

DE SCHIEDAMSE REKENTEST

- H. HEESEN, hoofd van de openbare l.o.m.-school te Schiedam.
D. STRELITSKI, remedial teacher van de Gemeente Schiedam. ¹⁾
DRS. A. V/D WISSEL, schoolpsycholoog van de Gemeente Schiedam. ²⁾

INHOUD

1. INLEIDING
2. DE KEUZE VAN DE OPGAVEN
 - 2.1 Opgavenserie 1-1; het optellen t/m 10.
 - 2.2 Opgavenserie 1-2; het aftrekken t/m 10.
 - 2.3 Opgavenserie 2-3; optellen t/m 20 met overschrijding van het tiental.
 - 2.4 Opgavenserie 2-4; aftrekken t/m 20 met aanbreken van het tiental.
 - 2.5 Opgavenserie 4-1 t/m 4-3; de tafels van vermenigvuldiging.
3. DE KEUZE VAN DE SCHOLEN
 - 3.1 Intelligentieniveau en intelligentiespreiding.
 - 3.2 Het sociaal niveau van de scholen.
4. DE KEUZE VAN DE KINDEREN
 - 4.1 De gelijkwaardigheid van de A en de B-groep.
 - 4.1.1 Rapportcijfers.
 - 4.1.2. Intelligentie.
 - 4.1.3. De rekentest van Schonell.
 - 4.2. Het testgewenningseffect.
5. DE NORMERING VAN DE S.R.T.
6. DE BETROUWBAARHEID VAN DE S.R.T.
7. RAPPORTCIJFERS, TESTRESULTATEN EN INTELLIGENTIE
8. DE MEER ALGEMENE GELDIGHEID VAN DE SCHIEDAMSE UITKOMSTEN
 - 8.1. Het onderzoek van Drs. Mommers in Brabant.
9. SAMENVATTING

¹ thans hoofd van de openbare l.o.m.-school te Gouda.

² thans hoofd van de psychologische afdeling in het district Alkmaar van de Stichting van Geestelijke Volksgezondheid in N.H.

I Inleiding

Door de auteurs werd een onderzoek verricht op drie Schiedamse scholen naar de mate van beheersing van de rekenstof in het G.L.O. De aanleiding daartoe vormde het probleem, om bij kinderen met leermoeilijkheden een duidelijke uitspraak te doen over de ernst van de achterstand in een bepaald vak.

In een vorig artikel¹ werd uiteengezet hoe de rekenstof van het G.L.O. werd geïnventariseerd aan de hand van leerplannen en leermethoden en vervolgens werd ingedeeld in leerstofeenheden. Hierbij werd nauwe aansluiting gezocht bij de gangbare indeling van de leerstof in het G.L.O. Vervolgens werd per leerstofeenheid een serie van 10 (soms 5) testopgaven samengesteld.

Op deze wijze werden 67 opgavenseries ontworpen, waarvan men een overzicht vindt in bovengenoemd artikel. Behalve op de diverse rekenbewerkingen, hebben de opgavenseries betrekking op nevenonderwerpen, zoals „het rekenen met geld”, „klokkijken” e.d. De Schiedamse Rekentest * bestaat derhalve uit 67 van elkaar onafhankelijke opgavenseries, waarvan er 64 werden geijkt.

Het is de bedoeling van de auteurs in dit artikel nader in te gaan op de psychometrische aspecten van deze test. Daar de S.R.T. in eerste instantie is ontstaan als een praktisch hulpmiddel voor de werkers in het Schiedamse veld en niet a priori als landelijk instrument dat aan alle voorwaarden van testconstructie en testijking zou moeten kunnen voldoen, krijgt dit artikel meer de betekenis van een psychometrisch commentaar achteraf ter adstructie van de praktische bruikbaarheid van de test, dan van een volledige methodologische verantwoording van opzet, uitvoering en resultaten.

Een onderzoek, volledig methodologisch verantwoord van opzet en uitvoering, kan onmogelijk ondernomen worden vanuit een praktische werksituatie als waarin de auteurs zich bevinden.

De representativiteit van de gekozen opgaven voor het onderdeel van de rekenstof, waarvoor de opgavenserie werd samengesteld, wordt in 2 aan de orde gesteld. De keuze van de scholen wordt in 3 nader verantwoord, de keuze van de kinderen in 4. In 5 geven wij een verslag van de normering in quintielen. De betrouwbaarheid van de testuitslagen wordt in 6 nader onderzocht, terwijl de samenhang tussen de resultaten op enkele testbladen onderling en hun verband met het rapportcijfer rekenen en de intelligentie in 7 aan bod komt.

* De Schiedamse Rekentest wordt binnenkort uitgegeven bij de fa. Wolters-Noordhoff.

Bij de beantwoording van de vragen in 4, 6 en 7 hebben wij ons beperkt tot de gegevens van klas III en de resultaten behaald bij de opgaven van blad 3 (optellen en aftrekken t.e.m. 100), blad 4 (tafels van vermenigvuldiging) en blad 16 (eenvoudige redactiesommen). Daar alle derde klas leerlingen de opgavenseries 3-1 (splitsen van getallen t.e.m. 100) en 4-1 (de tafels van 2, 5 en 10 door elkaar) vrijwel foutloos maken, werden deze twee opgavenseries hierbij buiten beschouwing gelaten.

Gewerkt is dus met de volgende opgavenseries:

- 3-2: Optellen en aftrekken zonder aanvullen, overschrijden of aanbreken van het tiental.
- 3-3: Optellen en aftrekken met aanvullen, overschrijden of aanbreken van het tiental met de tweede term kleiner dan 10.
- 3-4: Optellen en aftrekken zonder aanvullen, overschrijden of aanbreken van het tiental, met beide termen groter dan 10.
- 3-5: Optellen en aftrekken met aanvullen, overschrijden en aanbreken van het tiental met beide termen groter dan 10.
- 4-2: De tafels door elkaar, behalve de tafels van 2, 5 en 10, met de eerste factor kleiner dan 6.
- 4-3: De tafels door elkaar met de eerste factor groter dan 5.
- 4-4: Tafels als puntsommen, waarbij de eerste factor ontbreekt.
- 4-5: Tafels als puntsommen, waarbij de tweede factor ontbreekt.
- 16 : Eenvoudige redactiesommen.

De keuze viel op klas III, waar wij de 54 kinderen welke deelnamen aan het onderzoek in het najaar '63 opnieuw in het voorjaar '64 aan het onderzoek lieten deelnemen, te samen met een gelijkwaardige groep van 54 andere kinderen. Hierdoor werd het mogelijk de resultaten van de groep kinderen uit het najaar '63 te vergelijken met hun resultaten in het voorjaar '64, terwijl ook een vergelijking mogelijk wordt tussen de beide groepen, van de „oude” en de „nieuwe” kinderen die in '64 aan het onderzoek deelnamen. Bovendien werd in klas III een intelligentietest afgenomen, zodat het mogelijk was de testresultaten met het I.Q. te vergelijken.

Wij beperkten ons tot de bovengenoemde onderwerpen, omdat wij meenden hiermee de belangrijkste facetten van het rekenen getroffen te hebben, n.l. het kunnen werken met toegepaste automatiseringen (blad 3), het beheersen van de primaire automatiseringen (blad 4) en het kunnen oplossen van redactiesommen (blad 14).

Tenslotte bespreken wij in 8 de vraag in hoeverre aan de Schiedamse uitkomsten een meer algemene, landelijke betekenis mag worden toegekend.

II *De keuze van de opgaven*

De S.R.T. wil de onderzoeker in staat stellen te bepalen in hoever een kind de rekenstof van het G.L.O. beheerst. Het gaat primair om de schoolvorderingen hier en nu te bepalen. De predictieve mogelijkheden zijn hierbij wel interessant maar niet essentieel voor de waarde van de test. Wel is de vraag naar de inhoudsvaliditeit (content validity), zoals door De Groot² wordt onderstreept, van essentieel belang voor de deugdelijkheid van de test. Dit probleem komt voor de S.R.T. op de volgende twee vragen neer:

A. Hebben de gezamenlijke opgavenseries betrekking op alle onderdelen van de rekenstof van G.L.O.?

B. Is er binnen de opgaven van één serie voldoende differentiatie mogelijk en zijn de diverse nuanceringsen binnen dit onderdeel van de rekenstof adaequaat vertegenwoordigd?

Ad A.:

De indeling van de S.R.T. in een groot aantal opgavenseries is tot stand gekomen op basis van een analyse van een aantal leerplannen en methoden. Voor elk onderdeel van de rekenstof werd een opgavenserie samengesteld. Deze opzet waarborgt o.i., dat er inderdaad alle onderdelen van de rekenstof van het G.L.O. door de S.R.T. worden bestreken.

Hierbij wordt aangetekend, dat het aanvankelijk rekenen en de zesde klasstof welbewust buiten beschouwing werden gelaten.

Ad B.:

Dit vereist een inventarisatie van de verzameling van opgaven welke binnen een onderdeel van de rekenstof mogelijk zijn naar moeilijkheidsgraad (psychometrisch principe), terwijl bovendien de diverse moeilijkheden bij de verschillende stadia van het leerproces zichtbaar moeten kunnen worden (didactisch principe).

