattende lijst van kenmerken genoemd staat voor kleuters met een

bovengemiddelde intelligentie. Hiervan komt wel een gedeelte terug in de vragen van het verkennende gesprek met ouders, maar een gedeelte van deze kenmerken wordt in beide signaleringslijsten niet genoemd.

Ook de aanvaardbaarheid is met behulp van een rubric vastgesteld. Hiervoor zijn acht punten behaald. Doordat met deze items kinderen met een andere taal of cultuur dan de leerkracht benadeeld kunnen worden, scoorde de Sidi-R hier de helft van de punten. Ook was een aantal items niet eenduidig geformuleerd.

Tot slot is de bruikbaarheid op twee manieren gemeten. De Sidi-R is met behulp van de rubric gescoord en haalde het maximaal aantal punten. Dit kwam onder andere doordat een duidelijk onderscheid te maken is tussen al dan niet bovengemiddeld intelligente kinderen. Daarnaast hebben de leerkrachten een vragenlijst ingevuld om de bruikbaarheid te beoordelen. De leerkrachten gaven respectievelijk een vijf en een zes als eindcijfer. De leerkrachten geven vrijwel gelijke beoordelingen voor de verschillende vragen. Daarbij valt op dat beide leerkrachten wel het gevoel hadden na drie weken over voldoende informatie te beschikken om het instrument in te vullen. Daarentegen hadden de leerkrachten niet het gevoel een voordeel te hebben aan het invullen van het instrument. De leerkrachten die de Sidi-R hebben ingevuld, gaven beiden aan dat de vragen van het instrument weinig diepgang hebben en dat de vragen weinig onderscheid kunnen maken tussen verschillende intelligentieniveaus van kinderen.

## **DHH**

Het tweede instrument dat is onderzocht is de Quickscan van het Digitaal Handelingsplan Hoogbegaafdheid (DHH). Dit instrument werd door twee leerkrachten ingevuld. Het percentage overeenstemming tussen deze twee leerkrachten is 86,7%. De Cohen's Kappa is onvoldoende, namelijk  $K = 0$ .

Daarnaast is de validiteit gemeten met behulp van de rubric. Het instrument behaalde acht van de vierentwintig punten. Hierbij valt op dat de onderdelen *scoring*, *generalization* en *extrapolation* geen punten opleverden, omdat het merendeel van de kwalitatieve gedragsindicaties uit het theoretisch kader niet of niet eerlijk verdeeld terugkomen in het instrument. Uit de items van het instrument zijn elf van de achttien kwalitatieve gedragsindicaties af te leiden. Wel heeft het instrument duidelijke en eenduidige richtlijnen over aan welke norm een kind moet voldoen om als bovengemiddeld intelligent te worden beschouwd.

Naast de validiteit is ook de aanvaardbaarheid gemeten met behulp van de rubric. Hierbij behaalde het DHH zes van de twaalf punten. Opvallend is dat het instrument op het onderdeel *normhandhaving* geen punten heeft behaald, aangezien meer dan de helft van de items niet eenduidig en duidelijk gedefinieerd is.

Tot slot heeft het desbetreffende instrument acht van de twaalf punten ontvangen voor de kwaliteitseis bruikbaarheid. Hierbij valt op dat het instrument goed scoorde op de onderdelen *onderscheidend vermogen* en *kosten*, maar nul punten ontving voor *moeilijkheid*. De reden hiervoor is dat binnen het instrument geen ruimte is voor vergelijking met de capaciteiten van leeftijdsgenoten. Daarnaast is door de leerkrachten een vragenlijst ingevuld over de bruikbaarheid. Hierbij valt op dat beide leerkrachten op bijna alle vragen cijfers tussen de één en de zes gaven. Alleen bij de vraag over

het niveau van de items gaf leerkracht 2 een hoger cijfer, namelijk een tien. Als opmerking schreef zij echter dat de vragen niet te beantwoorden zijn na drie weken, omdat de leerkracht niet genoeg informatie heeft over het kind.

### **Vragenlijst Beginkenmerken**

Ten derde is de Vragenlijst Beginkenmerken onderzocht. Dit instrument is zowel door twee leerkrachten als door twee ouders ingevuld. Om de betrouwbaarheid te meten zijn de resultaten van de twee leerkrachten vergeleken, de resultaten van de twee ouders en de gemiddelde resultaten van de leerkrachten en de ouders samen. Het percentage overeenstemming tussen ouders was opvallend hoog, namelijk 82,1%. De leerkrachten hadden slechts een percentage overeenstemming van 32,1%. Hoewel dit een lage overeenstemming is, valt op dat het merendeel van de antwoorden wel dicht bij elkaar ligt. Op de vijfpuntschaal was de afstand tussen de antwoorden klein, het scheelde in de meeste gevallen slechts één punt. Verder is de Cohen's Kappa tussen de ouders als enige voldoende met een score van  $K=.61$ .

Verder heeft het instrument op de kwaliteitseis validiteit elf van de vierentwintig punten behaald. Vijftien van de achttien kwalitatieve gedragsindicaties zijn terug te vinden in de items. Daarentegen komen veel andere items niet overeen met de gedragsindicaties. De onderdelen *scoring* en *generalization* leverden daarom beide nul punten op.

