

leidse diagnostische test

experimentele versie

cognitieve ontwikkeling
leervermogen en schoolprestaties

J.J.F. Schroots

ERRATA

Pagina 31: Tussen de 6e en 7e regel van onderen is de volgende tekst weggefallen:

"Een extreem operationale opvatting doet echter tekort aan het psycho-"

Pagina 40: Tussen de 13e en 14e regel van onderen is de volgende tekst weggefallen:

"Plaatjes aanwijzen is de enige test waarbij intermodale informatiever-"

STELLINGEN

1. Het onderzoek naar de structuur van de intelligentie is voorlopig meer gediend met een a priori model van de cognitieve ontwikkeling van functies dan met het a posteriori resultaat van faktoranalyse op bestaande intelligentietests.
2. Indien bij grenzen-toetsen geen rekening wordt gehouden met het regressie-effekt, dient men zich van een uitspraak over het leervermogen te onthouden.
3. De sterke nadruk die subjectivistisch georiënteerde psychologen leggen op een kwalitatieve benadering van de diagnostiek verraaft niet alleen gebrek aan logisch inzicht, maar ook onwetendheid omtrent de kwalijke uitvloeisels van het dialektisch continuüm 'kwantitatief - kwalitatief'.

Barth, E.M. De 19e eeuwse 'idealistische' logica en haar uitwerkingen. Intermediair 14 (1978) 1-11

4. Het is veelzeggend dat onderwijsadviescentra in het kader van hun professionaliseringsactiviteiten geen enkele aandacht besteden aan de studie der jeugdgezondheidszorg.
5. Indien na een onderzoek met de Leidse Diagnostische Test (LDT) enkel het deviatie IQ gerapporteerd wordt, levert dit bij 32.5% van de leerlingen uit het kleuter- en gewoon lager onderwijs (4 tot 8 jaar) en bij 89% van de leerlingen uit het LOM-onderwijs (6 tot 8 jaar) een onvolledig en vertekend beeld op van hun cognitieve ontwikkeling.
6. De opvatting dat een gefalsificeerd model verlaten moet worden ten gunste van een ander model verdient pas dan aandacht, indien het andere model meer verschijnselen kan verklaren dan het gefalsificeerde model.
7. De toevoeging 'freuduleus' als nadere aanduiding voor Freud's persoonlijkheidstheorie is mogelijk een verschrijving, maar zeker geen freudiaanse.

Ref.: Bersoff, D.N. Silk purses into Sow's ears; the decline of psychological testing and a suggestion for its redemption. Amer. Psychol. 28 (1973) p. 892

8. Aangezien iedere psychologische test aanspraak kan maken op het epitheton ornans 'hersenorganisch', is de LDT naast intelligentie- en diagnostische test ook 'hersenorganiciteits'test.

9. Aan beleidsvoerders moet zowel op eigen verzoek als op dat van anderen de mogelijkheid worden geboden om zonder gezichtsverlies een stap terug te doen in de hiërarchie van een arbeidsorganisatie.
10. De instelling van een apart ordinariaat in de psychodiagnostiek, evenals de bundeling van her en der verspreide testontwikkelingsactiviteiten in één centraal instituut, zal de kwaliteit van het psychodiagnostisch onderzoek ten goede komen.
11. Voor het aankweken van 'goede smaak' bij de interpretatie van clavichordmuziek is kennis van vingerzettingen uit de renaissance en barok een belangrijk vereiste.
12. De stelselmatige vergrijzing van het wetenschappelijk corps in de sub-fakulteiten der psychologie kan een stimulans betekenen voor de studie der gerontologie.

Stellingen behorende bij het proefschrift van J.J.F. Schroots: Cognitieve ontwikkeling, leervermogen en schoolprestaties.

Amsterdam, 20 september 1979

VRIJE UNIVERSITEIT TE AMSTERDAM

COGNITIEVE ONTWIKKELING,
LEERVERMOGEN
EN SCHOOLPRESTATIES

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van
doctor in de sociale wetenschappen
aan de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
dr. H. Verheul,
hoogleraar in de faculteit der wiskunde en natuurwetenschappen,
in het openbaar te verdedigen
op donderdag 20 september 1979 te 13.30 uur
in het hoofdgebouw der universiteit,
De Boelelaan 1105

door

JOHANNES JACOBUS FREDERICUS T SCHROOTS

geboren te Rotterdam

Swets & Zeitlinger, Lisse
1979

Promotor : Prof.Dr. P.J.D. Drenth

Coreferent: Dr. G.J. Mellenbergh

It would be a waste of valuable time if the
psychologist's only goal were to produce an IQ.

Jerome M. Sattler

VOORWOORD

In de serie over de Leidse Diagnostische Test (LDT) zijn tot nog toe vier delen verschenen: deel I Handleiding, deel II Handleiding voor het stroomdiagram, deel III Stroomdiagram en deel IV Uitspraken. Dit proefschrift - handelend over een experimentele versie van deze test, de LDT-E - is deel V uit de serie. Deel V is op zijn beurt in twee delen gesplitst: A. Cognitieve ontwikkeling en schoolprestaties, en B. Leervermogen en schoolprestaties.

