

Het ontstaan van schoolproblemen (II)*

Gedragsobservatie door kleuterleidsters

J. HERMANN

Afdeling Ontwikkelingspsychologie, Rijksuniversiteit te Utrecht

Samenvatting

Een observatieschaal van 22 items werd voor ongeveer 100 kleuters door hun leerkrachten tweemaal met een tussenpoos van 14 maanden, ingevuld. Subschalen waren: taakspanning, zelfstandigheid, aanspreekbaarheid en sociaal contact. De resultaten werden vergeleken met die van tests voor de cognitieve ontwikkeling die tegelijkertijd afgenomen werden. Veranderingen in niet-verbaal IQ in de loop van het onderzoek bleken vooral samen te hangen met scores op de subschaal taakspanning [zelfstandigheid. Dit effect was duidelijker voor meisjes dan voor jongens. De stabiliteit van probleemgedrag bleek niet zeer hoog. Met name gold dit voor jongens en voor kinderen uit hogere socio-economische nivo's.

1. Inleiding

Gedragsobservatie in de kleuterschool als methode van probleemsignalering mag zich de afgelopen jaren in toenemende belangstelling van gedragswetenschappers verheugen. In Nederland hebben de laatste tijd zeker 3 schalen bekendheid gekregen: de Schoolgedrag Beoordelingslijst (Schobl, Zaal 1978), de observatieschalen voor het speel-werkgedrag van de oudste kinderen van de kleuterschool (Stevens en Kuyk, 1976) en de vragenlijst van de Nijmeegse Schoolbekwaamheidstest (NST, Mönks e.a. 1978). Het aantal vragen- en observatielijsten dat door de schoolbegeleidende instellingen voor eigen gebruik ontwikkeld werd is zeker een veelvoud hiervan. In de Angelsaksische landen kunnen genoemd worden: de Children's Behaviour Questionnaire (Rutter, 1967), de Pupil Behaviour Rating Scale (Gerhart, 1973), De Devereux Elementary School Rating Scales (ook geschikt voor kleuters, Spivac en Swift 1966), de Pupil Rating Scales (Myklebust, 1971), De

Kohn Problem Checklist en Kohn Social Competence Scale (Kohn, Parnes en Rosman, 1976), de Preschool Attainment Record (Doll, 1965), de Preschool Behaviour Questionnaire (Behar en Stringfield, 1974), de Teacher Temperament Questionnaire (Chess en Thomas, pers. comm.) de Rhode Island Pupil Identification Scale (Novack, Bonaventura en Merenda, 1973), de Student Rating Scale (Feshbach, Adelman en Fuller, 1977) en een naamloze schaal van Stevenson e.a. (1976a). Hoewel de auteurs een grote diversiteit aan namen voor subschalen, factoren en clusters gebruiken, zijn de gebieden van functioneren van de kleuter die in het algemeen aan de orde komen relatief beperkt. Dit zijn veelal aspecten van werkhouding, aandacht, zelfstandigheid, aangepast gedrag in de klas, taalvaardigheid, oriëntatie in tijd en ruimte, sociaal gedrag, teruggetrokkenheid dan wel extraversie, motorische coördinatie en waarneming.

Studies naar de validiteit van de schalen vallen in drie groepen uiteen. Op de eerste plaats kunnen onderzoeken genoemd worden die een bepaalde observatieschaal afzetten tegen een andere maat voor eenzelfde gedrag of gedragingen. Vergelijkingen tussen verschillende observatieschalen (bijv. Feshbach e.a., 1974), vergelijking van een schaal met klinisch diagnostisch onderzoek (bijv. Rutter, 1967) of met beoordeling door de eigen ouders (Stevenson e.a., 1976a) behoren hiertoe. Meestal blijkt dat de observaties een redelijke, maar zeker niet perfecte correlatie vertonen met de criteriummaten. Onderzoek waarin observaties door de leerkracht worden vergeleken met observaties door anderen in gestandaardiseerde situaties zijn mij niet bekend.

Een tweede groep validiteitsstudies heeft als onderwerp de prediktieve waarde van observaties bij kleuters ten aanzien van het later functioneren van het kind in het lager onderwijs. De prognostische en probleem-signalerende waarde van een observatielijst kan aldus reëel getoetst worden. Feshbach e.a. (1977) onderzochten de prediktieve validiteit van de Student Rating Scale, ingevuld voor 5 tot 6 jarigen,

* Deel I van dit artikel is verschenen in *Pedagogische Studiën*, 1979 (56) 348-357.

ten aanzien van leesvorderingen in de volgende 3 schooljaren. De korrelatie van de totale SRS-score (de subschalen waren aandacht en gedragsbeheersing, taalvaardigheid, perceptuele discriminatie, motorische coördinatie en geheugen) met lezen één jaar later was .49, met lezen twee en drie jaar later resp. .52 en .39.

Stevenson e.a. (1976a) onderzochten de prediktieve validiteit van een kleuterobservatielijst van 13 items, betrekking hebbend op werkhouding, aandacht, sociale vaardigheden en cognitieve vaardigheden zoals lezen en geheugen. De multipele R van de 4 best voorspellende items met lezen in klas 1 tot en met 3 was resp. .50, .54 en .45. Een van de nadelen van deze studies is echter dat de observatieschalen een mengeling zijn van motivationele/emotionele én cognitieve gedragsaspecten. De waarde van een observatielijst is naar mijn mening vooral gelegen in het feit dat deze informatie geeft over gedragsaspecten zoals werkhouding, concentratie, zelfstandigheid en dergelijke *als aanvulling* op hetgeen uit psychometrische tests voor de cognitieve ontwikkeling naar voren komt. Het lijkt bijvoorbeeld weinig zin te hebben de kleuterleidster te vragen naar perceptuele vaardigheden van een kind, als deze (eventueel door de leidster zelf) eenvoudig met gestandaardiseerde taakjes onderzocht kunnen worden.