Daarnaast is de aanvaardbaarheid gemeten met behulp van de rubric. De Vragenlijst Beginkenmerken ontving zes van de twaalf punten, waarbij elk onderdeel twee punten kreeg. Dit kwam onder andere doordat sommige items niet eenduidig gedefinieerd zijn, waardoor niet altijd exact duidelijk is wat met de kenmerken bedoeld wordt.

Tot slot is ook de bruikbaarheid gemeten van de Vragenlijst Beginkenmerken. Op de onderdelen uit de rubric heeft dit instrument het maximaal aantal punten behaald. Daarnaast zijn ook de meningen van de leerkrachten en ouders meegenomen. Uit de vragenlijsten kwam naar voren dat alle deelnemers de vragenlijst een zes of een zeven als eindcijfer gaven. Hierbij viel op dat de ouders de vragenlijst positief beoordelen, behalve op het onderdeel handzaamheid. Ouder 1 vond het lastig om de vragenlijst in te vullen, omdat zij weinig kinderen als vergelijkingsmateriaal heeft. Ouder 2 vond sommige vragen multi-interpretabel. De leerkrachten gaven de verschillende vragen cijfers tussen de vijf en de acht. Leerkracht 2 gaf aan dat ze drie weken na binnenkomst van de kleuter te vroeg vindt om de vragenlijst in te vullen. Verder vond ze de vragenlijst een goede start voor verdere observatie.

### **Surplus-model**

Tot slot is gekeken naar de schets van het Surplus-model. Dit instrument is digitaal door de ouders en leerkrachten ingevuld. Opvallend aan de betrouwbaarheid is dat het percentage overeenstemming tussen ouders redelijk hoog is, namelijk 75,8%. De Cohen's Kappa voor de ouders is  $K= .53$  en daarmee bijna voldoende. De Cohen's Kappa voor de leerkrachten heeft een waarde van  $K = .45$ .

Op validiteit scoorde het Surplus-model tien van de vierentwintig punten. Op *aggregation* scoorde dit instrument maximaal, aangezien alle items coherent zijn aan het bijbehorende onderdeel. Daarnaast komen de items overeen met vijftien van de achttien kwalitatieve gedragsindicaties.

Daarentegen kregen de onderdelen *generalization* en *scoring* geen punten, aangezien de kwalitatieve gedragsindicaties niet eerlijk verdeeld waren over de items en een te groot aantal items niet overeenkwam met de kwalitatieve gedragsindicaties.

De aanvaardbaarheid heeft een hogere waarde behaald, hier scoorde het Surplus-model acht van de twaalf punten. Dit komt met name doordat de *objectiviteit* hoog is, aangezien de resultaten slechts op één manier geïnterpreteerd kunnen worden.

Met behulp van de rubric kreeg het Surplus-model slechts de helft van de punten voor bruikbaarheid. Dit komt grotendeels doordat de items van het instrument niet altijd aansluiten bij de capaciteiten van een bovengemiddeld intelligent kind, zoals het item 'mijn kind is actief'. Ook het feit dat dit instrument jaarlijkse kosten met zich meebrengt, zorgt voor een lagere score op bruikbaarheid. Tot slot is gekeken naar de vragenlijst om de bruikbaarheid te meten. De eindcijfers liepen uiteen van een vijf tot een zes en een half. Opvallend hierbij was dat drie van de vier deelnemers een drie of een vier gaven voor de mogelijkheid van het instrument om een verschil aan te tonen tussen kinderen die al dan niet bovengemiddeld intelligent zijn.

### **Vergelijkende analyse**

Wanneer de verschillende vragenlijsten worden vergeleken, valt een aantal dingen op. Ten eerste zijn hoge percentages overeenstemming te vinden bij de leerkrachten die de groepssignalering van de Sidi-R en de quickscan van het DHH hebben ingevuld en lage percentages bij de leerkrachten die de Vragenlijst Beginkenmerken en het Surplus-model hebben ingevuld. Ten tweede valt op dat bij beide vragenlijsten die door zowel ouders als leerkrachten ingevuld moeten worden, de Vragenlijst Beginkenmerken en het Surplus-model, de ouders veel meer overeenstemming hebben bereikt dan de leerkrachten. De waarde van Cohen's Kappa is bij geen enkele vragenlijst voldoende, behalve bij de ouders die de Vragenlijst Beginkenmerken hebben ingevuld.

Op de kwaliteitseis validiteit scoort geen enkele vragenlijst meer dan vijftig procent van de punten. Opvallend is dat geen van de instrumenten punten heeft behaald op de onderdelen *scoring* en *generalization*, aangezien bij geen enkel instrument de kwalitatieve gedragsindicaties eerlijk verdeeld waren over de items en/of onderdelen. Het DHH scoort het laagst van alle vragenlijsten met 33,3%. Het DHH scoort samen met de Vragenlijst Beginkenmerken ook het laagst op aanvaardbaarheid met vijftig procent van de punten.

Tot slot is uit de resultaten gebleken dat op de kwaliteitseis bruikbaarheid wel een instrument het hoogste scoort, namelijk de Vragenlijst Beginkenmerken. Deze vragenlijst heeft de volledige score op de rubric gehaald en de hoogste waardering van de deelnemers ontvangen. De Sidi-R heeft ook de volledige score op de rubric gehaald, maar heeft daarentegen een duidelijk lagere waardering ontvangen van de deelnemers. De andere twee instrumenten scoorden minder goed. Het Surplus-model heeft op deze kwaliteitseis de laagste score op de rubric behaald. Daarnaast heeft het DHH de laagste waardering ontvangen van de deelnemers. Opvallend is dat de deelnemers de instrumenten nooit beoordeelden met een cijfer hoger dan een zeven, waarbij het gemiddelde een 5,4 was.

