



rijksuniversiteit  
groningen

faculteit gedrags- en  
maatschappijwetenschappen

# Voorstudie Convergente Validiteit LIJ

Iris J. L. Egberink

Rob R. Meijer

Faculteit Gedrags- en Maatschappijwetenschappen,

Rijksuniversiteit Groningen



## Voorwoord

Voor u ligt het eindrapport van het vooronderzoek naar de convergente validiteit van het Landelijk Instrumentarium Jeugdstrafrechtketen (LIJ). Dit onderzoek is uitgevoerd door de Faculteit Gedrags- en Maatschappijwetenschappen van de Rijksuniversiteit Groningen in opdracht van het Wetenschappelijk Onderzoeks- en Documentatiecentrum (WODC).

Vanaf deze plaats willen wij de leden van de begeleidingscommissie bedanken voor hun adviezen, aanvullende informatie en waardevolle discussies gedurende de bijeenkomsten. De betrokken leden zijn:

- De heer prof. dr. H.J.A. Hoijtink, Universiteit Utrecht, Faculteit Sociale Wetenschappen (voorzitter begeleidingscommissie)
- Mevrouw R.B. Bolt, MA, ministerie van Veiligheid en Justitie/Raad voor de Kinderbescherming (lid begeleidingscommissie tot 1 januari 2012)
- Mevrouw dr. A.G. Donker, Hogeschool Utrecht (lid begeleidingscommissie)
- Mevrouw drs. S. de Winkel, Ministerie van Veiligheid en Justitie (lid begeleidingscommissie vanaf 1 januari 2012)
- Mevrouw drs. E.M.H. van Dijk, ministerie van Veiligheid en Justitie/Wetenschappelijk Onderzoeks- en DocumentatieCentrum (opdrachtgever)

dr. Iris J.L. Egberink

prof. dr. Rob R. Meijer

Groningen, april 2012



## **Inhoudsopgave**

<b>1. Aanleiding .....</b>	<b>1</b>
<b>2. Validiteit.....</b>	<b>3</b>
2.1 Convergente Validiteit als Onderdeel van Constructvaliditeit .....	3
2.2 Predictieve Validiteit .....	4
<b>3. Beschrijving LIJ .....</b>	<b>5</b>
3.1 Pre-selectiefase .....	5
3.2 Selectiefase .....	6
3.3 Interventiefase .....	7
<b>4. Beschouwing Inhoud LIJ.....</b>	<b>9</b>
4.1 Beschouwing op Itemniveau .....	9
4.2 Beschouwing op Algemeen Niveau .....	13
<b>5. De SDQ en de SPsy Nader Bekeken .....</b>	<b>15</b>
5.1 Psychometrische Gegevens van de SDQ en de SPsy .....	15
5.2 Beschouwing Onderzoek Regioplan .....	17
5.3 Aanbevelingen m.b.t. het Gebruik van de SDQ en de SPsy.....	18
<b>6. Suggesties voor Validiteitsonderzoek LIJ .....</b>	<b>21</b>
6.1 Het LIJ en Andere Risicotaxatie Instrumenten.....	21
<b>7. Samenvatting en Conclusies .....</b>	<b>25</b>
7.1 Antwoorden op Onderzoeksvragen .....	25
7.2 Aanbevelingen voor Validiteitsonderzoek m.b.t. het LIJ als Geheel .....	26
7.3 Aanbevelingen voor Validiteitsonderzoek m.b.t. ‘Weging van Zorg’ en de SDQ en/of SPsy .....	27
<b>8. Tot Slot .....</b>	<b>29</b>
<b>Literatuur.....</b>	<b>33</b>



## 1. Aanleiding

Het landelijk instrumentarium jeugdstrafrechtketen (LIJ) is bedoeld voor jeugdigen, die in de leeftijd van 12 tot en met 17 jaar met de politie in aanraking komen in verband met een misdrijf. Het LIJ voorziet in een samenhangend systeem van instrumenten voor screening, selectie en risicotaxatie, dat ketenpartners in de jeugdstrafrechtketen in staat stelt om risico's bij deze minderjarigen te onderkennen en een daarop afgestemde effectieve sanctie te adviseren, op te leggen en uit te voeren.

Het LIJ is wetenschappelijk onderbouwd en van de delen van het LIJ die als test kunnen worden aangemerkt is de psychometrische kwaliteit deels onderzocht. Van belang is dat het onderzoek naar de psychometrische kwaliteit van de 'test-delen' van het LIJ - aan de hand van de COTAN-criteria - wordt gecompleteerd. Uiteindelijk wordt gestreefd naar een beoordeling van de psychometrische kwaliteit van het LIJ door de COTAN zelf.

Als eerste stap in deze richting worden in het huidige onderzoek de mogelijkheden om de convergente validiteit m.b.t. het LIJ te onderzoeken, in kaart gebracht. Er is reeds door Regioplan (Timmermans & Witvliet, 2011) een poging gedaan om de convergente validiteit van de zorgindicatoren in het LIJ te onderzoeken, door een vergelijking te maken tussen het LIJ en de Strengths & Difficulties Questionnaire (SDQ; Goodman, 2001; van Widenfelt, Goedhart, Treffers, & Goodman, 2003). Echter, tegelijkertijd werd nader onderzoek naar de SDQ afgerond. Boonekamp (2010) concludeerde dat "de betrouwbaarheid en validiteit van de SDQ niet toereikend lijken voor afname bij delinquente jongeren en late adolescenten". Deze onderzoeksresultaten maken dat onderzoek naar de convergente validiteit met betrekking tot het LIJ nog steeds noodzakelijk is.

De centrale onderzoeksvraag in dit rapport is dan ook: "Kan de convergente validiteit van het LIJ worden onderzocht?". Hierbij zijn de volgende twee deelvragen van belang: 'Welke delen van het LIJ zijn als test te beschouwen?' en 'Welke vragenlijsten convergeren met het LIJ?'

De belangrijkste onderwerpen van onderzoek zijn hierbij *convergente validiteit* en *het LIJ*. Dit rapport zal daarom beginnen met een beschrijving van het begrip convergente validiteit. Tijdens de eerste bijeenkomst met de begeleidingscommissie werd duidelijk dat het LIJ drie doelen dient: screening, diagnostiek en evaluatie. In de screening fase ligt de nadruk op het onderscheid maken tussen personen met een laag, midden en hoog risico op recidive. In de diagnostische fase ligt de nadruk op het toewijzen aan interventies. In de evaluatiefase ligt de

nadruk op de effectiviteit van deze toewijzing. Omdat het LIJ wordt gebruikt voor verschillende doelen, kunnen verschillende vormen van validiteit per doel van belang zijn. Daarom beschrijven wij naast het begrip convergente validiteit, tevens de begrippen constructvaliditeit en predictieve validiteit.

Voor het onderzoeken van de convergente validiteit is het belangrijk om in kaart te brengen welke constructen gemeten worden met het LIJ, en/of er delen zijn die beschouwd kunnen worden als een test om vervolgens andere instrumenten te selecteren waarmee het LIJ vergeleken kan worden. Vandaar dat na de sectie over validiteit een beschrijving van het LIJ zal worden gegeven en vervolgens een beschouwing op de inhoud van het LIJ met betrekking tot convergente validiteit.

Wanneer dit duidelijk omschreven is zal ingegaan worden op de vraag hoe nu verder en hoe een eventueel onderzoek naar de convergente validiteit is te realiseren.



## 2. Validiteit

*“..., reliability refers to test scores, not tests, and validity refers to accuracy and appropriateness of test score interpretations – again, not to tests. ... Consider the question, “is the Wechsler Adult Intelligence Scale (4th ed.; WAIS-IV) a valid test?”. Again, the question is inappropriate. Both researchers and clinicians ask such questions, but the questions must be clarified. In this instance, a superior and acceptable question would be, “Is the WAIS-IV a valid measure of intelligence for adults for clinical evaluation?”. Validity must refer to a context and a construct, not to a test, and is relevant to the interpretation of scores on a test (i.e., the attachment of meaning to performance on a measuring device in psychology).” (Reynolds, 2010, p. 3)*

Dit citaat onderstreept het belang van het correcte gebruik van psychometrische termen conform de *Standards for Educational and Psychological Testing (1999)*. Het sluit tevens aan bij de huidige opvatting over validiteitsonderzoek. In de laatste decennia van de vorige eeuw is deze namelijk verschoven van ‘verschillende vormen van validiteitsbepaling zijn verschillende vormen van validiteit’ naar ‘verschillende vormen van validiteitsbepaling zijn verschillende manieren om informatie over de validiteit te verzamelen en validiteit is een ondeelbaar begrip’ (Evers, Lucassen, Meijer, & Sijtsma, 2010). De nadruk ligt nu op het verzamelen van die validiteitsinformatie die aansluit bij het doel van de test, bijvoorbeeld predictie. Het doel van validatie is hierdoor het ondersteunen van een bepaalde interpretatie van een test op een wetenschappelijke wijze geworden, waarbij verschillende soorten ondersteuning niet automatisch even belangrijk zijn voor het doel (Borsboom, Mellenbergh, & van Heerden, 2004; ter Laak & de Goede, 2003).

