

# Interne validering als haalbare strategie om programma's te evalueren

JAN J. F. TER LAAK

Vakgroep Ontwikkelingspsychologie, R.U. te Leiden

## Samenvatting

*Programma's zijn gehelen van systematische activiteiten om van te voren min of meer gespecificeerde doelen te bereiken. Aan programma's liggen veronderstellingen ten grondslag m.b.t. gedrags(verandering) van leerlingen, leerkrachten en soms ouders. Aan conceptuele analyse van programma's kunnen valideringsprocedures gekoppeld worden. Dit is te definiëren als interne validering. De meer gevraagde externe validering vereist in strikte zin een experimentele procedure, waardoor het programma als oorzaak kan worden aangewezen voor de gedragsverandering. Dit is in de praktijk gewoonlijk niet haalbaar en onderzoeksmatig moeilijk te realiseren. De experimentele opzet voor dit type probleem wordt niet algemeen meer als nuttig gezien. Een benadering om van externe validiteit te spreken, is de korrelatieve studie. Hier wordt samenhang met toekomstige criteria verwacht. Dit is vaak onwaarschijnlijk middels de gebruikelijke korrelatieve technieken, soms ongewenst en gezien de vraag van onderwijsgevend om directe konstruktieve suggesties, niet zo relevant. Een enkel facet van de werkwijze bij interne validering wordt toegelicht aan de hand van twee programma's voor het kleuteronderwijs: het curriculum schoolrijpheid van Dumont en Kok en themaprojekten, zoals ontwikkeld door Van Calcar c.s.*

## 1. Inleiding

Het aanbod van programma's voor het kleuteronderwijs is relatief groot. Inhoud, werkwijze, doelstellingen en doelgroepen lopen uiteen. Compensatieprogramma's worden geacht niet-specifieke achterstand op te heffen door training op vooral de taal. Programma's voor functie ontwikkeling nemen meer 'bedrijfsmatig' de criteriumsituatie als uitgangspunt en geven voorbereiding op specifieke schoolse taken. Themaprojekten hebben als inspiratie, dat de dagelijkse leefsituatie en de schoolsituatie

bij vooral de arbeidersmilieus uiteengegroeid zijn. De projekten hebben als doel de diskontinuiteit terug te dringen.

Afgezien van de verschillen in probleemdefinitie en-probleemsituering kan men een programma opvatten als een leer- of veranderingsplan om gespecificeerde doelen te bereiken. De inhoud, vorm en werkwijze van een programma veronderstelt begingedrag en een proces van gedragsverandering bij leerlingen, leerkrachten en soms ouders. Conceptuele analyse van een programma geeft aanwijzingen m.b.t. verwacht begingedrag, het veronderstelde veranderingsproces en de te bereiken doelstellingen. Deze analyse is nodig, omdat programma's daarover niet altijd uitdrukkelijk zijn en veelal gericht zijn op het geven van konkrete activiteiten. De conceptuele analyse maakt empirische validering mogelijk, mits het gelukt de veronderstellingen m.b.t. begin- en eindgedrag te operationaliseren en in de veronderstelde veranderingsprocessen meetpunten te definiëren.

Dit wordt interne validering genoemd en is zo strikt mogelijk aan het programma zelf en de situatie, waarin het wordt uitgevoerd, gekoppeld. Interne validering als hier bedoeld, is wat Snow (1975, pag. 275) de 'central region of referent generality' en Campbell en Stanley (1963) 'internal validity' noemen. De term slaat bij genoemde auteurs op onderzoekopzetten. Hier wordt het begrip toegepast op programma-opzetten. Het begrip heeft hier een wat ruimere betekenis, omdat vaststelling van verwacht begingedrag en binnen het programma te definiëren doelen tot de interne valideringsprocedure gerekend worden. Dit laatste is voor onderzoekopzetten niet zo relevant, omdat proefpersonen, die niet voldoen aan verwacht begingedrag voor onderzoek uit de proefgroep verwijderd worden. Doelen tot de interne valideringsprocedure rekenen is in vergelijking met de onderzoekopzet wat ongewoon, omdat bij onderzoek de beantwoording van een vraag en bij een programma te bereiken doelen voorop staan. Dat het in de gewenste richting sturen van een antwoord op een vraag (bepaalde doelen willen

bereiken) een wetenschappelijk geldige werkwijze kan zijn onder bepaalde kondities geeft Van Strien (1975) aan. Externe validering verwijst in abstracte zin naar het invariant zijn van een (experimentele) behandeling voor populaties proefpersonen en situaties (Bracht en Glass, 1968). De werkwijze om in strikte zin aan deze invariantie-eis te voldoen is het experiment; een benadering is de korrelationele studie. Experimentele studies met programma's houden in, dat variatie in de onafhankelijke variabele (gewoonlijk wel of niet deelname aan een programma) de gedragsvariatie in afhankelijke variabele verklaart. Andere invloeden (rijping, effecten van de specifieke proefgroep, interactie tussen proefgroep en behandeling) zijn uitgesloten. Men kan middels deze procedure een antwoord geven op de vraag naar het effect van een programma, in de zin dat de gedragsverandering veroorzaakt is door het programma. In zoverre die gedragsverandering bedoeld is kan men in zijn algemeenheid van een programma zeggen, dat het voldoet. Deze opzet is in de praktijk gewoonlijk niet haalbaar, alleen al, omdat het niet gewenst is de populatie, waarvoor hulp gevraagd wordt, te verdelen in experimentele- en controle-groep. Onderzoeksmatig blijkt deze opzet ook moeilijk realiseerbaar. Campbell en Stanley (1963) wijzen er op, dat er veel meer quasi-experimentele dan experimentele opzetten vigeren. Afgezien van de praktische en onderzoeksmatige haalbaarheid van deze opzet is er de wetenschapstheoretische discussie over de vruchtbaarheid van deze aanpak voor de menswetenschappen. Het type kennisprodukt, dat het experiment moet opleveren is Cartesiaans. Aldus de Zeeuw en Soudijn (1975) wordt dit type kennisprodukt steeds gevraagd, omdat het voor de natuurwetenschappen succesvol is gebleken. Of men dit analoog ook moet verwachten voor de menswetenschappen is niet noodzakelijk. Korrelationele studies zijn een benadering om tot externe validiteit van een programma te besluiten.