Uit bovenstaande onderzoeken wordt niet duidelijk in hoeverre de aanwezige prediktieve validiteit van de observatielijsten gebaseerd is op cognitieve dan wel op bijvoorbeeld motivationele aspecten ervan. Dat de items die betrekking hebben op meer cognitieve taken een duidelijke rol spelen blijkt bijvoorbeeld uit het onderzoek van Stevenson e.a. (1976a). De vier best voorspellende items van hun vragenlijst waren 'effective learning' (snapt iets snel), 'vocabulary, following instructions' (begrijpt en onthoudt opdrachten) en 'retaining information' (geheugen voor details en begrippen). De prediktieve validiteit van een observatieschaal berust aldus voor een deel op de vaardigheid van de kleuterleidster om de intelligentie en functie-ontwikkeling van kinderen te beoordelen. Het is dan ook niet verwonderlijk dat het gezamenlijk gebruik deze observatielijsten met psychometrische 'paper en pencil'-test niet veel meer extra informatie geeft dan ieder afzonderlijk. In het onderzoek van Stevenson e.a. (1976b) steeg de multiple R van een aantal tests met latere leesprestaties niet meer dan .05 door toevoeging van de observatiescores. Ook in het onderzoek van Feshbach e.a. (1977) was de prediktie op grond van de observatieschaal in combinatie met I.Q. en andere tests niet veel beter

dan de observatie-scores en de testcores afzonderlijk.

De vermenging van cognitieve functies en sociale/emotionele/motivationele gedragingen is niet aanwezig in de Kohn Social Competence Scale en de Devereux Elementary School Behaviour Rating Scale. De beide subschalen van Kohns lijst (interest-participation en coöperation-defiance) korreleren matig met later lezen: tussen .12 en .33 (Feshbach e.a. 1974 en 1976).

Het meest interessant is de studie van Wallbrown, Wallbrown en Engin (1977) met betrekking tot de Devereux-schaal. Deze observatielijst omvat 44 vragen voor 11 motivationele-emotionele onderwerpen: storende invloed in de klas, ongeduld, geen respect voor regels, schuld buiten zichzelf zoeken, faalangst, afhankelijkheid, begrijpen wat er gezegd wordt, aandacht, opvallend gedrag, creativiteit, 'hangen' aan leerkracht.

Multipele regressie-analyse van de relatie tussen 11 sub-schalen en schoolvorderingen gaf voor rekenen en lezen een Multiple R van .39 tot .53. De subschaal 'begrijpen wat er gezegd wordt' (comprehension) bleek steeds van groot gewicht bij het voorspellen van schoolvorderingen. Voor het leren rekenen bleek 'faalangst' van belang te zijn.

Deze studies hebben overigens met alle andere bovenstaande het bezwaar dat het gedrag in de lagere school niet meer geobserveerd is. Dit betekent dat niet bekend is in hoeverre dit gedrag van het kind in de loop van het onderzoek veranderd is. Het zou bijvoorbeeld zo kunnen zijn dat er alleen voor de kinderen die óók in de lagere school gedragsproblemen hebben, een korrelatie tussen gedrag en schoolvorderingen aanwezig is. Voor kinderen die geen continuïteit in het sociale/emotionele/motivationele gedrag vertonen zou de relatie tussen kleutergedrag en lagere schoolprestaties wel eens onduidelijk kunnen zijn. Ook is een nadeel van vrijwel alle Amerikaanse studies op dit gebied dat tussen de 'Kindergarten' en de 'first grade elementary school' grote delen van de onderzoekspopulatie van school veranderen. Hierdoor is de (waarschijnlijk selektieve) uitval in alle genoemde onderzoeken 30 à 40%.

Een derde groep van validiteitsonderzoeken die onderscheiden kan worden tracht niet alleen de relatie tussen het gedrag van de kleuter in de klas en het cognitieve functioneren later in de lagere school vast te stellen, maar probeert ook na te gaan op welke wijze dit gedrag samenhangt met *veranderingen* in het cognitieve functioneren. Een van de weinige studies op dit gebied is uitgevoerd door Zaal (1978). Een groep van 90 kleuters werd door de

leerkracht beoordeeld op de Schobl, met als subschalen 'vrijmoedigheid, werkhouding, sociale omgang en emotionaaliteit'. Diezelfde kinderen werden 2 jaar later nogmaals geobserveerd. Op beide tijdstippen werd ook een intelligentietest afgenomen (helaas geen schoolvorderingstoets). Van belang zijn de bevindingen dat sommige sociale-emotionele gedragsaspecten bij jonge kleuters samenhangen met veranderingen in I.Q.-scores. De resultaten waren evenwel nogal complex en verschillend voor jongens en meisjes. In de verdere loop van dit artikel zal nog enkele malen op dit onderzoek worden teruggekomen.

Samenvattend kan gesteld worden dat er een toenemende belangstelling bestaat voor observatielijsten voor het kleuteronderwijs. De opzet van verreweg de meeste studies laat geen konklusies toe over de specifieke wijze waarop gedrag in de klas samenhangt met het latere functioneren in school. De waarde van de gedragsobservatiemethoden bij het vroegtijdig signaleren van problemen in het kleuteronderwijs blijft hierom vooralsnog hypothetisch. Een verder exploratie van dit onderzoeksgebied lijkt zinvol.

In dit artikel wordt een longitudinaal onderzoek beschreven waarin nagegaan wordt welke de relatie is tussen motivationele/emotionele gedragingen van kleuters zoals geobserveerd door kleuterleidsters enerzijds en scores op individuele psychometrische tests voor een aantal aspecten van de cognitieve ontwikkeling anderzijds. Onderzocht zal worden of er een samenhang is tussen gedragsobservaties en veranderingen in het cognitief functioneren in de loop van de kleuterschoolperiode. Bijzondere aandacht zal besteed worden aan kinderen die extra

problemen in het functioneren in school ondervinden. In volgende artikelen zal de voortzetting van het onderzoek in de eerste klas lagere school aan de orde komen.

2. Opzet van onderzoek

2.1. Algemeen

In een eerder artikel werd de opzet van het onderzoek beschreven (Hermanns 1979a). Kort samengevat is deze als volgt: Een groep van 117 kleuters die op 1 september 1975 tussen 4 en 5 jaar oud was, werd in de periode november tot mei van dat schooljaar individueel getest. Door de kleuterleidster werd een observatielijst ingevuld. Het psychologisch onderzoek en de observatie werden ruim een jaar later bij 101 kinderen nogmaals uitgevoerd. De kinderen waren toen in de laatste maanden van het kleuteronderwijs.