### 2.1 Convergente Validiteit als Onderdeel van Constructvaliditeit

Bij constructvaliditeit wordt onderzocht of en hoe goed de test het veronderstelde construct meet. Het aantonen van constructvaliditeit is een lang en complex proces, waarbij vaak verschillende methoden worden gecombineerd (Evers et al., 2010; Gregory, 2007). Methoden die gebruikt worden zijn:

- factor-analyse om te bepalen om de test items of subtests homogeen zijn en slechts één construct meten (unidimensionaliteit);

- het vergelijken van scores van groepen personen met verschillende achtergronden en eigenschappen waarvan op basis van de theorie verwacht kan worden dat ze verschillende scores zullen hebben;
- het berekenen van correlaties van de test met andere gerelateerde (d.w.z. *convergente validiteit*) en ongerelateerde tests (d.w.z. *divergente/discriminant validiteit*);
- het aantonen dat testcores veranderen in de richting zoals beschreven door de theorie en dat de omvang van die verandering teweeg wordt gebracht door de interventie.

De convergente validiteit is dus onderdeel van constructvaliditeit. Een veel gebruikte methode voor het aantonen van convergente en divergente/discriminant validiteit is de multitrait-multimethod matrix (Campbell & Fiske, 1959).

## **2.2 Predictieve Validiteit**

De predictieve validiteit is onderdeel van criteriumvaliditeit. Bij criteriumvaliditeit wordt de voorspellende waarde van testcores op een geschikte uitkomstmaat (criterium) onderzocht. Wanneer de criteriummetingen op hetzelfde moment worden verzameld als de testcores spreken we van concurrent validiteit. Wanneer de criteriummetingen op een later tijdstip (vaak maanden of jaren later) worden verzameld dan de testcores, spreken we van *predictieve validiteit*. Hierbij moet overigens duidelijk worden gespecificeerd met welk type criteria relaties worden verondersteld. Correlationeel onderzoek en regressie-analyse zijn veelgebruikte methoden

Gegevens over de criteriumvaliditeit dragen ook bij aan informatie over de constructvaliditeit, ze geven op een andere wijze informatie over wat door de test wordt gemeten.

### 3. Beschrijving LIJ

In het Programmaplan Jeugdcriminaliteit ‘de uitvoering maakt het verschil’ dat de Minister van Justitie in februari 2008 lanceerde, wordt gesproken over het verbeteren van een op de persoon toegesneden sanctie/interventieprogramma. Een verbetering van de diagnostiek in de jeugdstrafrechtketen is in dit kader essentieel. Hiertoe is het Landelijk Instrumentarium Jeugdstrafrechtketen (LIJ) ontwikkeld, waarmee informatie wordt verzameld over

- risico’s op onder andere recidive;
- veranderbare criminogene factoren;
- zorgsignalen;
- responsiviteit.

Op deze wijze worden enerzijds het recidiverisico en anderzijds de zorgsignalen in kaart gebracht, op basis daarvan kan bepaald worden welke zorg en/of strafrechtelijke aanpak de jeugdige in kwestie nodig heeft.

Op basis van het vergelijken van verschillende risicotaxatie instrumenten voor kinderen en adolescenten a.d.h.v. diverse criteria, is gekozen voor de Washington State Juvenile Court Pre-screen Assessment (WSJCPA) en het Basis RaadsOnderzoek (BARO) als basis voor het LIJ (Vogelvang, Persoon en Sondeijker, 2007). Het instrumentarium bestaat uit drie onderdelen: 1) pre-selectiefase, 2) selectiefase en 3) interventiefase. Deze worden hieronder toegelicht.

#### 3.1 Pre-selectiefase

Tijdens de pre-selectiefase schat de politie met behulp van instrument 1 in of een jeugdige een laag, midden of hoog algemeen recidiverisico heeft (op basis van statische criminogene factoren zoals geslacht, leeftijd op moment van eerste politiecontact en delictgeschiedenis). Ook wordt met dit instrument ingeschat wat het risico op geweld tegen personen is.

Als in deze fase blijkt dat de jongere een laag recidiverisico heeft, worden er geen verdere instrumenten afgenomen. In dat geval wordt de zaak door de politie afgedaan of doorverwezen naar het Justitieel Casuïstiek Overleg (JCO). Als er sprake is van een feit dat voldoet aan de Halt-criteria (ongeacht het algemeen recidiverisico (ARR) of risico op geweld tegen personen (RGP)) worden de jongeren doorverwezen naar Halt.

De jeugdige gaat naar de selectiefase als sprake is van:

- midden of hoog ARR en/of
- midden of hoog RGP en/of

- op verzoek van het OM
- bij in verzekeringstelling (IVS; vroeghulp)
- bij een negatieve Halt.

### **3.2 Selectiefase**

Tijdens de selectiefase neemt de Raad voor de kindbescherming (RvdK) instrument 2A af, waarin de risicotaxatie van de politie wordt aangevuld met informatie over enkele dynamische criminogene factoren. Raadsmedewerkers verzamelen informatie over deze risicofactoren door middel van analyse van informatie uit dossiers en het stellen van vragen aan de jeugdige en de ouders. Ook worden andere informanten geraadpleegd (vooral telefonisch maar ook wel schriftelijk en in ieder geval de school). Instrument 2A leidt tot hernieuwde scores (laag, midden, hoog) voor het Algemeen RecidiveRisico (ARR) en Risico op Geweld tegen Personen (RGP). Ook brengt de RvdK het Dynamisch RisicoProfiel (DRP) in kaart en worden zorgsignalen geïnventariseerd. Als sprake is van een midden of hoog DRP neemt de RvdK ook instrument 2B af (uitgebreide selectiefase). Instrument 2B wordt ook door de Jeugdreclassering (JR) gebruikt bij evaluatie van de JR begeleiding of bij recidive. In geval van jongeren met een licht verstandelijke beperking wordt dit uitgevoerd door Bureau Jeugdzorg, Leger des Heils of de William Schrikker Groep (WSG). Instrument 2B is een uitbreiding van 2A: de informatie uit 2A wordt aangevuld met een meer diepgaande inventarisatie van alle dynamische risico- en protectieve factoren. Alle vragen en items uit 2A maken onderdeel uit van 2B. Wanneer informatie uit 2A al verzameld is, wordt deze bij vervolgfname van 2B niet opnieuw verzameld. D.w.z. bij de berekening van de scores op Instrument 2B, worden de informatie en de scores op basis van Instrument 2A meegenomen en aangevuld. Vervolgens kan met de systematiek gedragsinterventies worden bepaald voor welke erkende gedragsinterventies de jeugdige in aanmerking komt. Instrument 2B wordt ook in de interventiefase gebruikt.

In het geval jongeren doorverwezen zijn naar Halt, wordt op basis van een speciaal hiervoor ontwikkeld signaleringsinstrument (Halt SI) bepaald of de jeugdige zorg nodig heeft. In dat geval verwijst Halt (ook) door naar Bureau Jeugdzorg of een Opvoedbureau. Omdat Halt SI een verkorte versie is van Instrument 2A, geldt hetgeen dat voor Instrument 2A geldt ook voor Halt SI. Het wordt daarom verder niet meer apart benoemd.

### **3.3 Interventiefase**

Als de Rechtercommissaris beslist tot een in bewaring stelling, gaat de jeugdige naar een Justitiële Jeugdinstelling (JJI) en start de interventiefase. Tijdens de inbewaringstelling neemt de RvdK of de JR instrument 2B af. In geval van jongeren met een licht verstandelijke beperking wordt dit uitgevoerd door Bureau Jeugdzorg, Leger des Heils of WSG. Ook de JJI zet een aantal instrumenten in voor het verkrijgen van een beeld van de jeugdige, zoals de Structured Assessment of Violence Risk in Youth (SAVRY).



## 4. Beschouwing Inhoud LIJ

Twee deelvragen in de opdrachtomschrijving zijn ‘Welke delen van het LIJ zijn als test te beschouwen?’ en ‘Welke vragenlijsten convergeren met het LIJ?’. We vertalen “test” in dit geval als delen van het LIJ die bestaan uit een aantal vragen die samen een (psychologisch) construct meten. Om deze vragen te beantwoorden is het LIJ op twee manieren bekeken. De eerste manier is het bekijken en evalueren van alle items van de verschillende onderdelen van het LIJ met de vraag of de combinatie van een aantal items of van sommige delen van het LIJ als test beschouwd zouden kunnen worden. De tweede manier is een meer algemene manier waarop afgevraagd en onderzocht is welk(e) construct(en) het LIJ nu precies pretendeert te meten om vervolgens te zoeken naar andere vragenlijsten die hetzelfde pretenderen te meten.

Instrument 1 wordt hierbij buiten beschouwing gelaten, aangezien dit instrument statische (niet veranderbare) risicofactoren bevraagt. De instrumenten 2A en 2B verzamelen tevens dynamische risicofactoren om tot een risicotaxatie te komen en worden daarom in dit onderzoek nader bekeken.

Bij dit onderzoek is uitgegaan van de aanname dat de keuze voor en totstandkoming van het LIJ goed is onderzocht en onderbouwd, dit geldt in het bijzonder voor de scoring van het LIJ. Dit is een cruciale voorwaarde om verder onderzoek naar onder andere de convergente validiteit van het LIJ uit te kunnen voeren. We komen hier in sectie 8. *Tot Slot* nog op terug.

### 4.1 Beschouwing op Itemniveau

Voor deze beschouwing is gebruik gemaakt van de indeling van het LIJ zoals deze beschreven is in de handleiding van het LIJ<sup>1</sup>. Voor de volledigheid hebben we die tabel overgenomen in Tabel 1.