Een gebruikelijke procedure is de variantie in resultaten zoals (naar wordt aangenomen) ontstaan door de behandeling te vergelijken met variantie op toekomstige criteria. Resultaten op programma's voorspellen prestaties in het basisonderwijs. Deze benadering om het programma extern valide te noemen stelt problemen. Hoge korrelaties zijn onwaarschijnlijk door snelle wisseling van de criteriumsituatie. Hoge korrelaties zijn ook onwaarschijnlijk, omdat niet aannemelijk is, dat de behandeling op de basisschool de rangorde van de leerlingen ongemoeid laat (interactie tussen proefpersonen en behandeling). Soms zijn hoge korrelaties ongewenst en gelden indien gevonden, als afwezigheid van

bedoelde effecten. Dit is het geval bij de opheffing van achterstandsituaties. Een gelijke startkans d.m.v. een programma realiseren betekent immers variantie in prestaties gekoppeld aan sociaal milieu tot nul reduceren. Tenslotte is er het probleem, dat onderwijsgeevenden niet om prediktieve informatie vragen, maar om informatie, die direkt zijn neerslag heeft op het onderwijsproces. De korrelationele benadering kan niet altijd een bijdrage leveren om iets te zeggen over de externe validiteit van het programma; het is wel informatief, maar levert niet de bedoelde informatie. De vlucht van de korrelationele benadering heeft wellicht te maken met de stelling, dat iedere wetenschappelijk verantwoorde aktie, iedere wetenschappelijke uitspraak tot voorspellingen moet leiden. (De Groot, 1961). Op programma's kan deze stelling niet in zijn algemeenheid van toepassing geacht worden.

Externe validering, experimenteel en korrelationeel is met programma's moeilijk haalbaar. Interne validering is bij een in de praktijk uitgevoerd programma haalbaar. Een enkel facet ervan wordt toegelicht aan de hand van de uitvoering van twee programma's. Achtereenvolgens komt aan de orde: een korte beschrijving van het programma, de konkrete uitvoering op kleuterscholen, de conceptuele analyse, vraagstellingen op basis van de conceptuele analyse en de empirische resultaten.

## *2. Het programma voor functie-ontwikkeling van Dumont en Kok*

Dumont en Kok (1970) hebben een programma ontwikkeld om 'schoolrijpheid' te bevorderen. Zij onderscheiden 63 'functies' of onder bepaalde naam samengenomen aktiviteiten, die, na geobserveerd te zijn, geoefend kunnen worden. De 63 functies zijn elk geleed in een vijftal stappen. Bij elke stap, op elke functie, worden trainingsaktiviteiten beschreven. Dit curriculum heeft als voordeel, dat diagnostiek (observatie door de leerkracht) en begeleiding (konkrete aktiviteiten bij elk niveau op elke functie) onmiddellijk op elkaar aansluiten. Het programma is uitgevoerd in een speciaal samengestelde 'overgangsklas' met voor dat doel geïnstrueerde leerkrachten. Dumont en Kok (1970) noemen het programma succesvol onder verwijzing naar de bereikte niveaus van de leerlingen na afloop van het programma in vergelijking tot het beginniveau.

Het programma is door de vele konkrete trainingsaktiviteiten ook in andere situaties bruikbaar.

### 2.1. *Uitvoering van gedeelten van het programma*

Gedeelten van het oorspronkelijke programma werden uitgevoerd op een 18-tal kleuterscholen\*) door 62 leidsters met die groep kleuters, die volgens de normering van de N.S.T. (Mönks, e.a., 1969) niet- of twijfelachtig schoolbekwaam waren. Het aantal te observeren en te oefenen functies werd beperkt tot 13. In de gewone kleuterschoolsituatie (geen speciale overgangsklas) werden de oorspronkelijke 63 door de leerkrachten niet haalbaar geacht. De selectie vond plaats op grond van waargenomen overlap met al geprogrammeerde activiteiten op de kleuterscholen. Aldus kwamen voor observatie en training door leerkrachten in aanmerking: visuele waarneming (konkreet, abstrakt, geheugen); akoestische waarneming (analyse, geheugen); ruimtelijke oriëntatie (situaties, abstrakt, konkreet); denken (kategoriaal ordenen, serieel ordenen); voorbereidend rekenen (hoeveelheid, konkreet, getalbegrip). De 13 functies werden aangeboden als 9-punts observatieschalen, waarbij de verbaal omschreven stappen bij elke functie als ankers dienden. Het programma werd uitgevoerd bij 131 kleuters op 18 scholen gedurende het gehele schooljaar te beginnen vanaf november. De N.S.T. werd aan het begin van het schooljaar afgenomen bij alle oudste kleuters door daartoe geïnstrueerde leerlingen van de hoogste klas van de opleiding tot kleuterleidster ( $n = 942$ ).

Aan het eind van het schooljaar werd de N.S.T. nogmaals afgenomen bij de 131 aanvankelijk niet- of twijfelachtig schoolbekwame kleuters. Gedurende het schooljaar werd 3 keer geobserveerd. De observaties werden besproken en trainingsactiviteiten vastgesteld. Bij hun eerste observatie waren de leerkrachten niet op de hoogte van de N.S.T. scores van de kleuters.

