Interbeoordelaarbetrouwbaarheid Standaard Taxatie Ernst Problematiek (STEP)

Karin Eijgenraam
Tom van Yperen
Machteld van der Pijll
Lianne Lekkerkerker
Femke Post
Marian de Graaf

Nederlands Jeugdinstituut / NJi
Utrecht, april 2008

Opdrachtgevers
Ministerie van Volksgezondheid, Welzijn en Sport (VWS), Ministerie voor Jeugd en Gezin en Ministerie van Justitie

Begeleidingscommissie
Wilma Vollebergh (voorzitter, Universiteit Utrecht)
Harry van den Bosch (MOgroep)
Corine Brekelmans (GGZ Nederland)
Els Deijkers (Ministerie van Justitie)
Ton Eijken (Ministerie van Justitie)
Nien Karelsen (cliëntenorganisatie)
Ivonne Keuzenkamp (Interprovinciaal Overleg/IPO)
Hans Matthaei (bureau jeugdzorg Agglomeratie Amsterdam/BJAA)
Wim Slot (PI Research)

Auteurs
Karin Eijgenraam
Tom van Yperen
Machteld van der Pijll
Lianne Lekkerkerker
Femke Post
Marian de Graaf

Nederlands Jeugdinstituut / NJi
Postbus 19221 (Catharijnesingel 47), 3501 DE Utrecht
Telefoon (030) 230 63 44
Fax (030) 230 63 12
Website www.nederlandsjeugdinstituut.nl

Inhoudsopgave

Managementsamenvatting

De STEP - In de periode 2001-2003 is door NIZW Jeugd (nu NJi) in opdracht van het Regionaal Orgaan Amsterdam (ROA) gewerkt aan de constructie van de zogeheten Standaard Taxatie Ernst van de Problematiek (STEP). Dit instrument bestaat uit zes schalen, waarbij de eerste vier de zogeheten *QUICK*STEP vormen: Functioneren Jeugdige (STEP-FJ), Kwaliteit Omgeving (STEP-KO), Zwaarte Zorg (STEP-ZZ), Urgentie Zorg (STEP-UZ), Risico Jeugdige (STEP-RJ), Risico Omgeving (STEP-RO). Per cliënt kost het invullen van het instrument (na enige ervaring) ongeveer vijf minuten. Uit een eerste proef in 2003 blijken de conceptuele indeling, interne consistentie, dekking en hanteerbaarheid van de eerste vier schalen voldoende tot goed.

Doel van dit onderzoek - In 2006 is een tweede onderzoekstraject gestart naar het gebruik van de STEP in de jeugdbescherming en de jeugdreclassering, de interbeoordelaar-betrouwbaarheid, de voorspellende en evaluatieve waarde en de trefzekerheid van de STEP. In dit rapport staat het onderzoek naar de interbeoordelaarbetrouwbaarheid centraal. Doel is na te gaan of de scoring op het instrument niet te zeer afhankelijk is van de hulpverlener die het invult. Indien er wel sprake is van die afhankelijkheid, dan zal het onderzoek inzicht moeten verschaffen in de punten waarop de STEP aangepast dient te worden om het instrument te verbeteren.

Onderzoeksopzet - Het onderzoek is volgens een geaccepteerde vorm opgezet. Er is een casusboek samengesteld met twintig gevalsbeschrijvingen, met verschillende juridische kaders van de hulp. Gelet is op een voldoende spreiding wat betreft de leeftijd, sekse en de mate van ernst op de schalen Functionering Jeugdige en Kwaliteit Omgeving. De vignetten zijn in een voorstudie op bruikbaarheid getoetst. Daarna hebben 80 hulpverleners elk vier vignetten beoordeeld. Voor elk vignet is door zestien hulpverleners een STEP ingevuld. Dat resulteerde in totaal tot 320 beoordelingen.

Resultaten - Op de schalen Functioneren Jeugdige, Kwaliteit Omgeving, Zwaarte Zorg, Risico Jeugdige en Risico Omgeving is er een voldoende mate van overeenstemming tussen de beoordelaars . Alleen op de schaal Urgentie Zorg is er matige overeenstemming. Een nadere analyse op itemniveau levert belangrijke aanwijzingen waarop de handleiding van de STEP te verbeteren valt, zodat gebrekkige overeenstemming minder zal voorkomen. Dit zal de betrouwbaarheid van de genoemde schalen verder doen stijgen. De betrouwbaarheid is ook per vignet bepaald. Met uitzondering van één vignet geldt dat de overeenstemming op schaalniveau goed is. Verder blijkt dat 20 codeurs (25%) een redelijk en 48 codeurs (60%) een voldoende tot goed betrouwbaarheidsniveau behalen.

Conclusies - Dit onderzoek laat zien dat de interbeoordelaarbetrouwbaarheid van de STEP voldoende is. Dit houdt in dat de scoring op het instrument niet te zeer afhankelijk is van de hulpverlener die het invult. Indien de handleiding van de STEP op een aantal punten verbeterd wordt, is de interbeoordelaarbetrouwbaarheid van de STEP nog verder te vergroten. Voorts is training en geregelde intervisie voor een goed hanteren van het instrument van belang.

1 Inleiding

1.1 Het project 'Ernst van de problematiek'

Hulpverleners beoordelen dagelijks de problematiek van jeugdigen en gezinnen. Om hen daarbij handvatten te geven en om meer inzicht te krijgen in de groep cliënten die zij bereiken, is een goede inschatting van de ernst van de problematiek een vereiste. Dat geldt ook als men wil bepalen of jeugdzorg effectief is: als het doel van de zorg is de problemen te verminderen, is een goed meetinstrument nodig om te meten of de ernst van de problematiek daadwerkelijk is afgenomen.

In de periode 2001-2003 is door NIZW Jeugd[1] in opdracht van het Regionaal Orgaan Amsterdam (ROA) gewerkt aan de constructie van de zogeheten Standaard Taxatie Ernst van de Problematiek (STEP) in het project *Registratie Ernst van de Problematiek* (Van Yperen, Van den Berg & Eijgenraam, 2002, 2003a, 2003b, 2003c). Dat project beoogde het volgende te bereiken:

- Er is een inhoudelijk kader voor de taxatie van de ernst van de problematiek door de hulpverlener van bureau jeugdzorg.
- Er is met de praktijk een inhoudelijk voorstel gemaakt voor de inrichting van een dergelijke taxatie.
- Het inhoudelijk voorstel is aan een eerste kwaliteitstoets onderworpen.

Voor de formulering van het inhoudelijk kader is in een eerste deelproject een uitgebreide literatuurstudie en praktijkoriëntatie verricht. Gezocht is naar een verbinding met bestaande theoretische kaders en instrumenten. Er zijn hieruit drie doelen naar voren gekomen die de taxatie voornamelijk moet dienen:

1. Het helder maken van het probleem op individueel niveau of een inschatting kunnen maken van hoe erg de problematiek is, onder meer om dit te kunnen vergelijken met andere instellingen.
2. Een inschatting kunnen maken van welke hulp er nodig is, inhoudelijk en in financieel opzicht.
3. Het kunnen evalueren van de hulp in termen van veranderingen in de ernst van de problematiek.

Op basis van een studie is een eerste opzet gemaakt van de zogeheten Standaard Taxatie Ernst Problematiek (STEP). Het instrument bestaat uit zes schalen die ingevuld kunnen worden zodra er voldoende informatie over de cliënt verzameld is[2].
De bureaus jeugdzorg in de agglomeratie Amsterdam en in Gouda hebben met deze eerste opzet van het instrument een proef uitgevoerd. Deze proef moest gegevens opleveren over de

[1] Vanaf 1 januari 2007 vormt NIZW Jeugd samen met de afdeling Jeugd van het NIZW International Centre het Nederlands Jeugdinstituut/NJi.

[2] De handleiding en het STEP-formulier zijn te downloaden van www.nji.nl (Producten > Publicaties bestellen/ downloaden > Jeugdzorg > Kwaliteit en effectiviteit). Het formulier is in dit rapport opgenomen als bijlage 1.

psychometrische kwaliteit van de STEP (met name de dekking en de interne consistentie) en er moest duidelijk worden wat de hanteerbaarheid van de STEP is in de dagelijkse praktijk.

Niet alle kwaliteitsaspecten zijn aan de orde gekomen in het onderzoek naar deze eerste proef. Vandaar dat in 2006 een tweede traject is gestart waarin onderzoek gedaan wordt naar het gebruik van de STEP in de jeugdbescherming en de jeugdreclassering, naar de voorspellende en evaluatieve waarde van het instrument en naar de interbeoordelaarbetrouwbaarheid van de STEP. Over het laatste aspect, de interbeoordelaarbetrouwbaarheid, doet dit rapport verslag. Alvorens op de studie naar dit kwaliteitsaspect in te gaan, staan we in dit hoofdstuk eerst stil bij de theoretische achtergrond van de STEP, de opbouw van het instrument en de relatie met andere vragenlijsten, de resultaten uit eerder onderzoek naar de kwaliteit en de geplande vervolgonderzoeken.

1.2 Het concept ernst

De literatuurstudie en de praktijkoriëntatie uit het project *Registratie Ernst van de Problematiek* lieten zien dat er verschillende definities bestaan van het begrip 'ernst'. In plaats van eindeloos op zoek te gaan naar de ultieme definitie van ernst is er gezocht naar kenmerken waarvan het idee bestaat dat die relevant zijn voor het concept[3]. Op basis daarvan is een theoretisch model ontwikkeld dat het fundament vormt voor de STEP.

Een begrip met verschillende facetten

Er zijn vier facetten gevonden die van belang zijn bij de ernst van de problematiek:

- *Abnormaliteit gedrag*: dit heeft betrekking op de mate waarin het gedrag afwijkt van wat als normaal wordt beschouwd;
- *Bijdragende factoren* in de jeugdige, gezin, opvoeding en wijdere omgeving: zoals risicofactoren die het probleem verzwaren en protectieve factoren die het probleem verlichten[4];
- *Gevolgen probleemgedrag*: zoals de lijdensdruk voor de jeugdige, gevolgen voor de jeugdige zelf en anderen (gezin en omgeving)
- *Kwaliteit van leven*: het algemene welbevinden, bepaald door objectieve indicatoren en subjectieve waardering van lichamelijk, materieel, sociaal en emotioneel welbevinden, alsmede van de ervaren competentie en dit alles gewogen aan de hand van de waarden die de persoon in kwestie erop na houdt (Felce & Perry, 1996).

In Figuur 1 op pagina 7 zijn de eerste drie facetten weergegeven.

3 Een uitgebreide beschrijving is opgenomen in het eerste deelrapport van het project *Registratie Ernst van de Problematiek* (Van Yperen e.a., 2002).Te downloaden van www.nji.nl.

4 Risicofactoren zijn factoren die bedreigend zijn voor de ontwikkeling van de jeugdige, protectieve factoren zijn factoren die de invloed van risicofactoren kunnen beperken.

Figuur 1. Beoordelingscriteria ernst van de problematiek Pelzer, Steerneman & De Bruyn (1999), aangevuld met factoren van andere auteurs (Bakker, 1999; Groenendaal & Van Yperen, 1994).

Abnormaliteit gedrag	Bijdragende factoren	Gevolgen probleemgedrag
▪ niet passend bij leeftijd ▪ niet passend bij sekse ▪ lange duur ▪ uitgebreidheid over de situaties van functioneren ▪ specificiteit van de symptomatologie ▪ frequentie optreden probleem ▪ gedragsverandering ▪ niet passend bij de socio-culturele context ▪ niet passend bij de levensomstandigheden	*Risicofactoren:* ▪ biologische kwetsbaarheid jeugdige ▪ pathogene gezinsrelaties ▪ incompetent opvoedingsklimaat ▪ factoren in wijdere omgeving, bijv. wonen in een achterstandsbuurt *Protectieve factoren:* ▪ jeugdige: sociale en probleemoplossende vaardigheden, gevoel voor humor, goede intelligentie ▪ gezin: goede relatie jeugdige-ouder(s); opvoedend handelen dat wordt gekenmerkt door warmte, disciplinering, responsiviteit en sensitiviteit ▪ sociale relaties wijdere omgeving: steunend netwerk, positieve schoolervaringen, goede relaties met leeftijdgenoten en leerkrachten ▪ sociaal-maatschappelijk: goede voorzieningen, goede woonomgeving, werkgelegenheid	▪ lijdensdruk jeugdige ▪ sociale belemmering jeugdige ▪ ontwikkeling jeugdige ▪ gevolgen voor anderen ▪ gevolgen voor behandeling

Theoretisch werkmodel

Nadat de verschillende facetten van het begrip 'ernst' op een rij zijn gezet, is er een theoretisch werkmodel uitgewerkt waarin deze verschillende bijdragen zijn gebruikt. Figuur 2 hieronder geeft dit werkmodel weer.

Figuur 2. 'Ernst van de problematiek': een theoretisch werkmodel

Historische en actuele problemen en risico's	Historische en actuele protecties
A. Gewicht van de stressfactoren en risicofactoren ('Draaglast') gelegen in jeugdige, opvoeding, gezin en wijdere omgeving • Abnormaliteit van de problematiek - Specificiteit, intensiteit, frequentie, duur - Passendheid bij leeftijd, ontwikkelingsstadium, sekse, context - Aantal terreinen / accumulatie van stressfactoren gelegen in jeugdige, opvoeding en gezin, wijdere omgeving • (Bijkomende) risicofactoren gelegen in jeugdige, opvoeding en gezin, wijdere omgeving **Taxatie ernst is:** Taxatie van de *zwaarte van de problematiek*, door een weging van de draaglast ten opzichte van draagkracht.	**B. Gewicht van de protectieve factoren ('Draagkracht') gelegen in jeugdige, opvoeding, gezin en wijdere omgeving** • Aard van deze factoren - Specificiteit, intensiteit, frequentie, duur - Type risicofactor die in werking geremd wordt - Aantal terreinen / accumulatie van protectieve factoren gelegen in jeugdige, opvoeding en gezin, wijdere omgeving • (Bijkomende) protectieve factoren gelegen in jeugdige, opvoeding en gezin, wijdere omgeving **Taxatie ernst is**: zie A.
C. Gewicht van negatieve gevolgen van een overgewicht aan draaglast of een tekort aan draagkracht • Historisch, actueel en verwacht: • Ervaren beperking in kwaliteit van leven jeugdige • 'Objectieve' sociale belemmeringen • 'Objectieve' risico's voor verdere ontwikkeling • Negatieve gevolgen voor anderen (gezin, samenleving) • Gevolgen voor behandeling (moeilijker, minder effect) **Taxatie ernst is:** Taxatie van de *gevolgen* van de negatieve onbalans.	**D. Mobiliserende, compenserende reacties op de 'negatieve onbalans'** • Hulpzoekgedrag • Mate van legitimering van zorg met een bepaalde urgentie, ingrijpendheid, duur, intensiteit ('zorgzwaarte'), ter vermindering van draaglast of versterking van draagkracht • Mate van legitimering voor civiele of strafrechtelijke maatregelen van de samenleving (jeugdbescherming, detentie) **Taxatie ernst is:** Taxatie van de *zorgzwaarte*.