Aan deel A gaat een algemene inleiding vooraf. Hierin wordt het kader geschetst waarbinnen de LDT - als één van de meetinstrumenten voor de individuele psychodiagnostiek van leermoeilijkheden - ontwikkeld werd. Deel A zelf behelst een uitvoeriger toelichting op het Mark-model dan in de handleiding voor de LDT (deel I) gegeven kon worden. In aansluiting op deze toelichting vindt de verslaggeving plaats van een exploratief onderzoek naar de betekenis en voorspellende waarde van de LDT-E. Aangezien het hier om een experimentele versie gaat, zijn de resultaten van dit onderzoek niet in alle opzichten exemplarisch voor de LDT. Wél vormen zij een hopelijk nuttige informatie- en inspiratiebron voor de praktiserend psycholoog die met de LDT werkt enerzijds en voor de onderzoeker-psycholoog die tot begripsvalidering van de LDT overgaat anderzijds.

Deel B heeft betrekking op 'grenzen-toetsen', een klinische onderzoeksprocedure die vaak als aanvulling op het traditionele standaard-testonderzoek wordt gevolgd. Door middel van grenzen-toetsen hoopt men in de praktijk van het testonderzoek onder meer het leervermogen van een kind te kunnen vaststellen. In dit deel wordt niet alleen een literatuuroverzicht gegeven van de achtergronden hiervan, maar ook wordt verslag gedaan van een vooronderzoek naar grenzen-toetsen op vier tests uit de LDT-E. De beschrijving van de verkregen resultaten wordt gevolgd door een uitgebreide discussie over de (on)mogelijkheden van grenzen-toetsen, alsmede door enkele suggesties voor verder onderzoek. De relatie tussen de cognitieve ontwikkelingsdiagnostiek en behandeling van leermoeilijkheden vormt het onderwerp van de slotbeschouwing uit dit proefschrift.

Veel mensen zijn - direkt of indirekt - behulpzaam geweest bij het totstandkomen van mijn proefschrift. Ofschoon het ondoenlijk is ieders bijdrage volledig tot zijn recht te doen komen, wil ik hier graag toch enkele mensen persoonlijk bedanken.

Mijn promotor en leermeester, Prof.Dr. P.J.D. Drenth, dank ik voor zijn

heldere en zakelijke begeleiding, en mijn coreferent, Dr. G.J. Mellenbergh, voor zijn kritisch stimulerend kommentaar. Hun invloed op dit proefschrift is naar ik hoop niet alleen voor mij duidelijk zichtbaar.

Ook mijn kollega-projectleiders, Dr. J.H.A. Groenendaal en Drs. R.M. van Waarden-Ramondt, ben ik dankbaar voor hun vriendschappelijke samenwerking binnen het kader van SVO-project 0274 'Vroegtijdige herkenning van en hulp aan kinderen (4-8 jaar) die op enigerlei wijze vastlopen in het onderwijs'. Zonder hun daadwerkelijke inzet bij het onderzoek zou deze publikatie niet verschenen zijn.

Prof.Dr. R. van der Vlist heeft als onderdirecteur van het NIPG-TNO bij de start van bovengenoemd project alle steun gegeven die nodig was voor het welslagen daarvan. Niet in de laatste plaats is dit welslagen bevorderd door de wetenschappelijke en bestuurlijke inbreng van Prof.Dr. J. de Wit, alsmede door de ruime faciliteiten die geboden werden door de heren W. van Dongen en G.A. van 't Hoff, respectievelijk voormalig en huidig directeur van het Gewestelijk Onderwijs Advies Centrum 's-Hertogenbosch e.o.

Drs. R.J. van Alphen de Veer heeft vrijwel van het begin af aan mee gewerkt aan de ontwikkeling van de LDT. Zijn enthousiaste hulp bij het oplossen van allerlei onderzoeksproblemen is voor mij heel belangrijk geweest. Mijn voormalige kollega's Drs. A.E. Akkerman en mw. Drs. A. de Groot zullen ongetwijfeld de uitslag van veel enerverende discussies in deel A van dit proefschrift herkennen. Met warmte denk ik terug aan 'ons' projectteam, waarin ook mw. E.G.M. Janssen en mw. J.W. Plomp al spoedig onmisbaar bleken.

Gaarne wil ik verder jegens de volgende medewerkers van het NIPG-TNO mijn dank uitspreken: dhr N.H. de Kleijn, voor de precieze, technische afwerking van het manuscript; dhr H.L.M. Nieuwland, voor de snelle en efficiënte wijze waarop hij zorg droeg voor de rekenmachinale verwerking van het gegevensbestand; dhr W.N. van Nooten voor zijn ironisch-geformuleerd, maar niettemin degelijk statistisch advies; en mw. D.M. Tuinstra-Williams voor de nauwkeurigheid waarmee zij als vanouds de literatuurlijst verzorgde. De liefdevolle zorg en grote nauwgezetheid waarmee mw H.G.S. Nooy van der Kolff het vele typewerk verrichte, heb ik in de loop der jaren steeds meer leren waarderen.

Zonder de hulp van Drs. A. Koet, M.A. zou de 'summary' vast veel korter, maar zeker niet beter zijn geworden.