### 2.3. *Konceptuele analyse*

Uit de vorm, inhoud en voorgeschreven werkwijzen van een programma kan men afleidingen maken m.b.t. de doelstellingen (de aard ervan bv. de gedifferentieerdheid), m.b.t. wat geleerd en onderwezen wordt en m.b.t. wat verondersteld wordt voor de begeleiding en het soort begeleiding. Dit afleiden is vereist om meetpunten te definiëren, zodat interne validering mogelijk wordt. Het afleiden bevat een zekere willekeur, die groter is naarmate het programma minder expliciet is.

\* Dit project werd uitgevoerd door P. Maes en auteur op kleuterscholen te Haarlem.

Onderstaande afleidingen hebben niet de pretentie volledig te zijn. Het is bedoeld als een eerste aanzet, waarbij duidelijk wordt, dat interne validering mogelijk en voor het onderwijs-leerproces nuttig is.

Dumont en Kok hebben een duidelijk doel met het programma: het schoolrijp maken van kleuters.

Deze eindterm is geoperationaliseerd als een bepaald niveau op een 63-tal functies. De vaststelling van het niveau is rationeel; d.w.z. op basis van ideeën over wat een 1e klasser basisonderwijs moet kunnen en niet empirisch, d.w.z. op basis van onderzoek over wat gemiddelde 1e klassers feitelijk behalen op de functies. De keuze voor 63 functies is er een geweest met het oog op volledigheid (hoge inhoudsvaliditeit) en een analytische beschrijving van de kleuters. Bruikbaarheid is, aldus de auteurs, een criterium geweest om niet nog meer functies te definiëren. De eindterm wordt geacht bereikt te worden, doordat de kleuter een aantal stappen doorloopt. Deze stappen zijn enerzijds didactisch; men weet op welk niveau er training moet worden gegeven en er is individualisering van training mogelijk. Anderzijds zijn er aanwijzingen, dat de stappen een ontwikkelings- of leerproces beschrijven: structurele verandering door differentiatie. Dit laatste komt in de opbouw van de stappen bij o.a. de waarneming naar voren. Overweegt de didactische benadering dan moet aangetoond worden, dat de gelede aanpak, gemiddeld genomen, hogere criteriumprestaties levert dan een niet-gelede (het gewoonlijk verwachte effect van individualisering). Overweegt het feit, dat de stappen een ontwikkelings- of leerproces beschrijven dan dient dit proces in de gegevens teruggevonden te worden. Beide benaderingen sluiten elkaar niet uit. Afhankelijk van de nadruk op een van beide zal de interne validering zich daarop richten.

Het uitvoeren van dit programma veronderstelt het kunnen diagnostiseren bij leidsters. Het is niet bij voorbaat noodzakelijk, dat de verbale ankers bij de niveaus door leidsters betrouwbaar waar te nemen zijn bij de kleuters. De omschrijvingen van de activiteiten bij elk niveau in het curriculum moeten uitvoerbaar zijn.

Het programma op papier behoeft niet samen te vallen met het programma in het gebruik. Verder dient het programma te passen in de (waargenomen) taak van het kleuteronderwijs.

### 2.3. *Vraagstellingen*

Dumont en Kok hebben gekozen voor een 63-tal functies. Dit samen met de 5 niveaus op elke functie staat een analytische beschrijving van iedere leerling toe in een profiel (matrix) van 63 bij 5. Op basis van

een beslissingsregel kan men dan kiezen voor trainingsactiviteiten, bv. trainen op alle functies, waarop het hoogste niveau niet gehaald wordt; trainen op alle functies, die het laagst gescoord worden, enz. De vraagstelling is of er sprake is van een 'true profile' (Cronbach, 1975), d.w.z. zijn de dimensies betrouwbaar en empirisch relatief onafhankelijk? Leveren de functies inderdaad analytische informatie of meet (en traint) men, extreem gesteld, steeds dezelfde dimensie? Er is gekozen voor een geleding van de functies in 5 stappen. De vraagstelling hierbij is, doet deze volgorde zich empirisch voor of zijn het slechts benamingen waaraan geen ontwikkelings-c.q. leerproces beantwoordt?

De leerkrachten dienen aan de hand van de labels bij elk niveau vast te stellen op welk niveau elke leerling thuishoort. Is gedifferentieerde schaling door relatief ongetrainde leidsters mogelijk aan de hand van de omschrijvingen? Het programma moet ingevoerd worden in de situatie van de kleuterschool. De vraag is of de bedoeling van het programma, waarbij systematische training vooropstaat, past in de doelstellingen waarvoor kleuterleidsters opteren m.b.t. het onderwijs?

Het programma invoeren in een gewone kleuterschool is nieuw. De implementeerbaarheid dient nagegaan te worden. Tenslotte is er de vraag of het programma beantwoordt aan zijn doel, d.w.z. doorlopen kleuters de schaal en neemt de schoolrijpheid toe. Dit is geen strikte vraag naar het effect en geen vraag naar samenhang met toekomstige criteria. De vooruitgang wordt binnen het programma vastgesteld.

#### 2.4. Procedure en Resultaten

Om na te gaan in hoeverre er sprake is van relatief onafhankelijke dimensies zouden de 63 functies gescoord moeten worden bij een grote groep kleuters. In dit onderzoek werden de functies beperkt tot 13. Bij 73 oudste kleuters, die volgens de normering van de N.S.T. niet- of twijfelachtig schoolbekwaam waren, werd de interkorrelatie-matrix berekend. Als scores dienden de waarden, die de leidsters op de dertien 9-puntsschalen invulden aan het begin van het schooljaar (zie tabel 1).

De scores op de 13 schalen uit dit voornamelijk cognitieve gebied blijken nogal hoog te correleren bij de hier gekozen vorm van observatie. Factoranalyse op de interkorrelatiematrix middels zwaartepuntmethode geeft hoge ladingen op de 1e faktor. De tweede faktor neemt relatief weinig variantie voor zijn rekening. (zie tabel 2).