Van deze onderdelen van het LIJ bevatten de 10 domeinen items met betrekking tot dynamische risicofactoren, de overige categorieën bevatten aanvullende (meer beschrijvende) informatie (o.a., weergave van feitelijk informatie, verwachtingen van verschillende betrokkenen en inschattingen door de interviewer). Vandaar dat met name gekeken is welke (items) van de 10 domeinen beschouwd zouden kunnen worden als een test. We hebben ons hierbij gericht op de items die meewegen in het DRP. In Tabel 2 is de beschouwing per domein weergegeven, tevens is aangegeven om welke items het gaat.

---

<sup>1</sup> Hoofdstuk 4 ‘Instrument 2A en 2B: Achtergrond en algemene instructie’

Tabel 1. Overzicht indeling LIJ met aantal items per onderdeel

	Aantal items in IVS-2A	Aantal extra items in 2A	Aantal extra items in 2B	Totaal
<b>Casusgegevens</b>				
Aanleiding onderzoek	4	0	0	4
Contacten	2	0	0	2
<b>Items voor interview</b>				
Delictbespreking	10	0	0	10
<i>Domeinen</i>				
1. School	10	0	4	14
2. Werk	1	3	5	9
3. Gezin	7	2	15	24
4. Vrije tijd	3	0	0	3
5. Relaties	3	0	5	8
6. Alcohol- en drugsgebruik en gokken	2	0	4	6
7. Geestelijke gezondheid	11	3	0	14
8. Attitude	3	0	9	12
9. Agressie	2	2	3	7
10. Vaardigheden	1	4	6	11
Totaal Domeinen	43	14	51	108
Hulpverlening en responsiviteit	7	2	0	9
<b>Totaal items voor interview</b>	<b>60</b>	<b>16</b>	<b>51</b>	<b>127</b>
<b>Analyse en Conclusie</b>				
Achtergrondinformatie jeugdige	7	0	0	7
Weging van zorg	11	0	0	11
Conclusie en advies	6	0	2	8
<b>Totaal Analyse en conclusie</b>	<b>24</b>	<b>0</b>	<b>2</b>	<b>26</b>

In het algemeen geldt voor elk domein dat de (meeste) items observaties en/of constateringen zijn van bepaald gedrag en/of de aanwezigheid van een bepaalde (risico)factor en dat de onderwerpen binnen elk domein uiteenlopend zijn. Hierdoor is het moeilijk om te spreken van één onderliggend (psychologisch) construct (unidimensionaliteit). Daarnaast zijn een aantal complicerende factoren voor het meten van een (psychologisch) construct aanwezig. Het gaat hierbij om:

- de aanwezigheid van zogeheten conditionele items: veel items zijn afhankelijk van het gegeven antwoord op een eerder item
- er zijn meerdere versies van het instrument beschikbaar (voor jeugdige, voor ouders, voor mentoren/leerkrachten) en de informatie van verschillende bronnen wordt gecompleteerd c.q. kan gecompleteerd worden.
- niet alle items bevatten een weging c.q. scoring
- sommige domeinen bevatten weinig items (bijv. Domein 4 Vrije Tijd)



Tabel 2. Beschouwing per domein op itemniveau

Domein	s	Beschouwing
1. School	1, 3, 4a, 7, 8, 9a, 10-12, 13a, 13b, 14	De items hebben betrekking op dagbesteding, onderwijs/opleiding, behaalde diploma('s) en gedrag op school. Bijna alle items in dit domein (m.u.v. items 7 en 14) zijn observaties en/of constatering. Item 7 be vraagt attitude jegens belang onderwijs/opleiding en item 14 be vraagt naar de inschatting van de mentor/leerkracht.
2. Werk	3-9	De items hebben betrekking op het arbeidsverleden en de huidige werksituatie. Alle items in dit domein zijn observaties en/of constatering.
3. Gezin	1a, 2, 3, 4a, 4b, 5-16, 19-24	De items hebben betrekking op de (eerdere en huidige) gezinssituatie, aanwezigheid van andere personen, ouderlijk gezag, conflicten in gezinssituatie, contact met Bureau Jeugdzorg, contact met justitie van andere personen, bereidheid tot steun van anderen en informatie over de ouders. Alle items in dit domein zijn observaties en/of constatering. Bij item 7 (detentieverleden van huidige gezinsleden), item 8 (verzorgers waarmee jeugdige een goede band heeft), item 13 (detentieverleden van eerdere gezinsleden), item 14 (gezinsleden waarmee jeugdige een goede band heeft), en item 20 (overige problemen van gezinsleden) zijn meerdere antwoorden mogelijk.
4. Vrije tijd	1, 2, 3	De items hebben betrekking op vrijetijdsbesteding en of deze als problematisch kan worden beschouwd. Alle items in dit domein zijn observaties en/of constatering.
5. Relaties	1-8	De items hebben betrekking op het type leeftijdgenoten waar de jeugdige mee omgaat en hoe hij met hen omgaat, de binding met de gemeenschap, pesten en gepest worden en een eventuele huidige relatie. Alle items in dit domein zijn observaties en/of constatering.
6. Alcohol- en drugsgebruik en gokken	1a, 1b, 1c, 2a, 2b, 2c, 3-5, 6a, 6b, 6c	De items hebben betrekking op alcohol-, drugsgebruik en gokken, en of dit tot problemen heeft geleid. Alle items in dit domein zijn observaties en/of constatering. Bij item 1c (problemen tgv alcoholgebruik), item 2c (problemen tgv drugsgebruik) en item 6c (problemen tgv gokken) zijn meerdere antwoorden mogelijk.

vervolg Tabel 2. Beschouwing per domein op itemniveau

Domein	Items	Beschouwing
7. Geestelijke gezondheid	1, 2b, 2c, 9, 10, 13	De items hebben betrekking op aanwezigheid van een verstandelijke beperking, type stoornis en of jeugdige slachtoffer is van verwaarlozing, mishandeling en/of seksueel misbruik. Alle items in dit domein zijn observaties en/of constateringen. Bij item 7b (type stoornis), item 10 (lichamelijk mishandeling) en item 13 (seksueel misbruik) zijn meerdere antwoorden mogelijk.
8. Attitude	1-12	De items hebben betrekking op acceptatie van verantwoordelijkheid, respect voor eigendom van anderen en autoriteitsfiguren, emoties tijdens delictpleging(en), denkfouten, gevoelens jegens het slachtoffer, problemen in morele ontwikkeling, mate waarin Nederlandse omgangsvormen voor jeugdige gelden, mate waarin jeugdige controle denkt te hebben over eigen gedrag, toekomstbeeld en openstaan voor gedragsverandering. Zoals de naam van het domein aangeeft gaan de items meer in op de (antisociale) attitude van de jeugdige, waarbij de items vragen naar indicatoren, observaties en constateringen van antisociaal gedrag en niet zozeer de attitude van de jeugdige zelf. Uitzonderingen zijn item 9 (mate waarin Nederlandse omgangsvormen voor jeugdige gelden), item 10 (mate van controle over eigen gedrag), item 11 (toekomstbeeld) en item 12 (openstaan voor gedragsverandering), hier wordt specifiek gevraagd naar de mening van de jeugdige zelf. De onderwerpen van deze items zijn divers (d.w.z. niet unidimensioneel).
9. Agressie	1-7	De items hebben betrekking op verbale en fysieke agressie als oplossing voor een conflict, meldingen van gewelddadig en/of seksueel grensoverschrijdend gedrag, problemen met frustratietolerantie, emotionele uitbarstingen en interpretatie van andermaats bedoelingen en gedrag. De meeste vragen in dit domein zijn observaties en/of constateringen. Uitzonderingen zijn item 1 (verbale agressie als oplossing voor conflict), item 2 (fysieke agressie als oplossing voor conflict) en item 9 (interpretatie van andermaats bedoelingen en gedrag). Bij item 3 (meldingen gewelddadig gedrag) en item 4 (meldingen seksueel grensoverschrijdend gedrag) zijn meerdere antwoorden mogelijk.
10. Vaardigheden	1-11	De items hebben betrekking op zelfstuuringsvaardigheden. Bij dit domein gaat het specifiek om de indruk en inschatting van de interviewer over het vaardigheidsniveau van de jeugdige.

NB. Alle overige vragen in elk domein die niet worden meegewogen in het DRP zijn constateringen, bijv. opleidingsniveau van de ouders of diagnose van psychische klachten.

Door de aard van de items (vnl. observaties en constatering) en de aanwezigheid van complicerende factoren voor het meten van een construct in de instrumenten 2A en 2B zijn geen (items uit) domeinen gevonden die beschouwd zouden kunnen worden als test. Dit wil overigens niet zeggen dat het LIJ als geheel niet bruikbaar is om in te zetten als instrument (zie sectie 6. *Suggesties voor Validiteitsonderzoek LIJ* verderop in dit rapport). Wel impliceert dit dat betrouwbaarheidschattingen zoals die veel worden gebruikt voor psychologische instrumenten, zoals bijvoorbeeld coëfficiënt alpha, niet geschikt zijn om de betrouwbaarheid van (onderdelen van) het LIJ te bepalen. Interbeoordelaarsbetrouwbaarheid lijkt beter geschikt. Regioplan heeft in hun onderzoek de inter- en intrabeoordelaarsbetrouwbaarheid van het LIJ onderzocht (Timmermans & Witvliet, 2011). In het algemeen wordt er een voldoende tot goede overeenstemming gevonden tussen en binnen beoordelaars. Echter, de resultaten moeten met voorzichtigheid worden geïnterpreteerd aangezien de interbeoordelaarsbetrouwbaarheid gebaseerd is op 21 casussen en de intrabeoordelaarsbetrouwbaarheid op 7 casussen<sup>2</sup>.