De S.R.T. is in eerste instantie uit een praktische behoefte ontstaan. Wetenschappelijke normen, zoals hierboven geformuleerd, moesten bij de opzet van de test derhalve wel buiten beschouwing blijven. Wel bestaat de mogelijkheid achteraf te onderzoeken of de respectievelijke

opgavenseries aan het onder B gestelde criterium voldoen. Wij hebben ons hierbij beperkt tot de opgavenseries 1-1 (optellen t.e.m. 10), 2-3 (optellen t.e.m. 20), 1-2 (aftrekken t.e.m. 10), 2-4 (aftrekken t.e.m. 20) en de opgavenseries 4-1 t.e.m. 4-3, welke betrekking hebben op de tafels. Deze onderdelen kozen wij, daar ze de basisstof vormen van het rekenen, terwijl daarnaast de overweging dat hierover elders reeds werd bericht, meespeelde.

2-1 *Opgavenserie 1-1; het optellen t.e.m. 10*

Uitgaande van de getallen van 0 t.e.m. 10 zijn er in het totaal 64 optellingen met twee termen mogelijk. Hiervan zijn er 19, waarbij één van beide getallen een 0 is. Van de overige opgave zijn er 9, waarbij de som 10 is ($4 + 6 = \dots$). Voorts zijn er 10 opgaven, waarbij de som gelijk aan of kleiner dan 5 is. Van de overige 26 opgaven is de som groter dan 5 en kleiner dan 10.

In de opgavenserie 1-1 hebben slechts 2 i.p.v. 3 opgaven betrekking op de 0 (n.l. $8 + 0 = \dots$ en $0 + 9 = \dots$). Eén opgave heeft als uitkomst 10 ($7 + 3 = \dots$), hetgeen een redelijke representatie van deze categorie lijkt. Geen enkele opgave heeft een uitkomst gelijk aan of kleiner dan 5, waarbij deze groep in de opgavenserie dus ontbreekt.

Schonell³ vermeldt twee onderzoeken in Schotland en Nieuw-Zeeland. Hierbij bleek het volgende:

1. Optellingen met een 0 worden als moeilijker beleefd dan opgaven waarin geen 0 voorkomt.
2. Optellingen, waarbij het kleinere getal voorop staat, b.v. $2 + 7 = \dots$ (type B) zijn moeilijker dan optelsommen, waarbij het grotere getal voorop staat, b.v. $7 + 2 = \dots$ (type A).

Uit 1 volgt, dat de opgaven $8 + 0 = \dots$ en $0 + 9 = \dots$ vermoedelijk de moeilijkste uit deze serie zijn. Mede gezien de aparte plaats, die de nul bij het rekenen blijkbaar inneemt, zou het bij een eventuele revisie van de test overwogen kunnen worden, de opgaven met een 0 in een aparte score te verwerken.

Wat 2 betreft zij opgemerkt, dat in de opgavenserie 1-1 vier opgaven van type A en drie opgaven van type B voorkomen, hetgeen een redelijk evenwichtige verdeling is.

2-2 *Opgavenserie 1-2; aftrekken t.e.m. 10*

Uitgaande van de getallen van 0 t.e.m. 10 zijn in het totaal 66 aftrekkin-

gen met twee termen mogelijk. Hiervan zijn er 11 met een 0 als aftrektaal, terwijl in 10 opgaven de uitkomst een 0 is ($6 - 6 = \dots$).

In de opgavenserie 1-2 komt één opgave met een 0 voor, terwijl één opgave 0 als uitkomst heeft. Qua evenredige vertegenwoordiging had er nog een opgave met een 0, of met 0 als uitkomst in moeten staan.

Uit de literatuur (zie Schonell) blijkt dat juist deze twee typen opgaven bij uitstek als moeilijk worden ervaren. Qua moeilijkheid vormen deze typen opgaven een categorie apart. Derhalve geldt ook hier de kritische opmerking, gemaakt bij opgavenserie 1-1, n.l. dat de te meten beheersing van het aftrekken t.e.m. 10 te eenzijdig beïnvloed wordt door de opgaven die betrekking hebben op de nulfunctie.

De overige opgaven vertonen een redelijke spreiding over het gehele bereik van de overgebleven 45 mogelijkheden.

2-3 *Opgavenserie 2-3; optellen t.e.m. 20 met overschrijding van het tiental*

Uitgaande van de getallen van 0 t.e.m. 9 zijn er in het totaal 36 opgaven mogelijk, waarbij het tiental wordt overschreden. Bij nadere analyse van de 10 opgaven van de serie 2-3 blijkt dat deze overwegend tot de minder moeilijke opgaven van dit totaal van 36 behoren. Dit betekent dat de opgaven van 1-1 in verhouding tot die van 2-3 moeilijker zijn uitgevallen.

Gaan wij bij deze opgavenserie na, hoe de verhouding is tussen het type A en het type B, dan zien wij dat slechts één opgave in de A-vorm en 7 opgaven in de B-vorm worden aangeboden. Ook dit vormt mede een factor waardoor de vergelijkbaarheid tussen de opgavenseries 1-1 en 2-3 niet optimaal is.

2-4 *Opgavenserie 2-4; aftrekken t.e.m. 20 met aanbreken van het tiental*

Analyse van deze serie wijst erop dat de 10 opgaven redelijk zijn gespreid over de 36 mogelijke combinaties. Dit impliceert tevens dat de vergelijkbaarheid met opgavenserie 1-2 vrij goed is, met uitzondering dan dat bij serie 1-2 de opgaven waarbij de nul-functie een rol speelt, op het totale resultaat een vertroebelend effect hebben.

2-5 *Opgavenserie 4-1 t.e.m. 4-3; de tafels van vermenigvuldiging*

De tafels van vermenigvuldiging zijn ondergebracht in de opgavenseries 4-1 (de tafels van 2, 5 en 10), de opgavenserie 4-2 (de overige tafels, met als vermenigvuldiger een factor gelijk aan of kleiner dan 5) en in 4-3

(de tafels met als vermenigvuldiger een factor groter dan 5 maar kleiner dan 10).

Het logische uitgangspunt van deze indeling is duidelijk; een indeling naar moeilijkheid.

Uit de literatuur³ blijkt dat het geen verschil maakt of de grootste of de kleinste factor voorop staat bij de tafels. Wel heeft elke factor zijn eigen moeilijkheidskarakter. Hiervan is een rangorde aan te geven. Deze rangorde is in het Schotse onderzoek als volgt:

Rangorde: 2 3 1 5 4 6 8 7 9 0
(van makkelijk naar moeilijk)

Vervangt men in de opgaven de factoren door hun rangordecijfer dan krijgt men een indicatie van de onderlinge moeilijkheidsgraad van de diverse vermenigvuldigingen. Hoe kleiner het product, hoe makkelijker de opgave. Toetsing van de opgavenseries 4-1, 4-2 en 4-3 aan dit criterium van relatieve moeilijkheid leidt tot de constatering, dat deze drie series inderdaad in moeilijkheid opklimmen. Wel komt er wat de afzonderlijke opgaven betreft overlapping voor tussen 4-1 en 4-2 enerzijds en tussen 4-2 en 4-3 anderzijds. Tussen 4-1 en 4-3 is geen overlapping. Hiermee is bevestigd dat wij in de opzet zijn geslaagd, n.l. om drie series samen te stellen betrekking hebbende op de tafels en opklimmend in moeilijkheid van de opgaven.

III *De keuze van de scholen*

Om toch enige spreiding in de ijkingsgroep te krijgen was het noodzakelijk kinderen van verschillende scholen in het onderzoek te betrekken. Besloten werd ons hierbij te beperken tot drie scholen. Op elke school werden per klas telkens 18 kinderen onderzocht. De keuze van deze drie scholen geschiedde niet op grond van exact geverifieerde criteria, maar op grond van de ervaring die de samenstellers in de loop van enkele jaren met deze drie scholen hadden verkregen.

De keuze viel op:

Een school, die van oudsher als opleidingsschool een goede naam had. Dit was vermoedelijk de school die qua milieu en begaafdheid boven de middelmaat lag, zonder hierin een extreme tendens te vertonen. (school 1);

Een school in een typisch oude stadsbuurt met een overwegende arbeidersbevolking, waarbinnen de categorie van de ongeschoolde arbeiders overheerst (school 3);

Een school in de nieuwbouw die qua milieu en begaafdheid een grotere spreiding zou vertonen dan de beide andere scholen en vermoedelijk een tussenpositie inneemt tussen beide (school 2).

Achteraf werd om enige controle op deze keuze te hebben een intelligentietest afgenomen in de derde klas en werd een overzicht gemaakt van de beroepen van de vader.

3.1 Intelligentieniveau en intelligentiespreiding

Het intelligentieonderzoek vond plaats in de derde klas. Gebruikt werd de Pintner-Durost Lager School Test. Wij gebruikten model M en hiervan zowel schaal 1 (plaatjesinhoud) als schaal 2 (leesinhoud). De I.Q.'s werden hierbij berekend volgens de formule:

$$I.Q. = 100 + \frac{15}{20} (\text{mediaanwaarde} - 120 - C),$$

waarbij C (correctie) = chronologische leeftijd in maanden - 103. In feite maakten wij dus van een spreidings' I.Q. met een gemiddelde van 100 en een standaardafwijking van 15.¹

Een overzicht van de resultaten vindt men in tabel I.

Tabel I.