De directeur van het NIPG-TNO, Dr. M.J. Hartgerink, ben ik zeer erkentelijk voor de ruimte die hij mij gaf om dit proefschrift te voltooien. Mw.

Flora van Laar, jeugdarts, ben ik niet minder erkentelijk voor de wijze waarop zij mij op cruciale momenten in het onderzoek van advies diende. Niet zonder tevredenheid vermeld ik verder de financiële steun die ik van de Stichting voor Onderzoek van het Onderwijs (SVO) mocht ontvangen. Aan Jolet Plomp tenslotte ben ik meer dank verschuldigd dan ik hier wil verwoorden.

Rest mij de wens uit te spreken dat ook deze uitgave zijn weg mag vinden in de praktijk van kinderpsychologen, onderwijsadviesdiensten, pedologische instituten, kinderklinieken en universitaire afdelingen voor onderwijs en onderzoek op het gebied van leermoeilijkheden.

NIPG-TNO, Leiden, mei 1979

J.J.F. Schroot's

INHOUD

	pag.
VOORWOORD	III
INHOUD	VI
LIJST VAN FIGUREN	VIII
LIJST VAN TABELLEN	IX
1. ALGEMENE INLEIDING	1
1.1 Achtergrond	1
1.2 Leerstoornissen	3
1.3 Uitgangspunten en eerste doel van het onderzoek	9
A. COGNITIEVE ONTWIKKELING EN SCHOOLPRESTATIES	13
2. MARK-MODEL	15
2.1 Inleiding tot de theorie van Mark	15
2.2 Mark's theorie over de ontwikkeling van leer- en denkprocessen	17
2.3 Model	22
3. LDT-E: SAMENSTELLING, BESCHRIJVING EN PLAATSBEPALING	31
3.1 Samenstelling	31
3.2 Beschrijving van de tests	37
3.3 Overeenkomst en verschil tussen LDT-E en LDT	42
3.4 Plaatsbepaling	43
4. TYPERING VAN HET ONDERZOEK	47
4.1 Inleiding	47
4.2 Frankforter Schoolrijpheidstest en Nijmeegse Schoolbekwaamheidstest	50
4.3 Steekproef	52
4.4 Schoolprestaties	57
5. BETROUWBAARHEID	61
5.1 Itemkenmerken	61
5.2 Interne consistentie en homogeniteit	63
6. STRUKTUUR	69
6.1 Invloed van leeftijd, geslacht en sociaal- ekonomisch niveau	69
6.2 Korrelaties	75
6.3 Faktorstructuur	79
6.4 Faktorstructuur naar leeftijd, geslacht en sociaal-ekonomisch niveau	85
6.5 Faktorstructuur in het licht van de FST en NST	93
7. PREDIKTIEVE VALIDITEIT	97
7.1 Korrelaties tussen LDT-E en criteriumvariabelen	97
7.2 Stapsgewijze multiële regressie	103
7.3 Canonische korrelaties tussen LDT-E en criterium- variabelen	105

	pag.
8. DISKUSSIE	111
8.1 Mark-model: terugblik en vooruitzicht	112
8.2 LDT-E als intelligentietest	119
8.3 LDT-E als diagnostische test	126
B. LEERVERMOGEN EN SCHOOLPRESTATIES	145
9. LEERVERMOGEN EN GRENZEN-TOETSEN	147
9.1 Psychodiagnostiek in beweging	147
9.2 Leervermogen	150
9.3 Grenzen-toetsen	152
9.4 Tweede doel van het onderzoek	157
10. METHODE	161
11. RESULTATEN	169
11.1 Betrouwbaarheid	169
11.2 Structuur	174
11.3 Bewerking van de schaalgegevens	179
11.4 Prediktieve validiteit	182
12. DISKUSSIE	191
12.1 Kanttekeningen bij de methode	191
12.2 Learning potential assessment	197
12.3 Suggesties voor verder onderzoek	201
SLOTBESCHOUWING	207
SUMMARY	211
NOTEN	231
LITERATUUR	239
BIJLAGE	255

LIJST VAN FIGUREN

	pag.
Figuur 1. Schema van de psychologische test als meetinstrument in een 'input-black box-output' model	16
Figuur 2. Kontrolesysteem	19
Figuur 3. Gedragsklassifikatieschema	20
Figuur 4. Dwarsdoorsnee van een kanaal	20
Figuur 5. Schema van één enkel leerkanaal waarin n functies onderscheiden worden op ontwikkelingstijdstip t_1	26
Figuur 6. Schema van één enkel leerkanaal waarin de functies $x_1, x_2, x_3 \dots x_n$ serieel en de functie x_{LTM} parallel geschakeld zijn	27
Figuur 7. Schema van informatieverwerking via n hiërarchisch gerangschikte functies	28
Figuur 8. Schema van aanbieding en scoring der items per test tijdens het standaard- en grenzentoetsend onderzoek . . .	160