Tabel 1. Interkorrelatiematrix van 13 observatieschalen bij 73 niet- of twijfelachtig schoolbekwame kleuters (zie voor de omschrijving van de variabelen tabel 2).

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	.47											
3	.68	.68										
4	.58	.42	.57									
5	.55	.65	.67	.55								
6	.53	.57	.64	.67	.66							
7	.49	.60	.41	.46	.60	.65						
8	.38	.42	.48	.40	.52	.60	.54					
9	.43	.52	.51	.55	.64	.51	.55	.50				
10	.44	.48	.46	.36	.46	.45	.54	.32	.47			
11	.69	.53	.58	.55	.53	.59	.54	.39	.47	.49		
12	.54	.39	.48	.49	.41	.56	.44	.42	.38	.57	.75	
13	.63	.56	.53	.63	.56	.63	.56	.48	.51	.57	.72	.74

Bij deze groep kleuters (niet- of twijfelachtig schoolbekwaam) wordt geen analytische informatie verkregen middels de door het programma voorgeschreven wijze van dataverzameling (observatie door kleuterleidsters) bij 13 functies.

Voor dit resultaat zijn meer oorzaken als verklaring mogelijk. De hoge ladingen op de eerste faktor kunnen een beoordelings- of observatie-effect zijn. In dit geval moet de informatie-verzameling gewijzigd worden. Verder kan men het toeschrijven aan kenmerken van de groep kleuters: bij deze groep lage scoorders op de N.S.T. is nog geen differentiatie opgetreden. Dit zou als implicatie kunnen hebben, dat het programma verschillend kan worden

Tabel 2: Faktorladingen ongeroteerd op de 1e en 2e faktor (zwaartepunt-methode) van de 13 observatieschalen bij 73 niet- of twijfelachtig schoolbekwame kleuters.

	a <sub>11</sub>	a <sub>12</sub>
1 Visuele waarneming (konkreet)	.74	-.24
2 Visuele waarneming (abstrakt)	.73	-.12
3 Visuele waarneming (geheugen)	.77	.18
4 Akoesties analyserende waarneming	.72	-.14
5 Akoesties geheugen	.78	.11
6 Ruimtelijke oriëntatie (situaties)	.81	.07
7 Ruimtelijke oriëntatie (abstrakt)	.73	.12
8 Ruimtelijke oriëntatie (konkreet)	.63	-.16
9 Denken (kategoriaal ordenen)	.70	-.20
10 Denken (serieel ordenen)	.66	-.14
11 Voorbereidend rekenen (hoeveelheid)	.80	.27
12 Voorbereidend rekenen (konkreet)	.72	-.20
13 Voorbereidend rekenen (getalbegrip)	.82	.09

uitgevoerd bij gedefinieerde groepen. Dit is de idee, welke ten grondslag ligt aan de Aptitude-Treatment-Interaction opzetten (Cronbach en Snow, 1969). Tenslotte is een mogelijke verklaring, dat, indien de schalen ontleend zijn aan de bestaande praktijk, deze praktijk weinig gedifferentieerd is en men vooral een algemene faktor oefent. Cronbach (1975) noemt deze werkwijze van observeren en trainen als start goed, maar er moet een schatting gemaakt worden van het 'true profile' en dit wijkt af van het geobserveerde, omdat 'outcomes are correlated and mastery tests are none too accurate' (pag. 9). Op dat 'true profile' dient de training aan te sluiten.

Op basis van de hoge samenhang tussen de 13 functies werd in dit projekt gekozen voor training op steeds 3 à 4 schalen, die het laagst gescoord werden.

Een volgend facet van interne validering is de status van de stappen of niveaus bij elke functie. Bij de eenmalige uitvoering was een vergelijking tussen een gelede en niet-gelede aanpak niet mogelijk. Wel is nagegaan of de opklimming een ontwikkelings-c.q. leerschaal vormde. Als model voor de ontwikkelingsschaal werd de perfecte schaal genomen in de zin van Guttman's scalogramanalyse (Guttman, 1944). Van de Hoogen-Lamers (1972) zette daartoe de schalen om op een wijze dat elke stap meetbaar werd en objectief te beslissen was of een kleuter de stap behaalde of niet. Alle 13 functies werden taken met scoringssystemen. Bij 20 kleuters (4-6 jaar) bleken de reproduceerbaarheidscoëfficiënten te liggen tussen de .80 en .90. De variantie over de schalen

per kleuter bleek gering. Of de Guttman-schaal zich ook voordoet bij de oorspronkelijke werkwijze: observatie op de functies, is niet nagegaan. Bij de aanpak met scorebare taken bij elke functie voldoet de schaal aan de eisen van een Guttmanschaal. De niveau-verdeling wordt empirisch ondersteund.

Het betrof een voor de leidsters nieuw programma. Het programma veronderstelt dat leidsters op basis van de verbale ankers bij de niveaus leerlingen kunnen schalen. Om deze veronderstelling te toetsen werden drie groepen samengesteld van 20 kleuters getrokken uit de totale groep ( $n = 942$ ). De groepen werden samengesteld op basis van de empirische verdeling van de N.S.T.-scores: groep  $\geq 120$ , groep  $\geq 104 \leq 109$  en groep  $\leq 92$ . Deze groepen werden door de leidsters geobserveerd op 9-punts observatieschalen bij de 13 functies. De leidsters waren niet op de hoogte van de N.S.T.-scores van de te observeren kleuters. Tabel 3 geeft de resultaten van de vergelijking. Hier is ook een vergelijking gemaakt tussen de niet- en twijfelachtige schoolbekwame groep. De observaties blijken in de verwachte richting te differentiëren vooral voor de 'gemiddelde' en de niet- en twijfelachtige schoolbekwame groep samen.