I.Q.	A	B	C	D	E	
	N ₁ =29 N ₂ =40 N ₃ =26			N ₁ + N ₂ + N ₃ N ₁ = N ₂ = N ₃ in percentages	normaal verdeling in percentages	
> 127,5					8	} 11,5
123—127	3	1	2	6,8	3,5	
118—122	3	3	0	5,8	4,4	} 20,5
113—117	2	9	1	11,0	5,6	
108—112	4	6	1	11,0	5,9	} 20,5
103—107	3	9	2	13,3	14,6	
98—102	8	6	6	22	16	} 20,5
93—97	3	4	7	15,8	14,6	
88—92	3	0	4	8,7	5,9	} 11,5
83—87	0	2	2	4,3	5,6	
78—82	0	0	0	0	4,4	} 11,5
73—77	0	0	1	11,3	3,5	
< 72,5					8	
Gemiddelde	107	107	95	103	100	
				> = 11,5	> = 12,5	

¹ Voor een verdere verantwoording wat betreft het gebruik van deze formule zij verwezen naar het artikel van Drs. A. v. d. Wissel „Spellingsmoeilijkheden, minus-variant of dysorthografie”, Nederlands Tijdschrift voor de Psychologie, Deel XVIII, 1963, aflevering 1, blz. 18.

In de kolommen A, B en C zijn de uitkomsten aangegeven van respectievelijk school 1, 2 en 3. Hieruit blijkt dat school 2 met een gemiddelde I.Q. van 107 hetzelfde niveau behaalt als school 1, terwijl de spreiding ook in grote lijnen de zelfde blijkt te zijn. Bij beide scholen valt ongeveer 50% van de kinderen in het I.Q.-gebied van 93 - 107, d.w.z. van de normale middelmaat; bij school 1 valt 10% onder I.Q. 93, bij school 2 slechts 5%. Hier staat tegenover dat er bij school 1 een grotere groep goed begaafden is met een I.Q. van 123 of hoger, n.l. 10% tegenover 2,5% bij school 2.

Uitgaande van onze opzet, n.l. om via deze I.Q.-bepaling een beeld te krijgen van het gemiddeld niveau en de begaafdheidsspreiding van de leerlingen, moet men de verschillen tussen school 1 en 2 als onbelangrijk zien bij de verdere grote mate van overeenkomst. De veronderstelde en bedoelde differentiëring in drie niveaus wordt door dit intelligentieonderzoek niet bevestigd. Veeleer moeten school 1 en 2 beide als representant van de school boven de middelmaat worden gezien, dit in contrast met school 3, waar het niveau duidelijk onder de middelmaat ligt.

Voegt men de drie scholen samen tot één groep en houdt men hierbij het aantal kinderen per school gelijk, zoals in het onderzoek inderdaad is gebeurd (18 kinderen per school per klas) dan krijgt men op grond van de gegevens in de kolommen A, B en C de I.Q.-verdeling die in kolom D in percentages is weergegeven. De standaardafwijking van deze verdeling is 11,5.

Ter vergelijking hebben wij in kolom E de normaalverdeling weergegeven bij een standaardafwijking van 12,5. Als sigma kozen wij 12,5 daar hiermee enerzijds de gevonden sigma van 11,5 het beste wordt benaderd, terwijl anderzijds de verhouding tot de gebruikte I.Q.-intervallen redelijk te hanteren blijft ($5 : 12\frac{1}{2} = 2 : 5$).

Vergelijking van kolom D en E leert ons dat de I.Q.'s in het bereik van 87,5 tot 112,5 sterker zijn vertegenwoordigd dan op grond van een normaalverdeling zou zijn te verwachten. Deze versterkte aanwezigheid is voornamelijk gegaan ten koste van de I.Q.'s onder de 87,5.

We moeten dan ook concluderen dat bij het rekenonderzoek gewerkt is met een qua gemiddelde begaafdheid enigszins geflatteerde groep.

3.2 *Het sociaal niveau van de scholen*

De indeling van de beroepen van de vaders volgens een sociaal maatschappelijk criterium hebben wij ontleend aan bijlage 1 van het artikel van Matthijsen: Sociaal-hiërarchische indeling van de ouders der leer-

lingen bij het voortgezet onderwijs.

De zes in ons onderzoek gebruikte categorieën zijn respectievelijk:

1. Vrije beroepen en topfuncties bij overheid of bedrijfsleven.
2. Leidinggevend kader en daarmee gelijk te stellen functies.
3. Middenkader en daarmee gelijk te stellen functies.
4. Lager bedrijfskader en uitvoerende hoofdarbeiders.
5. Hoger geklasseerde handarbeiders.
6. Lager geklasseerde handarbeiders.

In tabel II is een overzicht gegeven van de resultaten per school van deze indeling.

Tabel II.

<i>School 1.</i>	<i>School 2.</i>	<i>School 3.</i>
Categ. 1: 2 maal	Categ. 1: 2 maal	Categ. 5: 6 maal
Categ. 2: 4 maal	Categ. 2: 3 maal	Categ. 6: 13 maal
Categ. 3: 7 maal	Categ. 3: 4 maal	
Categ. 4: 8 maal	Categ. 4: 6 maal	
Categ. 5: 9 maal	Categ. 5: 4 maal	
Categ. 6: 4 maal	Categ. 6: 3 maal	

N = 34, Gem. 3,9

N = 22, Gem. 3,7

N = 19, Gem. 5,7

Ook hier blijkt weer duidelijk een uiteenvallen in twee niveaus. Enerzijds school 3 met een gemiddelde classificatiescore van 5,7 en anderzijds school 1 en 2 met een gemiddelde classificatiescore van respectievelijk 3,9 en 3,7. Bij school 3 vallen alle beroepen in de categorieën hoger en lager geklasseerde handarbeiders. Bij school 1 en 2 valt ongeveer 45% in de categorieën midden- en lagerkader, terwijl school 2 zelfs een wat hoger percentage heeft dan school 1, dat valt in de categorieën leidinggevend kader of vrij beroep. Als globale typering is dit verschil echter te verwaarlozen en kunnen wij opnieuw, thans op grond van het sociale milieu, stellen dat school 1 en 2 beide „de school voor de betere middelmaat” representeren.

Uit het oogpunt van de representativiteit van de ijkingsgroep betekent dit, dat „de betere middelmaat” te sterk zal zijn vertegenwoordigd ten koste van „de middelmaat”, kinderen dus met vaders die qua beroep in de categorie van het lagere bedrijfskader en de hoger geklasseerde handarbeiders zouden vallen.

Conclusie: Zowel op grond van het intelligentie onderzoek als het onderzoek naar het sociaal milieu blijkt dat de samenvoeging van de kinderen van de drie scholen tot één ijkingsgroep leidt tot een enigszins geflatteerde groep. Het verdient aanbeveling hierop bij de interpretatie

van de resultaten verdacht te zijn.

IV *De keuze van de kinderen*

In hoofdstuk 3 werd ingegaan op de keuze van de scholen, waarop het onderzoek werd verricht. In dit hoofdstuk gaan wij in op de wijze waarop de leerlingen voor het onderzoek werden uitgekozen.

Daar wij per kind wilden nagaan hoe lang dit over de opgavenseries deed, was het niet mogelijk meer dan drie kinderen tegelijkertijd te onderzoeken. Het totaal aantal kinderen dat wij per school per klas konden onderzoeken moest derhalve beperkt blijven tot 18, d.w.z. dat over de drie scholen samen in het totaal 54 kinderen per klas aan het onderzoek deelnamen.

De keuze van deze 18 kinderen geschiedde niet op grond van het toeval maar systematisch op grond van de rapportcijfers bij het vak rekenen. De gekozen 18 kinderen waren dus representatief voor de gehele klas wat gemiddelde en spreiding van het rapportcijfer rekenen betreft. De rapportcijfers waren ontleend aan het eindrapport van de cursus '62-'63. Daar de zittenblijvers bekend waren, was ook de samenstelling van de nieuwe klas III voor de cursus '63-'64 bekend. Er werd voor gezorgd dat de zittenblijvers in verhouding tot de totale klas in deze groep van 18 kinderen waren vertegenwoordigd.

De opzet van het onderzoek was de beschikking te krijgen over halfjaarlijkse normen wat betreft de prestaties van een representatieve groep G.L.O.-kinderen.

Omstreeks oktober '63 werd het onderzoek voor de eerste keer verricht en een half jaar later, omstreeks april '64 herhaald. Klas VI werd niet in het onderzoek betrokken, klas I slechts bij het onderzoek in april '64.

De vraag rijst of het verantwoord is bij de afname in april '64 de zelfde kinderen, die ook in oktober '63 meededen, in het onderzoek te betrekken. Hoe groot is de kans dat test- of onderzoekgewenning als storende factor optreedt, waardoor de vergelijkbaarheid van de onderzoekresultaten in '63 en '64 onbetrouwbaar wordt. In principe meenden wij het risico van zulk een gewenningseffect te moeten vermijden door bij het onderzoek in '64 een groep van 18 kinderen (per klas) te kiezen dan in '63. In de praktijk bleken alleen de klassen van school 2 groot genoeg te zijn om hieruit twee parallel groepen van 18 kinderen samen te stellen. De scholen 1 en 3, die zich in het weinig kinderrijke centrum bevinden, hebben meestal te kleine klassen om in twee groepen van 18 kinderen te splitsen.

Dit bezwaar losten wij op, deels door de groep kinderen die overbleef aan te vullen met kinderen uit de eerste groep, deels door tweemaal met de zelfde groep te werken. Indien achteraf zou blijken dat er inderdaad sprake is van een storend gewenningseffect, dan zouden de resultaten van het onderzoek in '64 bij de scholen 1 en 3 hierdoor minder betrouwbaar worden.