Dit model gaat uit van enerzijds problemen en risicofactoren die 'druk' uitoefenen op de ontwikkeling van de jeugdige. Onder deze factoren rekenen we problemen of risicofactoren in de jeugdige, het gezin en de omgeving. Het gaat hier zowel om actueel aanwezige factoren als om factoren die in historisch perspectief moeten meewegen. Anderzijds zijn er protectieve factoren (in de jeugdige, het gezin, de omgeving, actueel of in historisch perspectief) die in geval van de aanwezigheid van bepaalde risicofactoren 'verlichting' kunnen geven of compenserend kunnen werken. Verstoring van de balans (zwaarder gewicht van aanwezige problemen en risicofactoren ten opzichte van het gewicht van aanwezige protectieve factoren) uit zich onder meer in lijden, sociale belemmeringen, belemmeringen in de verdere ontwikkeling en gevolgen voor anderen. De onbalans moet zowel in de historische, actuele als prognostische betekenis worden beschouwd. Het model is voor elke afzonderlijke informant in te vullen (vanuit het perspectief van de jeugdige, de ouder, de hulpverlener, enz.).
Dit theoretische model vormt de basis voor de constructie van de STEP.

1.3 STEP en scoringshulpen

De Standaard Taxatie Ernst Problematiek (STEP) is een korte vragenlijst waarmee de medewerker zicht krijgt op de ernst van de problematiek van cliënten in de jeugdzorg (Van Yperen, Van den Berg, Eijgenraam & De Graaf, 2006). De STEP is níet bedoeld om informatie te verzamelen; met de STEP kan reeds verzamelde informatie worden gebruikt om een ernsttaxatie te doen. Omdat er al redelijk wat informatie verzameld moet zijn om de STEP te kunnen invullen, kan dit instrument niet heel vroeg in het proces worden gebruikt.
De STEP bestaat uit 21 vragen, verdeeld over zes schalen, waarop verschillende aspecten van ernst van de problematiek van een jeugdige gescoord kunnen worden:

1. Functioneren Jeugdige (STEP-FJ);
2. Kwaliteit Omgeving (STEP-KO);
3. Zwaarte Zorg (STEP-ZZ);
4. Urgentie Zorg (STEP-UZ);
5. Risico Jeugdige (STEP-RJ);
6. Risico Omgeving (STEP-RO).

Voor alle zes schalen is een scoringshulp ontwikkeld. Bij de schalen Functioneren Jeugdige en Kwaliteit Omgeving geeft de medewerker aan de hand van een aantal gedetailleerd omschreven criteria een ernsttaxatie per schaal. De criteria zijn grotendeels ontleend aan andere instrumenten (die vaak al uitgebreid getoetst zijn) en samengebracht tot een handzaam instrument: de Child Global Assessment Schedule (CGAS), onderdelen van de Strengths and Difficulties Questionnaire (SDQ), de Child and Adolescent Functional Assessment Scale (CAFAS), de criteria kindermishandeling van Willems, de Landelijke Heerlense Ernst Taxatie Schaal (L-HETS), de Vragenlijst Gezinsproblemen (VGP) en de Gezinsdimensieschalen (GDS)[5]. Met de twee risicoschalen wordt in beeld gebracht of de problematiek zal verergeren, gelijk blijven of verminderen bij het uitblijven van hulp.
De *QUICK*STEP is een verkorte versie van de STEP: alleen de eerste vier schalen worden ingevuld, de twee risicoschalen (STEP-RJ en STEP-RO) worden dan dus achterwege gelaten.
De scores op de vier (*QUICK*STEP) of zes (STEP) schalen worden overzichtelijk op een ernstprofiel weergegeven, zodat de medewerker snel een indruk krijgt van de ernst van de gezinssituatie. Een ervaren gebruiker heeft ongeveer vijf minuten nodig om de STEP in te vullen.

5 Voor bronnen zie Van Yperen, Van den Berg & Eijgenraam (2002).

De ernsttaxatie gaat uit van een gebruik van instrumenten op drie niveaus.

- Het eerste *niveau* gaat uit van een zeer globale taxatie door de hulpverlener op zes aandachtspunten: Functioneren Jeugdige (FJ); Kwaliteit Omgeving (KO); Risico Jeugdige (RJ); Risico Omgeving (RO); Zwaarte Zorg (ZZ); Urgentie Zorg (UZ). Voor elk aandachtspunt is een schaal geconstrueerd, variërend van 1=(Erg) goed functioneren tot 5=Zware tot extreme problemen in functioneren.
- Het *tweede niveau* biedt verdere operationalisaties van de schalen. Middelen die hier passen zijn bijvoorbeeld de L-HETS, de SDQ, de CBCL, de CAFAS, de GAF-schalen, de JIM-familie, de SAM, de VSPS en de RED. De bureaus jeugdzorg blijken deze instrumenten nauwelijks te gebruiken. Daarom is voor de STEP een simpele - 21 items tellende - scoringshulp gemaakt om de scoring op de zes schalen te kunnen bepalen. Voor de hantering van de STEP en de scoringshulp is een uitgebreide handleiding gemaakt.
- Het *derde niveau* beslaat een brede variëteit aan specifieke tests en vragenlijsten die voor de praktijk beschikbaar is (bijvoorbeeld NOSI, CBCL, SAS-K etc.) en in principe aan de STEP-standaard is te koppelen. Dat niveau viel buiten het bestek van het project *Registratie Ernst van de problematiek*.

1.4 Eerste kwaliteitstoets

De allereerste versies van de STEP zijn in 2003 in kleinschalige praktijkproeven op bruikbaarheid getoetst. In de laatste fase van het project is het onderhavige gebruiksonderzoek uitgevoerd. Daaruit bleek onder meer dat er op de werkvloer zorgen waren over de omvang van het instrument (6 schalen, totaal 21 items), die men te groot vond. De verdere kwaliteitstoets zou daarom onder meer uitwijzen of het aantal items van de scoringshulp wellicht nog kleiner kan. In een daarop volgend grootschaliger onderzoek is voor 551 jeugdigen en gezinnen de STEP ingevuld (Van Yperen, Van den Berg & Eijgenraam 2003b). Er deden 109 codeurs aan het onderzoek mee. De meeste daarvan vulden de STEP voor één tot vier cases in. Een kleine groep codeurs (13 personen) kwam tot een routinematig gebruik van het instrument. De onderzoeksgroep bestond uit de gebruikelijke clientèle van bureau jeugdzorg, zij het dat het aantal jeugdigen dat hulp in een niet-vrijwillig kader ontvangt klein is als gevolg van het moment waarop de STEP moest worden ingevuld. Van de opgestuurde formulieren bleken er 535 bruikbaar voor de statistische analyses.

Belangrijkste resultaten eerste kwaliteitstoets

Spreiding
Als eerste is onderzocht of de items en de schalen voldoende spreiding in de scores vertonen, zodat het instrument in staat is om cases te differentiëren. Op basis van deze analyse is de conclusie dat met name de schaal Risico Omgeving (RO) en de bijbehorende items ongunstige verdelingen laten zien: de scores hopen zich teveel aan één kant op. Dit vormt een aanwijzing dat de schaal verwijderd of verbouwd moet worden. Voor het overige laat het instrument voldoende tot goede verdelingen zien.

Interne consistentie en redundantie
Gekeken is of de schalen van het instrument en items in elke afzonderlijke schaal ook werkelijk bij elkaar horen (betrouwbaarheid in termen van interne consistentie) en of bepaalde items wellicht te weinig toevoegen (redundantie). Dat is onderzocht via een principale componentenanalyse en berekening van de alpha's van de schalen. De schalen en de daarbij behorende items blijken een goede interne consistentie te vertonen. De twee

risicoschalen blijken echter bijzonder weinig toe te voegen aan de schalen voor Functioneren Jeugdige (FJ) en Kwaliteit Omgeving (KO). Op basis van analyses is een verkorte versie van de STEP gemaakt (de zogeheten *QUICK*STEP), die vijftien items bevat, verdeeld over vier schalen:

- Functioneren Jeugdige (FJ; 6 items; alpha=.78, voldoende betrouwbaarheid);
- Kwaliteit Omgeving (KO; 5 items; alpha=.81, goede betrouwbaarheid);
- Zwaarte Zorg (ZZ; 3 items; alpha=.72, voldoende betrouwbaarheid);
- Urgentie Zorg (UZ; 1 item; alpha niet relevant).

Dekking
Bij de dekking gaat het om de vraag of het instrument goed bij de populatie van bureau jeugdzorg past. Uit het onderzoek blijkt dat het instrument over het algemeen goed dekkend is. Niettemin komen een paar problemen naar voren, waarop het instrument en/of de handleiding is aangepast. Voorts geldt als kanttekening dat de STEP voornamelijk getoetst is in de vrijwillige hulpverlening. Er zijn aanwijzingen dat het instrument ook past bij de groep jeugdigen met een OTS. Verdere toetsing van de dekking bij deze groep is echter van belang.

Validiteit
De validiteit betreft de mate waarin het instrument een concept adequaat operationaliseert, trefzeker is en de mogelijkheid biedt om te voorspellen welke zorgzwaarte de cliënten uiteindelijk krijgen. Uit de eerder genoemde principale componentenanalyse is gebleken dat de eerste vier theoretische deelconcepten van het begrip 'ernst' (Functioneren Jeugdige, Kwaliteit Omgeving, Zwaarte Zorg en Urgentie Zorg) ook empirisch op een consistente manier uit het onderzoek naar vormen komen. De andere validiteitsaspecten stonden in het eerste onderzoek niet op de agenda. In het voorhanden zijnde materiaal zijn niettemin aanwijzingen te vinden voor de trefzekerheid van de STEP. Die aanwijzingen vormen echter nog geen basis om de STEP een valide instrument te kunnen noemen. Daar moet nieuw onderzoek licht op werpen.

Hanteerbaarheid
Uit het onderzoek komt het beeld naar voren van een redelijk te hanteren instrument. De gemiddelde tijd om de STEP in te vullen bedraagt in het onderzoek veertien minuten. Naarmate de codeurs het instrument vaker gebruiken, neemt het tijdsbeslag af tot zo'n vijf tot tien minuten. Bij bijna driekwart van de cases vonden de codeurs de STEP gemakkelijk in te vullen. Ook hier geldt: hoe vaker de STEP door een codeur wordt ingevuld, hoe gemakkelijker dat gaat. Een aantal onderdelen van de STEP leverde vaak moeilijkheden op. De hulpverleners hadden wel moeite met de inschatting van het verloop van de problematiek op de twee risicoschalen (RJ en RO).

1.5 Vervolgonderzoek STEP (2006-2008)

Het onderzoek naar de kwaliteit van de STEP is nog niet compleet. Vandaar dat in 2006 een tweede traject is gestart. Daarin wordt onderzoek gedaan naar het gebruik van de STEP in de jeugdbescherming en de jeugdreclassering, de interbeoordelaarbetrouwbaarheid en de voorspellende en evaluatieve waarde van de STEP.

In dit rapport staat het onderzoek naar de interbeoordelaarbetrouwbaarheid van de STEP centraal. De interbeoordelaarbetrouwbaarheid of –overeenstemming gaat over de mate waarin verschillende beoordelaars bij eenzelfde casus met de STEP tot dezelfde ernsttaxatie komen. Is

dat onvoldoende, dan zijn de STEP-scores te afhankelijk van welke hulpverlener het instrument toevallig invult. Het instrument moet dan op dit punt worden verbeterd.
Voor dit deelonderzoek zijn twintig casusbeschrijvingen gemaakt. Verschillende hulpverleners vulden voor dezelfde beschrijvingen de STEP in, waarna de uitkomsten met elkaar zijn vergeleken om te beoordelen of deze voldoende met elkaar overeenkomen.

Naast het onderzoek naar de interbeoordelaarbetrouwbaarheid van de STEP lopen gelijktijdig twee andere onderzoekstrajecten.
Het eerste onderzoekstraject is bedoeld om na te gaan of de STEP bij cliënten van de gezinsvoogdij/voogdij (jeugdbescherming) en de jeugdreclassering andere uitkomsten op de ernst van de problematiek geeft dan bij de andere cliënten van bureau jeugdzorg. Speciaal aandachtspunt is de vraag of het scoreprofiel ook inhoudelijk voldoende dekkend is voor de maatschappelijke opdracht[6] van de jeugdreclassering en de (gezins)voogdij. Wij verwachten dat de STEP bij hulp in een gedwongen kader relatief hogere scores laat zien op de taxatieschalen Risico Jeugdige (RJ) en Risico Omgeving (RO) en op de schaal Kwaliteit van de Omgeving (KO). Het is de vraag of die hogere scores voldoende onderscheidend zijn voor deze groep. Als de resultaten van het onderzoek daartoe aanleiding geven, passen we de STEP aan om de bruikbaarheid van dit instrument in de (gezins)voogdij en de jeugdreclassering te vergroten. Daarnaast voeren we een verkenning uit naar de inzetbaarheid van de STEP bij het AMK en bij de Raad voor de Kinderbescherming.
Het andere onderzoekstraject richt zich in de eerste plaats op de predictieve validiteit van de STEP bij hulp in het vrijwillig kader en bij hulp in het gedwongen kader. Het gaat hierbij om de vraag of de STEP-scores te gebruiken zijn om te voorspellen of überhaupt jeugdzorg nodig is (beoordeling aanmelding bij bureau jeugdzorg). Voorts gaat het om de vraag of de scores kunnen voorspellen of lichte ambulante dan wel geïndiceerde jeugdzorg nodig is. Wanneer deze predictieve validiteit voldoende is, is het instrument goed bruikbaar bij de indicatiestelling door medewerkers van bureau jeugdzorg. Daarnaast kan de vraag worden beantwoord of vanuit STEP-scores een voorspelling kan worden gedaan voor de urgentie van de problematiek. Wie moet eerst worden geholpen wanneer er wachtlijsten bestaan?
Daarnaast gaan we in dit onderzoek na hoe de gevoeligheid van de STEP is voor verandering bij de cliënt gedurende en aan het einde van de hulpverlening (in vrijwillig en in gedwongen kader). Het instrument is erop gebouwd verandering vast te kunnen stellen: de items refereren aan kenmerken van de jeugdige en de omgeving die in principe met jeugdzorg te veranderen zijn (uitgezonderd het item dat betrekking heeft op de bestaansduur van de problematiek). De STEP dient voldoende gevoelig te zijn voor verandering, anders kan het instrument niet voor evaluatieve doeleinden worden ingezet.

In de toekomst zullen ook studies worden uitgevoerd naar de stabiliteit van het instrument (als een hulpverlener na enige tijd een geval opnieuw met de STEP beoordeeld – zonder dat de jeugdige of zijn situatie is veranderd – ziet de scoring er dan hetzelfde uit?) en de trefzekerheid (staat een lage of hoge score op de STEP ook in de werkelijkheid voor een échte late of hoge ernst van de problematiek?). Dit zijn lastige onderzoeken. Bij het meten van de stabiliteit is het nodig gevallen te hebben die niet in de tijd veranderen, zodat veranderingen in scores op de STEP aan het instrument op de codeur toe te schrijven zijn en niet aan de zich ontwikkelende problematiek. De onderhavige studie naar de interbeoordelaarbetrouwbaarheid levert daar mogelijk bruikbaar materiaal voor. Het onderzoek naar de trefzekerheid van de STEP gaat het

6 Voor de gezinsvoogdij is de maatschappelijke opdracht door de kinderrechter verstrekt: het kind beschermen door het realiseren en bestendigen van een aanvaardbare opvoedingssituatie. Voor de jeugdreclassering geldt als maatschappelijke opdracht: het voorkomen dan wel doen afnemen van een criminele carrière bij jongeren van 12-18 jaar.

erom te meten in welke mate het instrument ook 'echte ernstige gevallen' en 'echte niet-ernstige gevallen' terecht als zodanig identificeert (de zogeheten sensitiviteit en specificiteit). Dit vereist dat duidelijk is wat nu die 'echte gevallen' zijn. Zij vormen immers de standaard die de berekening van de sensitiviteit en specificiteit mogelijk moeten maken. Een probleem is nu dat er geen soortgelijke ernsttaxatieinstrumenten zijn waarmee we de scores op de *QUICK*STEP kunnen vergelijken[7]. Ook hier kan het interbeoordelaarbetrouwbaarheidsonderzoek materiaal voor leveren: als het instrument voldoende betrouwbaar blijkt, kunnen de consensusscores worden gebruikt als standaard waarmee de oordelen van nieuwe codeurs te vergelijken zijn.