#### 4.2 Beschouwing op Algemeen Niveau

Om de vraag “Wat ‘meet’ het LIJ?” te beantwoorden, is begonnen met het nader bekijken van het doel van het LIJ. Het volgende staat beschreven in de handleiding van het LIJ<sup>3</sup>: “Doelstelling van de instrumenten 1, 2A en 2B van het Landelijk Instrumentarium Jeugdstrafrecht (LIJ) is:

- een inschatting geven van
  - o het risico op herhaling van crimineel gedrag (*Algemeen Recidive Risico (ARR)*);
  - o het risico op herhaling van crimineel gedrag waarbij geweld tegen personen gebruikt wordt (*Risico op Geweld tegen Personen (RGP)*);
- het in beeld brengen van
  - o de dynamische beschermende en risicofactoren die samenhangen met de kans op recidive (*Dynamisch Risicoprofiel (DRP)*);
  - o signalen van psychosociale of psychische problemen (*zorgsignalen*).

---

<sup>2</sup> De COTAN noemt geen benodigde aantallen met betrekking tot inter- en intra-beoordelaarsbetrouwbaarheid, omdat het vaak afhankelijk is van de situatie en de context waarin de gegevens verzameld worden. In het onderzoek van Regioplan werd, rekening houdend met de praktische haalbaarheid en in overeenstemming met de begeleidingscommissie gestreefd naar 30 casussen voor de interbeoordelaarsanalyse en eveneens 30 voor de intrabeoordelaarsanalyse. In praktijk zijn dat er minder geworden.

<sup>3</sup> sectie 4.2 *Doel* in Hoofdstuk 4 ‘Instrument 2A en 2B: Achtergrond en algemene instructie’

- de motivatie / motiveerbaarheid, leerstijl en (on)mogelijkheden van de jeugdige en zijn omgeving om aan een bepaalde interventie deel te nemen en ervan te profiteren (*responsiviteit*).”

Het LIJ brengt dus enerzijds het recidiverisico en anderzijds de zorgsignalen in kaart om zodoende een passende interventie voor de jeugdige te vinden. Het LIJ is dus een informatieverzamelingsinstrument die door de verschillende ketenpartners gebruikt kan worden. Echter, het inschatten van de kans op recidivisme op basis van die informatie is geen eenduidig (psychologisch) construct. Zorgsignalen in de hoedanigheid van psychische problemen kunnen daarentegen wel beschouwd worden als psychologisch construct.

Een aantal items uit de 10 verschillende domeinen hebben betrekking op het signaleren van zorgsignalen. Voor het onderdeel ‘Weging van Zorg’ worden deze eerder ingevulde items bij elkaar gezet in één tabel. Vervolgens wordt aan de gebruiker gevraagd per domein de mate van zorg aan te geven. Aan het eind van de tabel wordt gevraagd om een totaal oordeel te geven over de mate van zorg en om het risico op schade aan de jeugdige zelf in te schatten. Echter, deze items zijn tezamen niet te beschouwen als test (zie sectie 4.1 *Beschouwing op Itemniveau* van dit rapport). Naast deze items met betrekking tot zorgsignalen, is een ander belangrijk onderdeel van ‘Weging van Zorg’ het signaleren van psychische problemen. Psychische problemen daarentegen zijn wel psychologische construct(en) en zouden afhankelijk van de gebruikte items en afnamevorm beschouwd kunnen worden als test. Binnen het LIJ is gekozen voor afname van de SDQ en/of het Screeningsinstrument Psychische stoornissen (SPsy; van Oort, van ’t Land, & de Ruiter, 2005).

Regioplan heeft in een eerder onderzoek de convergente validiteit van de zorgsignalen van het LIJ onderzocht (Timmermans & Witvliet, 2011) door deze indicatoren te vergelijken met de SDQ (Goodman, 2001; van Widenfelt, et al., 2003). Echter, de resultaten van dit onderzoek zijn niet geïnterpreteerd omdat tegelijkertijd in een ander onderzoek de betrouwbaarheid en validiteit van de SDQ voor het gebruik bij delinquente jongeren in twijfel werd getrokken (Boonekamp, 2010). Daarom zullen in het onderhavige onderzoek de SDQ en de SPsy en hun psychometrische kwaliteit nader worden bekeken. Dit om enerzijds weer te geven wat er bekend is over deze instrumenten en anderzijds om een afweging te kunnen maken met betrekking tot de inzetbaarheid van dit instrument voor het signaleren van psychische problemen. Tevens zal een beschouwing worden gegeven op het eerder door Regioplan uitgevoerde onderzoek.

## 5. De SDQ en de SPsy Nader Bekeken

Voor het signaleren van psychische problemen is binnen het LIJ gekozen voor afname van de SDQ en/of de SPsy. De SDQ is een screeningsinstrument voor psychosociale problemen bij jeugdigen. De vragenlijst bestaat uit 25 items, verdeeld over de volgende vijf schalen: Emotionele problemen, Gedragsproblemen, Hyperactiviteit/Aandachtsproblemen, Sociale problemen en Prosociaal gedrag. De eerste vier schalen vormen samen de totale probleemschaal. Tevens wordt gevraagd naar de invloed van eventuele problemen op de omgeving en/of het dagelijks leven van de jeugdige, de zogeheten ‘impactscore’. Er zijn drie versies, één voor ouders van 4-16 jarige jeugdigen, één voor leerkrachten van 4-16 jarige jeugdigen, en één voor jeugdigen zelf van 11-17 jaar. De SDQ is oorspronkelijk ontwikkeld als routinematig screeningsinstrument voor vroegtijdige opsporing van psychische problemen bij jongeren (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000).

De SPsy bestaat naast de vragen van de SDQ uit 23 aanvullende items over Problemen met alcohol, Problemen met drugs, Eetstoornissen, Zelfdestructief gedrag en Psychotische kenmerken. Dit instrument is ontwikkeld door het Trimbos Instituut en wordt gebruikt door de Bureaus Jeugdzorg bij jeugdigen van 4-18 jaar. Er is een versie voor ouders van jeugdigen van 4-11 jaar en van 12-18 jaar en een versie voor de jeugdige zelf van 12-18 jaar.

### 5.1 Psychometrische Gegevens van de SDQ en de SPsy

Informatie over de betrouwbaarheid (de mate waarin testresultaten kunnen worden gerepliceerd, de herhaalbaarheid) en validiteit van de SDQ met betrekking tot de Nederlandse versie bij gebruik in de ‘normale’ populatie jeugdigen is schaars (Boonekamp, 2010). Voor zover bij de onderzoekers bekend, zijn er slechts enkele onderzoeken uitgevoerd naar de psychometrische eigenschappen van de Nederlandse versie van de SDQ in de daarvoor bedoelde populatie en gepubliceerd in wetenschappelijke tijdschriften. Deze zullen hieronder worden besproken. In de handleiding van het LIJ<sup>4</sup> wordt voor een beschrijving en de handleiding van de SPsy verwezen naar van Oort, van ’t Land en de Ruiter (2006). Er is

---

<sup>4</sup> Op p. 18, sectie 4.8.2 *Signalering psychische problemen* in Hoofdstuk 4 ‘Instrument 2A en 2B: Achtergrond en algemene instructie’

getracht deze publicatie op te vragen, maar zonder succes<sup>5</sup>. Voor zover bij de onderzoekers bekend is geen aanvullend wetenschappelijk onderzoek uitgevoerd naar de psychometrische kwaliteit van aanvullende vragen in de SPsy. We richten ons daarom op de psychometrische eigenschappen van de SDQ, welke tevens een groot onderdeel is van de SPsy.

Van Widenfelt en anderen (2003) beschrijven de vertaling van de Nederlandse versie en onderzochten de psychometrische eigenschappen van de drie verschillende versies van de SDQ in de ‘normale’ populatie. Voor de leerkrachtversie was de betrouwbaarheid redelijk tot goed, het varieerde van  $\alpha = .74$  (Sociale problemen) tot  $\alpha = .89$  (Aandachtsproblemen). Voor de ouderversie was de betrouwbaarheid van de verschillende schalen minder goed, deze varieerde van  $\alpha = .57$  (Sociale problemen en Prosociaal gedrag) tot  $\alpha = .84$  (Aandachtsproblemen). Voor schalen van de versie van de jeugdige zelf is de betrouwbaarheid nog lager, deze varieerde van  $\alpha = .39$  (Sociale problemen) tot  $\alpha = .70$  (totale probleemschaal). Muris, Meesters, en van den Berg (2003) onderzochten de psychometrische eigenschappen van de ouder- en jeugd versie bij ‘normale’ jeugdigen van 9-15 jaar. Zij vonden dat voor de ouderversie de betrouwbaarheid van de schalen varieerde van  $\alpha = .55$  (Gedragsproblemen) tot  $\alpha = .80$  (totale probleemschaal) en voor de versie voor de jeugdige van  $\alpha = .45$  (Gedragsproblemen) tot  $\alpha = .78$  (totale probleemschaal). Muris, Meesters, Eijkelenboom, en Vincken (2004) namen de versie voor ‘normale’ jeugdigen van 11-17 jaar af bij ‘normale’ jeugdigen van 8-13 jaar om te onderzoeken of de psychometrische eigenschappen bij deze jongere groep vergelijkbaar zijn met die voor de oudere groep waarvoor de vragenlijst bedoeld is. Ook zij vonden een lage geschatte betrouwbaarheid voor verschillende schalen, variërend van  $\alpha = .41$  (Sociale problemen) tot  $\alpha = .76$  (totale probleemschaal). Deze onderzoeken laten zien dat de betrouwbaarheid van de totale probleemschaal redelijk tot goed is ( $\alpha$  rond .80) voor de drie versies. Echter, de betrouwbaarheid van de verschillende subschalen is lager en varieert per versie, waarbij de