LIJST VAN TABELLEN

	pag.
Tabel 1. Verkorte lijst van 24 kanalen	24
Tabel 2. Verkorte lijst van 14 systemen	24
Tabel 3. Kanuaal- en funktiekenmerken van de tests	36
Tabel 4. Overzicht van subtests en bijbehorende taken in de Frankforter Schoolrijpheidstest en de Nijmeegse Schoolbekwaamheidstest	51
Tabel 5. Steekproefverdeling naar leeftijd, sociaal-ekonomisch niveau en geslacht	56
Tabel 6. Moeilijkheidsgraad en discriminatie-index van de items, per test	60
Tabel 7. Maximumscore, gemiddelde, standaarddeviatie, koëfficiënt van interne consistentie en homogeniteits- koëfficiënt van de tests	64
Tabel 8. Drieweg variantie-analyse over de tests en het totaal met leeftijd, geslacht en sociaal-ekonomisch niveau als onafhankelijke variabelen	72
Tabel 9. A posteriori-analyse van SEN-effekten op enkele tests en het totaal	73
Tabel 10. Korrelaties tussen de tests en het totaal	74
Tabel 11. Korrelaties tussen de tests, gekorrigeerd voor 'attenuatie'	77
Tabel 12. Korrelaties tussen de tests	256
Tabel 13. Faktorladingen van de tests na varimaxrotatie over vier factoren	81
Tabel 14. Faktorladingen van de tests na varimaxrotatie over drie factoren	84
Tabel 15. Korrelaties tussen de tests, naar leeftijd	257
Tabel 16. Faktorladingen van de tests na varimaxrotatie over drie factoren, naar leeftijd	87
Tabel 17. Korrelaties tussen de tests, naar geslacht	258
Tabel 18. Faktorladingen van de tests na varimaxrotatie over drie factoren, naar geslacht	89
Tabel 19. Korrelaties tussen de tests, naar sociaal-ekonomisch niveau	259
Tabel 20. Faktorladingen van de tests na varimaxrotatie over drie factoren, naar sociaal-ekonomisch niveau	91
Tabel 21. Korrelaties tussen de LDT-E, FST en NST	260
Tabel 22. Faktorladingen van de (sub)tests uit de LDT-E, FST en NST na varimaxrotatie over drie factoren	94
Tabel 23. Korrelaties tussen acht criteriumvariabelen	98
Tabel 24. Korrelaties van de tests en het totaal met acht criteriumvariabelen	100

	pag.
Tabel 25. Kumulatieve multipele korrelaties van de tests met acht criteriumvariabelen	102
Tabel 26. Canonische korrelaties tussen prediktor- en criteriumvariabelen en de gewichten van deze variabelen in drie combinaties (I, II en III) voor twee canonische variaten per combinatie	106
Tabel 27. Moeilijkheidsgraad van de items uit Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste, tweede en derde aanbieding	168
Tabel 28. Discriminatie-index van de items uit Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste, tweede en derde aanbieding	170
Tabel 29. Maximumscore, gemiddelde, standaarddeviatie, koëfficiënt van interne consistentie en homogeniteitskoëfficiënt van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste, tweede en derde aanbieding	172
Tabel 30. Korrelaties tussen de tests bij eerste, tweede en derde aanbieding	176
Tabel 31. Relatieve frekwentieverdeling, gemiddelde en standaarddeviatie van de schaalwaarden van vijf schalen voor Blokpatronen, Vouwblaadjes, Natikken en Woordenspan na tweede en derde aanbieding	178
Tabel 32. Maximumscore, gemiddelde en standaarddeviatie van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste, tweede en derde aanbieding voor twee groepen: groep A 'zonder hulp' en groep B 'met hulp' . .	182
Tabel 33. Korrelaties van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste, tweede en derde aanbieding met acht criteriumvariabelen	184
Tabel 34. Vierweg kruistabel-gegevens voor Blokpatronen - Lezen/taal	262
Tabel 35. Vierweg kruistabel-gegevens voor Vouwblaadjes - Lezen/taal	263
Tabel 36. Vierweg kruistabel-gegevens voor Natikken - Lezen/taal	264
Tabel 37. Vierweg kruistabel-gegevens voor Woordenspan - Lezen/taal	265
Tabel 38. Chi-kwadraten aannemelijkheidsverhouding met aantal vrijheidsgraden voor vier modellen per test	189

1. ALGEMENE INLEIDING

1.1 Achtergrond

Kennis van lezen, rekenen, taal en schrijven is van wezenlijk belang in onze maatschappij. Ieder jaar weer blijken duizenden kinderen moeite te hebben met het verwerven van deze kennis. Cijfers van het Centraal Bureau voor de Statistiek (CBS, 1978) wijzen erop dat zo'n 5.5% ($\pm$ 75.000) van alle Nederlandse lagere schoolleerlingen het gewoon lager onderwijs niet kan volgen en aangewezen is op een van de vele vormen van buitengewoon onderwijs. Geen gegevens worden verstrekt over het percentage kinderen dat weliswaar het gewoon lager onderwijs volgt maar daar nogal moeite mee heeft.