1.6 De opbouw van dit rapport

Zoals gezegd, biedt dit rapport een verslag van de studie naar de interbeoordelaar-betrouwbaarheid van de STEP. Het rapport is als volgt opgebouwd: in hoofdstuk 2 komt een beschrijving en verantwoording van de onderzoeksopzet aan de orde, in hoofdstuk 3 worden de onderzoeksresultaten gepresenteerd en hoofdstuk 4 tenslotte bestaat uit een samenvatting en conclusies.

[7] Momenteel lopen er een paar studies waarbij STEP-scores worden vergeleken met resultaten op de Strengths and Difficulties Questionnaire (SDQ) en enkele gezinsinstrumenten, waaronder de verkorte Nijmeegse Opvoedings Stress Index (NOSI-K). Een probleem hierbij is dat instrumenten als SDQ en NOSI-K door een ander type informant wordt ingevuld (namelijk de ouders of de jeugdige) dan de STEP (ingevuld door de hulpverlener). Uit allerlei onderzoek is bekend dat typen informanten vaak sterk verschillen in hun oordeel over de problematiek. Een lage overeenkomst tussen de instrumenten hoeft daarom nog niet te wijzen op een lage trefzekerheid.

2 Opzet en uitvoering van het onderzoek

2.1 Doel van het onderzoek

Doel van dit deelonderzoek is na te gaan of de interbeoordelaarbetrouwbaarheid van de STEP hoog genoeg is om te kunnen spreken van een instrument dat op dit punt voldoende betrouwbaar is. In de studie in 2003 is reeds de betrouwbaarheid in termen van interne consistentie getoetst. Het onderhavige onderzoek gaat na in welke mate invullers van de STEP met elkaar overeenstemmen in de scores die ze toekennen bij eenzelfde casus.
Indien de interbeoordelaarbetrouwbaarheid niet hoog genoeg blijkt te zijn, dan zal het onderzoek inzicht moeten verschaffen in de punten waarop de STEP aangepast dient te worden om dat type betrouwbaarheid te verbeteren.

2.2 Verantwoording onderzoeksopzet

Bij de opzet van een interbeoordelaarbetrouwbaarheidsstudie heeft men te maken met vijf variantiebronnen die ertoe kunnen leiden dat invullers verschillende scores toekennen bij een casus (Spitzer, Endicott & Robins, 1975; zie ook Van Yperen, 1990):

A. *Subjectvariantie*. Als twee of meer codeurs elk op achtereenvolgende tijdstippen een cliënt spreken en observeren en daarna de STEP invullen, kan de psychische conditie of de omgeving van de casus tussentijds zijn veranderd. Dat zorgt ervoor dat de codeurs terecht verschillende scores op de STEP geven, omdat het beeld intussen in veranderd.

B. *Situatievariantie*. Bij het observeren of spreken op verschillende tijdstippen kan het probleem van de casus misschien niet zijn veranderd, maar kan het beeld verschillen door fluctuaties in de tijd. Ook dat kan er toe leiden dat er terecht verschillende scores op de STEP worden toegekend.

C. *Informatievariantie*. Codeurs kunnen uit verschillende informatiebronnen geput hebben. Dat levert ongewenste verschillen in scores op, omdat deze feitelijk niets te maken hebben met de conditie waarin de cliënt verkeert.

D. *Observatievariantie*. Codeurs kunnen uit dezelfde informatiebronnen putten, maar verschillen in wat hen opvalt of verschillende accenten leggen. Ook dit leidt tot uiteenlopende scores op de STEP die ongewenst zijn.

E. *Criteriumvariantie*. Codeurs kunnen verschillen in de eigen regels die ze toepassen bij het toekennen van scores.

Een instrument als de STEP is gemaakt om met name variantiebron E uit te schakelen. De regels van de codeurs om scores toe te kennen worden zoveel mogelijk geleid door een handleiding en door score-items. Om te onderzoeken of dat goed gebeurt, is het belangrijk de andere variantiebronnen zoveel mogelijk uit te schakelen. Een gebruikelijke manier om dat te doen is door te werken met vignetten: codeurs krijgen casusbeschrijvingen voorgelegd die ze onafhankelijk van elkaar moeten scoren op het instrument. Langs deze weg worden vooral de variantiebronnen A, B en C uitgeschakeld. Voor variantiebron D (de observatievariantie) is dit veel lastiger. Een manier om te zien of deze bron invloed heeft op de uiteindelijk gemeten

betrouwbaarheid is door typen codeurs te onderscheiden, waarvan het vermoeden is dat deze verschillen in de zaken waar ze op letten. In de literatuur over betrouwbaarheidsonderzoek wordt in dat verband wel onderscheid gemaakt in de mate van ervaring die codeurs in hun vak en/of met het instrument hebben.

In dit deelonderzoek is een opzet gehanteerd die de subject-, informatie- en situatievariantie (A, B en C) uitschakelt, waardoor de resultaten vooral door de invloed van observatievariantie (D) bepaald worden. Daartoe is een casusboek samengesteld met 20 vignetten. Van elk vignet is de leeftijd, sekse, cultuur, leefverband, gezinssituatie, dagbesteding en aard problematiek van het kind en het juridisch kader van de hulpverlening bekend (zie 2.3 voor een verantwoording van de constructie van de vignetten).

Met het oog op de generaliseerbaarheid van de resultaten hebben medewerkers van verschillende bureaus jeugdzorg (Limburg, Zeeland, Friesland en Agglomeratie Amsterdam) en de William Schrikker Groep (een landelijk werkende organisatie voor o.a. LVG-jeugd) deelgenomen. Volgens een geblokt design is aan meer dan tachtig hulpverleners van deze organisaties gevraagd elk vier vignetten met de STEP te scoren. Dit aantal is nodig om redelijk te kunnen spreiden over uiteenlopende kenmerken van de hulpverleners. De meeste hulpverleners hebben een training gevolgd in het gebruik van de STEP[8]. Van elke hulpverlener is de functie, de opleiding, het aantal jaren werkervaring en het aantal jaren ervaring met de STEP bekend. Uit de literatuur is bekend dat met name ervaring een factor is in de mate waarin codeurs met elkaar overeenstemmen. Uiteindelijk zijn gegevens van 80 hulpverleners gebruikt voor de studie. Elke casus is door 16 hulpverleners beoordeeld. Dat heeft geleid tot in totaal 320 beoordelingen over de vignetten[9]. De verzamelde gegevens zijn met behulp van SPSS ingevoerd, gecontroleerd, opgeschoond en geanalyseerd.

Keuze overeenstemmingsmaat

Bij de analyse van de gegevens gaan we uit van de veronderstelling dat de items en de daarop gebaseerde schalen minimaal een ordinaal meetniveau hebben. Gezien het meetniveau en het feit dat er per casus meer dan twee codeurs zijn, geldt de intraclass correlatiecoëfficiënt (ICC) als de meest geschikte en internationaal geaccepteerde overeenstemmingsmaat. Verschillende auteurs hebben laten zien dat de ICC in hoge mate equivalent is aan de zogeheten gewogen Kappa, een andere bekende index voor interbeoordelaarbetrouwbaarheid[10]. Van deze maat bestaan verschillende varianten, waarbij de keuze afhankelijk is van een aantal overwegingen. Hieronder lichten we de belangrijkste keuzes toe, met verwijzing naar de opties die in het gebruikte statistisch verwerkingspakket (SPSS) zijn geselecteerd.

---

8 Voor de invoering van de STEP ten behoeve van dit onderzoek is een trainingsprogramma ontwikkeld van één dagdeel. Het programma bestaat uit een inleiding, een deel informatie en voorlichting over achtergronden, nut en opbouw/inhoud van de STEP, en een deel oefening met de STEP plus nabespreking. De trainingen zijn over het algemeen per team/werksoort gegeven, enkele uitzonderingen daargelaten. Bij één locatie is de training niet gegeven, omdat men daar al langer werkte met de STEP. Analyses laten zien dat de betrouwbaarheid van coderen van deze locatie niet verschilt van de rest.

9 Oorspronkelijk hebben meer hulpverleners aan het onderzoek deelgenomen. Als zij minder dan vier vignetten hebben beoordeeld, zijn zij uit het bestand verwijderd. In een blok bleek er uiteindelijk één codeur teveel te zijn, waardoor enkele vignetten niet door 16, maar door 17 beoordelaars gescoord waren. Om dit te corrigeren is – zonder acht te slaan op de overeenstemming in STEP-scores van de codeur met anderen – een beoordelaar verwijderd waarvan de kenmerken (opleiding en ervaring) leken op die van de meerderheid van de andere codeurs in het bestand. Overigens is uit analyses met verschillende aantallen codeurs gebleken dat dit geen grote verschuivingen in de resultaten van de betrouwbaarheidsstudie tot gevolg heeft.

10 De gewogen Kappa is een variant van de bekende Cohen's Kappa en wordt gebruikt om overeenstemming van schalen met een ordinaal meetniveau te bepalen. De ICC en gewogen Kappa zijn onder bepaalde voorwaarden equivalent (voor een bespreking zie Fleiss & Cohen, 1973 en Van Yperen, 1990).

- *Hoeveel dimensies?* De vraag is hier hoeveel dimensies er in de analyse gelden waarop systematische verschillen te verwachten zijn. Besloten is in eerste instantie er van uit te gaan dat het vignet de enige systematische variantiebron is. De codeurs hebben we – binnen de beperkingen van het geblokte design – verder willekeurig toegewezen, waarbij elk vignet door een beperkt aantal codeurs uit een grotere pool van codeurs is gescoord. In termen van SPSS betekent dit dat in eerste instantie gekozen is voor het 'One-way Random'-model (zie ook Shrout & Fleiss, 1979). Bij de resulterende ICC's is vervolgens achteraf gekeken wat de invloed is van de typen codeurs. Voor de keuzes die in dat verband verder zijn gemaakt, verwijzen we naar de desbetreffende paragraaf.
- *Enkele codeur of gemiddelde scores van codeurs?* De volgende vraag is of bij het gebruik van de STEP in de alledaagse praktijk de score van een enkele codeur van belang is, of dat men gebruik maakt van een gemiddelde score van een aantal codeurs. In ons geval geldt het eerste: er zal één hulpverlener van bureau jeugdzorg zijn die voor 'zijn' cliënt de STEP invoert. Voor die situatie moeten we – in termen van SPSS - kiezen voor de zogheten 'Single'-variant van de ICC[11].
- *Absolute of relatieve overeenstemming?* De ICC is een echte correlatiemaat. Dat houdt in dat een ICC hoog kan zijn, terwijl de scores van de codeurs toch aanzienlijk verschillen. Dit doet zich bijvoorbeeld voor als codeur A op alle items een 3 scoort, codeur B op alle items een 5 en codeur C een 1. Bij dit soort systematische afwijkingen is de correlatie perfect (1.00), terwijl er in de praktijk een fors gebrek aan overeenstemming is. In ons geval vertekent de correlatie dus het beeld van de betrouwbaarheid. Als een codeur op een item bijvoorbeeld een 3 heeft gescoord, willen we weten in welke mate de andere codeurs ook precies deze score hebben gegeven. In termen van SPSS hebben we om die reden gekozen voor de 'Absolute agreement'-variant van ICC. Systematische afwijkingen van die score worden dan – afhankelijk van de hoogte van de andere score – in de berekening van de ICC-index meegewogen als niet-overeenstemming.

Interpretatiekader van de resultaten

Er bestaan geen internationale conventies over de interpretatie van de hoogte van de betrouwbaarheid. Dit hangt mede samen met het feit dat de hoogst haalbare betrouwbaarheid afhankelijk is van de spreiding van scores die door de codeurs zijn toegekend en het type betrouwbaarheidsmaat dat is gebruikt (voor een bespreking zie Van Yperen, 1990). Niettemin is het goed om voorafgaand aan de verzameling van de onderzoeksgegevens een kwalificatieschema op te stellen voor de interpretatie van de index. Het bekendste schema is dat van Landis en Koch (1977) voor de interpretatie van de Cohen's Kappa (zie Tabel 1 op pagina 16). De ICC staat echter bekend als een iets tolerantere index dan de gewone (ongewogen) Cohen's Kappa. Daarom moet het kwalificatieschema van de ICC wat strenger zijn. Tabel 1 geeft aan welke vuistregels we in dit onderzoek hebben opgesteld voor de kwalificatie van de interbeoordelaarbetrouwbaarheid (zie ook Van Yperen, 1990 en 1995; Van Yperen, Roosma & Veerman, in druk). Ter vergelijking is daarbij het gebruikelijke schema voor de Cohen's Kappa toegevoegd.

[11] Het alternatief, de 'Average'-variant, gebruikt men alleen als in de dagelijkse praktijk van bureau jeugdzorg er meerdere hulpverleners zijn die op moment X de STEP invullen en bij de verdere besluitvorming de gemiddelde score van die hulpverleners wordt gebruikt. De ICC die uitgaat van dit model levert dezelfde resultaten op als de Cronbach Alpha, en laat in ons geval doorgaans veel hogere waarden zien dan de 'single'-variant.

Tabel 1. Kwalificatie overeenstemmingsmaat

Kwalificatie	Overeenstemmingsmaat	
	ICC	Cohen's Kappa
Slecht	-1.00 - .30	-1.00 - .20
Matig	.31 - .50	.21 - .40
Voldoende[12]	.51 - .70	.41 - .60
Goed	.71 – 1.00	.61 – .80
Zeer goed		.81 – 1.00

De vraag die in onze studie voorts van belang is, luidt: wanneer mogen we nu tevreden zijn, als we de resultaten vergelijken met andere instrumenten? Met name in de jaren tachtig en negentig is er enig onderzoek gedaan naar de betrouwbaarheid van de oordelen over de aard en de zwaarte van de problematiek. Uiteenlopende studies wijzen erop dat de mate van overeenstemming in de oordelen over de ernst van de problematiek bij professionals nogal varieert. Bij jeugdigen blijkt er sprake van een matige tot voldoende betrouwbaarheid (voor een overzicht, zie Van Yperen, 1995; Kroes, 2006). Weinig onderzoek is er ten aanzien van de overeenstemming over de zwaarte en urgentie van de zorg. Er zijn aanwijzingen dat die overeenstemming over het algemeen slecht tot matig is (zie o.a. Berben, Konijn, Verheij, Donker, Steketee, Roede & De Savorin Lohman, 1997). Voor de waardering van de resultaten is door ons aan het begin van ons onderzoek aan de begeleidingscommissie voorgesteld dat we bij een ICC op de schalen en items van minsten .51 tevreden mogen zijn (i.e. een voldoende overeenstemming).

In een aantal analyses is tevens gekeken naar de 'betrouwbaarheid' van elke codeur. Daarbij is nagegaan in welke mate een codeur in zijn scoringspatronen afwijkt van het gemiddelde van de andere codeurs. Daarvoor is geen ICC gebruikt, maar een andere index. In paragraaf 3.5 volgt daarop een toelichting.

2.3 Verantwoording constructie vignetten

Voor dit deelonderzoek naar de interbeoordelaarbetrouwbaarheid van de STEP is een casusboek samengesteld met 20 gevalsbeschrijvingen. Er is deels gebruikgemaakt van bestaande beschrijvingen, aangevuld met dossiermateriaal van bureaus jeugdzorg. De beschrijvingen zijn bewerkt tot geanonimiseerde en niet tot personen herleidbare vignetten. De vignettenmethode wordt vooral gebruikt om de invloed van bepaalde factoren op een keuze te onderzoeken (Veenma, Batenburg & Breedveld, 2004). Zo kunnen systematische invloeden, zoals subjectvariantie (een casus kan veranderen als codeurs op verschillende tijdstippen observeren) en situatievariantie (de situatie kan veranderen als codeurs op verschillende tijdstippen observeren), met behulp van de vignettenmethode uitgeschakeld worden (Van Yperen, 1990).