In het onderstaande zullen de gegevens van de observaties en het individuele psychometrische onderzoek naar cognitieve functies in de kleuterschoolperiode aan de orde komen.

2.2. De tests voor de cognitieve ontwikkeling

Ook hier wordt volstaan met een kort overzicht. De tests die in beide onderzoeksfasen in de kleuterschool gebruikt werden, zijn in Tabel 1 weergegeven.

Uit de tabel kan worden afgelezen dat er 13 specifieke tests zijn en 3 samengestelde tests: het IQ., Utant en de totaal-score.

Tabel 1 *Individuele tests die bij kleuters op een gemiddelde leeftijd van 58 maanden en 72 maanden zijn afgenomen*

Sorteren	}	niet verbaal I.Q. (SON)	}	Totaal score
mozaïek				
puzzels				
geheugen				
figuren natekenen	}	Utrechtse Taalnivo Test (Utant)		
woordenschat				
analogieën/tegenstellingen				
grammatica				
onder woorden brengen	}			
woord-herkenning (visueel)				
ruimtelijke oriëntatie (Frostig)				
zinsgeheugen				
plaatjes benoemen				

2.3. De Kleuter Observatielijst

De in de inleiding genoemde observatielijsten werden op hun geschiktheid voor dit onderzoek beoordeeld. De belangrijkste criteria waren dat een breed gebied van functioneren in de klas aan de orde moest komen (in ieder geval taakhouding, zelfstandigheid, sociale ontwikkeling) en dat de items betrekking moesten hebben op zo concreet mogelijke, direct observeerbare gedragingen in de klas. Verder werd er op gelet dat invulling van de lijst niet meer dan een half uur per kind zou vergen. De Student Rating Scale van Feshbach e.a. (1974) kwam het dichtst bij de bedoeling (factoren waren: gedragsbeheersing en aandacht, taalvaardigheid, perceptuo-motorische coördinatie en geheugen). Aanvulling met items betreffende taakhouding, sociale ontwikkeling en motorische coördinatie bleek noodzakelijk. Ook de ingewikkelde wijze van skoren werd vervangen door een 5-punts schaal achter ieder item. Het resultaat was een lijst van 49 items, in willekeurige volgorde geplaatst. In de tweede onderzoeksfase is de lijst in wat uitgebreide vorm (56 items) nogmaals door de kleuterleidsters ingevuld. Factor-analyses ondersteunden de veronderstelde structuur van de lijst. Om de in dit artikel geformuleerde vraagstelling te kunnen beantwoorden werden slechts de volgende subschalen (totaal 22 items) bij de verdere analyses gebruikt.

a) *Taakspanning en zelfstandigheid* (het gekoncentreerd zelfstandig en gemotiveerd met iets bezig zijn).

Itemvoorbeelden: 'Kan het kind volhouden te werken aan een taak die het moeilijk vindt of waar het tegen op ziet?' en 'Laat het kind zich niet te snel afleiden als het met iets bezig is?'

b) *Aanspreekbaarheid* (het opletten, luisteren en zich aan de regels houden).

Itemvoorbeelden: 'Houdt het kind zich aan de regels van de klas?' en 'Kan het kind zijn beurt afwachten' en 'Heeft het kind zijn aandacht erbij als U de gehele klas iets gaat uitleggen?'

c) *Sociaal contact* (het samen met andere kinderen spelen en werken).

Itemvoorbeelden: 'Kan het kind in een groepje van drie of vier andere kinderen samenwerken?' en 'Beseft het kind dat als andere kinderen met bepaald materiaal spelen het er op dat moment niet mee kan spelen?'

De subschalen omvatten tussen 6 en 8 items.

De lijst werd door 11 verschillende kleuterleidsters in 5 scholen ingevuld. Bij het verwerken van de gegevens bleek dat de gemiddelde score per leidster

nogal wat verschilde. Sommige leidsters beoordeelden hun kinderen konsekvent 1 à 1½ punt hoger dan de andere. Zo varieerde het gemiddelde per schoolklas voor de subschaal 'aanspreekbaarheid' tussen 3.4 en 4.4. Het is uiteraard mogelijk dat dit verwijst naar werkelijke gedragsverschillen tussen kinderen. Waarschijnlijker is echter dat het hier gaat om verschillende antwoord-tendenties bij leidsters. Om deze reden werd de score van de kinderen op ieder afzonderlijk item uitgedrukt in termen van standaarddeviaties ten opzichte van hun klasgemiddelde, dat op 0 gesteld werd. De aldus resulterende Z-skores zijn dan een uitdrukking van de relatieve positie van een kind op dat item ten opzichte van zijn klas, waarbij de klassen onderling hetzelfde gemiddelde hebben. Verschillen tussen de verschillende schoolklassen werd zo verwaarloosd. Op de konsekventies die dit heeft voor de interpretatie van de resultaten zal later worden teruggekomen.

De skoreverdeling op de 22 items vertoont steeds een negatieve skewness en een positieve kurtosis; d.w.z. dat in vergelijking met de perfecte normaalverdeling er meer rechts van het gemiddelde geskoord wordt en dat de verdeling wat smaller en hoger is. Dit wijst erop dat de leerkrachten in het algemeen er toe neigen hun skores niet te ver van het gemiddelde te laten afwijken terwijl dit gemiddelde hier wat hoger ligt dan het midden van de 5-puntsschaal. Een duidelijke instructie vooraf waarbij gevraagd wordt alle categorieën te gebruiken rondom het midden van de schaal moet kennelijk veel meer nadruk krijgen. Een andere, maar voor de leerkracht nogal arbeidsintensieve oplossing, zou een Q-sort techniek kunnen zijn.