## Discussie

In dit onderzoek zijn verschillende signaleringsinstrumenten voor bovengemiddelde intelligentie bij startende kleuters vergeleken. De onderzoeksvraag die hierbij centraal stond was: In hoeverre zijn de quickscan van het Digitaal Handelingsprotocol Hoogbegaafdheid, de groepssignalering van de SIDI-R, de schets van het Surplus-model en de Vragenlijst Beginkenmerken geschikt voor het signaleren van bovengemiddelde intelligentie bij nieuwe kleuters op de St. Dominicussschool, wanneer gekeken wordt naar de betrouwbaarheid, validiteit, aanvaardbaarheid en bruikbaarheid? De verwachting was dat de quickscan van het DHH, de groepssignalering van de Sidi-R en de schets van het Surplus-model geschikt zouden zijn, aangezien deze drie modellen gangbaar zijn in Nederland (Van Gerven, 2009). Uit het onderzoek is gebleken dat de resultaten niet overeenkomen met de verwachting. De quickscan van het DHH scoorde over het algemeen het laagst op de vier kwaliteitseisen en is daarmee het minst geschikt. De andere instrumenten scoorden binnen dit onderzoek over het algemeen beter dan het DHH. Voordat een advies kan worden uitgebracht aan de school, zal eerst een aantal resultaten verklaard worden en zal worden ingegaan op eventuele beperkingen van het onderzoek.

### Verklaringen en beperkingen

De resultaten die behaald zijn op de kwaliteitseisen aanvaardbaarheid en bruikbaarheid geven een goed beeld van de instrumenten, aangezien geen factoren zijn gevonden die de resultaten kunnen beïnvloeden en die kunnen zorgen voor een foutieve interpretatie. De resultaten op de kwaliteitseisen betrouwbaarheid en validiteit zouden een vertekend beeld kunnen geven. Hiervoor is een aantal verklaringen te vinden.

Ten eerste valt bij de betrouwbaarheid op dat het percentage overeenstemming bij de Sidi-R en het DHH goed is, terwijl het percentage overeenstemming bij de Vragenlijst Beginkenmerken en het Surplus-model een stuk lager uitvalt. Dit is te verklaren doordat bij de eerste twee instrumenten steeds maar uit twee antwoordmogelijkheden te kiezen valt, terwijl de laatste twee instrumenten een ordinale schaal hebben met respectievelijk vijf en vier mogelijkheden. Hierdoor is bij de Vragenlijst Beginkenmerken en het Surplus-model de kans kleiner dat twee respondenten precies dezelfde antwoorden geven. Wat verder opvallend is aan het percentage overeenstemming, is dat deze veel hoger is tussen de ouders dan tussen de leerkrachten. Dit is mogelijk veroorzaakt doordat de ouders het kind al vier jaar kennen en de leerkracht het kind slechts drie weken kent. De meeste leerkrachten gaven zelf ook aan dat ze een periode van drie weken te kort vonden om een goed beeld van de leerling te vormen om het instrument in te vullen.

Om een vollediger beeld te verkrijgen van de betrouwbaarheid, is ook de Cohen's Kappa uitgerekend. Helaas is gebleken dat de Cohen's Kappa bij geen enkele vragenlijst voldoende is, behalve tussen de ouders die de Vragenlijst Beginkenmerken hebben ingevuld. Hierin spelen twee factoren een rol. Ten eerste het aantal waarnemingen. Bij een laag aantal deelnemers dat het instrument invult bestaat namelijk de mogelijkheid dat veel overeenkomsten tussen antwoorden op kans berusten, waardoor geen realistisch beeld van de betrouwbaarheid wordt weergegeven. Daarnaast speelt ook het aantal antwoordmogelijkheden dat elk instrument heeft een rol. Doordat bij de eerste twee vragenlijsten slechts twee antwoordmogelijkheden zijn, is de kans groot dat verschillende deelnemers gelijke antwoorden hebben. Wanneer de antwoorden niet gelijk zijn, heeft



dit een grote invloed op de Cohen's Kappa. Bij de instrumenten met meer antwoordmogelijkheden is de kans op overeenstemming kleiner en heeft een afwijkende waarde minder invloed op de Cohen's Kappa. Doordat deze instrumenten een ordinale schaal gebruiken, is hier gekozen voor een gewogen Kappa. Hierbij wordt rekening gehouden met kleine en grote verschillen tussen de antwoorden van deelnemers. Kleine verschillen wegen hierdoor minder zwaar dan grote verschillen, waardoor een beter beeld wordt gegeven van de betrouwbaarheid. Bij het interpreteren van de betrouwbaarheid is het dus belangrijk om rekening te houden met het instrument. De Sidi-R en het DHH hebben slechts twee antwoordmogelijkheden, waardoor het percentage overeenstemming een realistischer beeld geeft van de betrouwbaarheid. Daarentegen hebben de Vragenlijst Beginkenmerken en het Surplus-model meer antwoordmogelijkheden, waardoor bij deze instrumenten de Cohen's Kappa beter bekeken kan worden.