De vignettenmethode is eerder gebruikt in sociaalpsychologisch onderzoek (Rossi, Simpson & Miller in: Berben, 2000), maar ook bij onderzoek naar de interbeoordelaarbetrouwbaarheid van taxatie-instrumenten (Nasuti & Pecora, 2003) en classificatiesystemen zoals de DSM (American Psychiatric Association, 2000) en het MAC (World Health Organization, 1996; Mezzich, Mezzich & Coffman, 1985; Remschmidt, Schmidt, & Göbel, 1983; Rutter, Shaffer &

[12] In andere bronnen wordt dit niveau ook wel als 'Redelijk' ('Moderate') aangeduid. In de begeleidingscommissie is voorafgaand aan het onderzoek afgesproken dat we bij een interbeoordelaarbetrouwbaarheid op dit niveau tevreden mogen zijn. In dit rapport is dit aangeduid als 'Voldoende'.

Shepherd, 1975). Volgens Veenma, Batenburg en Breedveld (2004) kan met relatief weinig vignetten de problematiek praktijkdekkend in kaart worden gebracht.

Veertig dossiers

Voor het formuleren van de vignetten is gebruik gemaakt van veertig dossiers van bureau jeugdzorg Agglomeratie Amsterdam (BJAA), van zowel de toegang als de jeugdbescherming en de jeugdreclassering. Acht dossiers uit het bestand betreffen jeugdbeschermingdossiers, twee betreffen jeugdreclasseringdossiers. Bij vier dossiers heeft de cliënt zorg van de jeugd-GGZ toegewezen gekregen. Er is bij de selectie van dossiers gecontroleerd of er voldoende spreiding is wat betreft de leeftijd, sekse en de mate van ernst op de schalen Functioneren Jeugdige en Kwaliteit Omgeving.

Twintig vignetten

Aan de hand van de veertig dossiers zijn twintig vignetten geconstrueerd. Twee vignetten betreffen cliënten die zijn doorverwezen naar de GGZ en drie vignetten zijn zo geformuleerd dat de kwaliteit van de omgeving zo ernstig is dat het cliënten betreft die in aanmerking komen om door de Raad voor de Kinderbescherming te worden onderzocht en eventueel in de jeugdbescherming terecht te komen. De vignetten variëren op leeftijd (er is onderscheid gemaakt naar de leeftijdscategorieën 0 t/m 5 jaar, 6 t/m 11 jaar en 12 jaar en ouder) en op het onderscheid tussen lichte en zware problematiek.
Om ervoor te zorgen dat de vignetten voldoende variatie vertonen in de mate van functioneren van de jeugdige en de kwaliteit van de omgeving is er onderscheid (op schaalniveau én itemniveau) gemaakt tussen verschillende categorieën. Binnen het kenmerk functioneren jeugdige is onderscheid gemaakt tussen redelijk tot goed functioneren, matig functioneren en slecht tot zeer slecht functioneren. Binnen de kwaliteit van de omgeving wordt onderscheid gemaakt tussen redelijk tot goede kwaliteit, matige kwaliteit en slechte tot zeer slechte kwaliteit.

Om te voorkomen dat de vignetten een opsomming vormen van gegevens die voor de STEP gecodeerd moet worden (die als ware afgevinkt kan worden voor het coderen van de STEP), voorzien de vignetten in meer informatie dan nodig is voor het invullen van de STEP, zodat het invullen van de STEP op basis van het vignet niet makkelijker is dan in de praktijk. Uiteindelijk zijn er dus vier groepen vignetten. De vignetten in elke groep komen wat betreft de ernst van de problematiek met elkaar overeen. De eerste groep bevat ‘bijzondere vignetten’ (jeugdbescherming, Raad voor de Kinderbescherming of GGZ). De tweede groep bevat de lichtste gevallen en de vierde groep de zwaarste gevallen. Vignetten die tussen licht en zwaar inzitten komen in de derde groep. Omdat de gevalsbeschrijvingen in willekeurige volgorde staan berust de toewijzing van de vignetten aan beoordelaars volledig op toeval. Tevens is er een maximum van één bijzonder vignet (verwijzing naar de Jeugdbescherming/Raad voor de Kinderbescherming/GGZ) per beoordelaar.

Van de 20 vignetten zijn 20 combinaties van vier gemaakt. Elke beoordelaar (in totaal 80 beoordelaars) heeft vier vignetten beoordeeld. Dat resulteert erin dat iedere combinatie van vier vignetten door vier beoordelaars is beoordeeld en dat er in totaal 320 beoordelingen zijn.

Voorstudie bruikbaarheid vignetten

De vignetten zijn in een voorstudie op bruikbaarheid getoetst. Daarbij is met tien hulpverleners nagegaan of er in de vignetten belangrijke informatie ontbreekt, die het moeilijk maakt de STEP goed in te vullen. Bij twee casus zijn op basis hiervan onvolkomenheden weggewerkt.

2.4 Kenmerken vignetten

De vignetten zijn zodanig geconstrueerd dat er een verdeling is met betrekking tot het functioneren van de jeugdige en de kwaliteit van de omgeving (redelijk tot goed, matig en slecht tot zeer slecht). Alle combinaties komen twee keer voor. Daarmee zijn 18 van de 20 vignetten ingevuld. De overgebleven vignetten zijn vrij ingevuld, afhankelijk van de beschikbare dossierinformatie.
Eén vignet is gebaseerd op dossiers van cliënten waarvoor een raadsonderzoek is aangevraagd en twee vignetten zijn gebaseerd op dossiers van cliënten die zijn doorverwezen naar de GGZ. Drie vignetten zijn jeugdbeschermingszaken, de betreffende kinderen zijn onder toezicht gesteld. Eén van deze zaken betreft tevens een GGZ-verwijzing.

Omdat de gevalsbeschrijvingen nog in vervolgonderzoek gebruikt zullen worden, geven we in het onderhavige rapport geen gespecificeerd overzicht van de ingebouwde STEP-karakteristieken van de vignetten. Wel wordt in paragraaf 3.5 nagegaan of de ingebouwde mate van ernst op de schalen Functioneren Jeugdige en Kwaliteit Omgeving overeenkomt met de beoordelingen van de codeurs.

2.5 Kenmerken deelnemende codeurs

De deelnemende codeurs zijn afkomstig van vijf organisaties:
- Bureau jeugdzorg Agglomeratie Amsterdam (26 codeurs, 32,5%)
- Bureau jeugdzorg Friesland (9 codeurs, 11,3%)
- Bureau jeugdzorg Limburg (30 codeurs, 37,6%), waarvan 19 codeurs uit de regio Maastricht (23,8%) en 11 uit de regio Roermond (13,8%)
- Bureau jeugdzorg Zeeland (12 codeurs, 15%)
- William Schrikker Groep (3 codeurs, 3,8%)

Bij de deelnemende codeurs is nagegaan wat hun opleidingsniveau en leeftijd is, hoeveel jaren werkervaring zij hebben en hoeveel ervaring in het gebruik van de STEP.

Verreweg de meeste codeurs hebben een hogere beroepsopleiding (bijna 83%), meestal met een sociaal-agogische of pedagogische achtergrond (zoals jeugdwelzijnswerk, maatschappelijk werk, inrichtingswerk, SPH). Ruim 15% heeft een wetenschappelijke opleiding gevolgd en 2,5% heeft een post-HBO achtergrond.

Bij de leeftijdsopbouw is te zien dat de grootste groep tussen de 25 en 35 jaar oud is (40%). Ook de 35-45 jarigen en 45-55 jarigen zijn redelijk vertegenwoordigd (resp. 22,5 en 25%). Zie ook Tabel 2 op pagina 19.

Tabel 2. Leeftijd codeurs

Leeftijd	Frequentie	Percentage
tot 25 jaar	4	5
25 tot 35 jaar	32	40
35 tot 45 jaar	18	22,5
45 tot 55 jaar	20	25
55 tot 65 jaar	5	6,3
Missing	1	1,3
Totaal	80	100

Wat betreft het aantal jaren werkervaring is te zien dat de grootste groep tot 5 jaar werkervaring heeft (ruim 40%), ongeveer een kwart heeft 6 tot 10 jaar werkervaring, de groep die 11 tot 15 jaar ervaring heeft is de helft zo groot (12,5%) en nog weer eens een stuk kleiner is de groep die 16 tot 20 jaar ervaring heeft (6,3%). Opvallend is dat de groep die 21 jaar of langer werkervaring heeft toch nog 12,5% uitmaakt. Zie ook Tabel 3 hieronder.

Tabel 3. Aantal jaren werkzaam in de jeugdzorg

Aantal jaren werkervaring	Frequentie	Percentage
0- 5 jaar	34	42,5
6-10 jaar	21	26,3
11-15 jaar	10	12,5
16-20 jaar	5	6,3
21 jaar en langer	10	12,5
Totaal	80	100

Ruim tweederde van de deelnemende codeurs heeft geen ervaring met de STEP (afgezien van de STEP-training die ze hebben gevolgd), 12,5% heeft 1 jaar ervaring en nog eens 12,5% heeft 2 jaar ervaring met de STEP. Slechts enkele codeurs hebben 3 of meer jaar ervaring met de STEP. Zie ook Tabel 4 hieronder.

Tabel 4. Aantal jaren ervaring met de STEP

Aantal jaren STEP-ervaring	Frequentie	Percentage
Geen	55	69,1
1 jaar	10	12,5
2 jaar	10	12,5
3 jaar	2	2,5
4 jaar	1	1,3
5 jaar	2	2,5
Totaal	80	100

Wanneer we de gegevens met betrekking tot werkervaring en ervaring met de STEP met elkaar in verband brengen, dan zijn er vier typen codeurs te onderscheiden (zie ook Tabel 5 op pagina 20):

I. Minder ervaren in jeugdzorg en geen ervaring met de STEP (25 personen);
II. Ervaren in de jeugdzorg en geen ervaring met de STEP (30 personen);
III. Minder ervaren in jeugdzorg en ervaring met de STEP (9 personen);
IV. Ervaren in de jeugdzorg en ervaring met de STEP (16 personen).

Tabel 5. Typen codeurs

	Geen ervaring met de STEP	**Ervaring met de STEP**	
Tot en met 5 jaar werkervaring	(groep I) 25	(groep III) 9	34
Vanaf 6 jaar werkervaring	(groep II) 30	(groep IV) 16	46
Totaal	55	25	**80**

Aan de hand van deze indeling zal in hoofdstuk 3 worden nagegaan of er een verband is tussen (werk)ervaring en de mate van overeenkomst tussen beoordelaars.

3 Overeenstemming beoordelaars

3.1 Analyses

De volgende analyses zijn uitgevoerd:

- Per schaal is met behulp van de intraclass correlatiecoëfficiënt (ICC) de overeenstemming berekend over alle codeurs en over de vier subgroepen van de codeurs. Dit moet duidelijk maken of bepaalde schalen gemakkelijker dan wel moeilijker tot overeenstemming leiden.
- Er is ook gekeken naar de overeenstemming per item, om de 'zwakke plekken' van de interbeoordelaarbetrouwbaarheid van de STEP op dit niveau op te sporen. Dit gebeurt wederom met de ICC.
- Per vignet is met behulp van de intraclass correlatiecoëfficiënt (ICC) de algemene overeenstemming bepaald op alle schalen en voor alle codeurs tegelijk.
- Vervolgens is per codeur met een aparte index gekeken naar de mate waarin het scorepatroon afwijkt van het gemiddelde patroon van de andere codeurs. Dat laat zien in hoeverre een codeur ook als 'betrouwbaar' (i.e. overeenkomend met de rest) kan gelden.
- Daarnaast is er gekeken of de karakteristieken van de vignetten, met betrekking tot de mate van ernst op de schalen Functioneren Jeugdige en Kwaliteit Omgeving (zoals aangegeven in paragraaf 2.4), overeenkomen met de beoordelingen van de codeurs. De scores op de STEP-schalen zijn doorgelicht op de vraag of de mate van werkervaring en ervaring met de STEP ertoe leiden dat de codeur op een schaal significant hoger scoort.
- Tot slot is er gekeken of er een verband is tussen de typen codeurs (zoals aangegeven in Tabel 5 op pagina 20) en de mate van overeenkomst tussen beoordelaars.

Alle analyses zijn uitgevoerd op 320 cases en 80 codeurs.

3.2 Resultaten per schaal

Op de schalen Functioneren Jeugdige, Kwaliteit Omgeving, Zorgzwaarte, Risico Jeugdige en Risico Omgeving is er een voldoende mate van overeenstemming tussen de beoordelaars (zie Tabel 6 op pagina 22). De scores op deze schalen variëren van .54 tot .68[13]. Tijdens de trainingen wordt door cursisten vaak aangegeven dat men de twee risicoschalen het lastigst vindt om in te vullen. Dit vertaalt zich desondanks niet terug in een lage mate van overeenstemming; die is ook op deze twee schalen voldoende.

Op de schaal Urgentie Zorg is er minder overeenstemming. Deze schaal haalt een score van .45, wat gekwalificeerd wordt als matig. De in verhouding lagere score op deze schaal heeft mogelijk te maken met de zgn. *decision threshold*, oftewel de persoonlijke *beslisdrempel* van de beoordelaars (Dalgleish, 1997). Beslissers met een lage beslissingsdrempel gaan snel over tot actie, ook wanneer de geschatte bedreiging niet groot is. Beslissers met een hoge beslissingsdrempel zullen alleen besluiten om actie te ondernemen als zij van mening zijn dat er sprake is van een zeer bedreigende situatie. Ook als er overeenstemming is over wat er aan

[13] De scores worden als volgt gekwalificeerd: -1.00 - .30 is slecht, .31 - .50 is matig, .51 - .70 is voldoende en .71 - 1.00 is goed.

de hand is in een gezin, is het niet ongebruikelijk dat hulpverleners van mening verschillen over welke hulp er nodig is en hoe snel deze hulp geboden moet worden (Eijgenraam, 2006).

Tabel 6. Resultaten op schaalniveau

Beschrijving schaal	Kwalificatie	ICC *	Sign.	95% interval
Functioneren Jeugdige (FJ)	Voldoende	.68	.000	.54 - .82
Kwaliteit Omgeving (KO)	Voldoende	.58	.000	.42 - .75
Zorgzwaarte (ZZ)	Voldoende	.56	.000	.41 - .74
Urgentie Zorg (UZ)	Matig	.45	.000	.31 - .65
Risico Jeugdige (RJ)	Voldoende	.54	.000	.39 - .72
Risico Omgeving (RO)	Voldoende	.68	.000	.54 - .83

* ICC-model one-way random/absolute/single

3.3 Resultaten per item

De betrouwbaarheid van een schaal hangt samen met de items waarmee deze is gevuld. Die items zijn vaak (veel) minder betrouwbaar dan de schaal. Constructeurs van instrumenten vinden dat meestal niet erg, zolang de items bij elkaar maar een voldoende betrouwbare schaal opleveren. Twee beoordelaars kunnen bijvoorbeeld op alle items verschillend scoren, en bij optelling van de items toch allebei op een totaal van 30 punten uitkomen. Daarom zien we vaak dat afzonderlijke items onvoldoende betrouwbaar gescoord worden, terwijl dit voor de totale schaal veel gunstiger uitpakt. Dit is dan ook een belangrijk principe van schaalconstructie: hoe meer items eens schaal bevat, hoe groter de kans dat in de totaalscore de verschillen zijn 'uitgemiddeld', hoe groter de kans op een betrouwbare schaal.