---

<sup>5</sup> Er is gezocht naar contactgegevens van M. van Oort, maar die zijn op het internet niet gevonden. Daarop is contact gelegd met H. van 't Land. Zij verwees ons door naar C. de Ruiter. Deze liet per mail weten dat er geen officiële handleiding van de SPsy bestaat en dat zij de bedoelde publicatie niet in haar bezit heeft. Tevens is door het WODC getracht de publicatie op te vragen bij het Trimbosinstituut, ook hier zonder succes. Mevrouw De Winkel heeft contact gehad met Bureau Jeugdzorg Utrecht, maar ook daar was de publicatie niet verkrijgbaar. Wij hebben zelf ook nogmaals contact gehad met het Trimbosinstituut, maar daar ontstond na een lange zoektocht het vermoeden dat de literatuurverwijzing onjuist is. Daarop is contact gelegd met de heer Spanjaard, een van de auteurs van de handleiding van het LIJ. Hij gaf aan dat hij destijds in 2006 een concepthandleiding heeft ontvangen van een van de auteurs die nog afgerond zou worden. Hij heeft echter nooit de definitieve handleiding ontvangen.

betrouwbaarheden het hoogst zijn voor de ouderversie en het laagst voor de versie voor de jeugdige zelf.

In 2007 beoordeelde de COTAN de ouderversie van de SDQ voor ‘normale’ jeugdigen van 7-12 jaar (Evers, Braak, Frima, & van Vliet-Mulder, 2009-2011). De onderdelen *handleiding*, *normen* en *criteriumvaliditeit* werden als onvoldoende beoordeeld. De handleiding verschaft te weinig informatie, de normen zijn niet representatief en/of de representativiteit is niet te beoordelen en er is te weinig onderzoek verricht naar de criteriumvaliditeit. Verder wordt vermeld dat in de handleiding van de ouderversie geen gegevens verstrekt worden over de leerkracht- en jeugd versie.

Het afstudeeronderzoek van Boonekamp (2010) is mede naar aanleiding van deze resultaten uitgevoerd. Een andere reden was dat is gebleken dat de SDQ in de klinische praktijk bij delinquente jongeren en jongeren ouder dan 16 jaar wordt afgenomen. Er is echter nog geen onderzoek uitgevoerd naar de psychometrische eigenschappen van de SDQ binnen deze specifieke populaties. Zoals eerder beschreven, in dit onderzoek werd geconcludeerd dat er twijfels zijn met betrekking tot de betrouwbaarheid en validiteit van de SDQ voor deze specifieke doelgroepen.

## **5.2 Beschouwing Onderzoek Regioplan**

In het kader van de hiervoor beschreven onderzoeken is nader gekeken naar het eerder door Regioplan uitgevoerde onderzoek naar de convergente validiteit van de zorgindicatoren in het LIJ (Timmermans & Witvliet, 2011). Hiervoor zijn correlaties uitgerekend tussen de zorgsignalen (d.w.z. de items uit de 10 domeinen van het DRP m.b.t. het in kaart brengen van zorgsignalen) en de scores op de SDQ. Wat hierbij opvalt is dat veelal itemscores met betrekking tot zorgsignalen in het LIJ gecorreleerd worden met ofwel itemscores op corresponderende SDQ items ofwel met verschillende SDQ schaalscores. Het is vrij ongebruikelijk om itemscores te gebruiken voor het berekenen van correlaties, vanwege de vaak hoge(re) onbetrouwbaarheid van individuele items. Vermoedelijk hebben de onderzoekers van Regioplan hiervoor gekozen, omdat zij zich ook realiseerden dat de zorgsignalen in het LIJ niet beschouwd kunnen worden als een test die een (psychologisch) construct meet. Wat verder opvalt is dat de steekproefgrootte klein is; gemiddeld zijn 55 waarnemingen gebruikt.

Tevens is het ongebruikelijk dat de convergente validiteit van de zorgindicatoren van het LIJ onderzocht wordt met behulp van een instrument dat ook een onderdeel is van het LIJ.

### 5.3 Aanbevelingen m.b.t. het Gebruik van de SDQ en de SPsy

Voor zover wij kunnen overzien is er nog te weinig bekend over de psychometrische eigenschappen van de Nederlandse versie van de SDQ en de SPsy om deze zonder problemen in te kunnen zetten als screeningsinstrumenten voor psychische problemen en/of stoornissen bij de specifieke doelgroep jeugdige delinquenten. Naar aanleiding daarvan zou gekozen kunnen worden uit de volgende mogelijkheden:

1. er wordt verder onderzoek uitgevoerd naar de psychometrische eigenschappen van de SDQ/SPsy bij gebruik voor Nederlandse jeugdige delinquenten.
2. er wordt gekozen voor een ander bestaand instrument dat ook als doel heeft om psychische problemen bij jeugdigen te signaleren, waar de psychometrische eigenschappen goed zijn onderzocht en welke door de COTAN als voldoende en/of goed beoordeeld zijn. Dit instrument zou bij voorkeur bedoeld moeten zijn voor gebruik bij jeugdige delinquenten. Bij deze variant zou er ook gekozen kunnen worden om verschillende kortere vragenlijsten in te zetten die de verschillende domeinen van de SDQ en SPsy in kaart brengen. Voorbeelden van te overwegen instrumenten zijn:

- Sociaal-Emotionele Vragenlijst (SEV; Scholte & van der Ploeg, 2005b)

COTAN 2004: de criteria *uitgangspunten bij de testconstructie* en *kwaliteit van de handleiding* zijn als voldoende beoordeeld, de overige criteria als goed. Populatie: kinderen en adolescenten 4-18 jaar. “Het kan gebruikt worden voor screening van schoolgaande jeugd en in de jeugdzorg.” (Evers, et al., 2009-2011)

- Vragenlijst Sociale en Pedagogische Situatie (VSPS; Scholte, 1996)

COTAN 1999: de criteria *kwaliteit van het testmateriaal* en *kwaliteit van de handleiding* zijn als goed beoordeeld, de overige criteria als voldoende. De normen zijn verouderd, waardoor dit criterium als onvoldoende is beoordeeld. Populatie: jongeren met psychosociale (gedrags)problematiek, vanaf ca 6 jaar.

- Vragenlijst voor Gedragsproblemen bij Kinderen (VvGK; Oosterlaan, Scheres, Antrop, Roeyers, & Sergeant, 2000)

COTAN 2001: de criteria *uitgangspunten bij de testconstructie*, *normen* en *betrouwbaarheid* zijn als voldoende beoordeeld, de overige criteria als goed. Er is geen onderzoek verricht naar de criteriumvaliditeit, omdat deze test volgens de auteur(s)/uitgever niet bedoeld is voor voorspellend gebruik. Populatie: kinderen, 6 t/m 12 jaar. Nieuwe normen nodig.



- Child Behavior Checklist (CBCL; Verhulst, van der Ende, & Koot, 1996)  
COTAN 1999: de criteria *betrouwbaarheid* en *criteriumvaliditeit* zijn als voldoende beoordeeld, de overige criteria als goed. Bij het criterium *normen* wordt de aantekening gemaakt dat de normen verouderd zijn. Deze stammen uit 1993. Populatie: jongens en meisjes, 4 t/m 18 jaar. Deze beoordeling heeft alleen betrekking op de versie voor de ouders. Er is ook een versie voor de leerkracht, Teacher Report Form (TRF), en voor de jeugdige zelf, Youth Self-Report (YSR; zie hieronder).
- Teacher's Report Form (TRF; Verhulst, van der Ende, & Koot, 1997a)  
COTAN 1999: de criteria *begripsvaliditeit* en *criteriumvaliditeit* worden als voldoende beoordeeld, de overige criteria als goed. Bij het criterium *normen* wordt de kanttekening gemaakt dat de normen (uit 1993) verouderd zijn. Bij het criterium *betrouwbaarheid* wordt de kanttekening gemaakt dat de betrouwbaarheid van de totaalscores als goed beoordeeld is, maar dat het varieert voor de overige schalen en soms onvoldoende is. Populatie: jeugdigen 4-18 jaar. Er is ook een versie voor de ouders, CBCL, en voor de jeugdige zelf, YSR.
- Youth Self-Report (YSR; Verhulst, van der Ende, & Koot, 1997b)  
COTAN 1999: de criteria *begripsvaliditeit* en *criteriumvaliditeit* worden als voldoende beoordeeld, de overige criteria als goed. Bij het criterium *normen* wordt de kanttekening gemaakt dat de normen (uit 1993) verouderd zijn. Bij het criterium *betrouwbaarheid* wordt de kanttekening gemaakt dat de betrouwbaarheid van de totaalscores Internaliseren, Externaliseren en de Totale probleemscore als goed beoordeeld is, maar dat het varieert voor de overige schalen en is soms onvoldoende. Populatie: jeugdigen 4-18 jaar. Er is ook een versie voor de ouders, CBCL, en voor de leerkracht, TRF.
- ADHD vragenlijst (AVL; Scholte & van der Ploeg, 2005a)  
COTAN 2004: de criteria *normen* en *criteriumvaliditeit* zijn als voldoende beoordeeld, de overige criteria als goed. Populatie: jeugdigen, normaal en klinisch.