Om na te gaan of er zulk een testgewenningseffect optreedt, lieten wij de groep kinderen van klas III, die aan het eerste onderzoek hadden deelgenomen (groep A) ook het tweede onderzoek doen (groep A'). De resultaten van deze kinderen (groep A') konden vergeleken worden met de groep kinderen die bij het tweede onderzoek voor het eerst meededen (groep B). Wel moeten deze beide groepen geheel gelijkwaardig zijn, wil men uit een vergelijking van de resultaten mogen concluderen dat inderdaad met zulk een storend testgewenningseffect tekening gehouden moet worden.

4.1 *De gelijkwaardigheid van de A en B-groep*

De samenstelling van twee gelijkwaardige parallelgroepen, elk van 18 kinderen uit het totaal aantal leerlingen van klas III, geschiedde eveneens op grond van de rapportcijfers. Aan het eind van de cursus '63-'64, dus na afloop van het gehele onderzoek werd in deze klas nog een intelligentietest afgenomen. Dit diende er deels toe een begaafdheidsbeeld van de school als zodanig te krijgen (zie hoofdstuk 3), deels maakte dit het mogelijk de beide parallel groepen achteraf te vergelijken qua intelligentie. Naast een intelligentieonderzoek namen wij bovendien nog 5 rekestests af van Schonell⁴, eveneens om de vergelijkbaarheid van de beide groepen te kunnen nagaan. Vergelijking van de gelijkwaardigheid van de A en de B-groep geeft dan de volgende uitkomsten:

4.1.1 *Rapportcijfers: (klas III, school 2).*

	Groep A (A')	Groep B
4	1	—
5	—	2
5½	1	—
6	2	3
7	8	6
7½	—	1
8	6	6

Het gemiddelde rapportcijfer voor beide groepen is precies gelijk, n.l. 7 (Voor de klassen IV en V gingen wij dit eveneens na. Wij vonden

voor klas IV respectievelijk 6,7 en 6,5 voor klas V 7,1 en 7,2).

4.1.2 *De intelligentie*

Tabel III.

I.Q.	totaal	groep A	groep B
123—127	1	1	0
118—122	3	0	3
113—117	9	3	3
108—112	6	4	2
103—107	9	4	3
98—102	6	5	1
93—97	4	0	4
88—92	0	0	0
83—87	2	1	1
N	40	18	17
M	107	106½	106

Daar de indeling in twee groepen, respectievelijk groep A en groep B, tot stand kwam op grond van de rapportcijfers rekenen en niet op grond van de I.Q.'s, is het zeer wel mogelijk dat de I.Q.-gemiddelden en de I.Q.-spreiding tussen de A en de B-groep gaat verschillen.

In tabel III is de I.Q.-verdeling aangegeven voor klas III van school 2, respectievelijk van de totale klas, van de A en van de B-groep. De gevonden gemiddelden stemmen vrijwel overeen. Het minieme verschil blijkt dan ook in het geheel niet significant te zijn. De A en de B-groep blijken qua I.Q.-verdeling gelijkwaardig te zijn en tevens representatief voor de totale klas.

4.1.3 *De rekestests van Schonell*

Aan het eind van de cursus namen wij aan beide groepen de 5 eerste tests af van de door Schonell samengestelde „Diagnostic Arithmetic Tests”. Deze 5 tests hebben respectievelijk betrekking op 100 optelsommen (0 t.e.m. 18), de hiermee corresponderende 100 aftreksommen, 100 vermenigvuldigingen, 90 delingen en tenslotte 100 opgaven met optellen, aftrekken, vermenigvuldigen en delen door elkaar. Op grond van de Engelse ijking werd voor elke test een rekenleeftijd berekend. Als index voor het rekenniveau werd voor elk kind de middelste van deze vijf waarden genomen. De middelste rekenleeftijd in maanden is dan voor groep B 125 maanden, voor groep A' 130; de middelste chronologische leeftijden zijn respectievelijk 110 en 109. Omgerekend

in een rekenquotiënt krijgen we dan voor groep B 117, voor groep A' 119,1.

Ook in dit opzicht blijken de groepen redelijk parallel te lopen.

4.2 *Het testgewinningseffect*

Op grond van de voorafgaande toetsing op gelijkwaardigheid van de A (A')-groep en de B-groep, zijn wij verantwoord een verschil in resultaat tussen de A'- en de B-groep te interpreteren als een testgewinningseffect. Nemen wij aan dat zich zulk een effect voordoet, dan zullen de resultaten van groep A' in het algemeen beter zijn dan van groep B. (Hypothese $A' > B$).

Uitgaande van de opgavenseries van blad 2 t.e.m. blad 5, in totaal 18 rijtjes, vinden we dat bij 7 rijtjes de resultaten van groep A' beter, bij 11 rijtjes slechter zijn dan van groep B. Van een positief gewinningseffect is op grond van deze resultaten niets gebleken. Uitgaande van de resultaten bij de meer specifieke onderwerpen, zoals klokkijken, rekenen met geld e.d. zien we dat in 12 van de 19 gevallen de resultaten bij de A'-groep beter zijn dan bij de B-groep. Dit zou dus op een positief gewinningseffect kunnen wijzen.

Tenslotte hebben wij voor de bladen 3 (opgaveserie 3-2 t.e.m. 3-5), 4 (opgaveserie 4-2 t.e.m. 4-5) en 16 de verschillen in resultaten tussen groep A' en groep B statistisch getoetst op significantie. (Zie tabel IV).

Uit deze gevonden t-waarden blijkt dat de hypothese van een testgewinningseffect niet bevestigd kan worden.

De betekenis van deze uitkomsten moet geïnterpreteerd worden tegen de achtergrond van de prestatietoename gedurende het laatste halve jaar. Deze prestatietoename is gegeven voor de periode najaar '63 — voorjaar '64. Wij hebben deze prestatietoename statistisch getoetst op significantie. (Hypothesen $A' > A$, $B > A$).

Toetsing op significantie geschiedde eveneens voor de bladen 3, 4 en 16, voor de drie scholen afzonderlijk. (zie tabel IV)

Uit de gevonden t-waarden blijkt dat de hypothese van een aantoonbaar leereffect statistisch niet kon worden bevestigd.

¹ Aangezien het hier een Engelse ijking betreft kan men de hoge rekenquotiënten interpreteren als een aanwijzing dat school 2 in vergelijking met het gemiddelde Engelse niveau tot buitengewoon goede resultaten komt. Zelfs indien men rekening houdt met het gemiddelde I.Q. van 107, voor school 2, blijft deze conclusie geldig.

Tabel IV

4-5

	B-A	Blad 3 A'-A	A'-B	B-A	Blad 4 A'-A	A'-B	B-A	Blad 16 A'-A	A'-B
School 1	30-33=-3 t>50%	35-34=1 10<t<25	36-33=3 5<t<10	35-34=1 $\frac{1}{2}$ <t<1	37-34=3 10<t<25	37-35=2 t>50%	6-6=0	7-6=1 20<t<25	7-6=1 10<t<25
School 2	31-30=1 t>25%	33-30=3 10<t<25	33-31=2 t>25%	38-34=4 1<t<5	37-34=3 5<t<10	37-38=-1 20<t<50	6-5=1 5<t<10	7-5=2 t<0,0005	7-6=1 2 $\frac{1}{2}$ <t<5
School 3	31-30=1 t>25%	31-30=1 t>25%	31-31=0	31-29=2 10<t<25	32-29=3 10<t<25	32-31=1 t>50%	6-5=1 2 $\frac{1}{2}$ <t<5	6-5=1 5<t<10	6-6=0
	Leereffect	Testge- wennings- effect	Leereffect	Testge- wennings- effect	Leereffect	Testge- wennings- effect	Leereffect	Testge- wennings- effect	Leereffect

V De normering van de S.R.T.

Uit het oogpunt van een goede testconstructie, waarbij juist het principe van een geleidelijke opklimming in moeilijkheidsgraad van de items en de mogelijkheid van een voldoende spreiding naar boven en naar beneden zo belangrijk is, vertonen de opgavenseries van de S.R.T. ernstige tekortkomingen. Zij lenen zich dan ook niet tot een verfijnde schaaldifferentiatie. Differentiatie mogelijkheid biedt niet zozeer iedere opgavenserie afzonderlijk, maar de vergelijking van de resultaten op de diverse opgavenseries onderling.

Wij hebben derhalve volstaan met een normering in vijf categorieën, elk 20% van de kinderen bevattende. Bij de uitvoering van zulk een ijkings volgens de quintielwaarden stuiten wij echter op twee moeilijkheden. In de eerste plaats lenen zich de 54 kinderen van onze ijkingsgroep niet tot een exacte indeling in vijven. Wij zochten de oplossing in een verdeling van de 54 kinderen in vier groepen van 11 en één groep van 10 kinderen.

Dit betekent dus dat in elk van de vier onderste quintielgebieden telkens 20½% van de kinderen valt, in het vijfde quintielgebied, slechts 18½%.

Deze indeling is voor alle quintielbepalingen aangehouden. Daar waar door de uitval van één of meer kinderen de groep uit 53, 52 of 51 bestond, zijn respectievelijk de verdelingen 10-11-11-11-10, 10-11-11-10-10 en 10-10-11-10-10 aangehouden. Het betreft hier echter uitzonderingsgevallen.

De tweede moeilijkheid ontstaat door het geringe bereik van de schaal. Deze loopt bij een opgavenserie van 10 opgaven van 0 tot en met 10. Daar de opgavenseries in de meeste gevallen op een groep kinderen wordt afgenomen, die tot een zeer redelijke prestatie in staat zijn, blijven daardoor bovendien de laagste schaalwaarden vaak onbenut.