Hoewel uit de vorige paragraaf is gebleken dat de meeste STEP-schalen over een voldoende betrouwbaarheid beschikken (behalve de schaal Urgentie Zorg, die maar één item heeft!), is het toch nuttig ook de betrouwbaarheid per item te bestuderen. Immers, items waarbij beoordelaars het vaak met elkaar oneens zijn, kunnen de algemene betrouwbaarheid van de schaal negatief beïnvloeden. Om dit soort 'zwakke plekken' in de schalen op te sporen, is ook in dit onderzoek de overeenstemming tussen codeurs op itemniveau bepaald. Voor de schaal Urgentie Zorg is dat niet gedaan, omdat deze uit slechts één item bestaat.

Items op de schaal Functioneren Jeugdige (FJ)
Voor wat betreft de schaal Functioneren Jeugdige zien we dat de helft van de items voldoende scoort en de helft matig (zie Tabel 7 op pagina 23).
De items 'Hoe lang bestaan de problemen?' (FJ2), 'Hoe erg is de jeugdige van slag?' (FJ3) en 'Belemmering functioneren van de jeugdige op vier leefgebieden' (FJ4) scoren lager dan .51 en worden om die reden gekwalificeerd als matig. De overige drie items scoren hoger dan .50 en worden daarmee gekwalificeerd als voldoende.

Tabel 7. Schaal Functioneren Jeugdige (resultaten op itemniveau)

Item	Beschrijving	ICC	Sign.	95% interval
FJ1	Item 1: Problemen persoonlijk functioneren?	.52	.000	.36 - .70
FJ2	Item 2: Hoe lang bestaan de problemen?	.40	.000	.26 - .60
FJ3	Item 3: Hoe erg is de jeugdige van slag?	.43	.000	.28 - .63
FJ4	Item 4: Belemmering functioneren totaal	.48	.000	.33 - .68
FJ4a	Item 4a: Belemmering functioneren thuis	.51	.00	.36 - .70
FJ4b	Item 4b: Belemmering functioneren in relaties	.31	.00	.18 - .51
FJ4c	Item 4c: Belemmering functioneren in crèche, school, werk	.67	.00	.53 - .82
FJ4d	Item 4d: Belemmering functioneren overige omgeving	.44	.00	.29 - .64
FJ5	Item 5: Belasting voor overige leden thuis?	.67	.000	.52 - .81
FJ6	Item 6: Belasting voor overige omgeving?	.65	.000	.51 - .81

Uit de reacties die de trainers van de STEP van cursisten krijgen valt het één en ander te leren over de zwakke en sterke punten van de verschillende items. Dit kan zijn licht werpen op mogelijke verklaringen van de resultaten uit de analyse én het kan verbeterpunten aandragen voor de STEP. We laten ze hier de revue passeren.

Voor het beantwoorden van de vraag 'Heeft de jeugdige problemen op één of meer van de aspecten van het persoonlijk functioneren?' (FJ1), zijn in de handleiding gespecificeerde antwoordcategorieën opgenomen (gebaseerd op operationalisaties uit andere instrumenten)[14], waaruit de beoordelaar er één kan kiezen. Bij het maken van de handleiding is ervan uitgegaan dat de werker over voldoende deskundigheid beschikt om de afweging te kunnen maken wat als normaal en wat als problematisch wordt beschouwd. Met betrekking tot bepaalde groepen jeugdigen en/of problemen (zoals jongere kinderen, jongeren in de jeugdreclassering, verstandelijke en/of lichamelijke beperkingen) vindt een aantal hulpverleners het echter lastig om die beoordeling te maken. Er is behoefte aan meer gedetailleerde informatie in de handleiding om een adequate inschatting te kunnen maken. Daarnaast geven beoordelaars aan dat ze geneigd zijn om de kennis die ze hebben over de omgeving hier al mee te laten wegen. Dat is echter niet de bedoeling, omdat dit afzonderlijk wordt beoordeeld bij de schaal Kwaliteit Omgeving. Ondanks deze ervaren problemen, is de mate van overeenstemming tussen beoordelaars wel voldoende op dit item. Met enige aanpassingen aan de handleiding - die nog minder ruimte voor eigen interpretatie laten - zou de betrouwbaarheid van dit item nog verder verhoogd kunnen worden.

Het item 'Hoe lang bestaan de problemen?' (FJ2) scoort matig. Een mogelijke verklaring ligt in het feit dat de aanvang van de problemen mogelijk verschillend wordt opgevat. Sommige beoordelaars kiezen mogelijk als aanvangsmoment de allereerste tekenen van het ontstaan van een probleem, ook als dat probleem tussentijds opgelost was cq. hanteerbaar was voor de ouders of de jeugdige. Terwijl andere beoordelaars mogelijk het moment kiezen dat de actuele problemen begonnen op te spelen. De handleiding dient hierin meer duidelijkheid te geven.

Het item 'Hoe erg is de jeugdige door persoonlijke of omgevingsproblemen van slag?' (FJ3) scoort matig. Ook wat betreft dit item is de mate van overeenstemming dus voor verbetering vatbaar. Deze vraag heeft betrekking op de door de jeugdige ervaren lijdensdruk. Sommige beoordelaars willen hier mee laten wegen, dat wanneer de jeugdige geen lijdensdruk ervaart terwijl er wel ernstige problemen zijn, dit een ernstige situatie is, met mogelijk een ongunstige prognose voor de verdere ontwikkeling van de jeugdige. Ook hier geldt dat dit niet

[14] Zie paragraaf 1.3.

meegewogen dient te worden in de score op dit item: sec de lijdensdruk wordt ingeschat en gescoord. De lijdensdruk dient niet te worden beoordeeld in samenhang met andere aspecten zoals de kwaliteit van de omgeving of het risico voor de jeugdige of de omgeving. De ongunstige prognose komt vervolgens tot uiting in een hoge score op de schaal Risico Jeugdige en/of Risico Omgeving. In de handleiding zou dit nadrukkelijker geïnstrueerd kunnen worden, zodat beoordelaars niet meer in de verleiding komen de ongunstige prognose te verwerken in dit item.

Ook het vierde item, 'Belemmert het functioneren van de jeugdige zijn dagelijks leven op de vier leefgebieden?' (FJ4) scoort matig. Dit item is opgebouwd uit vier subvragen. Beoordelaars geven aan dat de vraag met betrekking tot het gedrag op de crèche/leren op school/werk verwarring geeft. De indruk zou kunnen bestaan dat met betrekking tot school alleen de schoolprestaties aan de orde zijn, terwijl ook het gedrag op school meegewogen dient te worden in de beoordeling. In de handleiding dient dit punt aangepast te worden, om duidelijk te maken dat het om het functioneren in verschillende settings gaat (dus gedrag en prestaties).

Items op de schaal Kwaliteit Omgeving (KO)

De items 'Zijn er problemen op de belangrijkste aspecten van de overige omgeving?' (KO2) en 'Hoe lang bestaan voorkomende problemen in de omgeving?' (KO3) scoren matig, de overige items op deze schaal scoren voldoende (zie Tabel 8 hieronder).

Tabel 8. Schaal Kwaliteit Omgeving (resultaten op itemniveau)

Item	Beschrijving	ICC	Sign.	95% interval
KO1	Item 1: Kwaliteit opvoedingsomgeving of netwerk?	.62	.000	.47 - .78
KO2	Item 2: Problemen overige omgeving?	.38	.000	.24 - .58
KO3	Item 3: Hoe lang bestaan problemen?	.11	.000	.04 - .26
KO4	Item 4: Is er iemand die steunt?	.52	.000	.37 - .71
KO5	Item 5: Hoezeer is omgeving risicofactor?	.56	.000	.41 - .74

Het item 'Zijn er problemen met de kwaliteit van de primaire opvoedingsomgeving?' (KO1) scoort voldoende, maar er zijn bij beoordelaars wel vragen over wat er precies bedoeld wordt met 'primaire opvoedingsomgeving'. Hiermee wordt bedoeld het samenlevingsverband met de biologische ouder(s) of stiefouder(s), ofwel het gezin. Wanneer de ouders niet de dagelijkse verantwoordelijkheid voor de opvoeding hebben en er geen perspectief is op terugkeer naar dat gezin, is de primaire opvoedingsomgeving de vervangende opvoedingssituatie (wanneer de jeugdige bijvoorbeeld in een residentiële instelling of een pleeggezin verblijft of bij een familielid woont). Het gaat dus niet om de juridische verantwoordelijkheid, maar de verantwoordelijkheid voor de dagelijkse uitvoering van de verzorging en opvoeding. Ook op dit punt dient de handleiding explicieter te worden.

Het item 'Hoe lang bestaan voorkomende problemen in de omgeving?' (KO3) scoort slechts .11, wat dit item de kwalificatie 'slecht' geeft. Voor dit item geldt hetzelfde als voor item FJ2: het aanvangsmoment is waarschijnlijk verschillend opgevat. De handleiding dient op dit punt verduidelijkt te worden.

Items op de schaal Zorgzwaarte (ZZ)

Op deze schaal laten alle items een matig niveau van interbeoordelaarbetrouwbaarheid zien: de ICC varieert van .43 tot .49. (zie Tabel 9 op pagina 25). In de vorige paragraaf is gebleken dat er op schaalniveau niettemin sprake is van een voldoende betrouwbaarheid (.56). We zien hier de wetten van de schaalconstructie in werking: verschillen tussen de scores van codeurs

op itemniveau worden kennelijk op schaalniveau 'uitgemiddeld'. Dat resulteert in een betere betrouwbaarheid.

Tabel 9. Schaal Zorgzwaarte (resultaten op itemniveau)

Item	Beschrijving	ICC	Sign.	95% interval
ZZ1	Item 1: Wat voor zorg is er nodig?	.47	.000	.32 - .67
ZZ2	Item 2: Hoe lang zal de hulp duren?	.49	.000	.34 - .68
ZZ3	Item 3: Hoe groot is de intensiteit contacten?	.43	.000	.29 - .63

Niettemin zijn ook hier bij de items verbeteringen mogelijk. Uit reacties bij de trainingen leiden we af dat de formuleringen bij de verschillende antwoordcategorieën van item ZZ1 ('Wat voor soort zorg is er nodig?') tot verwarring kunnen leiden. Bijvoorbeeld wanneer er combinaties nodig zijn van behandeling en verblijf. En wanneer er aanvullende diagnostiek (binnen het toegangstraject) nodig is, wordt dit soms opgevat als observatiediagnostiek. Terwijl de STEP wordt ingevuld op het moment dat het indicatiebesluit al dan niet opgesteld wordt. Dus op dat moment is de aanvullende diagnostiek al achter de rug (hiervoor wordt geen indicatie opgesteld, dus dit kan ook niet in de STEP aangegeven worden). Ook bij een herindicatie is niet altijd duidelijk wat men moet invullen. Beide punten dienen preciezer omschreven te worden in de handleiding.

Items op de schaal Risico Jeugdige (RJ)

Eén van de vier items scoort matig, de rest voldoende (zie Tabel 10 hieronder). Net als bij de schaal Functioneren Jeugdige, vindt men de lijdensdruk van de jeugdige lastig te beoordelen.

Tabel 10. Schaal Risico Jeugdige (resultaten op itemniveau)

Item	Beschrijving	ICC	Sign.	95% interval
RJ1	Item 1: Kans op problemen persoonlijk functioneren?	.51	.000	.36 - .70
RJ2	Item 2: Hoe erg zal jeugdige van slag zijn?	.40	.000	.26 - .60
RJ3	Item 3: Zal jeugdige op leefgebieden belemmerd zijn?	.51	.000	.36 - .70
RJ4	Item 4: Zal functioneren belasting voor thuis zijn?	.50	.000	.34 - .69

Items op de schaal Risico Omgeving (RO)

Beide items op deze schaal scoren voldoende (zie Tabel 11 hieronder).

Tabel 11. Schaal Risico Omgeving (resultaten op itemniveau)

Item	Beschrijving	ICC	Sign.	95% interval
RO1	Item 1: Zal functioneren belasting voor omgeving zijn?	.55	.000	.40 - .73
RO2	Item 2: Zal jeugdige (opnieuw) een strafbaar feit plegen?	.68	.000	.54 - .83

3.4 Resultaten per vignet

Per vignet is met behulp van de intraclass correlatiecoëfficiënt (ICC) de algemene overeenstemming bepaald op alle schalen en voor alle codeurs tegelijk (zie Tabel 12 op pagina 26).

Tabel 12. Resultaten per vignet

Vignet nr.	*QUICK*STEP schalen	*QUICK*STEP items	STEP risico-schalen	STEP risico-items
1	.85 = Goed	.47 = Matig	.94 = Goed	.62 = Voldoende
2	.87 = Goed	.56 = Voldoende	.92 = Goed	.56 = Voldoende
3	.68 = Voldoende	.46 = Matig	.81 = Goed	.25 = Slecht
4	.86 = Goed	.46 = Matig	.87 = Goed	.47 = Matig
5	.83 = Goed	.66 = Voldoende	.86 = Goed	.39 = Matig
6	.85 = Goed	.48 = Matig	.92 = Goed	.61 = Voldoende
7	.88 = Goed	.49 = Matig	.93 = Goed	.26 = Slecht
8	.83 = Goed	.53 = Voldoende	.89 = Goed	.50 = Matig
9	.87 = Goed	.49 = Matig	.78 = Goed	.25 = Slecht
10	.87 = Goed	.52 = Voldoende	.90 = Goed	.41 = Matig
11	.89 = Goed	.63 = Voldoende	.93 = Goed	.60 = Voldoende
12	.88 = Goed	.51 = Voldoende	.95 = Goed	.63 = Voldoende
13	.88 = Goed	.77 = Goed	.91 = Goed	.55 = Voldoende
14	.93 = Goed	.74 = Goed	.95 = Goed	.59 = Voldoende
15	.94 = Goed	.53 = Voldoende	.94 = Goed	.11 = Slecht
16	.93 = Goed	.69 = Voldoende	.89 = Goed	.09 = Slecht
17	.91 = Goed	.66 = Voldoende	.94 = Goed	.41 = Matig
18	.85 = Goed	.52 = Voldoende	.92 = Goed	.44 = Matig
19	.96 = Goed	.83 = Goed	.96 = Goed	.19 = Slecht
20	.93 = Goed	.57 = Voldoende	.95 = Goed	.39 = Matig

Voor vrijwel alle vignetten geldt dat de overeenstemming op schaalniveau goed is. Zowel wat betreft de vier *QUICK*STEPschalen als de twee risicoschalen. Uitzondering hierop vormt vignet 3, waarbij de overeenstemming op de vier QuickSTEPschalen voldoende is en die op de twee risicoschalen goed. Op itemniveau is de overeenstemming wat lager: de items van de vier *QUICK*STEPschalen scoren matig tot goed, de items van de risicoschalen slecht tot voldoende. Hieruit kunnen we concluderen dat eventuele verschillen in beoordeling op itemniveau, binnen de schaal worden gecompenseerd. Bij de vignetten 13 en 14 is de overeenstemming het grootst: de items van de risicoschalen scoren voldoende, de overige items én alle schalen scoren goed.