## 6. Suggesties voor Validiteitsonderzoek LIJ

De beschouwing van het LIJ op item en algemeen niveau heeft uitgewezen dat (de items van) de 10 domeinen van het DRP geen onderdelen bevatten die beschouwd kunnen worden als een test, maar dat het LIJ als geheel (d.w.z. het DRP) beschouwd zou kunnen worden als een risicotaxatie instrument. De convergente validiteit van het LIJ als geheel zou dan onderzocht kunnen worden door scores op het DRP te vergelijken met scores op een ander soortgelijk (gevalideerd) risicotaxatie instrument. In sectie *6.1 Het LIJ en Andere Risicotaxatie Instrumenten* zullen een aantal mogelijk bruikbare instrumenten worden besproken.

Daarnaast heeft de beschouwing op algemeen niveau uitgewezen dat (het signaleren van) psychische problemen als onderdeel van ‘Weging van Zorg’, afhankelijk van de gebruikte items en afnamevorm, wel psychologische construct(en) zijn en beschouwd kunnen worden als test. Binnen de huidige vorm van het LIJ is gekozen voor afname van de SDQ en/of SPsy. In sectie *5.3 Aanbevelingen m.b.t. het Gebruik van de SDQ en/of SPsy* wordt kort beschreven wat nodig is om het signaleren van psychische problemen in kaart te brengen binnen het LIJ. In het kader van het onderzoeken van de convergente validiteit van (onderdelen van) het LIJ zou bij variant 1 eigen onderzoek uitgevoerd moeten worden naar de convergente validiteit van de SDQ en/of SPsy, waarbij, nadat de psychometrische eigenschappen van voldoende kwaliteit zijn bevonden binnen de populatie jeugdige delinquenten, onderzocht zou moeten worden hoe de SDQ en/of SPsy convergeren met bijvoorbeeld een aantal instrumenten die genoemd zijn onder variant 2. Wanneer binnen variant 2 gekozen wordt voor een ander bestaand, gevalideerd instrument dat tevens geschikt is voor jeugdige delinquenten, hoeft er geen vervolgonderzoek te worden uitgevoerd naar de convergente validiteit. Wanneer binnen variant 2 gekozen wordt voor een gevalideerd instrument dat niet geschikt is voor jeugdige delinquenten, zou het oorspronkelijke onderzoek voor die andere populatie gerepliceerd moeten worden bij jeugdige delinquenten.

### 6.1 Het LIJ en Andere Risicotaxatie Instrumenten

Bij de keuze voor risicotaxatie instrumenten die geschikt zijn voor het onderzoeken van de convergente validiteit van het LIJ is het belangrijk dat de psychometrische eigenschappen (o.a. betrouwbaarheid en validiteit) van die instrumenten onderzocht en goed zijn. Door de COTAN beoordeelde instrumenten met minimaal een voldoende beoordeling op de verschillende criteria genieten dan ook de voorkeur. Tevens is het belangrijk dat de

instrumenten ontwikkeld en bedoeld zijn voor dezelfde populatie als waarvoor het LIJ bedoeld is, namelijk Nederlandse delinquente jeugdigen van 12 tot en met 17 jaar. Een ander belangrijk aspect is dat het DRP en het convergerende instrument gebaseerd moeten zijn op informatie van exact dezelfde respondent(en). Het feit dat het DRP tot stand komt op basis van informatie van de jeugdige zelf, de ouders en andere informanten en het niet altijd duidelijk aangegeven wordt op basis van welke informanten een DRP tot stand is gekomen is hierin een complicerende factor. De COTAN verlangt dat bij verschillende respondenttypen afzonderlijke normen en betrouwbaarheid worden gegeven. Ook de correlatie tussen de versies van de verschillende respondenttypen moet worden vermeld. Bij ander validiteitsonderzoek is het minder duidelijk of aparte gegevens vereist zijn.

Vogelvang et al. (2007) hebben in een eerdere fase van de ontwikkeling van het LIJ verschillende risicotaxatie-instrumenten voor kinderen en adolescenten op een rij gezet en getoetst aan de hand van verschillende criteria. Op basis hiervan is gekozen voor het combineren van de WSJCPA en het BARO als basis voor het LIJ. Deze twee vragenlijsten kunnen daarom niet gebruikt worden om de convergente validiteit van het LIJ te onderzoeken.

Van de door Vogelvang en anderen (2007) bekeken instrumenten voor jeugdigen is geen enkele beoordeeld door de COTAN. Verder valt op dat er zeer weinig soortgelijke risicotaxatie instrumenten zijn voor jeugdigen. De Structured Assessment Violence Risk in Youth (SAVRY; naar het Nederlands vertaald door Lodewijks, Doreleijers, de Ruiter, & de Wit-Grouls, 2001) is één van de weinige instrumenten die bedoeld is voor delinquente jongeren van 12 tot 18 jaar. Een nadeel van de SAVRY is dat er nog te weinig bekend is over de validiteit (Vogelvang, et al., 2007), alhoewel in de afgelopen jaren meer validiteitsonderzoek uitgevoerd is (Lodewijks, 2008; Lodewijks, Doreleijers, de Ruiter, & Borum, 2008). Een ander nadeel van de SAVRY is dat het geen gebruik maakt van numerieke scores (Lodewijks, de Ruiter, & Doreleijers, 2003).

Een risicotaxatie instrument voor volwassenen dat wel door de COTAN in 2007 beoordeeld is zijn de Forensische Profiellijsten (FP-40<sup>6</sup>; Brand & van Emmerik, 2006). Dit instrument is ontwikkeld voor forensische patiënten en Tbs-gestelden en wordt ingevuld door leden van het multidisciplinaire behandelteam van een Tbs-instelling. De COTAN criteria

---

<sup>6</sup> De Dienst Justitiële Inrichtingen heeft de Forensisch Profiel Justitiële Jeugdlijst (FPJ-lijst; Brand & van Heerde, 2004, 2010) ontwikkeld. Bij het ontwikkelen van deze lijst is gebruik gemaakt van zeven bestaande instrumenten voor diagnostiek en risicotaxatie, waaronder de FP-40 (Brand & van den Hurk, 2008). De FPJ-lijst is geen vragenlijst, maar wordt gebruikt als checklist om gegevens uit dossiers van jongeren met de justitiële maatregel 'plaatsing in een jeugdinrichting' (PIJ-maatregel) op systematische wijze te scoren zodat deze opgenomen kunnen worden in een database.

*uitgangspunten bij de testconstructie en kwaliteit van de handleiding* zijn als goed beoordeeld, de overige criteria als voldoende, met uitzondering van de *criteriumvaliditeit*. Deze is als onvoldoende beoordeeld, er is te weinig onderzoek verricht. Een nadeel van het gebruik van een risicotaxatie instrument voor volwassenen bij jeugdigen is dat het niet duidelijk is of het instrument ook dezelfde psychometrische eigenschappen heeft bij het gebruik in een andere populatie.

Andere risicotaxatie instrumenten voor volwassenen die eventueel ook overwogen zouden kunnen worden zijn Historisch Klinisch Toekomst-30 (HKT-30; Werkgroep risicotaxatie forensische psychiatrie, 2003) en de Recidive InschattingsSchalen (RISc; van der Knaap, Leenarts, & Nijssen, 2007). Met betrekking tot de HKT-30 en RISc moet nog meer normerings- en validiteitsonderzoek uitgevoerd worden.



## 7. Samenvatting en Conclusies

### 7.1 Antwoorden op Onderzoeksvragen

Het LIJ is een informatieverzamelingsinstrument bedoeld voor jeugdigen, die in de leeftijd van 12 tot en met 17 jaar met de politie in aanraking komen in verband met een misdrijf, en kan door de verschillende ketenpartners gebruikt worden. Het LIJ brengt enerzijds het recidiverisico en anderzijds de zorgsignalen in kaart om zodoende een passende interventie voor de jeugdige te vinden. Het LIJ is wetenschappelijk onderbouwd en het streven is om het LIJ door de COTAN te laten beoordelen. Hiervoor is het van belang dat (deels) eerder uitgevoerd onderzoek naar de psychometrische kwaliteit van de ‘test-delen’ van het LIJ aan de hand van de COTAN-criteria wordt gecompleteerd. Door Regioplan (Timmermans & Witvliet, 2011) is geprobeerd de convergente validiteit van de zorgindicatoren in het LIJ te onderzoeken, door het LIJ te vergelijken met de SDQ (Goodman, 2001; van Widenfelt, Goedhart, Treffers, & Goodman, 2003). Echter, tegelijkertijd concludeerde Boonekamp (2010) in zijn onderzoek naar de SDQ dat “de betrouwbaarheid en validiteit van de SDQ niet toereikend lijken voor afname bij delinquente jongeren en late adolescenten”. Deze onderzoeksresultaten zijn de aanleiding geweest voor het huidige vooronderzoek naar de convergente validiteit met betrekking tot het LIJ.