Daarnaast heeft dit onderzoek nog twee beperkingen die de betrouwbaarheid kunnen beïnvloeden. Ten eerste is een beperking van dit onderzoek dat elk instrument slechts bij één kind is afgenomen. Dit zou echter geen vertekende resultaten mogen opleveren, aangezien de kinderen random zijn geselecteerd uit de beschikbare groep startende kleuters. Ten tweede is een mogelijke factor die invloed kan hebben op de betrouwbaarheid de afnamesituatie. Het is mogelijk dat deze afnamesituatie niet bij iedereen gelijk was. Met behulp van een brief is echter toegelicht onder welke condities de deelnemers het instrument zouden moeten invullen, waardoor verwacht wordt dat deze factor geen invloed heeft gehad op de verschillen tussen de resultaten.

Aangezien de meeste scores van de betrouwbaarheid onvoldoende zijn, is het belangrijk om deze in de toekomst te verhogen. De leerkrachten gaven aan dat het lastig is om na drie weken een volledig beeld te kunnen vormen van de leerling om het instrument goed in te kunnen vullen. Dit zal in de toekomst echter minder invloed hebben op de betrouwbaarheid, aangezien de leerkrachten het instrument steeds makkelijker kunnen invullen wanneer ze het veelvuldig gebruiken. De leerkrachten raken door de ervaring namelijk meer bekend met de items en weten beter waar zij bij de startende kleuters op moeten letten. Door gericht te kijken zal het minder moeite kosten om in korte tijd een goed beeld te vormen van de capaciteiten van de kleuter. Verder gaven de ouders aan dat het instrument lastig in te vullen was. Het is belangrijk om hier rekening mee te houden en bijvoorbeeld ouders hierin te begeleiden. Van der Vleuten et al. (2012) geven namelijk aan dat het belangrijk is dat elke assessor expertise heeft in het gebruik van een instrument om de beoordeling betrouwbaar te laten verlopen.

Bij de kwaliteitseis validiteit is ook een aantal dingen opgevallen. Geen enkele vragenlijst behaalde meer dan 50% van het maximaal aantal punten voor validiteit. De validiteit wordt aan de ene kant vergroot doordat de kwalitatieve gedragsindicaties aansluiten bij de definitie van bovengemiddelde intelligentie die de school hanteert. Daarentegen kan dit tegelijkertijd ook een beperking zijn, omdat deze definitie zeer specifiek is. In de rubric worden hoge eisen gesteld aan het aantal kwalitatieve gedragsindicaties dat terug moet komen in een instrument. Met zo'n specifieke definitie is het in de praktijk lastig om een bestaand instrument te vinden dat hier precies bij aansluit en welke op alle onderdelen het maximaal aantal punten kan behalen. In dit onderzoek is dat in de rubric terug te zien bij de onderdelen *scoring* en *generalization*, waarop geen enkel instrument punten

heeft behaald. Deze validiteitsaspecten gaan in op de verdeling van de kwalitatieve gedragsindicaties over de verschillende items en onderdelen van een instrument. Dit betekent dat de rubric te strenge eisen stelt aan het aantal kwalitatieve gedragsindicaties dat een instrument moet bevatten. Doordat het in de praktijk onmogelijk is gebleken om op *scoring* en *generalization* punten te behalen, kunnen de instrumenten in plaats van vierentwintig punten slechts zestien punten scoren op de kwaliteitseis validiteit. Dit houdt in dat niet 100% de hoogst haalbare score is, maar 66,7%. Wanneer hier rekening mee wordt gehouden scoren de Sidi-R, Vragenlijst Beginkenmerken en het Surplus-model allemaal voldoende. Dat het DHH ook op deze manier onvoldoende scoort is opvallend, aangezien de kwalitatieve gedragsindicaties die de school hanteert, opgesteld zijn door Drent en Van Gerven (2002). Zij zijn tevens de ontwikkelaars van het DHH. Dit wijst erop dat de kwalitatieve gedragsindicaties die Drent en Van Gerven gebruiken bij het DHH niet overeenkomen met hun kwalitatieve gedragsindicaties die door de school gebruikt worden voor de definitie van bovengemiddelde intelligentie.

Een andere beperking is dat de intelligentie van het kind niet bekend is en daarom kan in dit onderzoek niet worden gecontroleerd of de vragenlijst een juiste uitslag geeft. Daarentegen is deze beperking geprobeerd te ondervangen door de validiteit te meten met behulp van een inhoudsanalyse, waardoor toch een uitspraak gedaan kan worden over de mate waarin een instrument bovengemiddelde intelligentie kan signaleren.

Tot slot is een zwak punt dat voor het gebruik van de rubric specifieke *content knowledge* nodig is op het gebied van kwaliteitseisen, bovengemiddelde intelligentie, de ontwikkeling van kleuters en cultuurverschillen, omdat in de rubric veel vaktermen zijn opgenomen. Van der Vleuten et al. (2012) stellen dat expertise op de gebieden waarop beoordeeld wordt een voorwaarde is om het proces goed te laten verlopen. Hierdoor kan dit onderzoek niet zomaar worden herhaald door andere onderzoekers. Daarentegen is de rubric wel een betrouwbare manier om de kwaliteit van de instrumenten te bepalen, aangezien dit zorgt voor een objectieve meting (Jonsson & Svingby, 2007).

### **Advies**

Op basis van de resultaten en de interpretatie hiervan zou de school een keuze kunnen maken uit de vier instrumenten. De quickscan van het DHH scoort op meer dan de helft van de beoordelingsaspecten onvoldoende, waardoor het minder geschikt is dan de Sidi-R, de Vragenlijst Beginkenmerken en het Surplus-model.