Zoals in paragraaf 2.4 is opgemerkt, zijn de vignetten zodanig geconstrueerd dat er een verdeling is met betrekking tot het functioneren van de jeugdige en de kwaliteit van de omgeving (redelijk tot goed, matig en slecht tot zeer slecht). Nagegaan is in hoeverre dit in het codeergedrag van de beoordelaars is terug te vinden. De vignetten zijn daarvoor per schaal (Functioneren Jeugdige en Kwaliteit Omgeving) ingedeeld in drie groepen: (A) Redelijk tot goed; (B) Matig; (C) Slecht tot zeer slecht. Binnen elke groep is vervolgens per vignet uitgerekend wat de gemiddelde score is van de 16 codeurs die het vignet hebben beoordeeld. Daarna is gekeken of de vignetten per groep op die gemiddelden verschillen.
Tabel 13 op pagina 27 toont de resultaten.

Tabel 13. Karakteristieken bij constructie van de vignetten en uiteindelijk toegekende scores*

Schaal	Groepsindeling vignetten bij constructie	Aantal vignetten	Gem. score	Spreiding (min.– max.)	
				Vignetten	Ruwe scores
Functioneren Jeugdige (FJ)	A. Redelijk tot goed	6	13,20	8,7 – 15,7	6 – 20
	B. Matig	7	17,6	15,3 – 19,9	6 – 25
	C. Slecht tot zeer slecht	7	21,4	18,9 – 24,2	9 – 28
	Totaal	20	17,6***	8,7 – 24,2	6 – 28
Kwaliteit Omgeving (KO)	A. Redelijk tot goed	7	12,9	9,7 – 17,8	5 – 22
	B. Matig	7	14,3	8,9 – 19,4	5 – 24
	C. Slecht tot zeer slecht	6	17,6	13,1 – 21,3	7 – 25
	Totaal	20	14,8*	8,9 – 21,3	5 - 25

Gem. score = Gemiddelde score. *** = Significantie F-test p < .000; * = Significantie F-test p < .05. De kolom onder 'Vignetten' toont het minimale en maximale gemiddelde van de scores die per vignet door de codeurs (16 personen) werden toegekend. De kolom onder 'Ruwe scores' toont de minimale en maximale score die codeurs binnen groep A, B of C aan de afzonderlijke vignetten toekenden.

Voor 'Functioneren Jeugdige' geldt dat de scores van de groep 'Redelijk tot goed' (A) gemiddeld genomen het laagst scoren op de betreffende schaal, de groep 'Matig' (B) in het midden en de groep 'Slecht tot zeer slecht' (C) het hoogst. Dat zien we ook bij 'Kwaliteit Omgeving'. Voor beide schalen is deze trend significant. Met een zogeheten Post Hoc test (methode Scheffé, $p < .05$) is per schaal nagegaan of de groepen ook onderling van elkaar verschillen. Bij de schaal 'Functioneren Jeugdige' bleek dit inderdaad het geval. Bij Kwaliteit Omgeving verschillen alleen de groepen 'Redelijk tot goed' (A) en 'Slecht tot zeer slecht' (C). In Tabel 13 is onder de kolom 'Spreiding' aangegeven welke minimale en maximale scores zijn aangetroffen. Te zien is dat bij de schaal 'Functioneren Jeugdige' de berekende gemiddelden per vignet (zie kolom 'Vignetten') elkaar weinig overlappen. Kijken we echter naar de ruwe scores die daaronder liggen (kolom 'Ruwe scores'), dan lijkt de spreiding in zowel groep A als B en C veel op elkaar. Het beeld bij de schaal 'Kwaliteit Omgeving' toont dat de berekende gemiddelden per vignet elkaar behoorlijk overlappen. Dat zien we ook als we kijken naar de ruwe scores die daaronder liggen (kolom 'Ruwe scores').

Het beeld dat uit deze analyses rijst is dat de codeurs over het algemeen geneigd zijn hoog op de schalen te scoren. Voor de vignetten die zijn geconstrueerd met het oog op een 'Redelijk tot goed' functioneren komen de schalen hoger uit dan we met de vignetten beoogden. Niettemin stemmen de codeurs goed met elkaar overeen. Daar staat tegenover dat de afwijkingen bij enkele codeurs fors kunnen zijn: er zijn zowel naar de lage als de hoge uiteinden van de schaal extreme 'uitbijters'. Dat maakt het interessant te kijken naar het aantal codeurs dat anders beoordeelt dan de meeste anderen. De volgende paragraaf gaat daarop in.

3.5 Resultaten per codeur

Tot slot is nagegaan hoe de codeurs scoren, vergeleken met andere codeurs die dezelfde vignetten hebben beoordeeld. Daarbij is als volgt te werk gegaan:

- Per vignet is berekend wat de gemiddelde scores op elke schaal zijn en vervolgens is per schaal gekeken hoever elke codeur van dat gemiddelde afwijkt.
- De afwijkingen zijn vertaald in zogeheten standaarddeviaties. Een negatieve of positieve deviatie van meer dan 1 levert voor een codeur op een schaal een punt op. Een codeur kan maximaal 24 punten krijgen (een punt voor elke schaal van de STEP bij elk van de vier vignetten die de codeur heeft beoordeeld). De mate waarin de codeur van het gemiddelde

scorepatroon van de rest afwijkt is dan in een afwijkingsindex uit te drukken als het aantal punten dat de codeur heeft gekregen, gedeeld door het maximum van 24 punten.
- Als index voor de mate waarin een codeur overeenkomt met het gemiddelde scorepatroon van de andere codeurs geldt: 1 – de afwijkingsindex.

Aan de hand van deze eenvoudige (niet voor kans gecorrigeerde) overeenstemmingsmaat zijn alle codeurs gerangschikt. De vuistregel is: hoe kleiner het verschil is met het gemiddelde van alle codeurs, hoe betrouwbaarder die codeur beoordelingen maakt. Elke codeur krijgt een betrouwbaarheidscode op basis van het aantal afwijkingen ten opzichte van andere codeurs. Hoe kleiner het aantal afwijkingen, hoe hoger de score. We hebben daarbij gemakshalve de volgende kwalificatie gebruikt: tot .50 is 'slecht', .50 tot en met .59 is 'matig', .60 tot en met .69 is 'redelijk', .70 tot en met .79 is 'voldoende' en vanaf .80 is 'goed'[15]. Uit Tabel 14 hieronder blijkt dat 48 codeurs (60%) een voldoende tot goed betrouwbaarheidsniveau behalen. Een kwart (20 codeurs) staat op een redelijk niveau.

Tabel 14. Betrouwbaarheidsniveau codeurs

Betrouwbaarheidsniveau		Frequentie	Percentage
Slecht	(< .50)	3	3,8
Matig	(.50-.59)	9	11,3
Redelijk	(.60-.69)	20	25,0
Voldoende	(.70-.89)	32	40,0
Goed	(> .79)	16	20,0
Totaal		**80**	**100%**

De tabel laat ook zien dat er codeurs zijn die matig tot slecht met de anderen overeenstemmen. Nadere inspectie van de gegevens laat zien dat enkele hulpverleners soms zeer afwijkend van de rest de STEP hebben ingevuld. De 'uitbijters' die we in de vorige paragraaf bij de schalen voor 'Functioneren Jeugdige' en 'Kwaliteit Omgeving' tegenkwamen blijken – hoewel niet significant - gemiddeld een lager betrouwbaarheidsniveau te hebben dan de rest. Met andere woorden, deze codeurs laten op hun hele scoregedrag op de STEP een enigszins afwijkend patroon zien.

De scores op de STEP-schalen zijn doorgelicht op de vraag of de mate van werkervaring en ervaring met de STEP ertoe leiden dat de codeur op een schaal significant hoger of juist lager bij een vignet scoren dan de gemiddelde scores van alle codeurs op een vignet. De volgende trends zijn daarbij waargenomen:

- Codeurs met weinig werkervaring (conform de indeling van tabel 4) neigen tot een iets hogere score op de STEP-schalen dan gemiddeld; codeurs met veel werkervaring neigen tot juist wat lagere scores. Deze trends zijn echter niet significant. Uitzondering hierop vormt de schaal 'Zwaarte Zorg'. Hier is een significante trend waarneembaar ($F = 2,9$; $df = 4$; $p = .026$), waarbij de groep met 11-15 jaren ervaring enigszins lager scoort dan de rest. Paarsgewijze vergelijking (post hoc volgens methode Scheffé) van de groepen met verschillende werkervaring brengen echter geen duidelijke verschillen aan het licht.
- Als de codeurs worden ingedeeld naar de typen zoals in Tabel 5 op pagina 20 (weinig of veel werkervaring en wel of geen ervaring met de STEP), dan zien we dat de hulpverleners met STEP-ervaring over het algemeen neigen om op de schalen 'Functioneren Jeugdige' en 'Kwaliteit Omgeving' iets lager te scoren van de rest. Voor de schalen 'Urgentie zorg' en 'Risico Jeugdige' is dit net andersom. Deze trends zijn echter niet significant.

[15] Omdat de index niet voor kans gecorrigeerd is, hanteren we hier .70 als ondergrens voor 'voldoende' en een extra niveau voor .60 'redelijk'.

Alle gegevens zijn ook geïnspecteerd aan de hand van de vraag of het betrouwbaarheidsniveau van de codeurs samenhangt met werkervaring of eerdere ervaring met de STEP. Daarin zijn geen significante trends waarneembaar.

Afwijkingen in het scorepatroon zijn dus vooralsnog niet toe te schrijven aan de ervaring van de codeurs. Basismaatregel om onbetrouwbaarheid te voorkomen is het aanscherpen van instructies in de handleiding en verdere training van codeurs. Ook is een regelmatige intervisie met betrekking tot de scoring van het instrument sterk aan te bevelen. Dat kan bijvoorbeeld helder maken of sommige codeurs de neiging hebben erg afwijkend te scoren. Het voorkomt dat beoordelaars teveel hun eigen interpretaties in het materiaal en het scoren van de items gaan leggen. Maar dat geldt voor elk instrument.

4 Samenvatting en conclusies

4.1 Ontwikkeling STEP en eerder onderzoek

In de periode 2001-2003 is door NIZW Jeugd (nu NJi) in opdracht van het Regionaal Orgaan Amsterdam (ROA) gewerkt aan de constructie van de zogeheten Standaard Taxatie Ernst van de Problematiek (STEP) (Van Yperen, Van den Berg & Eijgenraam, 2002, 2003a, 2003b, 2003c).

In een eerste deelproject zijn een uitgebreide literatuurstudie en praktijkoriëntatie verricht. Hieruit is gebleken dat er verschillende definities bestaan van het begrip 'ernst'. Er zijn vier facetten gevonden die van belang zijn bij de ernst van de problematiek namelijk de *abnormaliteit van het gedrag*, *bijdragende factoren* in de jeugdige, gezin, opvoeding en wijdere omgeving, *probleemgedrag* en *kwaliteit van leven*. Op basis van deze facetten is een theoretisch werkmodel ontwikkeld, dit model gaat enerzijds uit van problemen en risicofactoren en anderzijds van protectieve factoren. Verstoring van de balans uit zich onder meer in lijden en de onbalans moet zowel in de historische, actuele als prognostische betekenis worden beschouwd. Dit model vormt de basis voor de constructie van de STEP.

De STEP bestaat uit zes schalen, waarbij de eerste vier de zogeheten *QUICK*STEP vormen: Functioneren Jeugdige (STEP-FJ), Kwaliteit Omgeving (STEP-KO), Zwaarte Zorg (STEP-ZZ), Urgentie Zorg (STEP-UZ), Risico Jeugdige (STEP-RJ), Risico Omgeving (STEP-RO).
Voor het gebruik van de *QUICK*STEP is een scoringshulp van vijftien items gemaakt. De laatste twee schalen zijn nog experimenteel, ook hierbij is een scoringshulp gemaakt. Per cliënt kost het invullen van het instrument (na enige ervaring) ongeveer vijf minuten.

Uit een eerste proef binnen de bureaus jeugdzorg in de agglomeratie Amsterdam en in Gouda naar de psychometrische kwaliteit en de hanteerbaarheid van de STEP blijkt de interne consistentie en dekking van de eerste vier schalen van het instrument binnen bureau jeugdzorg voldoende tot goed. Voorts is gebleken dat de conceptuele indeling van het begrip 'Ernst' in de vier subschalen van de *QUICK*STEP empirisch goed is te reconstrueren. Naar aanleiding van de uitkomsten van het onderzoek naar de hanteerbaarheid van het instrument is het aantal items van de eerste vier schalen van de STEP ingekort van 21 tot 15.

4.2 Opzet en uitvoering van het onderzoek

In 2006 is een tweede onderzoekstraject gestart waarin onderzoek gedaan wordt naar het gebruik van de STEP in de jeugdbescherming en de jeugdreclassering, de interbeoordelaarbetrouwbaarheid en de voorspellende en evaluatieve waarde van de STEP. In dit rapport staat het onderzoek naar de interbeoordelaarbetrouwbaarheid van de STEP centraal.

Doel van het onderzoek
Doel van dit deelonderzoek is na te gaan of de interbeoordelaarbetrouwbaarheid van de STEP hoog genoeg is om te kunnen spreken van een instrument dat op dit punt voldoende betrouwbaar is. Het onderhavige onderzoek gaat na in welke mate invullers van de STEP met elkaar overeenstemmen in de scores die ze toekennen bij eenzelfde casus. Indien de interbeoordelaarbetrouwbaarheid niet hoog genoeg blijkt te zijn, dan zal het onderzoek inzicht moeten verschaffen in de punten waarop de STEP aangepast dient te worden om het instrument te verbeteren.

Verantwoording onderzoeksdesign
In dit deelonderzoek is een opzet gehanteerd die de subject-, informatie- en situatievariantie uitschakelt waardoor de resultaten vooral bepaald worden door de invloed van observatievariantie (codeurs kunnen uit dezelfde informatiebronnen putten, maar verschillen in wat hen opvalt of leggen verschillende accenten). Daartoe is een casusboek samengesteld met twintig vignetten. Er hebben medewerkers van verschillende bureaus jeugdzorg (Limburg, Zeeland, Friesland en Agglomeratie Amsterdam) en de William Schrikker Groep (een landelijk werkende organisatie voor o.a. LVG-jeugd) deelgenomen aan het onderzoek. Volgens een geblokt design is aan 81 hulpverleners gevraagd elk vier vignetten met de STEP te scoren. Van elke hulpverlener is de functie, de opleiding en het aantal ervaringsjaren bekend. Uit de literatuur is bekend dat met name ervaring een factor is in de mate waarin codeurs met elkaar overeenstemmen. Alle hulpverleners hebben een training gevolgd in het gebruik van de STEP. De verzamelde gegevens worden met behulp van SPSS ingevoerd, gecontroleerd, opgeschoond en geanalyseerd. Als indexen voor de mate van overeenstemming op items en schalen wordt een internationaal geaccepteerde maat gebruikt (intraclass-correlatie).

Verantwoording constructie vignetten
Zoals gezegd, is er een casusboek samengesteld met twintig gevalsbeschrijvingen. Daarbij is deels gebruikgemaakt van bestaande beschrijvingen, aangevuld met dossiermateriaal van bureaus jeugdzorg. De beschrijvingen zijn bewerkt tot geanonimiseerde en niet tot personen herleidbare vignetten. Voor het formuleren van de vignetten is gebruik gemaakt van veertig dossiers van bureau jeugdzorg Agglomeratie Amsterdam (BJAA), van zowel de toegang als de jeugdbescherming en de jeugdreclassering. Van elk vignet is de leeftijd, sekse, cultuur, leefverband, gezinssituatie, dagbesteding en aard problematiek van het kind en het juridisch kader van de hulpverlening bekend. Er is bij de selectie van dossiers gecontroleerd of er voldoende spreiding is wat betreft de leeftijd, sekse en de mate van ernst op de schalen Functionering Jeugdige en Kwaliteit Omgeving. De vignetten zijn in een voorstudie op bruikbaarheid getoetst. Daarbij is met een aantal codeurs nagegaan of er in de vignetten belangrijke informatie ontbreekt, die het moeilijk maakt de STEP goed in te vullen. Eventuele onvolkomenheden in de cases zijn op basis hiervan weggewerkt.