De centrale vraag van dit vooronderzoek is “Kan de convergente validiteit van het LIJ worden onderzocht?”. Om deze centrale vraag te beantwoorden is begonnen met het bestuderen van de verscheidene beschikbare documenten en onderzoeksrapporten met betrekking tot de ontwikkeling van en eerder uitgevoerd onderzoek naar het LIJ. Op basis daarvan zijn additionele documenten verzameld en bestudeerd. Verder zijn alle items van de verschillende onderdelen van het LIJ bekeken en geëvalueerd met de vraag of de combinatie van een aantal items of van sommige delen van het LIJ als test beschouwd zouden kunnen worden. Het LIJ is ook op een algemeen niveau beschouwd waarbij onderzocht is welk(e) construct(en) het LIJ nu precies pretendeert te meten. Deze informatie is gebruikt om vervolgens literatuuronderzoek uit te voeren naar andere vragenlijsten die hetzelfde pretenderen te meten cq. in kaart brengen. Tevens is gekeken naar de huidige stand van zaken in de literatuur betreffende het begrip validiteit en met name het begrip convergente validiteit. Bij dit alles is geprobeerd zoveel mogelijk rekening te houden met hoe dit onderzoek past binnen de richtlijnen van de COTAN. Na het in overweging nemen van alle beschikbare informatie lijkt het erop dat de centrale onderzoeksvraag niet met een simpele ‘ja’ of ‘nee’ beantwoord kan worden.

Het antwoord op de eerste deelvraag “Welke delen van het LIJ zijn als test te beschouwen?” is dat het LIJ c.q. het DRP geen delen bevat die als test te beschouwen zijn. Echter, (het signaleren van) psychische problemen als onderdeel van ‘Weging van Zorg’ kan onder bepaalde voorwaarden wel beschouwd worden als test (voor een nuancering, zie sectie *7.3 Aanbevelingen voor Validiteitsonderzoek m.b.t. ‘Weging van Zorg’ en de SDQ en/of SPsy* van dit rapport). Het LIJ als geheel (het DRP in het bijzonder) is een risicotaxatie instrument.

Het antwoord op de daarop aansluitende deelvraag “Welke vragenlijsten convergeren met het LIJ?” is gecompliceerder. Onderzoek naar de (convergente) validiteit is dus alleen relevant voor het LIJ als geheel en het onderdeel ‘Weging van zorg’. Dit wordt nader besproken in de secties *7.2 Aanbevelingen voor Validiteitsonderzoek m.b.t. het LIJ als Geheel* en *7.3 Aanbevelingen voor Validiteitsonderzoek m.b.t. ‘Weging van Zorg’ en de SDQ en/of SPsy*.

## **7.2 Aanbevelingen voor Validiteitsonderzoek m.b.t. het LIJ als Geheel**

Hoewel er verschillende risicotaxatie instrumenten in Nederland beschikbaar zijn, er is nog te weinig validiteitsonderzoek (en soms ook normeringsonderzoek) bij deze instrumenten uitgevoerd en/of ze zijn niet bedoeld voor delinquente jeugdigen van 12 tot en met 17 jaar. Onderzoek naar de convergente validiteit m.b.t. het LIJ als geheel lijkt op dit moment dus niet mogelijk. Gezien de huidige visie met betrekking tot validiteit is het wellicht zinvoller om validiteitsinformatie te verzamelen die aansluit bij het doel van het instrument, namelijk op basis van het LIJ goed onderscheid maken tussen een laag, midden en hoog risico op recidive (screening), het toewijzen van jeugdigen aan de voor hen meest geschikte interventie (diagnostiek) en de effectiviteit van die toewijzing evalueren (evaluatie). Omdat dit buiten de range van dit vooronderzoek valt, maar een belangrijk onderwerp is zullen hier toch een aantal suggesties worden aangedragen voor een dergelijk onderzoek.

De recidive monitor van het WODC zou bij dit type validiteitsonderzoek een belangrijke rol kunnen spelen, met name bij de doelen screening en evaluatie. Echter, omdat alleen de zaken die bij het Openbaar Ministerie komen in deze monitor worden bijgehouden, moet men alert zijn op het zogeheten ‘restriction of range’ probleem. Het is daarom belangrijk om te zorgen voor zoveel mogelijk spreiding in de ernst en frequentie van de delicten. De gefaseerde invoering van het LIJ zou gebruikt kunnen worden voor het onderzoeken van de effectiviteit. Bijvoorbeeld door te vergelijken hoeveel en/of hoe vaak jeugdigen, waarbij het LIJ is ingezet, opnieuw de fout zijn ingegaan ten opzichte van jeugdigen waarbij het LIJ niet



is ingezet. Tevens zou het onderzoek van van der Knaap en Alberda (2009) met betrekking tot de predictieve validiteit van de RISc als leidraad kunnen dienen.

### **7.3 Aanbevelingen voor Validiteitsonderzoek m.b.t. ‘Weging van Zorg’ en de SDQ en/of SPsy**

Voor het onderdeel ‘Weging van Zorg’ worden de antwoorden op de items ten aanzien van de psychosociale situatie van de jeugdige uit de tien domeinen bij elkaar gezet. De gebruiker wordt gevraagd op basis van deze antwoorden een klinisch oordeel te geven over de mate van zorg per domein, over de algehele psychosociale situatie van de jeugdige en het risico op schade aan zichzelf. Tevens worden de SDQ en/of de SPsy afgenomen bij de jeugdige en/of zijn/haar ouder(s) voor het signaleren van psychische problemen.

Standaardisatie is een belangrijk onderdeel bij gegevensverzameling. Standaardisatie is soms lastig bij klinische oordelen, vandaar dat gestandaardiseerde instrumenten zonder klinische oordelen de voorkeur genieten boven gestandaardiseerde instrumenten met klinische oordelen (Bosker, 2008). Verder moet bij klinische oordelen onderzocht worden of de verschillende beoordelaars op dezelfde wijze beoordelen (Regioplan heeft de inter- en intrabeoordelaarsbetrouwbaarheid onderzocht, zie sectie *4.1 Beschouwing op Itemniveau*).

Omdat naast deze items m.b.t. zorgsignalen ook de SDQ en/of SPsy afgenomen worden voor het signaleren van psychische problemen en er overlap is tussen deze onderdelen, zou een alternatief kunnen zijn om deze klinische oordelen te vervangen door alleen een vragenlijst voor het in kaart brengen van psychosociale problemen. Dat is immers het doel van de afname van de SDQ en/of de SPsy en de items uit het LIJ t.b.v. de ‘Weging van Zorg’. Enerzijds zou de validiteit van zowel SDQ/SPsy verder onderzocht kunnen worden, waarbij het probleem van de overlap met de ‘Weging van Zorg’ wordt opgelost. Anderzijds zou gekozen kunnen worden voor een ander, (deels) al gevalideerde vragenlijst. Voor het signaleren van psychische problemen zijn meerdere instrumenten beschikbaar, waarvan de betrouwbaarheid en validiteit onderzocht zijn en die tevens door de COTAN beoordeeld zijn (bijvoorbeeld de Sociaal-Emotionele Vragenlijst). Echter, veel van deze vragenlijsten zijn niet gevalideerd voor de groep delinquente jongeren. In dat geval zou gekozen kunnen worden voor een door de COTAN beoordeeld instrument dat het dichtst in de buurt komt van de SDQ en/of SPsy en dat al goed en uitgebreid is onderzocht. Het onderzoek bij de originele doelgroep zou dan gerepliceerd moeten worden voor de doelgroep delinquente jongeren.



## 8. Tot Slot

Tot slot zullen met het oog op het streven naar een beoordeling van het LIJ door de COTAN per beoordelingscriterium een aantal adviezen en voorwaarden die noodzakelijk zijn voor vervolgonderzoek worden beschreven. Hierbij wordt benadrukt dat onderdelen waarover in dit rapport geen advies gegeven wordt niet automatisch voldoen aan de COTAN criteria.

### 1. Uitgangspunten van de testconstructie

Met name het onderdeel gebruiksdoel en meetpretentie zal duidelijker omschreven moeten worden. Het was voor de onderzoekers en voor de leden van de begeleidingscommissie te vaak onduidelijk wat nu precies het doel en de functie van het LIJ zijn en als wat voor soort instrument het beschouwd moet worden.

### 2. Kwaliteit van het testmateriaal

In dit onderzoek is aangenomen dat de scoring en weging van het LIJ goed onderzocht en onderbouwd is, omdat het onderzoeken van het scoringsysteem buiten dit onderzoek valt. Echter, gedurende het onderzoek viel op dat er veel vraagtekens zijn rondom de gebruikte scoring en weging en dat deze van cruciaal belang zijn voor het daadwerkelijk uitvoeren van een mogelijk onderzoek naar de validiteit van het LIJ. Het gaat hier specifiek om de weging van verschillende antwoordmogelijkheden bij de items van de 10 domeinen die meetellen in het DRP (dus de scoring). De totstandkoming van en verantwoording voor deze weging moet duidelijk op papier komen en beschikbaar worden gemaakt voor iedereen die met het LIJ werkt. De informatie die op dit moment in de handleiding van het LIJ<sup>7</sup> staat is te summier en onvolledig. Zo is bijvoorbeeld niet duidelijk hoe men gekomen is tot de verschillende puntentoekenning per antwoordmogelijkheid. Voor een mogelijke beoordeling door de COTAN moet per item onder andere verantwoord worden hoe gekomen is tot de puntentoekenning, welke (representatieve) steekproef daarvoor gebruikt is en hoe men tot deze steekproef gekomen is om op die manier te kunnen beoordelen of de scoring op een eerlijke en betrouwbare manier uitgevoerd wordt.