De interne consistentie van de sub-schalen taakspanning, aanspreekbaarheid en sociaal contact zoals gemeten met Cronbach's alpha was respectievelijk .89, .89 en .70 voor de eerste onderzoeksfase en .89, .87 en .78 voor de tweede. De correlaties tussen de schalen en de stabiliteit van de gedragsbeoordeling over de periode van ruim een jaar zijn in Tabel 2 weergegeven. Uit de correlaties tussen de subschalen blijkt dat de drie gedragsaspecten zeker niet onafhankelijk van elkaar zijn. Er blijft echter voldoende 'onafhankelijkheid' over om verdere analyses met de subschalen uit te voeren. De stabiliteit van de beoordelingen over de periode van één jaar is redelijk hoog voor 'aanspreekbaarheid' en 'sociaal contact' en ligt in dezelfde orde van grootte als Zaai (1978) voor de Schobl rapporteert. Het oordeel van (vrijwel steeds dezelfde) leerkracht over de 'taakspanning en zelfstandigheid' verandert nogal

Tabel 2 Pearson korrelaties tussen subschalen en stabiliteit van de subschalen over de periode van één jaar. De korrelaties in de eerste onderzoeksfase zijn boven de diagonaal geplaatst; die in de tweede fase eronder. De stabiliteitsscores (ook Pearson korrelaties) zijn op de diagonaal geplaatst.

	Taakspanning	Aanspreekbaarheid	Sociaal contact
Taakspanning	.36	.66	.59
Aanspreekbaarheid	.60	.50	.46
Sociaal contact	.50	.52	.56

gezien de koëfficiënt van .36. Wellicht dat kinderen juist op dit gebied in de loop van het kleuteronderwijs zich verschillend ontwikkelen.

Gegevens over de validiteit van de observatieschaal zullen in de volgende paragrafen aan de orde komen. Met name betreft het de criteriumvaliditeit (de mate van associatie tussen de testmeting en een of andere observeerbare meting, De Zeeuw, 1978).

Refererend aan de in de inleiding gemaakte onderverdeling in validiteitsonderzoek valt dit artikel in de groepen twee en drie (het voorspellen van een extern criterium en het voorspellen van verandering in zo'n criterium). In hoeverre de observatieschaal begripsvaliditeit heeft, d.w.z. in hoeverre bijvoorbeeld de items uit de subschaal 'taakspanning' verwijzen naar feitelijk bij het kind aanwezige taakspanning komt niet aan de orde. Van belang is derhalve steeds te bedenken dat het gaat om gedragsaspecten zoals de kleuterleidsters deze zien.

3. Resultaten

3.1. Inleiding

Enkele psychometrische gegevens met betrekking tot de observatieschaal zijn reeds gegeven. Nu zullen korrelatieve verbanden tussen de schaal, sexe, socio-ekonomische status en de individuele tests voor de kognitieve ontwikkeling besproken worden. Vervolgens wordt nagegaan in hoeverre de observaties in het begin van het onderzoek samenhangen met veranderingen in de testcores in de loop van het kleuteronderwijs. Tenslotte zal aandacht besteed worden aan kinderen die blijkens de observaties problemen ondervinden in het functioneren in de klas.

3.2. Enkele korrelatieve verbanden tussen socio-ekonomische status (SES), leeftijd en de observatieschalen

In Tabel 3 zijn de korrelaties tussen de drie subschalen in beide onderzoeksfasen met de leeftijd van het kind en het socio-ekonomisch nivo (ITS-klassificatie) van het gezin weergegeven. Er blijken significante korrelaties tussen de beoordeling door de kleuterleidster en het socio-ekonomische nivo van het gezin te bestaan (behalve voor 'aanspreekbaarheid' in de tweede onderzoeksfase). Van belang is hierbij op te merken dat het gaat om verschillen tussen kinderen in iedere klas afzonderlijk. De mogelijke verschillen tussen de klassen of scholen zijn vooraf zoveel mogelijk geëlimineerd door per item Z-scores per kleuterleidster te berekenen. In feite zullen de gegeven koëfficiënten hierdoor een onderschatting van de werkelijke gedragsverschillen

Tabel 3 Pearson korrelaties tussen de observatieschalen, leeftijd en socio-ekonomisch nivo. Het bovenste getal is de korrelatie in de eerste onderzoeksfase, het onderste die in de tweede fase

Subschaal		SES	leeftijd
Taakhouding	I	.30*	-.02
	II	.18*	-.02
aanspreekbaarheid	I	.16*	-.04
	II	.15	.00
sociaal contact	I	.19*	.19*
	II	.17*	.20*
totaal	I	.25*	.04*
	II	.20*	.07

* signifikant op minstens $p < .05$

zijn als men er van uit gaat dat de leerkrachten een referentiekader gebruiken dat zich niet alleen tot de eigen klas beperkt. De gevonden correlaties tussen SES en gedragsbeoordeling zijn in de eerste onderzoeksfase vergelijkbaar met die welke tussen SES en de tests voor cognitieve ontwikkeling gevonden werden. De correlaties met SON-I.Q., Utant en Totaalscore bijvoorbeeld zijn resp. .18, .30 en .21.

Anders is het in de tweede onderzoeksfase. Enkele maanden vóór de overgang naar de lagere school bleken de correlaties tussen de cognitieve tests en SES toegenomen tot .34, .39 en .40). Het verband tussen SES en de observatieschaal is echter niet sterker geworden; met name de correlatie SES-werkhouding is juist lager dan in de eerste onderzoeksfase gevonden werd. Men zou kunnen veronderstellen dat het kleuteronderwijs geen, of een afzwakkende invloed heeft op de relatie tussen motivationeel/emotioneel functioneren en SES. Dit in tegenstelling tot de invloed op de relatie cognitief functioneren-SES, die juist versterkt wordt. Een interessante hypothese wordt dan dat de kleuterschool sociale ongelijkheid in cognitieve functies versterkt, terwijl verschillen in motivationeel/emotioneel functioneren niet versterkt of misschien zelfs gekompenseerd worden.

Zaal (1978) vond géén relatie tussen de 5 schalen van zijn observatielijst en SES bij kleuters. Beide studies zijn op dit punt echter moeilijk vergelijkbaar. Zaal gebruikt een driedeling voor SES; in dit onderzoek zijn 5 categorieën gebruikt waardoor meer differentiatie, met name in de middelbare en hogere beroepen mogelijk was.