De groepssignalering van de Sidi-R zou op basis van deze data potentie hebben om in de praktijk te worden gebruikt. De leerkrachten gaven echter een negatief oordeel over dit instrument. Deze meningen zijn belangrijk, aangezien de weerstand tegen een instrument niet te groot mag zijn om dit in de toekomst in de praktijk in te voeren. Wanneer voor dit instrument zou worden gekozen, zou de school erover na kunnen denken om hierbij aanvullend het verkennend gesprek met ouders te voeren, aangezien op deze manier naar meer kenmerken van bovengemiddelde intelligentie wordt gekeken en leerkrachten naar verwachting meer tevreden zullen zijn over het instrument.

De Vragenlijst Beginkenmerken kreeg in tegenstelling tot de Sidi-R wel een positief oordeel van leerkrachten en komt daarnaast positief naar voren uit de resultaten. Op de kwaliteitseis

aanvaardbaarheid scoorde de Vragenlijst Beginkenmerken onvoldoende, aangezien niet alle items eenduidig waren gedefinieerd. Hieraan kan de school tegemoetkomen door met elkaar in gesprek te gaan over de betekenis van deze items. Op deze manier kunnen de deelnemers het instrument op dezelfde manier kunnen scoren.

Het Surplus-model scoort over het algemeen voldoende. De bruikbaarheid is echter onvoldoende aangezien het Surplus-model jaarlijkse kosten met zich meebrengt. Daarnaast zijn de waarden van de Cohen's Kappa voor het Surplus-model niet voldoende. Zoals al eerder benoemd is, moet bij alle drie de instrumenten rekening worden gehouden met de opmerking van de leerkrachten dat drie weken een korte periode is. Wanneer leerkrachten inhoudelijk beter op de hoogte zijn van het instrument, zullen zij makkelijker een volledig beeld kunnen vormen van de capaciteiten van de kleuter. Bij de Vragenlijst Beginkenmerken en het Surplus-model is het verder van belang dat ouders begeleid worden bij het invullen van de vragenlijst. Deze interventies zullen naar verwachting de betrouwbaarheid verhogen.

Wanneer alle voor- en nadelen tegen elkaar afgewogen worden, kan geconcludeerd worden dat het DHH niet geschikt is voor het signaleren van bovengemiddelde intelligentie bij startende kleuters op de St. Dominicusschool. Daarnaast zouden de Sidi-R, de Vragenlijst Beginkenmerken en het Surplus-model wel in aanmerking komen, waarbij de Sidi-R en de Vragenlijst Beginkenmerken het meest geschikt zijn, aangezien deze de hoogste waarde voor de validiteit hebben. Volgens Crooks et al. (1996) is de validiteit van een instrument het belangrijkste kenmerk om de kwaliteit te meten. Daarbij krijgt de Vragenlijst Beginkenmerken de hoogste beoordeling van de deelnemers, waardoor dit instrument het meest voor de hand ligt om op de St. Dominicusschool te gebruiken.

### **Suggesties voor vervolgonderzoek**

Dit onderzoek is een eerste stap geweest naar het uitbreiden van kennis over signaleringsmethoden van bovengemiddelde intelligentie bij startende kleuters. In vervolgonderzoek zou dit onderzoek op grotere schaal met meer deelnemers uitgevoerd kunnen worden zodat de betrouwbaarheid verhoogd wordt. Verder zal de rubric op de onderdelen *scoring* en *generalization* minder strenge eisen moeten bevatten, waardoor het voor bestaande instrumenten mogelijk wordt om hier punten op te behalen. Bovendien is het huidige onderzoek moeilijk te generaliseren naar andere scholen, aangezien in dit onderzoek rekening is gehouden met de specifieke behoeftes van de St. Dominicusschool. De onderzoeksopzet is daarentegen wel te gebruiken door andere scholen wanneer de kwalitatieve gedragsindicaties worden aangepast, waardoor deze aansluiten bij de definitie van bovengemiddelde intelligentie die de betreffende school hanteert. Op deze manier kan in de toekomst op verschillende scholen een gefundeerde keuze worden gemaakt voor een signaleringsinstrument dat past bij de behoeftes van de school. Uit de literatuur is gebleken dat scholen hiervoor de laatste jaren veel aandacht hebben (Keeman, 2013). Daarnaast wordt het belang van vroegtijdig signaleren van bovengemiddelde intelligentie steeds meer erkend om onderpresteren te voorkomen (Gagné, 2004, in Snyder & Linnenbrink-Garcia, 2013). Verder blijft het volgens Crooks et al. (1996) niet alleen bij signaleren, maar zijn ook de stappen *decision* en *impact* heel belangrijk. Hierbij gaat het ten eerste om wat de school met de resultaten doet. Hiervoor zou de school onderzoek moeten doen naar de

mogelijkheden om kinderen die als bovengemiddeld intelligent zijn gesignaleerd, in de schoolpraktijk te begeleiden. Ten tweede is het belangrijk om in vervolgonderzoek te evalueren of het gebruik van het gekozen signaleringsinstrument uiteindelijk een positief effect heeft op de ontwikkeling van bovengemiddeld intelligente leerlingen. Met behulp van de resultaten van het huidige onderzoek kunnen de volgende stappen worden gezet naar het uitbreiden van kennis over signaleringsmethoden van bovengemiddelde intelligentie bij startende kleuters.