In het Canadese CELDIC Report (1970) worden schattingen gegeven van het percentage leermoeilijkheden voor landen als Canada (10-16%), Engeland (14%), Frankrijk (12-14%) en de Verenigde Staten (10-15%). In deze percentages zijn ook de kinderen begrepen die een of andere vorm van buitengewoon onderwijs volgen. Indien we aannemen dat het onderwijs in deze landen niet al te zeer verschilt van dat in Nederland, dan leert een eenvoudige rekensom dat zo'n 5 à 10% van de kinderen moeite heeft met het volgen van gewoon lager onderwijs.

Wat ook de precieze percentages mogen zijn, de vraag rijst onmiddellijk of het niet mogelijk is zodanige maatregelen te treffen dat althans een aantal van deze moeilijkheden voorkomen kan worden. Dergelijke maatregelen zijn om verschillende redenen belangrijk.

Ten eerste: het stellen van een korrekte diagnose levert in geval van leermoeilijkheden veelal méér problemen op naarmate deze láter ontdekt worden, doorgaans pas in de tweede of derde klas van de lagere school. Tegen die tijd is het vaak bijzonder moeilijk uit te maken wat nu oorzaak of gevolg is van de aanvankelijke leermoeilijkheden. De primaire stoornissen - leesproblemen bijvoorbeeld - worden afgeschermd door allerlei sekundaire, emotionele stoornissen ten gevolge van slechte schoolprestaties in vakken die een goede leesvaardigheid vóóronderstellen. De kans dat aansluitende behandelingsmaatregelen hoofdzakelijk gericht zijn op het symptoomgedrag van het kind, in plaats van op de oorzaak van de leermoeilijkheden, wordt hierdoor aanzienlijk vergroot.

Ten tweede: het inhalen van een grote leerachterstand kost méér pijn en moeite, zowel voor het kind als zijn begeleider, dan het inhalen of mogelijk zelfs voorkómen van een kleine leerachterstand. Een tamelijk vanzelfsprekend argument, maar nog te vaak wordt uit het oog verloren

dat kinderen na een reeks van faalervaringen nauwelijks gemotiveerd zijn om mee te werken aan een onderwijsprogramma dat bedoeld is om de ontstane achterstand weg te werken.

Ten derde: afgaande op de jaarlijkse cijfers van het CBS (cf. Nota Speciaal Onderwijs, 1977) worden ieder jaar steeds méér kinderen verwezen naar het buitengewoon onderwijs. Zowel om redenen van menselijke als geldelijke aard achten wij deze ontwikkeling onwenselijk.

Voorkomen is om meer redenen dan de hier genoemde beter dan genezen. Duidelijk moge echter zijn dat de preventie van leermoeilijkheden de moeite van onderzoek waard is.

Sinds 1972 wordt op de afdeling Jeugdgezondheidszorg van het Nederlands Instituut voor Praeventieve Gezondheidszorg (NIPG-TNO) een dergelijk type onderzoek verricht onder de titel 'Vroegtijdige herkenning van leerstoornissen'¹. Doel van het onderzoek is de ontwikkeling van instrumenten en technieken voor de vroegtijdige signalering en diagnostiek van kinderen (4-8 jaar) met potentiële of feitelijke leermoeilijkheden.

Aanvankelijk richtten wij onze aandacht op de mogelijke toepassing van neuropsychologische tests voor de vroegtijdige opsporing van kinderen met het zogenaamde 'Minimal Brain Damage' (MBD)-syndroom. Deze kinderen zouden een variëteit aan leer- en gedragsmoeilijkheden vertonen ten gevolge van lichte beschadigingen, stoornissen of dysfuncties van de hersenen of - ruimer - het centrale zenuwstelsel. Al spoedig bleek echter dat neuropsychologische tests als meetinstrumenten voor afwijkend gedrag ten gevolge van lichte hersenbeschadigingen niet over de vereiste predictieve validiteit beschikken en vooralsnog ook niet zouden kunnen beschikken zolang de problemen rondom het onbetrouwbare, klinisch-neurologische criterium 'lichte hersenbeschadiging' nog niet waren opgelost (Schroots, 1972). Drie jaar later heeft Kalverboer (1975) vanuit een andere invalshoek laten zien welke feilen er nog meer aan het 'MBD'-concept en het daaraan gerelateerde onderzoek kleefden.

Gezien het onvruchtbare karakter van 'MBD'-onderzoek door middel van tests, besloten wij ons onderzoeksterrein te verleggen naar de vroegtijdige opsporing van leermoeilijkheden zelf. Dat ook hiermee het criteriumprobleem niet meteen was opgelost bleek pas toen we het begrip 'leermoeilijkheden' of 'leerstoornissen' nader onderzochten aan de hand van een aantal definities (cf. par. 1.2). Onze speurtocht langs de definities leverde echter wel enkele belangrijke uitgangspunten op voor de ontwikkeling van een instrument dat later de naam Leidse Diagnostische Test (LDT) zou krijgen.