Een relatie met de leeftijd van de kinderen blijkt alleen te bestaan voor de subschaal 'sociaal contact' (in beide onderzoeksfasen). Dit wijst erop dat de leidster vragen over werkhouding en aanspreekbaarheid steeds beantwoord heeft in relatie tot 'wat men van een kind van die leeftijd mag verwachten'. Voor de vragen over het sociale gedrag bleek echter de leeftijdsfactor moeilijker onder controle te houden. Wellicht waren de verschillen tussen de oudere en jongere kinderen op dit gebied te groot om ze te kunnen negeren. Ook kan het zo zijn dat verschillen in sociaal gedrag juist opvallen omdat bijvoorbeeld in samenwerkingssituaties vergelijking van oudere en jongere kinderen direct mogelijk is. Overigens zijn ook hier de gevonden correlaties weliswaar significant, maar van geringe betekenis. Zaal (1978) vond dat geen van de vier subschalen van de Schobl bij kleuters met leeftijd correleerde. Het feit dat de schaal 'sociale omgang' van de Schobl niet, maar onze schaal 'sociaal contact' wel met

leeftijd correleert is moeilijk te verklaren. Wellicht dat de specifieke inhoud van de items hier een rol speelt. In de Schobl hebben de vragen vooral betrekking op de kwaliteit van het persoonlijk contact tussen kinderen; in onze schaal betreffen ze vooral samenwerkingsrelaties.

De verschillen in de beoordeling van *jongens en meisjes* werden getoetst door middel van T-tests. Op de jongere leeftijd werden verschillen gevonden in taakhouding, aanspreekbaarheid en de totale score ($p < .05$). Meisjes werden (niet onverwacht) gunstiger beoordeeld dan jongens. Een jaar later werden geen verschillen meer aangetroffen. Wellicht dat ten aanzien van de jongens een 'opvoedende' werking van het kleuteronderwijs uitgaat.

3.3. Samenhang met de cognitieve ontwikkeling

In Tabel 4 zijn de correlaties tussen de 3 subschalen en de 3 samengestelde tests weergegeven voor beide onderzoeksfasen. Er kan op grond van deze getallen gesproken worden van een consistente relatie tussen de verschillende gedragsaspecten en de maten voor de cognitieve ontwikkeling. Deze relatie is echter weinig geprononceerd. De hoogste koëfficiënt is .37. Nog geen 14% variantie wordt hierdoor verklaard. Ook blijkt dat de relatie a-specifiek is. De diverse aspecten van de observatielijst correleren ongeveer even hoog met de verschillende tests. De hoogste correlaties zijn die tussen de totaalscores van zowel observatielijst als testcores. Gekonkludeerd kan worden dat er een tendens te konstaten valt bij de kleuterleidsters om kinderen die een relatief hoger nivo van functioneren op verstandelijk gebied

Tabel 4 Pearson correlaties tussen de subschalen van de observatielijst, I.Q., Utant en totaalscore voor beide onderzoeksfasen. Het bovenste getal is de correlatie in de eerste onderzoeksfase, het onderste die in de tweede fase

Subschaal	tests			Totaal score
	SON-I.Q.	Utant		
Taakhouding	I	.29	.27	.31
	II	.25	.18	.23
aanspreekbaarheid	I	.34	.28	.35
	II	.26	.25	.32
sociaal contact	I	.30	.22	.28
	II	.25	.23	.29
totaal	I	.36	.30	.37
	II	.30	.26	.34

Alle correlaties zijn significant op minstens $p < .05$

hebben, een iets betere taakhouding, grotere aanspreekbaarheid en beter sociaal contact met leeftijdgenoten toe te kennen.

Men zou in eerste instantie kunnen denken aan een diffuus 'halo-effekt'. Kinderen die goed overweg kunnen met cognitieve taken zoals ze ook vaak in het kleuteronderwijs aan de orde komen, zullen niet snel als niet aanspreekbaar, met een slechte taakhouding of minder sociaal vaardig aangemerkt worden. Het zou dan echter zo moeten zijn dat taalvaardigheid, die in de klas direkt en vaak aan de orde komt, meer zou moeten samenhangen met de gedragsbeoordeling dan ruimtelijke-abstrakte vaardigheden. Utant-skores korreleren echter minder met de observatielijst dan de niet-verbale I.Q.-skores. Waarschijnlijker wordt dan ook de verklaring dat een relatief hoog nivo van cognitieve ontwikkeling bij de kinderen inderdaad samenhangt met de genoemde gedragsaspecten. Beide zouden bijvoorbeeld het gevolg kunnen zijn van een soort 'algemene competentie'-faktor. Ook zouden de cognitieve en motivationele/emotionele ontwikkeling elkaar kunnen beïnvloeden. Op deze samenhang zal nu verder ingegaan worden.

3.4. *Gedragsobservaties en veranderingen in het cognitieve functioneren*

In de vorige paragraaf werd gezinspeeld op een interactie tussen cognitieve en motivationele/emotionele gedragskenmerken. In de ontwikkelingspsychologische literatuur treft men de opvatting aan dat factoren als competentiemotief (White, 1973), Locus of control (Mischel, 1974), prestatie-motivatie (Hermans, 1971) samenhangen met aspecten van de cognitieve ontwikkeling. Zaal (1978) vond dat voor een periode van 2 jaar kleuters die vrijmoedig, ondernemend zijn en initiatief vertonen vooruitgaan in I.Q., terwijl kinderen die zich terugtrekken uit het sociaal contact juist achteruitgaan. In ons onderzoek werden de gedragskenmerken 'taakspanning, zelfstandigheid, aanspreekbaarheid en sociale omgang' ingebracht onder meer omdat vermoed werd dat deze van belang zijn voor de cognitieve ontwikkeling van de kleuter. De veronderstelling was in de eerste plaats dat er een verband is tussen tests voor de cognitieve ontwikkeling en gedragsbeoordelingen in het begin van de kleuterschoolperiode: zelfstandige kinderen die zich goed op een taak of spel kunnen concentreren, goed hun aandacht kunnen houden bij wat de juffrouw zegt, zullen relatief hoog skoren op verschillende tests voor cognitieve functies. De resulterende correlatie zou beschouwd kunnen worden als een

'eindprodukt' van de voorschoolse periode.