Uitvoering van het onderzoek
Elke beoordelaar (in totaal 80) heeft vier vignetten beoordeeld. Dat resulteert erin dat elke casus door zestien beoordelaars is beoordeeld en dat er in totaal 320 beoordelingen zijn.
De volgende analyses zijn uitgevoerd:

- Per schaal is met behulp van de intraclass correlatiecoëfficiënt (ICC) de overeenstemming berekend over alle codeurs en over de vier subgroepen van de codeurs. Dit moet duidelijk maken of bepaalde schalen gemakkelijker dan wel moeilijker tot overeenstemming leiden.
- Er is ook gekeken naar de overeenstemming per item, om de 'zwakke plekken' van de interbeoordelaarbetrouwbaarheid van de STEP op dit niveau op te sporen. Dit gebeurt wederom met de ICC.

- Per vignet is met behulp van de ICC de algemene overeenstemming bepaald op alle schalen en voor alle codeurs tegelijk.
- Bekeken is in welke mate codeurs afwijken van het gemiddelde scorepatroon bij de vignetten waarover zij beoordelingen hebben gegeven. Dit laat zien of er veel 'betrouwbare' codeurs zijn. Dit is met een zelfgeconstrueerde index bepaald.
- Daarnaast is er gekeken of de mate van ernst op de schalen Functioneren Jeugdige en Kwaliteit Omgeving die bij de constructie van de vignetten is ingebouwd overeenkomt met de beoordelingen van de codeurs.
- Tot slot is er gekeken of er een verband is tussen de typen codeurs en de mate van overeenkomst tussen beoordelaars.

4.3 Belangrijkste resultaten

Resultaten per schaal

Op de schalen Functioneren Jeugdige, Kwaliteit Omgeving, Zorgzwaarte, Risico Jeugdige en Risico Omgeving is er een voldoende mate van overeenstemming tussen de beoordelaars. De scores op deze schalen variëren van .54 tot .68 (zie Tabel 15 hieronder)[16]. Op de schaal Urgentie Zorg is er minder overeenstemming. Deze schaal haalt een score van .45, wat gekwalificeerd wordt als matig. De in verhouding lagere score op deze schaal heeft mogelijk te maken met de zgn. *decision threshold*, oftewel de persoonlijke *beslisdrempel* van de beoordelaars (Dalgleish, 1997).

Tabel 15. Resultaten op schaalniveau

Beschrijving schaal	Kwalificatie	ICC *	Sign.	95% interval
Functioneren Jeugdige (FJ)	Voldoende	.68	.000	.54 - .82
Kwaliteit Omgeving (KO)	Voldoende	.58	.000	.42 - .75
Zorgzwaarte (ZZ)	Voldoende	.56	.000	.41 - .74
Urgentie Zorg (UZ)	Matig	.45	.000	.31 - .65
Risico Jeugdige (RJ)	Voldoende	.54	.000	.39 - .72
Risico Omgeving (RO)	Voldoende	.68	.000	.54 - .83

* ICC-model one-way random/absolute/single

Resultaten per item

Volgens de wetten van de schaalconstructie hoeft het geen probleem te zijn als items onvoldoende betrouwbaarheid laten zien. Belangrijker is dat de schalen op voldoende niveau scoren. Niettemin is aandacht voor de betrouwbaarheid van de items zinnig om 'zwakke plekken' in het instrument op het spoor te komen. Als we naar het itemniveau kijken, zien we dat de helft van de items van de schaal 'Functioneren Jeugdige' voldoende scoort en de helft matig. De items 'Hoe lang bestaan de problemen?' (FJ2), 'Hoe erg is de jeugdige van slag?' (FJ3) en 'Belemmering functioneren van de jeugdige op vier leefgebieden' (FJ4) scoren lager dan .51 en worden om die reden gekwalificeerd als matig. De overige drie items scoren hoger dan .50 en worden daarmee gekwalificeerd als voldoende. De items 'Zijn er problemen op de belangrijkste aspecten van de overige omgeving?' (KO2) en 'Hoe lang bestaan voorkomende problemen in de omgeving?' (KO3) scoren matig, de overige items op de schaal Kwaliteit Omgeving (KO) scoren voldoende. De items op de schaal Zorgzwaarte (ZZ) scoren allemaal onder de .51 (dus matig). Eén van de vier items op de schaal Risico Jeugdige (RJ) scoort matig, de rest voldoende. Net als bij de schaal Functioneren Jeugdige, vindt men de

[16] De scores worden als volgt gekwalificeerd: -1.00 - .30 is slecht, .31 - .50 is matig, .51 - .70 is voldoende en .71 - 1.00 is goed.

lijdensdruk van de jeugdige lastig te beoordelen. Beide items op de schaal Risico Omgeving (RO) scoren voldoende. De analyse op itemniveau levert aanwijzingen waarop de handleiding van de STEP te verbeteren valt, zodat gebrekkige overeenstemming minder zal voorkomen. Dit zal de betrouwbaarheid van het instrument op schaalniveau verder doen stijgen.

Resultaten per vignet
Per vignet is met behulp van de intraclass correlatiecoëfficiënt (ICC) de algemene overeenstemming bepaald op alle schalen en voor alle codeurs tegelijk. Voor vrijwel alle vignetten geldt (met uitzondering van één vignet), dat de overeenstemming op schaalniveau goed is. Zowel wat betreft de vier *QUICK*STEPschalen als de twee risicoschalen. Op itemniveau is de overeenstemming wat lager: de items van de vier *QUICK*STEPschalen scoren matig tot goed, de items van de risicoschalen slecht tot voldoende. Hieruit kunnen we concluderen dat eventuele verschillen in beoordeling op itemniveau binnen de schaal worden gecompenseerd.

Resultaten per codeur
Tot slot is nagegaan hoe de codeurs scoren, vergeleken met andere codeurs die dezelfde vignetten hebben beoordeeld. Per vignet is berekend wat de gemiddelde scores zijn, vervolgens is gekeken hoever elke codeur van die gemiddelde score afwijkt. De vuistregel is: hoe kleiner het verschil is met het gemiddelde van alle codeurs, hoe betrouwbaarder die codeur beoordelingen maakt. Elke codeur krijgt een betrouwbaarheidscode op basis van het aantal afwijkingen ten opzichte van andere codeurs. Hoe kleiner het aantal afwijkingen, hoe hoger de score. Uit de analyses blijkt dat 20 codeurs op een redelijk betrouwbaarheidsniveau staan en 48 codeurs (60%) een voldoende tot goed betrouwbaarheidsniveau behalen. Er zijn echter ook codeurs zijn die matig tot slecht met de anderen overeenstemmen. Nadere analyses waarin is nagegaan of hulpverleners die afwijkend scoren zich kenmerken met een bepaalde mate van ervaring in de jeugdzorg of met de STEP laten geen duidelijke trends zien. Verdere inspectie van de gegevens laat zien dat enkele hulpverleners soms zeer afwijkend van de rest de STEP hebben ingevuld. Om dat te voorkomen is een goede training in combinatie met een regelmatige intervisie met betrekking tot de scoring van het instrument sterk aan te bevelen. Het voorkomt dat beoordelaars teveel hun eigen interpretaties in het materiaal en het scoren van de items gaan leggen.

4.4 Conclusies

De resultaten van dit onderzoek laten zien dat de interbeoordelaarbetrouwbaarheid van de STEP voldoende is. Dit houdt in dat de scoring op het instrument niet te zeer afhankelijk is van de hulpverlener die het invult. Indien de handleiding van de STEP op een aantal punten verbeterd wordt, kan de interbeoordelaarbetrouwbaarheid van de STEP nog verder vergroot worden. Voorts is – zoals geldt voor het gebruik van alle instrumenten - training en geregelde intervisie voor een goed hanteren van het instrument van belang.

Literatuur

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition, Text Revision*. Washington DC: American Psychiatric Association.

Bakker, K. (1999). Sociale kwetsbaarheid en sociale competentie: een kaderstelling. In: K. Bakker, M. Pannebakker & J. Snijders (Red.). *Kwetsbaar en competent. Sociale participatie van kwetsbare jeugd. Theorie, beleid en praktijk*. Utrecht: NIZW.

Berben, E., Konijn, C., Verheij, F., Donker, M., Steketee, M. Roede, E. & Savorin Lohman, J. de (1997). *Grensvlakproblematiek in de jeugdzorg*. Rotterdam / Amsterdam: Erasmus Universiteit/SCO Kohnstamm Instituut.

Berben, E.G.M.J. (2000*). Als iedereen hetzelfde was… Indicatiestelling in de jeugdzorg*. Maastricht: Shaker Publishing B.V.

Dalgleish, L.I. (1997). *Risk assessment and decision making in child protection*. Brisbane, Australia: The University of Queensland, Department of Psychology.

Eijgenraam, K. (2006). *Beslissen is een werkwoord. Handreikingen voor het besluitvormingsproces in bureau jeugdzorg*. Utrecht: NIZW Jeugd.

Felce, D. & Perry, J. (1996) Assessment of Quality of Life. In: Schalock, R.L. & Siperstein, G.N. (ed.) *Quality of Life. Conceptualization and Measurement. American Association on Mental Retardation* (AAMR).

Fleiss, J.L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613-619.

Groenendaal, J.H.A. & Yperen, T.A. van (1994). Beschermende en bedreigende factoren. In: Rispens, J., Goudena, P.P., & Groenendaal, J.J.M. (Red.). *Preventie van psychosociale problemen bij kinderen en jeugdigen* (pag. 90-118). Houten: Bohn Stafleu Van Loghum.

Kroes, G., (2006). *The perception of child problem behavior. The role of informant personality and context*. Academisch proefschrift. Nijmegen: Radboud Universiteit Nijmegen.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Mezzich A.C., Mezzich J.E. & Coffman, G.A.(1985). Reliability of DSM III vs. DSM II in child psychopathology. *Journal of the American Academy Of Child Psychiatry*, 24, p.273-280.

Nasuti, J.P. & Pecora P.J. (1993). Risk assessment scales in child protection: a test of the internal consistency and interrater reliability of one statewide system. *Social Work Research and Abstracts, 29*, 28-35.

Pelzer, H.J., Steerneman, W.J.P.J.M., & Bruyn, E.E.J. de (1999). De ernst van het probleemgedrag: een conceptuele verkenning. In: Pelzer, H. & Steerneman, P. (Red.). *De taxatie van de ernst van de problematiek bij kinderen en jeugdigen: de ontwikkeling van een praktijkinstrument. Academisch proefschrift*. Nijmegen: Uitgeverij KU Nijmegen.

Remschmidt, H., Schmidt, M. & Göbel, D. (1983). Erprobungs- und Reliabilitätsstudie zum multiaxialen Klassifikationsschema für psychiatrische Erkrankungen im Kindes- und Jugendalter. In: Remschmidt, H., Schmidt, M. (Hrsg.). *Multiaxiale Diagnostik in der Kinder- und Jugendpsychiatrie. Ergebnisse empirischer Untersuchungen*. Bern: Huber.

Rutter, M., Shaffer, D., & Shepherd, M. (1975). *A multi-axial classification of child psychiatric disorders*. Geneva: World Health Organization.

Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 2, 420-428.

Spitzer, R.L., Endicott, J. & Robins, E. (1975). Clinical criteria for psychiatric diagnosis and DSM-III. *American Journal of Psychiatry*, 132, 1187-1192.

Veenma, K., Batenburg, R. & Breedveld, E. (2004). *De vignetmethode. Een praktische handreiking bij beleidsonderzoek*. Tilburg/Utrecht: IVA.

World Health Organization (1996). *Multiaxial classification of child and adolescent psychiatric disorders*. Cambridge: Cambridge University Press.

Yperen, T.A. van (1990). *Multi-axiale classificatie van specifieke ontwikkelingsstoornissen. Een studie over as II van het MAC.* Proefschrift, Rijksuniversiteit Leiden.

Yperen, T.A. van (1995). *Het gebruik van instrumenten. Registratie in de jeugdzorg*. Utrecht: NIZW Uitgeverij.

Yperen, T. van, Berg, G. van den & Eijgenraam, K. (2002). *Project 'Registratie ernst van de problematiek'. Eerste deelrapport: begrippen, doelen en instrumenten*. Utrecht: NIZW Jeugd.

Yperen, T. van, Berg, G. van den & Eijgenraam, K. (2003a). *Standaard Taxatie Ernst Problematiek (STEP). Handleiding en Formulieren. Tweede deelrapport in het project 'Registratie ernst van de problematiek'.* Utrecht: NIZW Jeugd.

Yperen, T. van, Berg, G. van den & Eijgenraam, K. (2003b). *Standaard Taxatie Ernst Problematiek (STEP). Derde deelrapport in het project 'Registratie ernst van de problematiek'.* Utrecht: NIZW Jeugd.

Yperen, T. van, Berg, G. van den & Eijgenraam, K. (2003c). *QUICKSTEP. Snelle Standaard Taxatie Ernst Problematiek. Handleiding*. Utrecht: NIZW Jeugd.

Yperen, T. van, Berg, G. van den, Eijgenraam, K. & Graaf, M. de (2006). *(QUICK)Step: Snelle Standaard Taxatie Ernst Problematiek. Handleiding*. Utrecht: Nederlands Jeugdinstituut/NJi.

Yperen, T.A. van, Roosma, D. & Veerman, J.W. (In druk). Instrumenten voor meten van uitkomsten en uitvoering van de zorg. In: T.A. van Yperen & J.W. Veerman (Redactie). Zicht op effectiviteit. *Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg*. Utrecht/Nijmegen: NJi/Praktikon.

Woord van dank

Het betrouwbaarheidsonderzoek is tot stand gekomen met medewerking van een groot aantal personen. Wij danken daarvoor managers en hulpverleners van de bureaus jeugdzorg Limburg, Zeeland, Friesland en Agglomeratie Amsterdam en die van de William Schrikker Groep.

Voorts is dit onderzoek mogelijk gemaakt door financiering via het Kennisprogramma Jeugd van het ministerie voor Jeugd en Gezin, het ministerie van VWS en het ministerie van Justitie.

Bijlage: STEP-formulier

Op de volgende pagina's is het STEP-formulier als bijlage opgenomen (in de vorm zoals het in de verschillende deelonderzoeken wordt gebruikt). Dit formulier bestaat uit:

- Vragenlijst Achtergrondgegevens
- Scoringshulpen bij de zes schalen
- STEP Ernstprofiel

STANDAARD TAXATIE ERNST PROBLEMATIEK (STEP)

Vragenlijst achtergrondgegevens

t.b.v. onderzoek 2006-2007

Vul onderstaande gegevens **zorgvuldig** en **volledig** in.

Deze gegevens zijn van belang voor het onderzoek: **alléén** formulieren met **volledig ingevulde achtergrondgegevens** kunnen in de kwaliteitsanalyse worden verwerkt (indien er achtergrondgegevens ontbreken, zijn de ingevulde STEP-formulieren onbruikbaar voor bepaalde analyses).