Het gevolg is dat het slechts zelden lukt bij kind no. 12, no. 23, no. 34 en no. 45 een exacte schaalwaarde te vinden. De kinderen met deze rangnummers bevinden zich meestal te midden van een groepje andere leerlingen dat de zelfde prestatie behaalde. Het zou dan ook meer voor de hand hebben gelegen bij de schaalwaarden een quintiel, eventueel deciel- op percentielwaarde te zoeken dan omgekeerd.

Dit laatste heeft echter het bezwaar dat er per half jaar steeds andere en daardoor moeilijk vergelijkbare percentielwaarden gevonden zullen worden, waardoor de vergelijkbaarheid van half jaar tot half jaar ernstig wordt bemoeilijkt. Wij hebben derhalve vastgehouden aan de quintielindeling zoals hierboven beschreven en bij de bepaling van de quin-

tielwaarden een eenvoudige interpolatie toegepast, daar waar het gezochte rangnummer zich te midden van een groepje bevindt.

Een voorbeeld moge het bovenstaande nader verduidelijken. Wij kiezen daarvoor de frequentieverdeling van de opgavenserie 2-5 bij de afname in het begin van klas III. (Het gaat hierbij om optellen en aftrekken t/m 20, aangeboden in de vorm van puntsommen).

Opgavenserie 2-5, afname III-a.

Schaalwaarde:	10	9	8	7	6	5	4	3	2	1	0		
frequentie		2	6	12	11	6	7	2	5	3	-	-	(N = 54)

Het kind met rangnummer 12 van onder af aan gerekend, bevindt zich in het groepje van 7 kinderen met uitslag 5. Indien deze 7 kinderen telkens zouden vallen op 5, 5, 51-7, 53-7, 54-7, 55-7, en 56-7; — dien de schaal zulk een verfijnde verdeling zou kennen —, dan zou bij no. 12 de schaalwaarde $5\frac{1}{7}$ horen. Dit is afgerond tot één decimaal 5,1 m.a.w. $Q_1 = 5,1$. De waarde van Q_2 , d.i. de uitslag van nummer 23 wordt dan op analoge wijze bepaald 6,8. Q_3 en Q_4 komen op respectievelijk 7,9 en 8,8. Het voordeel van de op deze wijze berekende quintielwaarden is, dat ook kleine verschillen tussen successieve halfjaarlijkse uitkomsten geregistreerd kunnen worden.

Op deze wijze ontstaat een 4-tal lijnen (Q_1 , Q_2 , Q_3 en Q_4), die bij de afname in het begin van de lagere school in principe in het lagere schaalbereik moeten beginnen om geleidelijk aan elk half jaar hoger te worden en te eindigen in het hogere bereik van de schaal. Een mooi voorbeeld van zulk een geleidelijke opklimming beginnend bij de lagere en eindigend bij de hogere schaalwaarden geven de quintiellijnen van blad 16 (eenvoudige redactiesommen).

Zoals uit figuur A is te zien werd deze opgavenserie afgenomen vanaf begin klas II (II-a) met als laatste afname eind klas V (V-b). De quintielwaarden bij de eerste afname zijn successievelijk 1,9, 3,1, 4,2 en 5,4. Deze quintielwaarden stijgen ieder half jaar geleidelijk om bij de afname aan het eind van klas V (V-b) als volgt te eindigen: 7,4, 8,6, 9,4 en 9,9.

fig. A

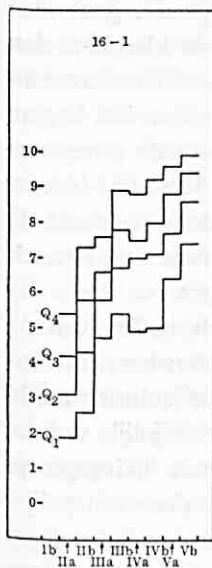


fig. B

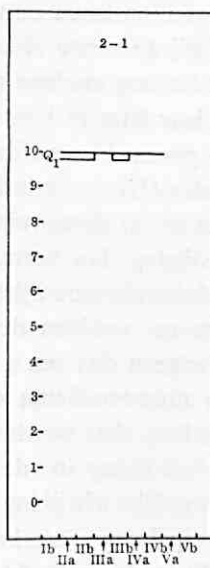
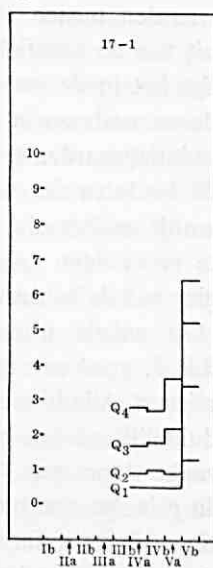


fig. C



Daar het onderzoek bij de afname gericht was op een groep, waarvan vrij goede prestaties konden worden verwacht, vertonen de meeste normenschema's niet zulk een evenwichtige verdeling over de gehele schaal, maar zijn meer beperkt tot het hogere schaalbereik. Een extreem voorbeeld hiervan, waarbij in feite iedere differentiëring wegvalt, vormt de opgavenserie 2-1 (optellen t/m 20). Het overzicht van de resultaten in figuur B spreekt in deze voor zichzelf. Het enige gegeven dat aan dit resultaat valt te ontleen is, dat vanaf II-a de overgrote meerderheid der kinderen deze opgaven vrijwel foutloos maakt.

Slechts daar waar de verwachting en de empirische uitkomst in het geheel niet met elkaar in overeenstemming zijn, ziet men de quintiellijnen laag inzetten en ook bij de hoogste klas waar de opgavenserie werd afgenomen, nog betrekkelijk laag eindigen. Een voorbeeld hiervan is de opgavenserie 17-1 (moeilijke redactiesommen). In figuur C ziet men dat in het begin van klas IV Q_4 op 2,8 ligt om aan het eind van klas V te eindigen met 6,4. Q_1 ligt begin klas IV op 0,5 en eindigt eind klas V op 2,8.

De resultaten bij de meeste opgavenseries liggen tussen het extreme beeld van 2-1 en het optimale beeld van 16-1.

De vraag rijst thans hoe betrouwbaar de gegevens zijn die door de respectievelijke quintielwaarden worden aangegeven. Deze quintielwaarden berusten immers op de resultaten van telkens 11 kinderen! Een

sterk argument voor de vrij goede betrouwbaarheid van deze quintielwaarden menen wij te mogen ontlenu aan het volgende: gaat men uit van de quintielwaarden bij de eerste afname, b.v. in klas II-a, dan ligt het in de verwachting dat een analoge groep van 54 andere kinderen, onderzocht een half jaar later in hun schoolcarrière, iets hogere quintielwaarden te zien zal geven. Het zelfde geldt voor de groep van 54 kinderen die een jaar later (III-a), anderhalf jaar later (III-b) enz. wordt onderzocht. Vertonen nu al deze quintielwaarden inderdaad de te verwachten geleidelijke stijging dan vormt dit in feite een bevestiging van de betrouwbaarheid der afzonderlijke bepalingen.

Op enkele uitzonderingen na voldoen de quintielwaarden van de S.R.T. goed aan deze eis, hetgeen dus een goede betrouwbaarheid impliceert. Vindt men in een normenschema een enkele quintielwaarde duidelijk uit het patroon vallen, dan wordt dit waarschijnlijk veroorzaakt door een toevallige afwijking in de betreffende ijkingsgroep. In principe zou men een dergelijke afwijking het beste door een polijsting van de quintiellijnen weg kunnen werken. Mede met het oog op de replicerbaarheid van dit onderzoek elders, hebben wij er echter de voorkeur aangegeven dit niet te doen en steeds de exacte empirische waarden te geven, berekend volgens de methode in het begin van deze paragraaf beschreven. Het is n.l. toch niet uitgesloten dat de afwijking een meer principiële verklaring heeft. Dit laatste wordt des te waarschijnlijker naarmate meerdere quintiellijnen van een halfjaarlijkse ijkingsgroep een afwijking in de zelfde richting vertoont. Zo blijken b.v. de quintielwaarden van alle opgavenseries van blad 3 (optellen en aftrekken t/m 100) bij de overgang van klas III (III-b) naar klas IV (IV-a) iets terug te vallen.

Zoekt men hierin een systematische factor, dan is de meest aannemelijke verklaring dat in klas IV overgeschakeld wordt naar de reken technieken, (optellen, aftrekken en vermenigvuldigen onder elkaar, staartdelingen) waardoor het hoofdrekenen, waarop de opgavenseries van blad 3 een beroep doen, eerder geremd dan gestimuleerd wordt. Op deze wijze kunnen de afwijkingen van het normale verloop van de quintiellijnen leiden tot zinvolle hypothesen met betrekking tot het leerproces.

Op grond van het bovenstaande mag dan ook worden geconcludeerd dat het beeld van de beheersing van de rekenstof, zoals dit door de quintiellijnen wordt weergegeven, inderdaad juist is en niet slechts de toevallige uitkomst van de 54 kinderen die „toevallig” voor het onderzoek werden uitgekozen.

VI *De betrouwbaarheid van de S.R.T.*

Het vorige hoofdstuk mochten we besluiten met de conclusie dat de normering in quintielen door een bevredigende stabiliteit wordt gekenmerkt. Daarnaast rijst de vraag of aan de individuele prestaties van de kinderen op de diverse opgavenseries eveneens een zekere constantie en daarmee een voldoende betrouwbaarheid kan worden toegerekend. M.a.w. blijkt een goede leerling bij herhaling van de betreffende opgavenserie opnieuw goed, een middelmatige middelmatig en een slechte leerling slecht te zijn?