### Literatuur

- Admiraal, W., Hoeksma, M., Kamp, M. van de, & Duin, G. van (2011). Assessment of teacher competence using video portfolios: Reliability, construct validity, and consequential validity. *Teaching and Teacher Education, 27*, 1019-1028. doi: 10.1016/j.tate.2011.04.002
- Al-Shabatat, A., Abbas, M., & Ismail, H. (2011). The direct and indirect effects of environmental factors on nurturing intellectual giftedness. *International Journal of Special Education, 26(2)*, 18-28. doi: xxx
- Bélanger, J., & Gagné, F. (2006). Estimating the size of the gifted/talented population from multiple identification criteria. *Journal for the Education of the Gifted, 30(2)*, 131-163. doi: 10.4219/jeg-2006-258
- Berk, L. (2009). *Child development* (8<sup>th</sup> ed.). Boston, M.A.: Allyn & Bacon.
- Betts, T.G. & Neihart, M. (1988). Profiles of the gifted and talented. *Gifted Child Quarterly, 32(2)*, 248-253. doi: 10.1177/001698628803200202
- Betts, T.G. & Neihart, M. (2010). Revised Profiles of the Gifted & Talented. Verkregen van: <http://www.ingeniosus.net/wp-content/uploads/2010/11/PROFILES-BEST-REVISED-MATRIX-2010.pdf>
- Borland, J. H. (2009). Myth 2: The gifted constitute 3% to 5% of the population. Moreover, giftedness equals high IQ, which is a stable measure of aptitude-spinal tap psychometrics in gifted education. *Gifted Child Quarterly, 53(4)*, 236-238. doi 10.1177/0016986209346825
- Bossink, L. (2004). *Wat is de diagnostische waarde van de menstekening toegepast bij psychologisch onderzoek met kinderen?* Verkregen van <http://dare.uva.nl/document/11262>
- Bruin - de Boer, A. de, & Kuipers, J. (2004). *Sidi-R*. Leeuwarden: Eduforce.
- Calero, M. D., Belen, G. M., & Robles, A. (2011). Learning potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children. *Learning and Individual Differences, 21*, 176-181. doi: 10.1016/j.lindif.2010.11.025
- Crooks, T.J., Kane, M.T., & Cohen, A.S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice, 3(3)*, 265-285. doi:xxx
- Dochy, F., & Janssens, S. (2009). *Evaluatie en assessment*. In *Onderwijskunde van Verloop, N., & Lowyck, J. (red.)*. pp: 402-403. Groningen/Houten: Noordhoff Uitgevers.
- Drent, S., & Gerven, E. van (2002). *Professioneel omgaan met hoogbegaafde leerlingen in het basisonderwijs*. Utrecht: Lemma.
- Gagné, F. (1998). A proposal for subcategories within gifted or talented populations. *Gifted Child Quarterly, 42(2)*, 87-95. doi: 10.1177/001698629804200203.
- Gerven, E. van (2009). *Handboek hoogbegaafdheid*. Assen: Van Gorcum.

- Gerven, E. van, & Drent, S. (2004). *Digitaal Handelingsprotocol Hoogbegaafdheid*. Utrecht: Lemma.
- Gravetter, F. J. & Forzano, L. B. (2009). *Research methods for the behavioral sciences* (3rd edition). Belmont: Wadsworth.
- Heller, K.A. (2007). Scientific ability and creativity. *High Ability Studies*, 18, 209-234. doi: 10.1080/13598130701709541
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. doi: 10.1016/j.edurev.2007.05.002
- Keeman, P. (2013). Pientere Peuters: hoe vind je ze? Praktische signaleringsmethode voor slimme peuters. *Talent*, 15(6), 8-10. doi: xxx
- Kort, H., & Dijk, G. van (2012). *Dol-fijn excellent waar praktijk en theorie samengaan: Gids voor duurzame begeleiding van excellente leerlingen in het basisonderwijs uitgaande van het Surplus-model*. Gouda: MHR.
- Kuo, C., Maker, J., Su, F., & Hu, C. (2010). Identifying young gifted children and cultivating problem solving abilities and multiple intelligences. *Learning and Individual Differences*, 20, 365-379. doi: 10.1016/j.lindif.2010.05.005
- Lichtenberger, E. O., & Kaufman, A. S. (2010). Kaufman Assessment Battery for Children—Second Edition (KABC-II). In *Encyclopedia of Cross-Cultural School Psychology* (pp. 557-560). Springer US.
- Mooij, T. (2011). Verschil tussen hoogbegaafd en excellent. Naar optimaal onderwijs. Verkregen van [dspace.learningnetworks.org](http://dspace.learningnetworks.org)
- Lubbe, M. van der, & Hollenberg, J. (2011). *De kwaliteit van toetsen in het primair onderwijs*. In Hollenberg, J., Van der Lubbe, M., & Sanders, P. (Red.). *Toetsen op School, Primair Onderwijs*. pp: 88. Arnhem: Cito.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2014). *Onderwijs Bewijs*. Verkregen van [www.onderwijsbewijs.nl](http://www.onderwijsbewijs.nl)
- Mooij, T. (2000). Screening Children's Entry Characteristics in Kindergarten, *Early Child Development and Care*, 165(1), 23-40, doi: 10.1080/0300443001650103
- Mooij, T. (2011). Verschil tussen hoogbegaafd en excellent. Naar optimaal onderwijs. Verkregen van [dspace.learningnetworks.org](http://dspace.learningnetworks.org)
- Morrissey, A. (2011). Maternal scaffolding of analogy and metacognition in the early pretence of gifted children. *Exceptional Children*, 77(3), 351-366. doi: 10.1177/001440291107700306
- Pieters, S., Dek, J.E., & Kooij, A.P. (2013). *Rakit-2 psychometrische eigenschappen deel 2* (White paper). Verkregen van: <http://www.pearsonclinical.nl/rakit-2-revisie-amsterdamse-kinder-intelligentietest>
- Pfeiffer, S.I. (2012). Current perspectives on the identification and assessment of gifted students. *Journal of Psychoeducational Assessment*, 30(1), 3-9. doi: 10.1177/0734282911428192
- Pfeiffer, S. I., & Petscher, Y. (2008). Identifying young gifted children using the gifted rating scales—Preschool/kindergarten form. *Gifted Child Quarterly*, 52(1), 19-29. doi: 10.1177/0016986207311055