Bij de 'gemiddelde' en 'hoge' groep blijkt op 8 schalen geen verschil, bij de niet schoolbekwamen, t.o.v. de twijfelachtige schoolbekwame groep blijkt op 5 schalen geen verschil.

De veronderstelling m.b.t. de diagnostische activiteiten van de leidsters blijkt een redelijke.

Tabel 3: t-toets voor verschillen tussen gemiddelden van de scores op de observatieschalen bij groepen samengesteld op grond van NST-scores ( $\pm 20$  aselekt getrokken ppn. uit elke groep).

	gemid- delde NST $\geq 89$ $\leq 92$	gemid- delde NST $\leq 88$	t	gemid- delde NST 104-109	gemid- delde NST $\leq 92$	t	gemid- delde NST $\geq 120$	gemid- delde NST 104-109	t
1	4.58	3.15	3.246 ++	6.53	3.91	4.226 ++	7.39	6.53	1.075 NS
2	6.28	4.28	3.223 ++	7.89	5.62	4.127 ++	8.50	7.89	1.974 NS
3	5.49	3.69	3.750 ++	6.10	4.63	2.227 +	7.83	6.10	3.460 ++
4	4.42	3.98	0.957 NS	7.58	4.20	5.086 ++	7.33	7.58	- .410 NS
5	6.11	5.37	1.762 NS	8.13	5.79	4.406 ++	8.71	8.13	1.381 NS
6	5.02	3.85	2.433 +	7.00	4.50	4.032 ++	8.00	7.00	2.273 +
7	4.17	2.59	3.681 ++	6.87	3.36	5.725 ++	7.76	6.87	1.589 NS
8	3.34	2.88	0.917 NS	6.13	3.12	5.137 ++	6.54	6.13	0.621 NS
9	5.55	4.32	3.199 ++	7.26	4.46	5.458 ++	7.73	7.26	1.068 NS
10	4.06	2.94	2.894 +	6.15	3.32	5.442 ++	8.24	6.15	3.167 ++
11	5.02	3.35	3.431 ++	6.63	4.09	3.908 ++	8.22	6.63	5.113 ++
12	3.44	2.56	1.852 NS	5.50	2.98	3.993 ++	7.17	5.50	1.853 NS
13	3.84	3.10	1.691 NS	5.29	3.46	4.186 ++	6.67	5.29	2.091 +

\* $p \leq .05$ ; ++ $p \leq .01$

De meest direkte interpretatie van deze verschillen, samengenomen met het feit, dat alle 13 schalen hoog laden op de eerste faktor (zie tabel 2) is, dat leidsters op één faktor de kleuters onderscheiden; waarbij de fijnheid het grootst is tussen 'gemiddelde-' en niet-twijfelachtig schoolbekwame groep samen, minder bij de groepen niet- t.o.v. twijfelachtig schoolbekwaam en nog minder bij de 'gemiddelde' t.o.v. de 'hoge' groep. Andere interpretaties van de verschillen zijn niet uitgesloten: differentiële betrouwbaarheid van de schalen en een differentieel ontwikkelingsverloop van de functies.

Het programma schrijft observatie en begeleiding door de leidsters voor. Het curriculum geeft concrete activiteiten bij elk niveau op elke functie. Het geheel dient als bedoeld uitvoerbaar te zijn in de gewone kleuterschoolsituatie. De meest juiste procedure om dit na te gaan zou zijn het verzamelen van observatiegegevens tijdens de uitvoering door de leidsters. Deze tijdrovende procedure is in dit project niet gebruikt.

Een meer verwijderde procedure is een vragenlijst na afloop van het project aan de leidsters. Het bezwaar van deze vragenlijst is, dat men geen gedrag tijdens uitvoering meet, maar beoordelingen achteraf van gedrag. De vragenlijst werd na afloop van het project toegezonden aan de 62 deelnemende leidsters; 56 werden geretourneerd. Enkele resultaten m.b.t. de achteraf waargenomen of beoordeelde uitvoerbaarheid: 53 % vond het aantal (13) schalen voldoende, 47 % te veel; 86 % oordeelde de niveaus bij de functies voldoende onderscheiden; 67 % stelde d.m.v. de voorgeschreven observaties meer informatie m.b.t. de kleuters te krijgen dan die waarover ze op basis van ongewapende waarneming van gedrag in de klas al beschikten. De gemiddelde tijd, besteed aan het invullen van het niveau op alle 13 functies bedroeg volgens opgave 1½ uur, 76 % achtte dit gezien de informatie, die dit m.b.t. de kleuters opleverde niet te veel. 93 % gaf als mening, dat de concrete activiteiten uit het curriculum voldoende aansloten bij de geobserveerde niveaus op de functies, 87 % nam vooruitgang waar bij de kleuters, die getraind werden d.m.v. de activiteiten als beschreven in het curriculum. 72 % van de respondenten wilde observatie en training in deze vorm handhaven, 90 % wilde na lichte herziening een soortgelijke werkwijze voor het volgende schooljaar.

De mediaan van de opgegeven tijd, besteed aan concrete activiteiten uit het curriculum bedroeg 1½ uur per week. Leidsters, die opgaven meer dan 2½ uur per week besteden, hadden kleuters in de klas,

waarbij de verschillscore van de tweede t.o.v. de eerste N.S.T. gemiddeld 14.52 punten bedroeg. Zij, die opgaven minder dan 2½ uur te besteden, kwamen tot verschillscores met een gemiddelde van 10.31 punten. Dit verschil is significant:  $t = 2,0147$ ;  $p \leq .05$ . Dit resultaat is betrekkelijk vanwege de onbetrouwbaarheid van verschillscores (Cronbach en Furby, 1970) en vanwege het feit, dat het beoordelingsgegevens betreft. Het curriculum definieert de kleuterschool als een fase, waarin systematische voorbereiding op het basisonderwijs plaatsvindt. Het is niet bij voorbaat bekend of deze opvatting past in de opvattingen, die leerkrachten hebben van het kleuteronderwijs. Om dit na te gaan zijn een 4-tal doelstellingen voor het kleuteronderwijs geformuleerd, waaraan de leidsters ( $n = 56$ ) rangorden van belangrijkheid toekenden (1: belangrijkste-, 4: minst belangrijke doelstelling). Tabel 4 geeft de percentages over leidsters van de keuzen.