Een test zal als regel des te betrouwbaarder zijn met betrekking tot de individuele prestaties die ermee worden onderzocht, naarmate de opgaven, waaruit de test is samengesteld meer gelegenheid geven tot differentiatie. De diverse opgaven moeten dus een opklimmend verschil in moeilijkheid te zien geven, beginnend bij zeer gemakkelijk en eindigend bij zeer moeilijk. Daarnaast is het aantal opgaven waaruit de test bestaat van invloed op de betrouwbaarheid. Een test zal als regel een hogere betrouwbaarheid bezitten, naarmate het aantal opgaven groter is. Ook zal de betrouwbaarheid toenemen, naarmate de ijkingssteekproef groter was.

Aan deze voorwaarden voor het vinden van een optimale betrouwbaarheids-coëfficiënt voldoen de opgavenseries van de S.R.T. slecht. De S.R.T. bestaat n.l. uit opgavenseries van niet meer dan 10 opgaven. Daar deze 10 opgaven een zeer duidelijk omschreven onderdeel van de rekenstof representeren, werd meer gestreefd naar homogeniteit, dan naar grote verschillen in moeilijkheid tussen de opgaven onderling. Verschil in moeilijkheid zal men bij de S.R.T. ook eerder aantreffen tussen de opgavenseries onderling, dan binnen de opgaven van een zelfde opgavenserie.

Bij de grote hoeveelheid opgavenseries waaruit de S.R.T. is samengesteld, is het niet mogelijk voor elke opgavenserie apart de betrouwbaarheid te berekenen.

Wij hebben ons, zoals wij reeds in de inleiding hebben vermeld, derhalve beperkt tot een steekproef van enkele opgavenseries in klas III, n.l. „het optellen en aftrekken t.e.m. 100” (de opgavenseries van blad 3), „de tafels van vermenigvuldiging” (de opgavenseries van blad 4) en „de eenvoudige redactiesommen” (de 10 opgaven van blad 16).

Om aan het bezwaar van de korthed van de opgavenseries tegemoet te komen, combineerden wij de opgavenseries van blad 3 met elkaar (met uitzondering van 3-1), evenals die van blad 4 (met uitzondering van 4-1).

Op deze wijze beschikten wij voor het „optellen en aftrekken t.e.m. 100” en „de tafels van vermenigvuldiging” telkens over een serie van 40 sommen. Daar blad 16, „eenvoudige redactiesommen” slechts uit één serie van 10 opgaven bestaat, was hierbij geen verdere combinatie mogelijk.

De betrouwbaarheid van de individuele testresultaten wordt als regel zelden bepaald aan de hand van een vrijwel gelijktijdig plaatsvindende tweede afname. De invloed van een leereffect kan dan n.l. van storende invloed zijn. Meestal gebruikt men als indicatie voor de synchrone betrouwbaarheid de correlatie van de ene helft van de testopgaven met de andere helft (split-half methode). Daar wij per kind slechts beschikten over de uitslag per opgavenserie, behielpen wij ons met een variant van deze split-half methode: van blad 3 combineerden wij enerzijds de opgavenserie 3-2 (optellen en aftrekken zonder aanvullen, overschrijden of aanbreken van het tiental) met 3-5 (optellen en aftrekken met aanvullen, overschrijden of aanbreken van het tiental, waarbij de beide termen groter zijn dan 10) en anderzijds de opgavenserie 3-3 (optellen en aftrekken met aanvullen, overschrijden of aanbreken van het tiental, waarbij de tweede term steeds kleiner is dan 10) met 3-4 (optellen en aftrekken zonder aanvullen, overschrijden of aanbreken van het tiental, waarbij beide termen groter zijn dan 10). Op analoge wijze combineerden wij van blad 4 de opgavenserie 4-2 (tafels door elkaar maar met de eerste factor kleiner dan 6) met 4-4 (puntsommen waarbij de eerste factor ontbreekt) en anderzijds 4-3 (tafels door elkaar met de eerste factor groter dan 5) met 4-5 (puntsommen waarbij de tweede factor ontbreekt). De correlaties werden berekend aan de hand van de onderzoekresultaten in het voorjaar '64, waaraan de 18 kinderen van het onderzoek in het najaar '63 te samen met 18 andere kinderen voor wie dit de eerste afname was, deelnamen.

In totaal dus een groep van 108 kinderen.

De uitkomst was als volgt:

Voor het optellen en aftrekken t.e.m. 100 : $r = .72$ ($N = 108$).

Voor de tafels van vermenigvuldiging : $r = .84$ ($N = 96$).

(Hierbij werd geen correctie toegepast voor een twee maal zo lange testserie).

Naast de synchrone betrouwbaarheid kent de psychometrie het begrip predictieve betrouwbaarheid: hoe betrouwbaar kan men op grond van de huidige testuitslagen de toekomstige resultaten op dezelfde test voorspellen?

Het is echter de vraag of dit begrip predictieve betrouwbaarheid

t.a.v. een schoolvorderingen test, zoals de S.R.T. is, van toepassing is. Het onderwijs is n.l. voortdurend bezig de kennis en de vaardigheid van de kinderen wat het rekenen betreft te beïnvloeden. De procedure van het doubleren illustreert dat het hierbij met name gaat om het bijbrengen van een minimum aan kennis, dat het kind in staat stelt het onderwijs in een volgende klas te volgen. De konsekventie hiervan zou moeten zijn dat het onderwijs voortdurend bezig is de zwakke leerlingen, de achterblijvers dus, er weer bij te trekken. (Dit is met name de specifieke opdracht aan remedial teachers). Zou men in deze opzet optimaal slagen, dan zou dit testtechnisch moeten leiden tot een vermindering van de prediceerbaarheid van toekomstige rekenprestaties op grond van de huidige: hoe effectiever dus dit erbij halen van de achterblijvers is, hoe lager de predictieve waarde van de test! Een hoge predictieve waarde zou dan een index zijn voor de rigiditeit van een gegeven „knapheids” of „domheids” rangorde. Het is als criterium voor de deugdelijkheid van de S.R.T. in feite niet relevant.

Daar de groep van 54 derde klas leerlingen, die aan het onderzoek in het najaar '63 deelnamen, een half jaar later opnieuw werden onderzocht, berekenden wij de correlatie tussen de resultaten van deze groep kinderen in het najaar '63 en in het voorjaar '64. Wij deden dit voor de testbladen 3, 4 en 16. Daar men in de praktijk afgaat op het resultaat per opgavenserie en niet per testblad, berekenden wij bovendien deze correlatie voor de opgavenseries 3-5 en 4-5 apart.

Daar deze correlaties gemakkelijk geflatteerd kunnen worden onder invloed van het vrij grote intelligentieverschil tussen school 1 en 2 enerzijds en school 3 anderzijds, berekenden wij deze correlatie nog eens afzonderlijk voor school 2.

De uitkomsten van deze berekeningen vindt men in tabel V.

<i>Tabel V.</i>	<i>school 1 + 2 + 3</i>	<i>school 2</i>
optellen en aftrekken t.e.m. 100 (opgavenseries 3—2 t.e.m. 3—5)	$r = .68$ (N = 51)	$r = .69$
tafels van vermenigvuldiging (opgavenseries 4—2 t.e.m. 4—5)	$r = .86$ (N = 50)	$r = .66$
eenvoudige redactiesommen (opgavenserie 16—1)	$r = .62$ (N = 52)	$r = .47$
optellen en aftrekken met aanvullen, overschrijden of aanbreken van het tiental; beide termen groter dan 10 (opgavenserie 3—5)	$r = .70$ (N = 51)	
puntsommen bij tafels van vermenig- vuldiging, waarbij de tweede factor ontbreekt (opgavenserie 4—5)	$r = .60$ (N = 51)	

Vergelijken wij deze uitkomsten voor blad 3 en 4 met de synchrone betrouwbaarheidsbepaling, dan treft de grote mate van overeenstemming. Bij de tafels van vermenigvuldiging respectievelijk .84 en .86, bij het optellen en aftrekken t.e.m. 100 .72 en .68. De opgavenserie 4-5 apart genomen geeft een correlatie te zien van .60. Dit ligt in de lijn van de normale verwachting wanneer men van 40 opgaven tot 10 opgaven teruggaat. Meer opvallend is de uitkomst bij de opgavenserie 3-5. Deze geeft, apart genomen de zelfde uitslag als de samengevoegde opgavenseries van blad 3!

Een interpretatie van deze verschillende uitkomsten is moeilijk te geven. Wel is het duidelijk dat de tafels van vermenigvuldiging een gefixeerder resultaat te zien geven dan het optellen en aftrekken t.e.m. 100. Men vraagt zich af of dit met de aard van de opgaven of met andere factoren samenhangt. De tafels immers zijn primaire automatiseringen, terwijl blad 3 het zelfstandig toepassen van primaire automatiseringen onderzoekt. Hier ligt een suggestie voor nader onderzoek.

Thans willen wij op grond van het bovenstaande de vraag beantwoorden of de S.R.T. met betrekking tot de individuele testuitslagen een voldoende betrouwbaarheid bezit. Gaan wij uit van de aparte opgavenserie en nemen wij aan dat de uitkomsten van .60 en .70 indicatief zijn voor het merendeel van de opgavenseries, dan laat deze betrouwbaarheid nogal wat te wensen over.