- Snyder, K.E., & Linnenbrink-Garcia, L. (2013). A developmental, person-centered approach to exploring multiple motivational pathways in gifted underachievement. *Educational Psychologist, 48*(4), 209-228. doi:10.1080/00461520.2013.835597
- Sternberg, R.J. (2003). A broad view of intelligence: The theory of successful intelligence. *Consulting Psychology Journal: Practice and Research, 55*, 139-154. doi: 10.1037/1061-4087.55.3.139
- Stokking, K.M. (2010). *Reader ALPO-cursus Toetsen en Beoordelen (1&2)*. Cursusjaar 2010-2011. Utrecht: Universiteit Utrecht.
- Stokking, K. M., & Schaaf, M. F. van der (1999). *Beoordelen van onderzoeksvaardigheden van leerlingen*. Tilburg: MesoConsult.
- Vleuten, C. P. M. van der, Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Tartwijk, J. van (2012). A model for programmatic assessment fit for purpose. *Medical teacher, 34*(3), 205-214. doi: 10.3109/0142159X.2012.652239
- Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- Wicks-Nelson, R., & Israël, A. C. (2009). *Abnormal child and adolescent psychology* (7th ed.). New Jersey: Pearson Education.
- Wijnekus, M., & Pluymakers, M. (2007). Begaafde leerlingen. In Verschueren & Koomen (Red.), *Handboek diagnostiek in de leerlingbegeleiding* (pp. 290). Antwerpen: Garant.
- Winkel, M., & Tellegen, P.J. (2001). Intelligentietests voor jonge kinderen: de SON-R 2.5-7 en andere intelligentietests. *Kind en Adolescent, 22*(3), 141-151. doi: xxx

**Bijlage 1. Rubric voor de kwaliteitseisen, validiteit, aanvaardbaarheid en bruikbaarheid**

		Validiteit	
	Onvoldoende (0 pt.)	Voldoende (2 pt.)	Goed (4 pt.)
<i>Administration</i>	De condities waaronder het instrument moeten worden afgenomen en het doel waartoe het instrument dient, zijn niet in de handleiding weergegeven.	De condities (1pt.) waaronder het instrument moeten worden afgenomen en het doel (1pt.) waartoe het instrument dient, zijn in de handleiding weergegeven.	De condities (2pt.) waaronder het instrument moeten worden afgenomen en het doel (2pt.) waartoe het instrument dient, zijn overzichtelijk en concreet in de handleiding weergegeven.
<i>Scoring</i>	Wanneer sprake is van verschillende onderdelen zijn de verschillende scores die worden toegekend niet gelijkmatig verdeeld over de kwalitatieve gedragsindicaties, waarbij de afwijking van het gemiddelde meer dan 25% is.	Wanneer sprake is van verschillende onderdelen zijn de verschillende scores die worden toegekend redelijk gelijkmatig verdeeld over de kwalitatieve gedragsindicaties, waarbij een maximale afwijking van 25% van het gemiddelde is toegestaan.	Wanneer sprake is van verschillende onderdelen zijn de verschillende scores die worden toegekend volledig gelijkmatig verdeeld over de kwalitatieve gedragsindicaties.
<i>Aggregation</i>	Wanneer sprake is van verschillende onderdelen, zijn minder dan de helft van de items behorende bij een bepaald onderdeel coherent.	Wanneer sprake is van verschillende onderdelen, zijn meer dan de helft van de items behorende bij een bepaald onderdeel coherent.	Wanneer sprake is van verschillende onderdelen, zijn alle items behorende bij een bepaald onderdeel coherent.
<i>Generalization</i>	Minder dan de helft van de kwalitatieve gedragsindicaties die in het theoretisch kader staan beschreven, komen eerlijk verdeeld terug in de items. Daarnaast komen veel items (meer dan 25%) niet overeen met de kwalitatieve gedragsindicaties.	Meer dan de helft van de kwalitatieve gedragsindicaties die in het theoretisch kader staan beschreven, komen eerlijk verdeeld terug in de items. Daarnaast komen weinig items (maximaal 25%) niet overeen met de kwalitatieve gedragsindicaties.	Alle kwalitatieve gedragsindicaties die in het theoretisch kader staan beschreven, komen eerlijk verdeeld terug in de items. Daarnaast komen weinig items (maximaal 25%) niet overeen met de kwalitatieve gedragsindicaties.