Tabel 4: Rangorde (belangrijkst: 1, minst belangrijk: 4) van doelstellingen van het kleuteronderwijs als waargenomen door kleuterleidsters ( $n = 56$ ), in procenten afgerond.

Doelstellingen kleuteronderwijs	Rangorde				
	1	2	3	4	
	↓				
	→				
	in %				
Spel en creatief bezig zijn	a	15	40	38	8
Sturen tot rustig en ordelijk gedrag	b	0	13	23	65
Zelfstandig worden en aangepast gedrag	c	68	20	10	0
Systematische voorbereiding d.m.v. programma's op het basisonderwijs	d	17	27	29	27

Uit tabel 4 blijkt, dat systematische voorbereiding op het basisonderwijs d.m.v. programma's, over leidsters genomen, niet als belangrijkste wordt waargenomen. Het zelfstandig worden en het vertonen en aanleren van aangepast gedrag blijkt in de waarneming van leidsters centraler. Op categorie 'd' zijn de responsen regelmatig over de rangorden verdeeld. Dit is niet eenzinnig te interpreteren. Als alternatieven zijn denkbaar: er zijn duidelijke verschillen in opvatting m.b.t. categorie d; vervolgens, gezien de bijna kansniveau responsen in de 4 rangorden kan men interpreteren, dat er m.b.t. 'd' nauwelijks een mening gevormd is. Dit laatste analoog aan de kansverdeling van responsen over kennisvragen in een meerkeuze-test. Leerkrachten, die categorie 'd' rangorde 1 of 2 gaven blijken kleuters te hebben met hogere ruwe winstcores op de twee N.S.T.-

afnames dan leerkrachten, die dezelfde categorie rangorde 3 of 4 toekenden ( $t = 2.315, p \leq .05$ ). Dit maakt de laatste interpretatie minder aannemelijk. D.m.v. conceptuele analyse zijn enkele afleidingen uit het programma gemaakt, waardoor een enkel facet van interne validering mogelijk werd. De validering heeft betrekking op een gedeelte van het curriculum (13 schalen). De observatie blijkt onvoldoende valide als manier om analytische informatie m.b.t. kleuters te verkrijgen. De geleiding in stappen blijkt empirisch een pendant te hebben in gedrag van kleuters. Hiertoe zijn evenwel niet observaties gehanteerd, maar gestandaardiseerde taken. Leerkrachten blijken gedifferentieerd waar te nemen met de stappen, als verbaal omschreven in het curriculum, als ankers. Deze differentiatie is dan wel op één faktor en de verdeling van de fijnheid van diskriminatie ligt vooral in het onderscheiden van 'gemiddelde' - 'zwakke' kleuters, als gemeten met de N.S.T. De uitvoerbaarheid in de kleuterschool blijkt voor ongeveer de helft een behoorlijke belasting. De door het programma aangenomen kijk op de doelstellingen van het kleuteronderwijs wordt niet door de overgrote meerderheid van de leidsters gedeeld. Het curriculum maakt veronderstellingen m.b.t. gedrag van kleuters, van leidsters en veranderingsprocessen in het gedrag van kleuters als gedefinieerd door het curriculum maar ten dele waar. De interne valideringsprocedure spoort dit op en kan een begin zijn om het curriculum bij te sturen, of gedragsvoorwaarden m.b.t. het curriculum aan te brengen.

### 3. Themaprojekten

De begeleidingsvorm en -inhoud, die men kiest hangt samen met de optiek die men heeft op school en maatschappij. De doelstelling van themaprojekten is de dagelijkse werkelijkheid meer bij de school (en omgekeerd) te doen aansluiten, vooral voor die groep kinderen bij wie de diskontinuiteit het grootst is: de arbeiderskinderen. Deze themaprojekten zijn ontwikkeld door Van Calcar c.s. Er zijn een aantal projecten zoals 'Winkel', 'Post', 'Eten' uitgegeven.

#### 3.1. Uitvoering van twee themaprojekten

Op een 10-tal kleuterscholen met voornamelijk arbeiderskinderen werden door resp. 33 en 28 leidsters twee projecten ('Winkel' en 'Post') uitgevoerd. \*) De

\*) Dit project werd uitgevoerd door E. van Doremaele, H. Kuijpers, M. Reynders en auteur op kleuterscholen in Breda.

projecten werden omschreven aangeboden aan de leerkrachten. Dit betrof de inhoud (welke activiteiten met de kleuters uit te voeren) en de werkwijze (verzamelen van ideeën bij de kleuters, informeren en rapporteren over het project aan de ouders). Een themaprojekt neemt ongeveer 2½ week in beslag. Na afloop van een projekt werd door de leidsters, na instructie over afname en scoring, een schriftelijke meerkeuze-toets afgenomen bij de oudste kleuters. Op grond van informatie uit de toets (per leerling en per klas) werden activiteiten gesuggereerd aan de leidsters.