Op de tweede plaats kan echter een hypothese geformuleerd worden over een dynamisch verband tussen het gedrag in de klas en de cognitieve ontwikkeling: kinderen die op genoemde gedragsaspecten positief beoordeeld worden in het begin van de kleuterschoolperiode zullen in het daaropvolgende jaar zich sneller ontwikkelen op verstandelijk gebied dan andere kinderen. Aan de orde is hier niet het nivo van cognitieve functies, maar de relatieve verandering hierin.

Alvorens deze hypothese te vergelijken met de bevindingen van het onderzoek is het goed op het volgende methodologische punt te wijzen. De hierboven geformuleerde hypothese veronderstelt een oorzakelijk verband tussen het ene gedragsaspect (bijvoorbeeld motivatie) en verandering in een ander gedragsaspect (bijvoorbeeld intelligentie). Een dergelijk oorzakelijk verband is in beschrijvend, korrelationeel onderzoek niet vast te stellen. Beschreven kan slechts worden hoe het een met het ander samenhangt. Of er sprake is van invloed van het een op het ander, en in welke richting deze invloed gaat, kan slechts in experimentele studies onderzocht worden. Het zal hierom dus niet mogelijk zijn de genoemde hypothesen te toetsen. Wel kan gezien worden of er in de verzamelde gegevens een patroon te onderkennen is op grond waarvan een verder onderzoek naar de hypothesen al dan niet als zinvol bestempeld kan worden. In een eerder verslag over een deel van de gegevens van dit onderzoek werd bijvoorbeeld gevonden dat er een zeer duidelijke samenhang was tussen SES en intelligentieveranderingen in de loop van één jaar kleuteronderwijs (Hermanns, 1979b). Interpretatie van deze samenhang als oorzakelijk van aard is echter niet toegestaan. Het gaat hier om een exploratief beschrijvend en niet om een experimenteel manipulerend onderzoek.

Bekijken we nu in dit licht de vraag of gedragsbeoordeling in het begin van het kleuteronderwijs samenhangt met veranderingen in skores op de tests voor de cognitieve ontwikkeling. De techniek die gebruikt is om dit verband te onderzoeken is de stapsgewijze multiële regressie analyse. De afhankelijke (of te voorspellen) variabele is steeds één van de tests (I.Q., Utant of totaalskore) in de tweede onderzoeksfase. De onafhankelijke (of voorspellende) variabele is in eerste instantie de beginskore van de afhankelijke variabele (dus in de eerste onderzoeksfase). Aldus wordt gecontroleerd voor individuele verschillen op de variabele die voorspeld moet worden. Op gelijke wijze wordt voor 'SES' en 'leeftijd' gecontroleerd.

Vervolgens worden een of meer voorspellende variabelen zoals de scores op de observatieschalen ingebracht in de regressievergelijking. Zo kan nagegaan worden in hoeverre een bepaalde variabele samenhangt met veranderingen in een andere terwijl één of meer variabelen konstant gehouden kunnen worden. Deze methode verdient om een aantal statistische redenen de voorkeur boven berekeningen met verschilskores tussen eerste en tweede meting.

In Tabel 5 zijn de resultaten van deze procedure weergegeven met de I.Q.-score aan het eind van de kleuterschoolperiode als afhankelijke (of te voorspellen variabele). Konstant gehouden werden individuele verschillen in het beginnivo (I.Q.-score in de eerste onderzoeksfase), leeftijd en SES. De totaalscore over de drie observatieschalen is als onafhankelijke (of voorspellende) variabele het laatst in de regressievergelijking gebracht. In feite hadden ook de 3 subschalen afzonderlijk ingebracht kunnen worden. Aangezien er echter een aanzienlijke korrelatie bestaat tussen de 3 schalen (multicollinearity) zouden de resulterende uitkomsten onbetrouwbaar kunnen zijn. De resultaten met betrekking tot de totaalscore geven aldus een conservatieve, algemene indruk in hoeverre gedrag in de klas samenhangt met veranderingen in I.Q.-scores. Welke van de drie subschalen hierin het belangrijkste is komt later aan de orde. Eerder is ingegaan op de relatie tussen leeftijd en SES enerzijds en de verandering in I.Q. anderzijds (Hermanns, 1979b). Nu blijkt ook dat, onafhankelijk van deze twee variabelen, de beoordeling door de kleuterleidsters samenhangt met een verandering in I.Q.-scores. Kinderen die gunstiger werden beoordeeld op een totaal van 22 items, betrekking hebbend op taakhouding, aanspreekbaarheid en sociaal contact

Tabel 5 *Stapsgewijs Multiple Regressie Analyse met I.Q. op gemiddeld 6,0 jaar als afhankelijke variabele; I.Q. op gemiddeld 4;10 jaar, leeftijd en SES als konstant gehouden variabelen en de totaalscore op de observatieschaal als onafhankelijke variabelen (N = 96).*

Variabelen	MR met tweede I.Q.	signifikantie van de verandering
eerste I.Q.	.61	
SES	.65	p < .007
leeftijd	.67	p < .05
totaalscore		
observatielijst	.69	p < .04

gaan relatief meer vooruit op een I.Q. test. Aangezien juist in deze leeftijdsperioden nogal eens gerapporteerd werd over verschillende ontwikkelingspatronen voor jongens en meisjes (Zaal, 1978, Van der Laan, 1973) is deze analyse nogmaals uitgevoerd, maar nu voor beide sexen afzonderlijk. Het beschreven effect bleek voor jongens niet, voor meisjes wel op te treden. (MR = .72) Bij een verdere analyse met de subschalen van de observatielijst kwam naar voren dat 'taakspanning' als afzonderlijke subschaal ook bij jongens samenhangt met I.Q. veranderingen. Bij meisjes bleek bovendien 'sociaal contact' een significant effect op te leveren. Ten aanzien van de Utant-scores en de totaal testskores werd geen effect gevonden. Een uitgebreidere discussie van sexe-verschillen en een vergelijking van de resultaten met die van Zaal (1978) zou in dit kader te ver voeren. Volstaan wordt met de konklusie dat gedragskenmerken van kinderen zoals gezien door kleuterleidsters significant samenhangen met veranderingen in intelligentie. Het belang van het observeren van gedrag in de klas wordt hierdoor nogmaals duidelijk.