<i>Extrapolation</i>	Minder dan 75% van de kwalitatieve gedragsindicaties uit het theoretisch kader komen terug in de items van het instrument.	Tenminste 75% van de kwalitatieve gedragsindicaties uit het theoretisch kader komen terug in de items van het instrument.	Alle kwalitatieve gedragsindicaties uit het theoretisch kader komen terug in de items van het instrument.
<i>Evaluation</i>	Het instrument heeft geen richtlijnen over aan welke norm een kind moet voldoen om als bovengemiddeld intelligent te worden beschouwd.	Het instrument heeft richtlijnen over aan welke norm een kind moet voldoen om als bovengemiddeld intelligent te worden beschouwd.	Het instrument heeft duidelijke en eenduidige richtlijnen over aan welke norm een kind moet voldoen om als bovengemiddeld intelligent te worden beschouwd.
		aanvaardbaarheid	
Objectiviteit	De resultaten kunnen op meerdere manieren geïnterpreteerd worden.	Over de resultaten kan enige twijfel bestaan.	De resultaten kunnen slechts op één manier geïnterpreteerd worden.
Vraagzuiverheid	Minder dan de helft van de vragen is zo geformuleerd dat er een verschil bestaat tussen resultaten van kinderen die dezelfde taal en conventies hebben als de persoon die de afname uitvoert en kinderen die een andere taal en conventie hebben.	Tenminste de helft van de vragen is zo geformuleerd dat er geen verschil bestaat tussen resultaten van kinderen die dezelfde taal en conventies hebben als de persoon die de afname uitvoert en kinderen die een andere taal en conventie hebben.	De vragen zijn zo geformuleerd dat er geen verschil bestaat tussen resultaten van kinderen die dezelfde taal en conventies hebben als de persoon die de afname uitvoert en kinderen die een andere taal en conventie hebben.
Normhandhaving	Meer dan de helft van de items is niet eenduidig en duidelijk gedefinieerd, waardoor ze op meerdere manieren geïnterpreteerd kunnen worden.	Meer dan de helft van de items is eenduidig en duidelijk gedefinieerd, zodat ze slechts op één manier geïnterpreteerd kunnen worden.	Alle items zijn eenduidig en duidelijk gedefinieerd, zodat ze slechts op één manier geïnterpreteerd kunnen worden.
		bruikbaarheid	
Moeilijkheid	Minder dan 75% van alle items van het instrument sluiten aan bij de capaciteiten van een bovengemiddeld	Tenminste 75% van alle items van het instrument sluiten aan bij de capaciteiten van een bovengemiddeld intelligent	Alle items van het instrument sluiten aan bij de capaciteiten van een bovengemiddeld intelligent kind in vergelijking met



	intelligent kind in vergelijking met leeftijdsgenoten.	kind in vergelijking met leeftijdsgenoten.	leeftijdsgenoten.
Onderscheidend vermogen	Er is een slecht onderscheid te maken tussen leerlingen die wel of niet bovengemiddeld intelligent zijn aan de hand van een kwantitatieve score.	Er is een redelijk onderscheid te maken tussen leerlingen die wel of niet bovengemiddeld intelligent zijn aan de hand van een kwantitatieve score.	Er is een duidelijk onderscheid te maken tussen leerlingen die wel of niet bovengemiddeld intelligent zijn aan de hand van een kwantitatieve score.
Kosten	De kosten voor het gebruik van het instrument bedragen jaarlijks meer dan 150 euro.	De kosten voor het gebruik van het instrument bedragen jaarlijks maximaal 150 euro.	Er zijn geen kosten verbonden aan het gebruik van het instrument.

## Bijlage 2. Vragenlijst bruikbaarheid voor leerkrachten

Vragenlijst leerkrachten	Naam: Leeftijd:	Welke opleiding heeft u gevolgd: Aantal jaar onderwijservaring:										
			niet					volledig				
1.	In hoeverre sluit het niveau van de vragen aan bij specifieke gedragingen van een bovengemiddeld intelligent kind? (Bijv. Zijn de vragen zo gesteld dat niet alle kinderen op deze school als bovengemiddeld intelligent kunnen worden beschouwd?)	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
2.	In hoeverre kan dit instrument het verschil aantonen tussen een kind dat wel en een kind dat niet bovengemiddeld intelligent is?	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
3.	In hoeverre voelt u zich als leerkracht in staat om met dit instrument bovengemiddeld intelligente kinderen te signaleren?	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
4.	In hoeverre vond u alle vragen duidelijk en eenduidig gesteld?	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
5.	In hoeverre is dit instrument voor u handzaam om in de praktijk te gebruiken?	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
6.	Had u het idee dat u na drie weken voldoende informatie had over de leerling om de vragenlijst in te vullen?	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
7.	Hoeveel tijd heeft u besteed aan het invullen van het instrument? En vond u dat weinig of veel?	..... minuten	weinig					veel				
			1	2	3	4	5	6	7	8	9	10
8.	In hoeverre vindt u dat de voordelen van de verkregen informatie opwegen tegen de inspanning die u heeft geleverd?	1 2 3 4 5 6 7 8 9 10	geen voordeel					veel voordeel				
			1	2	3	4	5	6	7	8	9	10
9.	Wanneer u de voor- en nadelen van het instrument tegen elkaar afweegt, tot welk eindcijfer komt u dan?	1 2 3 4 5 6 7 8 9 10	1	2	3	4	5	6	7	8	9	10
Heeft u nog op- en/of aanmerkingen?												