#### 3.2. Conceptuele analyse

De doelstelling van de themaprojekten is het doorbreken van de diskontinuiteit tussen school en dagelijkse leefsituatie. Het operationaliseren van deze eindterm is door de ongebruikelijkheid ervan niet eenvoudig. Gebeurtenissen, die geacht worden bij te dragen aan het bereiken van deze eindterm worden genoemd of zijn af te leiden: wijziging in de rol van de leerkracht (ideeën en materialen van de kleuters zelf worden uitgangspunt voor het onderwijs) wijziging in de rol van de ouders (actieve deelname aan het onderwijs, direct of indirect); bij de kleuters worden een aantal begrippen en taalrelaties geoefend, die belangrijk geacht worden voor het schoolbegin. De inhoud van deze begrippen is niet gedefinieerd: de projektboekjes geven wel aanwijzingen. De afleidingen voor interne validering zijn in behoorlijke mate arbitrair, omdat de eindtermen weinig geoperationaliseerd zijn, de werkwijzen niet uitdrukkelijk omschreven zijn en de 'pool' van begrippen niet vastgesteld is. Wel is het van belang, dat het zover komt. Onderstaande procedure is een allereerste aanzet om tot interne validiteitsbepaling van dit programma voor het kleuteronderwijs te komen.

#### 3.3. Procedure en resultaten

Onderwijs geven in de vorm van themaprojekten veronderstelt, dat een aantal begrippen en taalrelaties, van het belang geacht voor het basisschoolbegin, aangeleerd worden. De werkwijze dient de kans te vergroten dat de begrippen en taalrelaties bekliven. Door de begrippen in verband met het dagelijkse leven te brengen, wordt geacht een bezwaar van compensatie-programma's (het verdwijnen van het aangeleerde na afloop van het programma) te zijn ondervangen.

Op basis van lezing van het projektboekje bij het betreffende projekt werd een 40-tal begrippen vast-

gesteld. De begrippen betroffen rekentaal, ruimtelijke posities, kleuren, eigenschappen van objecten enz. Deze begrippen werden vooraf aan de leidsters doorgegeven, zodat ze uitdrukkelijk in het project aan de orde zouden komen. Na afloop van het project werd een 4-keuze toets afgenomen bij de oudste kleuters door de leidsters. Begrippen van het eerste project werden herhaald in het tweede project wanneer 30 % van de kleuters een bepaald begrip niet beheerste. De toetsen zijn bedoeld als 'formatieve' studietoetsen. Beide toetsen bevatten 72 vragen. De mediaan ruwe score voor resp. 'Winkel' en 'Post' was 64, 41 en 64,00. De hoogste en laagste gemiddelden per klas waren bij de respectievelijke projecten: 68,17 en 48,38; 69,00 en 43,00. Over klassen blijkt 88 % van de aangeboden begrippen gekend te zijn. De range van de gemiddelden maakt duidelijk, dat m.b.t. de hier gevraagde taalrelaties arbeiderskinderen geen homogene groep vormen. Dit resultaat is uitdrukkelijk geen bewijs voor het feit dat het project de oorzaak is voor het aanleren van de begrippen. De hier gevolgde procedure (het vooraf geven van de begrippen en de schriftelijke toetsing daarvan) is niet met het programma gegeven. Het is een poging wat verondersteld wordt geleerd te zijn te toetsen als onderdeel van interne validering. De nog belangrijkere veronderstelling van het programma is dat de begrippen bekliven. Dit is ook toetsbaar op deze wijze. Dit is in dit project niet gebeurd. Om stringenter afleidingen uit het programma te maken zou de fundering en structuur van te leren begrippen nodig zijn (keuze voor een theorie over taalontwikkeling).

Het programma veronderstelt een bepaalde werkwijze van de leerkracht: divergente produktie van ideeën door de kleuters dient gestimuleerd en gehonoreerd te worden. Materialen en begrippen uit het dagelijks leefmilieu dienen functioneel in het onderwijs gebruikt te worden. De leidsters worden geacht de ouders te stimuleren tot betrokkenheid en deelname. Het operationaliseren van uit het programma af te leiden gedragswijzen (doceerstijlen) van leidsters moet grotendeels nog gebeuren. Observatie op bovengenoemde dimensies zou daartoe een eerste aanzet kunnen zijn. In dit project is voor de zwakere aanpak gekozen: een vragenlijst, waarmee enkele beoordelingen door leidsters en feitelijke gebeurtenissen werden vastgelegd. Er werden 28 'Winkel' en 33 'Post' vragenlijsten verzonden met elk een respons van 82 %. Aan het verzoek een op schrift gesteld draaiboek voor de introductieperiode en een lesschema voor de uitwerkingsfase van het project te maken werd door 56 % van de leidsters

voldaan. Alle leidsters rapporteerden de ouders vooraf geïnformeerd te hebben; 0 % middels huisbezoek en 10 % d.m.v. individuele gesprekken met ouders. Terugrapportage aan de ouders na het project vond volgens opgaaf in 76 % van de gevallen plaats, waarvan de helft d.m.v. ouderbijeenkomsten. Het aantal ouders dat aan de organisatie van het project deelnam door materialen en ideeën werd op 80 % geschat. De deelname van de kleuters aan het project 'Winkel' werd tussen 75 % en 100 % geschat. Deelname door kleuters aan het project 'Post' werd minder geschat (50 %), aldus de leidsters was hiervoor de reden, dat dit project te 'kognitief' was voor met name de jongste kleuters. Alle kleuters brachten materialen mee, die zinvol waren voor het project. In het geval er ouderbijeenkomsten werden gehouden, verschenen volgens opgaaf tussen 50 en 75 % van de ouders. Middels de vragenlijst kwam naar voren, dat de verschillen tussen scholen groot en binnen scholen gering waren m.b.t. de wijze van uitvoering. Het onderdeel van de veronderstelde werkwijze van de leerkracht vereist verdere uitwerking. Uitwerking van een model voortvloeiend uit het programma is hiertoe een vereiste. Dit zal een procesmodel dienen te zijn, hetgeen meer is dan een multidimensionale beschrijving van leerkrachtgedrag. Een model om een beschrijving te hebben voor betrokkenheid van ouders bij het onderwijs kan ontleend worden aan de sociaal-psychologische literatuur over attitude-verandering (zie bijv. Van de Meer, e.a., 1976). Het werk van Kratochvil e.a. (1964) met betrekking tot affektieve leerdoelen kan de werkwijze beschrijven en empirisch toetsbaar maken.