Houden wij echter rekening met de factoren die zulk een betrouwbaarheidscoëfficiënt drukken, zoals deze aan het begin van dit hoofdstuk werden besproken, dan moet men voor een opgavenserie van 10 vrij homogene sommen deze uitkomsten van .60 en .70 zeer redelijk achten. Mede dankzij de ruime mogelijkheid van de S.R.T., de uitslag op een bepaalde opgavenserie aan de uitslagen bij andere opgavenseries te toetsen, is het de gebruiker zeker mogelijk tot een betrouwbare evaluatie te komen van de mate waarin een kind de rekenstof in zijn diverse onderdelen beheerst.

Tenslotte is een vergelijking met de betrouwbaarheidscoëfficiënten welke Schonell⁵ in zijn boek „Diagnosis and Remedial Teaching in Arithmetic” geeft, nuttig om georiënteerd te zijn, welke waarden elders werden gevonden.

Voor de tests „optellen”, „aftrekken” en „vermenigvuldigen”, elk bestaande uit 100 opgaven en afgenomen op een groep van 100 negenjarigen, geeft Schonell respectievelijk .97, .98 en .88 als betrouwbaarheidscoëfficiënten (berekend volgens de split-half methode en gecorrigeerd voor een tweemaal zo lange testserie). Deze tests hebben echter

betrekking op het cijferend rekenen. Het gaat hierbij om optellen, aftrekken en vermenigvuldigen onder elkaar. De tests zijn dus niet zonder meer te vergelijken met de opgaven van blad 3 en 4 van de S.R.T. Bovendien vertonen de tests een zeer grote spreiding in moeilijkheid tussen de eerste en de laatste sommen. Zo begint de test optellen met $14 + 3$ (onder elkaar) en eindigt met $951 + 382 + 467 + 539 + 196$ (onder elkaar). De stof heeft dus op meerdere leerjaren betrekking. De veel hogere betrouwbaarheids-coëfficiënten bij de tests „optellen” en „aftrekken” van Schonell vinden dan ook zeer waarschijnlijk een verklaring in het verschil in opzet van deze tests en de S.R.T. Combinatie van opgavenserie door de diverse leerjaren heen bij de S.R.T. zal vermoedelijk leiden tot waarden in de zelfde orde van grootte als door Schonell werden gevonden.

VII *Rapportcijfers, testresultaten en intelligentie*

Daar in het voorjaar '64 naast 54 nieuwe kinderen ook de 54 kinderen welke in het najaar '63 aan het onderzoek hadden meegedaan, opnieuw werden onderzocht, beschikten wij dus over een totaal van 108 kinderen, d.i. 36 per school. Bij deze groep kinderen hebben wij de correlaties berekend tussen de uitkomsten bij de opgavenseries 3-2 t.e.m. 3-5, de opgavenseries 4-2 t.e.m. 4-5, de opgavenserie 16, het I.Q. op de Pintner-Durost test en het rapportcijfer rekenen. De resultaten van deze correlatieberekeningen vindt men in tabel VI. Wij berekenden deze correlaties voor elke school apart. De uitkomsten van school 1, 2 en 3 vindt men in deze tabel onder elkaar vermeld.

Tabel VI.

	Automatiseringen	Toepassing van automatiseringen	Redactiesommen	Intelligentie	Rapportcijfers	
Automatiseringen	—	.65	.63	.46	.57	school 1
3—2 t.e.m. 3—5	—	.40	.46	.43	.48	school 2
	—	.83	.37	.46	.44	school 3
Toepassing van automatiseringen	.65	—	.53	.50	.54	school 1
4—2 t.e.m. 4—5	.40	—	.49	.46	.75	school 2
	.83	—	.41	.42	.33	school 3
Redactiesommen	.63	.53	—	.64	.61	school 1
16—1	.46	.49	—	.58	.46	school 2
	.37	.41	—	.31	.05	school 3
Intelligentie	.46	.50	.64	—	.36	school 1
	.43	.46	.58	—	.56	school 2
	.46	.42	.31	—	.65	school 3
Rapportcijfers	.57	.54	.61	.36	—	school 1
	.48	.75	.46	.56	—	school 2
	.44	.33	.05	.65	—	school 3

Uit het overzicht van de gevonden correlaties in tabel VI blijkt:

1. dat intelligentie, rapportcijfers rekenen en de diverse facetten van het rekenen een zeer duidelijke onderlinge samenhang vertonen, zonder dat één van deze criteria duidelijk een aparte positie inneemt. Op een enkele uitzondering na variëren alle correlaties tussen de .30 en de .65.
2. dat de verschillende criteria niet met elkaar gelijk gesteld kunnen worden, daarvoor zijn de correlaties weer te laag. Het onderscheid tussen verschillende facetten bij het rekenen lijkt derhalve relevant.
3. dat er duidelijke verschillen tussen de scholen onderling zijn te constateren, verschillen die meer relevant lijken dan tussen de aspecten van het rekenen, intelligentie en rapportcijfer.

Zo zien we dat het verband tussen I.Q. en rapportcijfer duidelijk toeneemt al naar gelang men te maken heeft met school 1 ($r = .36$) of met school 2 ($r = .56$) of 3 ($r = .65$). Het rapportcijfer is bij deze drie scholen echter op verschillende wijze tot stand gekomen. Bij school 3 ontbreekt elke correlatie met redactiesommen ($r = .05$), terwijl ook de

correlatie met „automatiseringen” en „toepassing van automatiseringen” in vergelijking met de beide andere scholen aan de lage kant blijft.

Trouwens bij school 3 lijkt het verschil tussen „automatiseringen” en „toepassing van automatiseringen” het minst relevant, gezien de correlatie van .83 welke wij tussen beide criteria vonden. Het rapportcijfer van school 3 houdt in feite meer verband met het I.Q. dan met het testresultaat! Het rapportcijfer bij school 2 wordt primair bepaald door het resultaat bij de toepassing van automatiseringen. Redactiesommen en primaire automatiseringen spelen daarnaast een duidelijk aanwijsbare rol. Bij school 1 valt primair op het grotere onderlinge verband tussen de drie rekenfacetten, respectievelijk .65, .63 en .53! Ook het rapportcijfer sluit zich hierbij aan, zonder dat één der rekenfacetten een opvallend grotere bijdrage levert. (.57, .54 en .61).

Opvallend is echter dat de intelligentie enerzijds een redelijke samenhang vertoont met de diverse rekenfacetten (.46, .50 en .64), maar daarentegen zwak samenhangt met de rapportcijfers, hoewel deze zoals wij zagen juist wel een duidelijke samenhang met de testresultaten tonen. Rapportcijfers en intelligentie, beide samenhangend met de testresultaten, waardenen deze blijkbaar vanuit een ander gezichtspunt.

De vraag rijst in hoeverre de gevonden verschillen tussen de scholen het gevolg zijn van het verschil in sociaal niveau. Zou dit n.l. op deze factor te herleiden zijn, dan opent dit een zicht op een sociologische benadering van het onderwijs(bedrijf). Deze vraag is in zijn algemeenheid moeilijk te beantwoorden. Het feit echter dat de resultaten bij de redactiesommen bij school 3 in het geheel niet meewegen bij het rapportcijfer, wat bij school 2 en zeker bij school 1 wel het geval is, zou in genoemde richting wijzen. Ook de samenhang tussen I.Q. en redactiesommen die vrij groot is bij de scholen 1 en 2 (respectievelijk .58 en .64), en vrij laag bij school 3 (.31) zou hieruit verklaard kunnen worden. Het kunnen lezen en het kunnen begrijpen van het gelezene speelt uiteraard in de sociaal hoger geklasseerde bevolkinglagen een meer belangrijke rol.

Uit het bovenstaande tekenen zich geen definitieve lijnen af. Wel wordt duidelijk dat de samenhang tussen de drie rekenfacetten, intelligentie en rapportcijfer bij elk van de drie scholen een andere is. Samenhang tussen I.Q. en rapportcijfer is afhankelijk van de betrokken school. Maar ook is de samenhang tussen de testfacetten en het rapportcijfer sterk afhankelijk van de betrokken school. De rapportcijfers verschillen dus niet alleen wat de onderlinge vergelijkbaarheid betreft qua gemiddelde en spreiding, zoals De Groot⁶ aantoonde, maar

ook inhoudelijk. Het rapportcijfer rekenen op school 3 heeft ook inhoudelijk een andere betekenis dan voor school 2 of 1. Wij menen dat dit de belangrijkste conclusie moet zijn van deze verkenning met betrekking tot de samenhang van onze gebruikte criteria. Deze conclusie impliceert dat bij grote landelijke ijkingen het gevaar bestaat dergelijke verschillen van school tot school, die niet slechts het niveau maar met name de structuur betreffen, te overdekken. Het lijkt met name voor plaatselijke schoolpsychologische of schoolpedagogische diensten van belang juist deze structuurverschillen bloot te leggen.

VIII *De meer algemene geldigheid van de Schiedamse uitkomsten*

De vraag in hoeverre de resultaten ook buiten Schiedam geldigheid hebben, is moeilijk te beantwoorden. Een definitief antwoord op deze vraag kan slechts gegeven worden op grond van een uitgebreid landelijk onderzoek. Zulk een onderzoek was binnen de mogelijkheden, die de Schiedamse situatie bood, niet te realiseren. Toch gaat het hier om een vraag van buitengewoon groot belang. Om toch enigermate georiënteerd te zijn in hoeverre aan de Schiedamse resultaten een meer algemene geldigheid kan worden toegekend, werd een beknopt verkennend onderzoek ingesteld op enkele scholen elders in het land.