Sinds oktober 1972 zijn vier verschillende versies van dit instrument getoetst op hun waarde. Eén versie van de LDT, afgenomen in februari/maart 1973 op een groep van 411 schoolverlatende kleuters (5- en 6-jarigen) is veruit de belangrijkste omdat tegelijk met de LDT - maar ook nadien - vele andere gegevens over dezelfde groep kleuters verzameld zijn. Deze gegevens kunnen een helderder inzicht verschaffen in de mogelijkheden en beperkingen van de LDT als meetinstrument voor de vroegtijdige herkenning van leerstoornissen. Een globale typering van dit onderzoek vindt plaats in paragraaf 1.3. Gezien het feit dat de hier beschreven versie van de LDT niet identiek is met de definitieve versie zoals beschreven in de handleiding (deel I; Schroots & Van Alphen de Veer, 1976), zullen wij deze in het vervolg aanduiden met 'experimenteel', afgekort tot LDT-E.

1.2 Leerstoornissen

Tot nog toe zijn de termen 'leermoeilijkheden' en 'leerstoornissen' afwisselend gebruikt. Deze afwisseling in terminologie geeft al enigszins aan welke verwarring er nog steeds op dit gebied bestaat. Opheldering van de verwarring lijkt het eenvoudigst door enkele definities van leerstoornissen te bespreken. Dumont (1976) bijvoorbeeld maakt onderscheid tussen twee typen leerstoornissen, primaire en sekundaire:

"Van een primaire leerstoornis is sprake als het kind op school leermoeilijkheden vertoont, tegen alle verwachting in en er geen motorische, zintuigelijke, intellektuele, emotionele of sociale factoren aanwijsbaar zijn waartoe die leerstoornis te herleiden is en van waaruit zij begrijpelijk te maken is" (p. 21).

Van sekundaire leerstoornissen - ook wel leermoeilijkheden genaamd - is volgens Dumont sprake indien de leermoeilijkheden het resultaat of een bijverschijnsel zijn van motorische stoornissen, zintuigelijke handicaps, intellektuele deficiënties, emotionele ontwikkelingsstoornissen en/of sociaal-educatieve ontwikkelingsbelemmeringen.

Met een indeling in primaire en sekundaire leerstoornissen lijkt in ieder geval één terminologische kwestie uit de weg geruimd: primaire leerstoornissen zijn de echte leerstoornissen en sekundaire leerstoornissen of leermoeilijkheden zijn de afgeleiden daarvan. Dat het gemaakte onderscheid misschien wel praktisch bruikbaar maar toch minder hard is dan het lijkt, willen wij illustreren aan het volgende.

Om te bepalen of een kind tegen alle verwachtingen in tot de primair leergestoorden behoort, maakt men gebruik van meetinstrumenten als

schoolvorderingentoetsen en intelligentietests. Met behulp van schoolvorderingentoetsen worden de schoolprestaties gemeten en met behulp van een intelligentietest de intellectuele capaciteiten van het kind. Volgens wordt een zogenaamd leerkotiënt bepaald (Myklebust, 1968), waarin de relatie tussen capaciteiten en prestaties als volgt wordt uitgedrukt:

$$\text{leerkotiënt} = \frac{\text{schoolvorderingenleeftijd}}{\text{verwachte leeftijd}} \times 100$$

waarin:

$$\text{verwachte leeftijd} = \frac{\text{intelligentielftd} + \text{kalenderlftd} + \text{klasselftd}}{3}$$

Het leerkotiënt of LQ - als meeteenheid te vergelijken met het IQ - vormt nu een maat voor de leergestoordheid van het kind. Als vuistregel wordt in de praktijk een kritische score van 90 gehanteerd (cf. Myklebust & Boshes, 1969). Kinderen met een LQ van 90 of hoger worden tot de groep leerlingen gerekend waarmee niets aan de hand is. Kinderen met een LQ van 89 of lager worden beschouwd als leergestoord met betrekking tot lezen, rekenen, taal of schrijven, afhankelijk van de gebruikte schoolvorderingentoets.

Ofschoon bovenstaande formule een nadere beschouwing verdient, gaat het ons in dit voorbeeld om het gegeven dat bij de bepaling of het kind al dan niet leergestoord is, gebruik wordt gemaakt van naar leeftijd genormeerde toetsen en tests. Welke toetsen en welke tests luidt dan de vraag, aangezien men in de praktijk nogal wat verschillen kan constateren tussen meetinstrumenten die tot een en dezelfde categorie behoren, bijvoorbeeld intelligentietests. Zo levert de Stanford-Binet, Form L-M (Terman & Merrill, 1960) hogere gemiddelde IQ's op dan de WPPSI (Wechsler, 1967), met verschillen die variëren van +1 tot +17 IQ punten (Sattler, 1974; p. 209). Bij toetsen uit eenzelfde categorie - leestoetsen bijvoorbeeld - mogen nog grotere verschillen verwacht worden, aangezien de vaardigheden die zij meten sterk van elkaar kunnen verschillen (bijvoorbeeld technisch lezen versus begrijpend lezen) en ook de leesprogramma's van school tot school variëren.

Genoemde verschillen hebben uiteraard consequenties voor het LQ dat voor een kind gevonden wordt en dus ook voor de vraag of het kind in de rubriek 'normaal', 'leerstoornissen' of 'leermoeilijkheden' thuishoort. De tamelijk arbitraire bepaling van de kritische score - 90, 85 of 80? - laten we dan nog maar buiten beschouwing (cf. Gaddes, 1976).