Bovenstaande is een zeer bescheiden aanzet tot interne validering voor themaprojekten. Na vastleggen van aan te leren begrippen kon nagegaan worden of de begrippen inderdaad gekend zijn. De spreiding over klassen m.b.t. de gekende begrippen was groot, hetgeen niet de bedoeling is van de projecten. Het programma werd standaard aangeboden; er was evenwel veel variantie in uitvoering tussen scholen, niet binnen scholen. Dit staat het programma toe. Men kan evenwel spekuleren, dat de variantie in uitvoering aan kleuters gebonden kan worden en niet zo zeer aan scholen. Iets meer dan de helft van de leidsters maakte een draaiboek en een lesschema, zoals bedoeld. Blijkbaar pasten de programmavoorschriften niet als bedoeld in de werkwijze van de overgrote meerderheid van de leidsters.

#### 4. *Diskussie*

Wanneer men programma's opvat als beschrijvingen



van leerling- leerkracht- en oudergedrag, dan zijn daaraan procedures te verbinden om het programma intern te valideren. De leerling wordt in het programma van Dumont en Kok beschreven als een 'mastery-learner' naar het criterium van de 1e klas basisschool. Hij wordt geacht dit op een bepaalde manier te doen. Aan de leerkracht stelt het bepaalde eisen van observatie en begeleiding. Deze opvattingen over leerling en leerkracht dienen empirisch steun te hebben. In themaprojekten wordt een leerling beschreven als iemand, die ervaringen uit het dagelijks leven in de schoolsituatie structureert en organiseert. Deze opvatting vereist een uitgewerkter theorie dan nu binnen de themaprojekten aangegeven is. De interne validering is op dit moment dan ook nog zeer beperkt. Ook stelt de beschrijving eisen aan de leerkracht en de ouders.

Voorlopig is dit gedefinieerd als een herdefinitie van de rol, waardoor het toegankelijk kan worden voor toetsing.

Programma's worden gewoonlijk ingezet als 'experiment' hetgeen in tegenstelling tot de oorspronkelijke betekenis is gaan inhouden, dat men een aansprekend idee in nauwelijks omschreven acties omzet, waarvan het resultaat bij gevolg alleen beoordeeld kan worden. Interne validering kan een bijdrage zijn om 'de aansprekende idee' in gedrags termen te definiëren en te toetsen. De werkwijze vereist een uitdrukkelijke vaststelling van wat en hoe geleerd en onderwezen wordt. Dit kan door conceptuele analyse afgeleid worden. In deze zin is de meest centrale 'region of referent generality' (Snow, 1975) van een programma te toetsen.

#### Literatuur

Bracht, G. H. & G. V. Glass, The external validity of comparative experiments in educational and the social sciences, *American Educational Research Journal*, 1968, 5, 437-474.

Calcar, C. van, *Evaluatierapport*, RITP, Amsterdam, 1974.

Campbell, D. & J. Stanley, Experimental and quasi experimental design for research on teaching. In: N. L. Gage (Ed.), *Handbook of Research on Teaching*, Rand McNally, 1963, pp. 171-246.

Cronbach, L. J., On the design of educational measures,

*Paper presented to the second international symposium on educational testing, Montreux, July, 1975.*

Cronbach, L. J. & L. Furby, How should we measure change, or should we? *Psychol. Bull.*, 1970, 71, pp. 68-80.

Cronbach, L. J. & R. E. Snow, *Individual differences in learning ability as a function of instructional variables*, Final report to USOE, March 1969, Stanford University, School of Education, Contract N DEC 061, 4-6-061, pp. 269-1217.

Dumont, J. & J. Kok, *Curriculum Schoolrijpheid*, Malmberg, Den Bosch, 1970.

Groot, A. D. de, *Methodologie*, Mouton, Den Haag, 1961.

Guttman, L., A basis for scaling qualitative data, *Am. Sociol. Rev.*, 1944, pp. 139-150.

Hoogen-Lamers, A. van de, Toetsopgaven bij 13 kognitieve schalen gebaseerd op curriculum schoolrijpheid van Dumont & Kok. Niet gepubliceerde skriptie, Nijmegen, 1972.

Krathwohl, D., B. Bloom & B. Masia, *Taxonomy of educational objectives, Handbook II, Affective Domain*, New York, 1964.

Meer, F. van der & A. Berghuis, Attitudes en instrumentaaliteit van gedrag, *Gedrag*, 4, 1976, pp. 24-39.

Mönks, F., H. Rost & N. Coffie, *Nijmeegse Schoolbekwaamheidstest*, Monografie en Handleiding, Berkhout, Nijmegen, 1969, 4e druk, 1972.

Snow, R. E., Representative and quasi-representative designs for research on teaching, *Review of Educational Research*, 1975, 12, pp. 265-291.

Strien, P. J. van, Naar een methodologie van het praktijkdenken in de sociale wetenschappen, *Ned. Tijdschr. Psychologie*, 1975, 30, pp. 601-621.

Zeeuw, G. de en K. A. Soudijn, Over rechtvaardiging van onderzoeksvorschriften, *Ned. Tijdschr. Psychologie*, 1975, 30, pp. 621-647.

#### Curriculum vitae

Jan J. F. ter Laak (1944) studeerde ontwikkelingspsychologie aan de K.U. te Nijmegen. Was werkzaam op de K.U. te Nijmegen, Hogeschool te Tilburg en de schoolbegeleidingsdienst te Breda. Is momenteel werkzaam bij de vakgroep ontwikkelingspsychologie R.U. te Leiden. Adres: Kwikstaarhoek 31, Leiden