Hiervoor werd een school in Groningen, in Amsterdam, in Wijk en in Middelbeers uitgekozen. Het onderzoek bleef beperkt tot enkele opgavenseries en werd slechts verricht in de klassen II, III en IV. De afname viel omstreeks de herfstvakantie, waardoor de resultaten vergelijkbaar werden met de Schiedamse uitkomsten (respectievelijk II-a, III-a en IV-a).

In tabel VII geven wij een overzicht van de uitkomsten in klas III voor respectievelijk de Schiedamse en de landelijke scholen. Hierbij werd echter niet de absolute uitkomst maar de onderlinge rangorde vermeld, waarbij 1 het hoogste en 7 het laagste niveau aangeeft. Naast dit gemiddelde rangcijfer, berekend voor de 10 afgenomen opgavenseries, bepaalden wij, analoog aan de wijze die bij de Schiedamse scholen gevolgd werd, per school een gemiddeld I.Q. en een index voor het sociaal milieu (zie 3.1 en 3.2).

Tabel VII.

	Schiedam		A'dam		Gron.	W'wijk	M'beers
	sch. 1	sch. 2	sch. 3	sch. 4	sch. 5	sch. 6	sch. 7
Gem. rangcijfer :	5,3	4,9	6,2	3,9	2,6	3,9	1,3
I.Q. :	107	107	95	108	110	109	102
Soc. Index :	3,8	4,0	5,7	4,4	3,4	4,6	5,0

In dit overzicht valt op dat school 3, zowel qua intelligentie, qua sociale index als qua leerresultaat op de laatste plaats komt. Geheel in strijd met de veronderstelling dat er een logisch verband zou zijn tussen leerresultaten, intelligentieniveau en sociale index zijn de uitkomsten bij school 7. Deze school staat bovenaan wat de leerresultaten betreft maar komt qua intelligentie en sociale index op de zesde plaats. Duidelijk komt uit tabel VII naar voren, dat de resultaten te Schiedam in vergelijking met de vier scholen elders slechter zijn. Mogelijk is dit een systematisch gegeven, mogelijk een toevallige uitkomst welke bij een gedegen vergelijkend onderzoek niet bevestigd zou worden. Om hierin enig inzicht te krijgen, hebben wij de resultaten van school 1 en 2 (Schiedam) vergeleken met de resultaten van school 4 (Amsterdam) en 5 (Groningen). Deze scholen zijn als stadsscholen beter te vergelijken met de Schiedamse scholen. Wij hebben hier tevens een vergelijkbare groep qua intelligentie (gemiddeld I.Q. van 107 naast een gemiddeld I.Q. van 109) als ook qua gemiddeld sociaal milieu. Voor de opgavenseries, afgenomen in de eerste helft van klas III bepaalden wij de kans dat het verschil in uitkomst tussen Schiedam (school 1 + 2) en Amsterdam-Groningen (school 4 + 5) aan het toeval kan worden toegeschreven. De berekening vond plaats volgens de X^2 -methode in een 2 maal 2 tabel: aantal kinderen dat de betreffende opgavenseries beheerst (8, 9 of 10 opgaven goed) tegenover het aantal kinderen dat de opgavenseries niet beheerst (minder dan 8 van de 10 opgaven goed).

De gevonden verschillen blijken bij de 10 opgavenseries, afgenomen in het begin van klas III nergens significant te zijn op het 1% niveau. Alleen de opgavenserie 4-4 is significant op het 5% niveau, terwijl in drie gevallen (2-4, 3-5 en 4-3) het verschil praktisch nihil is en daardoor geheel als toevallig kan worden gekenmerkt.

Conclusie:

Hoewel nergens statistisch significant, blijken de resultaten bij de combinatie Groningen-Amsterdam als regel beter uit te vallen dan bij de twee Schiedamse scholen. Wij menen hieruit te kunnen concluderen dat de Schiedamse situatie leidt tot een te pessimistisch beeld van de leerstofbeheersing. Het feit echter dat anderzijds de uitkomsten van het Schiedamse onderzoek vermoedelijk een te optimistische schatting geven van de werkelijke situatie ter plaatse (het gemiddeld I.Q. van de onderzoekgroep is 103) impliceert dat de feitelijke onderzoeksresultaten de landelijke situatie, voor zover het de steden betreft, beter

benaderen dan wanneer te Schiedam met een zuiver representatieve steekproef was gewerkt.

8.1 *Het onderzoek van Drs. Mommers in Brabant*

In zijn boekje „Naar een meer objectieve benadering van leerprestaties” van Frater M. Cesarius Mommers⁷ geeft deze de resultaten weer op een serie redactiesommen, afgenomen in klas IV en V van een groep Brabantse scholen. Het betrof hier scholen uit Tilburg en uit de dorpen van Midden en Oost-Brabant, waarbij ruim 2000 leerlingen waren betrokken.

Wij hebben de 25 sommen van Mommers, bestemd voor klas IV tegelijkertijd met de eenvoudige redactiesommen uit het Schiedamse onderzoek in een vierde klas lagere school afgenomen. Hierdoor werd het mogelijk de beide opgavensreeksen ongeveer te vergelijken wat moeilijkheid betreft.

Uitgaande van dit gebleken verschil in moeilijkheid maakten wij een schatting ter beantwoording van de vraag hoe de Brabantse resultaten geweest zouden zijn, indien de eenvoudige sommen van het Schiedamse onderzoek afgenomen waren. Onze conclusie moet zijn dat de Brabantse kinderen de Schiedamse opgaven beter zouden hebben gemaakt. Globaal geschat moet verwacht worden dat de Brabantse kinderen zich gemiddeld één quintiel hoger zouden klasseren dan de Schiedamse (d.w.z. de 20 % middelmatigen volgens de uitkomsten van het Brabantse onderzoek behoren dan volgens de Schiedamse uitkomsten al tot de beteren). Dit zou goed aansluiten bij het feit dat de Brabantse dorpsschool van de vier landelijke scholen, ondanks de middelmatige begaafdheid van de leerlingen als de beste naar voren kwam.

Een verdere speculatie leidt tot de vraag of deze betere leerresultaten speciaal een Brabants gegeven zijn, of dat men hierin een aanwijzing moet zien, dat de leerresultaten op het platteland beter zijn dan in de grote steden, waarbij wij dan met name aan Schiedam denken als deel van de groter Rotterdamse agglomeratie.

In ieder geval manen deze gegevens tot voorzichtigheid de Schiedamse uitkomsten zonder meer te generaliseren.

IX *Samenvatting*

In dit artikel werd een verslag gegeven over de diverse aspecten met betrekking tot de ijking van de Schiedamse Rekentest.

Zoals in de inleiding werd uiteengezet, was het niet mogelijk bij de opzet en de uitvoering hiervan aan alle methodologische eisen, welke in principe aan een dergelijk onderzoek gesteld moeten worden, te

voldoen. Als zodanig houdt dit onderzoek in meer strikt wetenschappelijke zin een voorlopig karakter. Het hier gegeven verslag moet men dan ook meer zien als een kritische toetsing achteraf, dan als de verantwoording van een van te voren in alle opzichten wetenschappelijk gefundeerde opzet. De auteurs beperkten zich bij deze controle achteraf voornamelijk tot de gegevens van de derde klas.

Uit deze gegevens bleek enerzijds de wenselijkheid van een aantal verbeteringen, zowel t.a.v. de samenstelling van de opgavenseries als t.a.v. de keuze van de scholen, anderzijds bleken deze bezwaren niet van dien aard, dat de waarde van de test voor de Schiedamse situatie hierdoor ernstig wordt aangetast.

Daarnaast werd belangrijke informatie verkregen over de toename van de beheersing van de rekenstof door de jaren heen. Het opvallende was, dat over de periode oktober-april geen significant verschil kon worden aangetoond.

Bij een vergelijking van rapportcijfers, intelligentieniveau en de resultaten op de S.R.T. bleek verder, hoezeer het onderlinge verband tussen deze gegevens van school tot school verschilt.

Hiermee is opnieuw een aanwijzing verkregen, hoe individualistisch de respectievelijke scholen hun onderwijs en hun waardering geven en hoe belangrijk het is over objectieve criteria te beschikken om de resultaten onderling vergelijkbaar te maken.

Tenslotte deden wij een poging enig houvast te krijgen, hoe de Schiedamse resultaten landelijk gesitueerd zouden moeten zijn. Uit de verschillen tussen plattelands- en stadsscholen bleek dat dit probleem gecompliceerder ligt dan wij aanvankelijk meenden. Onze indruk, dat de Schiedamse resultaten de stedelijke situatie in ons land redelijk benaderen, blijft dan ook een vrij speculatieve.

Literatuur

1. H. HEESSEN, D. STRELITSKI *en* A. VAN DER WISSEL: Een onderzoek naar de beheersing van de rekenstof in het G.L.O. (Paedagogische Studiën, 44e jaargang, nr. 11).
2. A. D. DE GROOT: Methodologie (Mouton, Den Haag, 1961).
3. FRED J. SCHONELL: Practice in Basic Arithmetic. (Oliver and Boyd Ltd. Edinburgh-London).
4. FRED J. SCHONELL: The Diagnostic Arithmetic Tests. (Oliver and Boyd Ltd. Edinburgh-London).
5. FRED J. SCHONELL *en* C. ELEANOR SCHONELL: Diagnosis and Remedial Teaching in Arithmetic. (Oliver and Boyd Ltd. Edinburgh-London).
6. A. D. DE GROOT: Vijven en zessen. Wolters-Noordhoff-Groningen).
7. Frater M. CESARIUS MOMMERS: Naar een meer objectieve benadering van leerprestaties. (Zwijssen-Tilburg).