GEGEVENS INVULLER

Naam invuller	**Instelling**	**Datum van invullen** (dd-mm-jjjj)
		__ __ - __ __ - __ __ __ __

GEGEVENS JEUGDIGE

Dossiernummer jeugdige	**Geboortedatum jeugdige** (dd-mm-jjjj)	**Sekse jeugdige***
	__ __ - __ __ - __ __ __ __	☐ jongen ☐ meisje

Culturele achtergrond van de jeugdige*	**Leefverband***
☐ Nederlands ☐ Antilliaans ☐ Turks ☐ Anders, nl. ..... ☐ Marokkaans ☐ Gemengd, nl. ..... ☐ Surinaams en	☐ Thuis ☐ Bij familie / vrienden ☐ Alleen wonend ☐ Anders, nl. .....
Huidige gezinssituatie*	**Onderwijs / dagbesteding***
☐ Tweeoudergezin ☐ Eénoudergezin ☐ Niet (meer) relevant vanwege leeftijd en leefverband jeugdige	☐ School ☐ Werkend ☐ Anders, nl.....
Aard problematiek Aankruisen wat van toepassing is. Hier kunnen *meerdere* hokjes aangekruist worden! De rubrieken corresponderen met de domeinen uit ISIS-tabel ☐ Psychosociaal functioneren jeugdige ☐ Lichamelijke gezondheid, aan lichaam gebonden functioneren ☐ Vaardigheden en verstandelijke ontwikkeling ☐ Gezin en opvoeding ☐ Omgeving jeugdige ☐ Overige problemen ☐ Niet gespecificeerde problematiek	**Juridisch kader*** ☐ Vrijwillige hulp ☐ OTS / VOTS ☐ Voogdij / voorlopige voogdij ☐ Jeugdreclassering ☐ Anders, nl. .....

* Aankruisen wat van toepassing is

Informatie over de STEP is te verkrijgen bij:
Nederlands Jeugdinstituut / NJi, Postbus 19221, 3501 DE Utrecht
Telefoon (030) 230 66 34, E-mail: r.schouten@nji.nl
Kijk voor achtergrondinformatie ook op www.jeugdzorg.nl

Nederlands Jeugd instituut

Zelf dit formulier kopiëren?
Kopieer pagina 1 t/m 3 dubbelzijdig op A3-formaat (pagina 1 aan één kant, pagina 2 en 3 naast elkaar op de andere kant) en vouw dit papier dubbel.
Kopieer pagina 4 (ernstprofiel) op A4-formaat en voeg dit los toe.

*QUICK*STEP - *SNELLE* STANDAARD TAXATIE ERNST PROBLEMATIEK
SCORINGSHULP

t.b.v. onderzoek 2007-2007

1. Heeft de jeugdige problemen op één of meer van de aspecten van persoonlijk functioneren?

[1] Geen of hoogstens normale problemen | **[2]** Kleine of lichte problemen | **[3]** Matige problemen | **[4]** Zware problemen | **[5]** Zeer zware tot extreme problemen

2. Hoe lang bestaan voorkomende problemen van de jeugdige?

[1] N.v.t. of hoogstens 1 week | **[2]** 2 tot 4 weken | **[3]** 5 weken tot 5 maanden | **[4]** 6 tot 12 maanden | **[5]** Meer dan 12 maanden

3. Hoe erg is de jeugdige door persoonlijke of omgevingsproblemen van slag?

[1] N.v.t. / niet van slag | **[2]** Een beetje van slag | **[3]** Tamelijk van slag | **[4]** Erg van slag | **[5]** Totaal van slag

4. a. Belemmert het functioneren van de jeugdige zijn dagelijks leven op de volgende vier leefgebieden?

Functioneren jeugdige thuis:	**[0]** *Helemaal niet*	**[5]** *Een beetje*	**[10]** *Tamelijk veel*	**[25]** *Heel erg*
Functioneren jeugdige in betekenisvolle relaties:	**[0]** *Helemaal niet*	**[5]** *Een beetje*	**[10]** *Tamelijk veel*	**[25]** *Heel erg*
Gedrag op crèche / leren op school / werk:	**[0]** *Helemaal niet*	**[5]** *Een beetje*	**[10]** *Tamelijk veel*	**[25]** *Heel erg*
Functioneren in overige omgeving*:	**[0]** *Helemaal niet*	**[5]** *Een beetje*	**[10]** *Tamelijk veel*	**[25]** *Heel erg*

* - functioneren buiten bekende personen of situaties (bijv. ongewone reacties op vreemden of op andere omgeving) en/of
- maatschappelijk functioneren (vrije tijd, sociaal netwerk, wonen, omgang met normen/geld/instanties/autoriteiten)

b. Tel de gescoorde punten bij vraag 4a op en geef hieronder aan in welke categorie de totaalscore valt.

[1] N.v.t. of 0–9 punten | **[2]** 10-19 punten | **[3]** 20-50 punten | **[4]** 51-80 punten | **[5]** 81-100 punten

5. Vormt het functioneren van de jeugdige een belasting voor de overige leden van de thuissituatie?

[1] N.v.t. of geen belasting | **[2]** Een beetje een belasting | **[3]** Tamelijk grote belasting | **[4]** Een erge belasting | **[5]** Een ondraaglijke belasting

6. Vormt het functioneren van de jeugdige een belasting voor de omgeving buiten de thuissituatie (bijvoorbeeld voor de chrèche, de school, de werksituatie, de maatschappij)?

[1] N.v.t. of geen belasting | **[2]** Een beetje een belasting | **[3]** Tamelijk grote belasting | **[4]** Een erge belasting | **[5]** Een ondraaglijke belasting

Functioneren Jeugdige (FJ) totaal: tel de punten van de aangekruiste hokjes [ ] bij de vragen 1-3, 4b, 5 en 6 op →

7. Zijn er problemen met de kwaliteit van de primaire opvoedingsomgeving óf (indien een opvoedingsomgeving niet aan de orde is) zijn er problemen met de kwaliteit van het primaire sociale netwerk van de jeugdige?

[1] Hoogstens normale problemen | **[2]** Kleine of lichte problemen | **[3]** Matige problemen | **[4]** Zware problemen | **[5]** Zeer zware tot extreme problemen

8. Zijn er problemen op de belangrijkste aspecten van de overige omgeving?

[1] Hoogstens normale problemen | **[2]** Kleine of lichte problemen | **[3]** Matige problemen | **[4]** Zware problemen | **[5]** Zeer zware tot extreme problemen

9. Hoe lang bestaan voorkomende problemen in de omgeving?

[1] N.v.t. of hoogstens 1 week | **[2]** 2 tot 4 weken | **[3]** 5 weken tot 5 maanden | **[4]** 6 tot 12 maanden | **[5]** Meer dan 12 maanden

10. Is er iemand die de jeugdige bij problemen steunt (opvangt, problemen helpt oplossen)?

[1] Er is goede steun | **[2]** Er is redelijke steun | **[3]** Er is matige steun | **[4]** Er is weinig steun | **[5]** Er is geheel geen steun

11. Hoezeer vormt de omgeving (alles bij elkaar) een risicofactor voor de jeugdige?

[1] Geen risicofactor | **[2]** Een beetje een risicofactor | **[3]** Tamelijk grote risicofactor | **[4]** Grote risicofactor | **[5]** Zeer grote risicofactor

Kwaliteit Omgeving (KO) totaal: tel de punten van de aangekruiste hokjes [ ] bij vraag 7-11 op →

12. Wat voor soort zorg is er nodig?

[1] Geen of hoogstens enkele contacten via telefoon of internet | **[3]** Vrij toegankelijke, ambulante zorg | **[6]** Geïndiceerde ambulante zorg | **[8]** Diagnostiek of behandeling in combinatie met partieel verblijf | **[10]** Diagnostiek of behandeling met 24-uurs verblijf (pleegzorg of residentiële zorg)

13. Hoe lang zal het hulpverleningstraject naar schatting duren?

[1] N.v.t. of zeer kort (hoogstens een maand) | **[2]** Kort (hoogstens 3 maanden) | **[3]** Matig lang (hoogstens 6 maanden) | **[4]** Lang (7 tot 12 maanden) | **[5]** Zeer lang (meer dan 12 maanden)

14. Hoe groot is de intensiteit van de contacten?

[1] N.v.t. of zeer licht (hoogstens 1 contact per maand) | **[2]** Licht (hoogstens 2-4 contacten per maand) | **[3]** Matig zwaar (gemiddeld meer dan 1 contact per week) | **[4]** Zwaar (1 tot 5 dagen per week) | **[5]** Zeer zwaar (6-7 dagen per week)

Zwaarte Zorg (ZZ) totaal: tel de punten van de aangekruiste hokjes [ ] bij vraag 12-14 op →

15. Hoe zwaar schat u de urgentie van de in vraag 12-14 beschreven zorg in?

[1] De zorg kan zonder nadere tijdsbepaling uitgesteld worden | **[2]** De interventie kan zeker tot 12 weken (3 maanden) uitgesteld worden | **[3]** Interventie binnen 4 weken vereist | **[4]** Interventie binnen 5 dagen vereist | **[5]** Interventie binnen 24 uur vereist

Urgentie Zorg (UZ): neem de punten van vraag 15 over →

RISICO JEUGDIGE (STEP RJ)
SCORINGSHULP

1. **Hoe zwaar zullen na zes maanden de problemen van de jeugdige op één of meer van de aspecten van persoonlijk functioneren zijn als interventie(s) zou(den) uitblijven?**

[1] Grote kans op hoogstens normale problemen | **[2]** Grote kans op slechts lichte problemen | **[3]** Grote kans op matige problemen | **[4]** Grote kans op zware problemen | **[5]** Grote kans op zeer zware tot extreme problemen

2. **Hoe erg zal de jeugdige na zes maanden van slag zijn als interventie(s) zou(den) uitblijven?**

[1] N.v.t. of niet van slag | **[2]** Een beetje van slag | **[3]** Tamelijk van slag | **[4]** Erg van slag | **[5]** Totaal van slag

3. **Als interventie(s) zou(den) uitblijven, zal het functioneren van de jeugdige dan na zes maanden zijn dagelijks leven op één of meer van de volgende leefgebieden verslechteren?**
- functioneren thuis
- functioneren in betekenisvolle relaties
- gedrag op crèche / leren op school / presteren op werk
- maatschappelijk functioneren

[1] N.v.t. of in het geheel niet | **[2]** Hoogstens een beetje | **[3]** Hoogstens tamelijk veel op een of meerdere gebieden | **[4]** Op enkele (maar niet alle) heel erg | **[5]** Op alle vier heel erg

4. **Als interventie(s) zou(den) uitblijven, in welke mate zal dan na zes maanden het functioneren van de jeugdige een belasting vormen voor de overige leden van de thuissituatie?**

[1] N.v.t. of geen belasting | **[2]** Een beetje een belasting | **[3]** Tamelijk grote belasting | **[4]** Een erge belasting | **[5]** Een ondraaglijke belasting

Risico Jeugdige (RJ) totaal: tel de punten van de aangekruiste hokjes [ ] bij vraag 1-4 op →

RISICO OMGEVING (STEP RO)
SCORINGSHULP

5. **Als interventie(s) zou(den) uitblijven, in welke mate zal dan na zes maanden het functioneren van de jeugdige een belasting vormen voor de omgeving buiten de thuissituatie (bijvoorbeeld voor de crèche, de school, de werksituatie, de maatschappij)?**

[1] N.v.t. of geen belasting | **[2]** Een beetje een belasting | **[3]** Tamelijk grote belasting | **[4]** Een erge belasting | **[5]** Een ondraaglijke belasting

6. **Hoe groot is het risico dat de jeugdige na zes maanden (opnieuw) een strafbaar feit pleegt als interventie(s) zou(den) uitblijven?**

[1] Geen of normaal risico | **[2]** Licht risico | **[3]** Matig risico | **[4]** Aanzienlijk risico | **[5]** Ernstig tot extreem risico

Risico Omgeving (RO) totaal: tel de punten van de aangekruiste hokjes [ ] bij vraag 5 en 6 op →

*(QUICK)*STEP - ERNSTPROFIEL
INLEGVEL

t.b.v. onderzoek 2006-2007

Omzetting totaalscores Scoringshulpen naar STEP-niveaus

Naam invuller	**Naam / Dossiernummer / code jeugdige**	**Datum van invullen** (dd-mm-jjjj) __ __ - __ __ - __ __ __ __

In het onderstaande schema kunt u de totaalscores van de Scoringshulpen omzetten naar STEP-niveaus.

Omcirkel bij elke schaal onder de pijl (▼) de totaalscore van de bijbehorende scoringshulp. U ziet dan bij welk standaard-ernstniveau van de STEP elke totaalscore ongeveer past. Als u de omcirkelde scores met een lijn verbindt, ontstaat er een 'ernstprofiel'.

De onderstaande omzettingstabel is experimenteel en wordt op grond van onderzoek geijkt. Soortgelijke tabellen worden voorbereid voor de omzetting van scores op andere instrumenten naar de STEP-standaard.

Zwaarte problematiek		**Zwaarte en urgentie zorg**		**Risicotaxatie (na 6 maanden, *zonder* interventie)**	
Functioneren Jeugdige (STEP FJ)	**Kwaliteit Omgeving (STEP KO)**	**Zwaarte Zorg (STEP ZZ)**	**Urgentie Zorg (STEP UZ)**	**Risico Jeugdige (STEP RJ)**	**Risico Omgeving (STEP RO)**
Omcirkel hier totaalscore Scoringshulp ▼	Omcirkel hier totaalscore Scoringshulp ▼	Omcirkel hier totaalscore Scoringshulp ▼	Omcirkel hier totaalscore Scoringshulp ▼	Omcirkel hier totaalscore Scoringshulp ▼	Omcirkel hier totaalscore Scoringshulp ▼
Jeugdige 6 **1** functioneert 7 | (erg) goed 8 | 9 | 10 | 11 | Functioneren 12 **2** is redelijk 13 | goed 14 | 15 | 16 | 17 | Matige 18 **3** problemen in 19 | functioneren 20 | 21 | 22 | 23 | Aanzienlijke 24 **4** problemen in 25 | functioneren 26 | 27 | 28 | 29 | Zware tot 30 **5** extreme problemen in functioneren **0** Onvoldoende informatie	(Zeer) goede 5 **1** kwaliteit 6 | omgeving 7 | | 8 | 9 | Redelijk 10 **2** goede 11 | kwaliteit 12 | | 13 | 14 | Matige 15 **3** kwaliteit 16 | omgeving 17 | | 18 | 19 | Aanzienlijke 20 **4** problemen in 21 | kwaliteit 22 | | 23 | 24 | Zware tot 25 **5** extreme problemen in kwaliteit **0** Onvoldoende informatie	Geen zorg of 3 **1** minimale zorg- 4 | zwaarte 5 | 6 | | 7 | 8 **2** Geringe 9 | zorgzwaarte | 10 | | 11 | 12 **3** Matige 13 | zorgzwaarte | 14 | | 15 | Aanzienlijke 16 **4** zorgzwaarte 17 | | 18 | | 19 | Zeer hoge 20 **5** zorgzwaarte **0** Onvoldoende informatie	Geen zorg of 1 **1** geen urgentie | van zorg | | | | Weinig 2 **2** urgentie | zorg | | | | Matige 3 **3** urgentie | zorg | | | | Aanzienlijke 4 **4** urgentie zorg | | | | Zeer hoge tot 5 **5** extreme urgentie zorg **0** Onvoldoende informatie	Prognose van 4 **1** functioneren 5 | (erg) goed | 6 | | 7 | Prognose 8 **2** functioneren 9 | redelijk goed | 10 | | 11 | Prognose van 12 **3** matige 13 | problemen | 14 | | 15 | Prognose van 16 **4** aanzienlijke 17 | problemen | 18 | | 19 | Prognose van 20 **5** zware tot extreme problemen **0** Onvoldoende informatie	Geen/normaal 2 **1** risico/last | omgeving | 3 | | | Lichte last of 4 **2** risico voor | omgeving | 5 | | | Matige last of 6 **3** risico voor | omgeving | 7 | | | Aanzienlijke last 8 **4** of risico | omgeving | 9 | | | Zware tot 10 **5** extreme last of risico omgeving **0** Onvoldoende informatie