3.5 Kinderen met een minder gunstige gedragsbeoordeling

In een eerder artikel over dit onderzoek werd beschreven hoe een zogenaamde risikogroep werd samengesteld. Dit waren de kinderen met de zwakste scores op de verschillende tests voor het cognitieve functioneren. Als definitie van 'zwakste scores' werd genomen: 'op minstens 3 van de 13 tests bij de laagste 10% skoren'. Dit waren in de eerste onderzoeksfase 19, in de tweede fase 18 kinderen¹. Analoog hieraan is ook op grond van de gedragsobservatie een probleemgroep samengesteld. Om vergelijking van 'risikogroep' en 'probleemgroep' mogelijk te maken is gezocht naar even grote groepen. Indien als criterium gehanteerd wordt: 'op minstens één van de drie subschalen bij de laagste 10% skoren' resulteert in de eerste onderzoeksfase een probleemgroep van 18 kinderen en in de tweede een van 16¹. Een eerste vraag betreft nu de stabiliteit van de probleemgroep. Nagegaan is in hoeverre in de loop van ruim één jaar juist de minder gunstige gedragsobservaties aan-verandering onderhevig waren. De stabiliteitscores van de observatieschaal als geheel waren niet zeer hoog (zie Tabel 2). Ditzelfde blijkt ook te gelden voor de probleemgroep: van de 18 kinderen verdwenen er 8 uit deze groep. Opvallend hierbij is een geslachts-specifiek effect. In Tabel 6 is het verband tussen sexe en verandering in de probleemgroep weergege-

Tabel 6 *Veranderingen in de probleemgroep in de loop van één jaar in relatie met geslacht*

	jongens	meisjes	totaal
stabiel probleem	5	5	10
ten gunste veranderd	7	1	8
totaal	12	6	18

ven. Het wordt in de eerste plaats duidelijk dat er meer jongens dan meisjes negatief beoordeeld worden. De meerderheid van de jongens blijkt één jaar later een meer gewaardeerd gedrag te vertonen tegenover slechts 1 van de 6 meisjes. In de tweede onderzoeksfase komen er echter weer 5 jongens bij, zodat er uiteindelijk in de probleemgroep 6 meisjes en 10 jongens zijn. Het aantal jongens dat hiermee in één van beide jaren binnen de probleemgroep valt is zo 17, het aantal meisjes is 7. Dit beeld van een frekwenter probleemgedrag en een grotere instabiliteit bij jongens is volgens de verwachtingen. Wellicht dat hier een verklaring ligt voor de resultaten die in de vorige paragraaf aan de orde kwamen. Gevonden werd dat de relatie tussen motivationeel/emotioneel gedrag en veranderingen in de intelligentie voor meisjes wel, maar voor jongens niet aanwezig was. De wisselvalligheid of instabiliteit bij jongens is hiervoor waarschijnlijk verantwoordelijk.

Bezien we nu de overlap die bestaat tussen een 'risikogroep' (op grond van cognitieve tests) en de 'probleemgroep' (op grond van observaties). In beide onderzoeksfases blijkt de overlap 7 van de 18, respectievelijk 16 kinderen te zijn. De observatie door de kleuterleidster signaleert kennelijk een grotendeels andere groep dan door middel van tests gebeurt. Interessanter is de vraag of de kinderen die in cognitieve functies zich zo ontwikkelen dat ze niet meer tot de risicogroep behoren door de leidsters positiever beoordeeld werden dan de kinderen die zwak bleven functioneren. Het zou dan zo moeten zijn dat kinderen in de risicogroep die bovendien op grond van de observatiescores in de probleemgroep geplaatst werden (totaal 7 kinderen) bijzonder slechte ontwikkelingsmogelijkheden zouden moeten hebben. Dit is echter niet zo duidelijk. Twee van deze 7 kinderen bleken een jaar later op de cognitieve tests niet meer tot de risicogroep te behoren. Zelfs binnen deze extreme groep is de stabiliteit (en daarmee de voorspelbaarheid) niet optimaal. Als konsekwentie hiervan kan gezien worden dat labeling van kinderen als 'probleemkind' bijzonder ongewenst is. Niet alleen heeft de labeling

als zodanig negatieve bij-effecten, maar ook dient er rekening mee gehouden te worden dat ze na een tijd onjuist kan blijken te zijn. Zoals in de vorige paragraaf (Tabel 5) duidelijk wordt is er een relatie tussen SES en verandering in het cognitieve functioneren. In het vorige artikel werd beschreven hoe dit ook in de risicogroep het geval was: 20% van de kinderen uit de laagste helft van de SES-verdeling veranderen ten gunste, tegenover 80% van de kinderen uit de hoogste helft. Een dergelijk patroon blijkt zich ook voor de gedragsprobleemgroep voor te doen: Tabel 7 toont de gegevens. De Fisher's Exact Test levert een significantienivo op van $p < .05$. Ook ten aanzien van motivationeel/emotioneel gedrag blijken de socio-economisch minder bedeeden vaker en meer stabiel in de probleemgroep aanwezig te zijn.

Tabel 7 *Veranderingen in de probleemgroep in de loop van een jaar in relatie met SES*

observaties	SES		totaal
	laagste helft	hoogste helft	
stabiel negatief	8	2	10
voortuitgegaan	4	4	8
totaal	12	6	18

4. Slot

De drie meest in het oog springende resultaten ten aanzien van het onderzoek met de observatielijst zijn de volgende:

1. De intelligentieontwikkeling tussen het 5e en 6e levensjaar blijkt samen te hangen met geobserveerde motivationele/emotionele gedragsaspecten. Met name een goede taakspanning en zelfstandigheid is van belang. Ook in de studie van Zaal (1978) kwam dit naar voren. Voorlopig lijkt de beste verklaring hiervoor dat een goede competentie motivatie bij kinderen zorgt voor een gerichte en adequate informatieverwerking en exploratie als een basis voor de cognitieve ontwikkeling. Dit effect hoeft niet alleen in school op te treden, maar kan ook als een gegeneraliseerde basishouding gezien worden. Onderzoek van Riksen - Walraven (1977) laat bij ongeveer één jaar oude kinderen zien dat stimulering van een dergelijke motivatie in een periode van enkele maanden kan leiden tot een snellere ontwikkeling. Dat de subschaal 'aanspreekbaar-

heid' van onze observatielijst niet samenhangt met intelligentie-veranderingen is in overeenstemming met bovenstaande veronderstellingen. Dat de ontwikkeling van het taalnivo niet bleek samen te hangen met het geobserveerde gedrag is achteraf moeilijk te interpreteren. Een mogelijke verklaring is dat het taalnivo vooral gekonditioneerd wordt door omgevingsfactoren. In een eerdere analyse van de gegevens (Hermanns 1979 b) kwam naar voren dat veranderingen in Utant-skores duidelijk samenhangen met de specifieke school waarop het kind zat. Ten aanzien van IQ-veranderingen werd deze relatie in het geheel niet aangetroffen.

2. Er bleken in de loop van de onderzoeksperiode nogal wat wisselingen op te treden in de groep kinderen die op grond van de observaties 'probleemgroep' werd genoemd (8 van de 18). Vooral jongens gaven vaak verbetering, dan wel verandering ten ongunste te zien. Dit betekent dat men voorzichtig moet zijn bij het onderkennen en behandelen van problemen bij kinderen van deze leeftijd. Anderzijds is het zo dat de meerderheid der problemen die aan het eind van de kleuterschoolperiode aanwezig bleken, ruim een jaar eerder ook reeds konstateerbaar waren. Vroegtijdige signalering en begeleiding waren in beginsel dus mogelijk geweest.
3. In een eerder artikel werd gerapporteerd over de volledige afwezigheid van sexe-verschillen in de resultaten van de tests voor de cognitieve ontwikkeling. Ten aanzien van de gedragsobservaties bleken er echter significante verschillen aanwezig te zijn in de eerste onderzoeksfase (meisjes werden gunstiger beoordeeld dan jongens). Intrigerend is dat de onder het eerste punt genoemde relatie tussen motivationeel/emotioneel gedrag en veranderingen in de intelligentie veel duidelijker en uitgebreider was voor meisjes dan voor jongens. Zaal (1978) beschrijft soortgelijke bevindingen. Een verder onderzoek naar sexe-typische ontwikkelingspatronen (in plaats van vergelijking en van nivo-verschillen) is noodzakelijk.

Noot

1. Nu en dan worden andere getallen vermeld indien van een kind geen complete gegevens aanwezig zijn in beide onderzoeksfasen.

Literatuur

Behar, L. & S. Stringfield, A behavior rating scale for the preschool, *Developmental Psychology*, 1974 (10), 601-610.

Doll, E., *Preschool attainment record*, research edition. Minnesota, Am. Guidance Service Inc., 1965.

Feshbach, S., H. Adelman, & W. Fuller, Early identification of children with high risk of reading failure, *Journal of learning disabilities*, 1974 (7), 639-644.

Feshbach, S., H. Adelman, & W. Fuller, Prediction of reading and related academic problems, *Journal of Educational Psychology*, 1977 (69), 299-308.

Gearhart, B., *Learning disabilities, Educational strategies*, St. Louis, Mosby, 1973.

Hermanns, J., Het ontstaan van schoolproblemen. Een longitudinaal onderzoek in kleuter en lagere school, *Pedagogische Studiën*, september 1979a, 348-357.

Hermanns, J., *On the stability and predictability of cognitive development in the preschool period*. Paper presented at the fifth biennial congress of the ISSBD, Sweden, 1979b.

Hermans, H., *Prestatiemotief en faalangst in gezin en onderwijs*. Amsterdam, Swets en Zeitlinger, 1971.

Kohn, M., B. Parnes, & B. Rosman, *A rating and scoring manual for the Kohn problem checklist and Kohn social competence scale*, Unpublished, 1976.

Laan, H. v.d., *Leren lezen, schrijven en rekenen*. Groningen, Tjeenk Willink, 1973.

Mischel, W., R. Zeiss, & A. Zeiss, Internal - external control and persistence, *Journal of personality and social psychology*, 1974 (29), 265-278.

Mönks, F., H. Rost, & N. Coffie, *Nijmeegse schoolbe-kwaamheidstest*. Nijmegen, Berkhout, 1978.

Myklbust, H. *The pupil rating scale: screening for learning disabilities*. New York, Grune & Stratton, 1971.

Novack, H., E. Bonaventura, & P. Merenda, A scale for the early detection of children with learning problems, *Exceptional Children*, oktober 1973, 98-105.

Rutter, M., A children's behaviour questionnaire for completion by teachers: preliminary findings, *Journal of child psychology and psychiatry*, 1967 (8), 1-11.

Spivack, G., & M. Swift, The Devereux elementary school rating scales. A study of the nature and organization of achievement related disturbed classroom behaviour, *Journal of special education*, 1966 (1), 71-91.

Stevens, L. & J. van Kuyk, *Handleiding bij de observatieschalen voor het speel- werkgedrag van de oudste kinderen van de kleuterschool*. Ongepubliceerd, 1976.

Stevenson, H., T. Parker, A. Wilkinson, A. Hegion, & E. Fish, Predictive value of teachers' rating of young children, *Journal of Educational Psychology*, 1976a (68) 507-517.

Stevenson, H., T. Parker, A. Wilkinson, A. Hegion, & E. Fish, Longitudinal Study of individual differences in cognitive development and scholastic achievement, *Journal of Educational Psychology*, 1976b (68), 377-400.

White, B. & J. Watts, *Experience and environment*. New

Jersey, Prentice Hall, 1973.
Zaal, J., *Sociaal Emotioneel gedrag in de klas*. Groningen,
Wolters-Noordhoff, 1978.
Zeeuw, J. de, *Algemene psychodiagnostiek II, testtheorie*.
Lisse, Swets en Zeitlinger, 1978.

Curriculum vitae

Zie *Pedagogische Studiën*, 1979 (56) 357
Adres: Heidelberglaan 1, 3584 CS Utrecht