Ook in de Verenigde Staten heeft men zich met het definitieprobleem beziggehouden. In het kader van een 'National Project on Learning Disabilities in Children' (PHS Publ. 2015, 1969) inventariseerde een commissie onder leiding van Haring de bestaande definities, maar kon ten aanzien van geen enkele definitie tot overeenstemming komen. De commissie ging vervolgens zelf aan het werk en uit de vele definities die geformuleerd werden hield zij er twee over, welke voor het merendeel van haar leden acceptabel waren.

De twee definities luiden als volgt:

- I. "Children with learning disabilities are those
 - a) who have educationally significant discrepancies among their sensory motor, perceptual, cognitive, academic or related developmental levels which interfere with the performance of educational tasks;
 - b) who may or may not show demonstrable deviation in central nervous system functioning; and
 - c) whose disabilities are not secondary to general mental retardation, sensory deprivation or serious emotional disturbance" (p. 2 en 3).
- II. "Children with learning disabilities are those
 - a) who manifest an educationally significant discrepancy between estimated academic potential and actual level of academic functioning as related to dysfunctioning in the learning process;
 - b) who may or may not show demonstrable deviation in central nervous system functioning; and
 - c) whose disabilities are not secondary to general mental retardation, cultural, sensory and/or educational deprivation or environmentally produced serious emotional disturbance" (p. 3).

Ook deze definities geven aanleiding tot enkele opmerkingen. Het belangrijkste verschil treft men aan onder punt a). Bij de eerste definitie gaat het om discrepanties tussen cognitieve, perceptuele niveau's, etc. als zodanig, zij het dat de onderscheiden psychologische functies op enigerlei wijze van belang moeten zijn voor schooltaken. Dit in tegenstelling tot de tweede definitie, waarbij het gaat om discrepanties tussen schoolprestaties (actual level of academic functioning) en intellectuele capaciteiten (estimated academic potential). Het verschil tussen de twee definities is daarom zo belangrijk, omdat in de tweede definitie automatisch een grote groep kinderen - namelijk de kleuters - uitgesloten wordt van de groep (potentieel) leergestoorde kinderen. Aan gezien de kleuter in schoolvakken als lezen, rekenen en schrijven nog niets hoeft te presteren, behoort het vaststellen van significante discrepanties tussen schoolprestaties en intellectuele capaciteiten voor deze groep tot de onmogelijkheden. Recentelijk is ook in ons land door

Stevens (1975) hierop gewezen.

Punt b) luidt in beide definities letterlijk hetzelfde: aan leerstoornissen kan een aantoonbare dysfunctie van het centrale zenuwstelsel ten grondslag liggen. Niet duidelijk is waarom de kommissie dit punt in genoemde definities heeft opgenomen, temeer waar zij zelf stelt dat:

"Effective educational identification and specifications for remediation of learning disabilities are functional without any reference to associations with functional deviations of the central nervous system. Identification of an educational deficiency is adequate for remediation plans with or without positive neurological signs. Further requirements of positive neurological signs might preclude or delay necessary remediation" (p. 2).

Met de vrijblijvende formulering dat al dan niet aantoonbare dysfuncties van het centrale zenuwstelsel gekoppeld zijn aan het vóórkomen van leerstoornissen, wordt onzes inziens zowel de betekenis van het medisch-neurologisch onderzoek als het doel van de medisch-neurologische diagnostiek onderschat, ten gunste van het psychologisch-didaktisch onderzoek. In 1966 gaf een kommissie onder leiding van Clements reeds aan wat men van het medisch-neurologisch onderzoek mag verwachten (PHS Publ. 1415, 1966):

"The medical evaluation is essential to prevent the development or continuation of unsuspected disease processes" (p. 15).

Ten aanzien van het doel van de diagnostiek stelde zij:

"Medical diagnosis is designed to investigate or demonstrate the existence of causative factors of disease or injury capable of ameliorating or prevention" (p. 12).

Waar in het onderwijs aan leergestoorde kinderen nog steeds een tweetal fundamenteel verschillende adagia gehanteerd wordt, respektievelijk het 'teach to his weakness and improve it' (o.a. Kirk, 1966) en het 'teach to his strength' (o.a. Cohn, 1964) mag het verwonderlijk heten dat door Haring's kommissie zo weinig aandacht besteed wordt aan de bijdrage van medisch-neurologische zijde bij de definiëring en het onderzoek van leergestoorde kinderen. Dit laatste geldt des te sterker voor de aanhangers van het 'teach to his weakness'-adagium. Immers, het is weinig zinvol de remediale aandacht op niet-intakte gebieden van het centraal zenuwstelsel te richten wanneer bekend is dat beschadigingen of dysfuncties van het centrale zenuwstelsel - aan te tonen via medisch-neurologisch onderzoek - in het algemeen irreversibel zijn.

Punt c) in beide definities luidt min of meer hetzelfde en omvat de groepen kinderen met leermoeilijkheden of sekundaire leerstoornissen, zoals wij die reeds eerder aanduiden bij de definitie van Dumont. Aangezien waterdichte definities van kinderen met leermoeilijkheden niet bestaan,

wordt het geven van een definitie van leerstoornissen nog eens extra bemoeilijkt.