---

<sup>7</sup> p. 11-13, sectie 4.5 *Weging van verschillende antwoordmogelijkheden* in Hoofdstuk 4 'Instrument 2A en 2B: Achtergrond en algemene instructie'

### 3. Kwaliteit van de handleiding

Het is zeer aan te bevelen om de veelheid aan rapporten m.b.t het LIJ terug te brengen naar twee handleidingen; één gebruikershandleiding m.b.t. afname, scoring en interpretatie en één ‘technische’ handleiding m.b.t. onderzoeksinformatie omtrent het LIJ.

### 4. Normen

In de handleiding van het LIJ<sup>8</sup> wordt uitgelegd hoe m.b.v. CHAID-analyses de grenswaarden voor de domeinscores bepaald zijn. Deze informatie moet uitgebreid beschreven worden en tevens per domein beschikbaar zijn. Er moet ook verantwoord worden waarom de uitkomsten van deze analyses gebruikt zouden kunnen worden als normen.

### 5. Betrouwbaarheid

Het onderzoek naar de inter- en intrabeoordelaarsbetrouwbaarheid, zoals uitgevoerd door Regioplan (Timmermans & Witvliet, 2011) zal uitgebreid moeten worden met meer casussen om een nauwkeuriger beeld te krijgen van deze vorm van betrouwbaarheid.

### 6. Begripsvaliditeit

De conclusies en aanbevelingen van dit rapport kunnen gebruikt worden om relevante informatie te verzamelen omtrent begripsvaliditeit. Samengevat dient het onderdeel m.b.t (het signaleren van) psychische problemen (‘Weging van Zorg’ en SDQ/SPsy) verder te worden onderzocht dan wel te worden vervangen door een ander instrument dat psychische problemen meet. M.b.t. het LIJ als geheel kan onderzocht worden of de doelen (met name screening en evaluatie) worden bereikt (zie hiervoor het volgende criterium *Criteriumvaliditeit*).

---

<sup>8</sup> p. 13-14, sectie 4.6 *Grenswaarden en domeinscores* in Hoofdstuk 4 ‘Instrument 2A en 2B: Achtergrond en algemene instructie’

## 7. Criteriumvaliditeit

Er zal eerst duidelijk in kaart moeten worden gebracht wat de doelen van het LIJ zijn en wat ermee voorspeld zou kunnen worden. Het is aan te bevelen om van tevoren goed na te denken over het juiste onderzoeksdesign binnen de mogelijkheden die er zijn om onderzoek naar de criteriumvaliditeit uit te voeren. Wanneer bepaalde 'standaardmethoden' van onderzoek binnen het werkveld zeer moeilijk uitvoerbaar zijn, moet goed verantwoord en beschreven worden waarom gekozen is voor een andere methode.

Tevens wordt hier nogmaals benadrukt dat de COTAN verlangt dat bij verschillende versies (ook als het om een verkorte versie gaat) en/of verschillende respondenttypen afzonderlijke normen en betrouwbaarheid worden gegeven. Ook de correlatie tussen de versies van de verschillende respondenttypen moet worden vermeld. Bij ander validiteitsonderzoek is het minder duidelijk of aparte gegevens vereist zijn. Dit geldt dus voor Halt SI, Instrument 2A IVS, Instrument 2A, Instrument 2B en voor de versies voor de jeugdige zelf, de ouder(s) en de mentor/leerkracht.



## Literatuur

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Boonekamp, T.J. (2010). *Een onderzoek naar de betrouwbaarheid en validiteit van de SDQ*. Masterscriptie Forensische Orthopedagogiek, Faculteit der Maatschappij- en Gedragswetenschappen; Pedagogische en Onderwijskundige Wetenschappen, Universiteit van Amsterdam.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061-1071.
- Bosker, J. (2009). Gestructureerd beslissen over reclasseringsinterventies. *Proces*, *88*, 169-173.
- Brand, E. F. J. M., & van den Hurk, A. A. (2008). *10 Jaargangen PIJ-ers: Kenmerken en veranderingen*. Den Haag: Dienst Justitiële Inrichtingen.
- Brand, E. F. J. M., & van Emmerik, J. L. (2006). *Handleiding FP40. Forensische Profiel Lijsten*. Den Haag: Dienst Justitiële Inrichtingen.
- Brand, E. F. J. M., & van Heerde, W. K. (2004). *Handleiding FPJ-lijst. Forensisch Profiel Justitiële Jeugdigen*. Den Haag: Dienst Justitiële Inrichtingen.
- Brand, E. F. J. M., & van Heerde, W. K. (2010). *Handleiding FPJ*. Den Haag: Dienst Justitiële Inrichtingen.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Evers, A., Braak, M. S. L., Frima, R. M., & van Vliet-Mulder, J. C. (2009-2011). *COTAN Documentatie*. Amsterdam: Boom test uitgevers
- Evers, A., Lucassen, W. I., Meijer, R. R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests (gewijzigde herdruk)*. Amsterdam: NIP
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*, 534-539.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1337-1345.

- Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5<sup>th</sup> ed.). Boston: Pearson Education, Inc.
- Lodewijks, H. P. B. (2008). *Violence risk assessment in adolescents in the Dutch juvenile justice system: Studies on the reliability and predictive accuracy of the SAVRY*. Proefschrift, Vrije Universiteit Amsterdam.
- Lodewijks, H. P. B., de Ruiter, C., & Doreleijers, T. (2003). Risicotaxatie en risicohantering van gewelddadig gedrag bij adolescenten. *Tijdschrift voor Directieve Therapie*, 1, 25-42.
- Lodewijks, H. P. B., Doreleijers, T. A. H., de Ruiter, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, 31, 263-271.
- Lodewijks, H. P. B., Doreleijers, T., de Ruiter, C., & de Wit-Grouls, H. (2001). *Gestructureerde taxatie van geweldsrisico bij jongeren (geautoriseerde vertaling en onderzoeksversie van de SAVRY)*. Eefde: Rentray.
- Reynolds, C. R. (2010). Measurement and assessment: An editorial view. *Psychological Assessment*, 22, 1-4.
- ter Laak, J. J. F. & de Goede, M. P. M. (2003). *Psychologische diagnostiek. Inhoudelijke en methodologische grondslagen*. Lisse: Swets & Zeitlinger.
- Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, 12, 1-8.
- Muris, P., Meesters, C., Eijkelenboom, A., & Vincken, M. (2004). The self-report version of the Strengths and Difficulties Questionnaire: Its psychometric properties in 8- to 13-year-old non-clinical children. *British Journal of Clinical Psychology*, 43, 437-448.
- Oosterlaan, J., Scheres, A., Antrop, I., Roeyers, H. & Sergeant, J. A. (2000). *Handleiding bij de Vragenlijst voor Gedragsproblemen bij Kinderen, VvGK*. Lisse: Harcourt Test Publishers.
- Scholte, E. M. (1996). *De Vragenlijst Sociale en Pedagogische Situatie. Handleiding bij het gebruik van de VSPS Versie 4.0*. Delft: Eburon.
- Scholte, E. M. & van der Ploeg, J. D. (2005a). *ADHD Vragenlijst, AVL, Handleiding*. Houten: Bohn Stafleu Van Loghum.
- Scholte, E. M. & van der Ploeg, J. D. (2005b). *SEV. Sociaal-Emotionele Vragenlijst*. Houten: Bohn Stafleu Van Loghum.



- Timmermans, M., & Witvliet, M. (2011). *Screenen en signaleren in de jeugdstrafrechtketen: Een onderzoek naar de validiteit van psychisch disfunctioneren en de betrouwbaarheid van het Landelijk Instrumentarium Jeugdstrafrechtketen (LIJ)*. Amsterdam: Regioplan Beleidsonderzoek.
- van der Knaap, L. M., Leenarts, L. E. W., & Nijssen, L. T. J. (2007). *Psychometrische kwaliteiten van de Recidive Inschattingsschalen (RISc): Interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit*. Den Haag: WODC. Cahier 2007-05.
- van der Knaap, L. M., & Alberda, D. L. (2009). *De predictieve validiteit van de Recidive Inschattingsschalen (RISc)*. Den Haag: WODC. Cahier 2009-12.
- van Oort, M., van 't Land, H., & de Ruiters, C. (2005). *Screeningsinstrument psychische stoornissen voor Bureau Jeugdzorg: Schriftelijk of mondeling?* Utrecht: Trimbos instituut.
- van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, 12, 281-289.
- Verhulst, F. C., van der Ende, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4-18*. Rotterdam: EUR/AZR/Sophia Kinderziekenhuis, afd. Kinder- en Jeugdpsychiatrie.
- Verhulst, F. C., van der Ende, J., & Koot, H. M. (1997a). *Handleiding voor de Teachers's Report Form (TRF)*. Rotterdam: afd. Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/ AZR/ EUR.
- Verhulst, F.C., van der Ende, J., & Koot, H.M. (1997b). *Handleiding voor de Youth Self-Report (YSR)*. Rotterdam: afd. Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/ AZR/ EUR.
- Vogelvang, B., Persoon, A., & Sondejker, F. (2007) *Landelijk Kader Instrumentarium Jeugdstrafrecht. Fase 2: Beoordeling en aanpassing van het instrumentarium en van de informatieoverdracht in de jeugdstrafketen*. Woerden: Adviesbureau Van Montfoort.
- Werkgroep risicotaxatie forensische psychiatrie (2003). *Handleiding HKT-30 versie 2002: Risicotaxatie in de forensische psychiatrie*. Den Haag: Ministerie van Justitie, Dienst Justitiële Inrichtingen.