In dit korte overzicht mag tenslotte een wat oudere definitie van leerstoornissen, afkomstig van het 'National Advisory Committee on Handicapped Children' (1967), niet ontbreken, aangezien deze definitie in de Verenigde Staten de meeste bekendheid geniet en sinds april 1970 in niet minder dan 49 staten wettelijk van kracht is (Vaughan & Hodges, 1973).

Deze definitie luidt als volgt:

"Children with Specific Learning Disabilities exhibit a disorder in one or more of the basic psychological processes involved in understanding or in using spoken or written language. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling or arithmetic. They include conditions which have been referred to as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, developmental aphasia, etc. They do not include learning problems which are due primarily to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to environmental disadvantage" (Gaddes, 1976; p. 20).

Opvallend is het kompromis-karakter van deze definitie. Een ieder die zich bezighoudt met grotere of kleinere deelgroepen van leergestoorde kinderen kan zich via de vertrouwde maar weinig zeggende termen als 'minimal brain dysfunction', 'dyslexia', etc. in deze definitie terugvinden. De term 'Specific Learning Disabilities' verwijst verder niet naar iets specifiek, maar is de oude benaming voor 'learning disabilities'; het onderscheid loopt ongeveer parallel met de in Nederland nog gebruikelijke indeling in 'leerstoornissen' en 'leermoeilijkheden' (cf. Dumont, 1976; p. 21 e.v.).

Naast grote overeenkomst met de beide voorafgaande definities, laat deze definitie tenslotte een belangrijk verschil zien: de term 'educationally significant discrepancies' is vervangen door 'disorders in one or more of the basic psychological processes involved (etc.)'. Hieruit zou men kunnen afleiden dat het aksent in het onderzoek naar leerstoornissen dient te liggen op basale psychologische processen, waarmee de weg wordt vrijgemaakt voor het gebruik van psychologische tests.

Welke konklusies mogen wij nu uit het voorgaande trekken?

In de eerste plaats: onduidelijk blijft wat men nu precies onder de term 'leerstoornissen' of 'leermoeilijkheden' moet verstaan. Zoals Vaughan & Hodges (1973) het uitdrukken en, in andere bewoordingen, velen met hen: "Leading authorities in the field have been unable to agree upon a definition" (p. 68).

Ten tweede: de term 'leerstoornissen' of 'leermoeilijkheden' verkeert

nog vrijwel in het stadium van de operationele definitie; dat wil zeggen de gebruikte meetinstrumenten of identifikatietechnieken bepalen groten-deels welke betekenis men aan de term 'leerstoornis' moet hechten. Dit komt heel duidelijk tot uiting in de definitie van Dumont en de tweede definitie van Haring c.s.; minder duidelijk is dit het geval bij de eerste definitie van Haring c.s. en de laatst gegeven definitie van het 'National Advisory Committee on Handicapped Children'. Ogenschijnlijk biedt een operationele definitie - zeker waar geen eenstemmigheid heerst - grote voordelen voor de individuele onderzoeker: hij kan immers bij de ontwikkeling en keuze van zijn instrumenten of technieken nog alle kanten op. De nadelen doen zich echter snel gelden indien hij zijn onderzoek ook theoretisch wil funderen. De bestaande definities geven in dit opzicht weinig aanknopingspunten.

Ten derde: ook al bieden de bestaande definities theoretisch weinig houvast, toch geven zij wel enkele aanwijzingen over de gebieden waarop men het onderzoek dient te richten. Gemakshalve duiden wij deze gebieden aan met de termen: medisch-biologisch, psychologisch-cognitief en sociaal-educatief. Medisch-biologisch, omdat deze term - in de meest ruime zin van het woord - het terrein bestrijkt waarop jeugdartsen, neurologen, kinderartsen etc. zich bezighouden met onderzoek naar het lichamelijk (dys)funktioneren van kinderen. Psychologisch-cognitief, omdat deze term verwijst naar de onderliggende psychologische processen met betrekking tot de kennisverwerving, welke van belang zijn voor het leren op school; en sociaal-educatief, omdat in deze term het opvoedings- en omgevings-klimaat tot uitdrukking wordt gebracht.

Ten vierde: uit de verschillende definities blijkt duidelijk dat het tijdstip waarop tot opsporing wordt overgegaan - ná de entr  e in het lager onderwijs of daarv  r - in hoge mate bepalend is voor het operationele karakter van de definitie. In het   ne geval - v  r de entr  e op de lagere school - ligt het aksent op 'discrepanties tussen cognitieve niveau's'; in het andere geval - ná de entr  e op de lagere school - ligt het aksent op 'discrepanties tussen schoolprestaties en intellectuele capaciteiten'. De bepaling van het tijdstip waarop tot opsporing wordt overgegaan is een kwestie van keuze, welke in de volgende paragraaf nader gemotiveerd wordt.

Tenslotte: onze poging om aan de hand van enkele definities het begrip leerstoornissen te verhelderen, blijkt bij nader inzien een tamelijk lastige opgave. Geen enkele definitie geeft bovendien een bevredigende oplossing voor ons criteriumprobleem. De verleiding is dan ook groot om

- in navolging van vele anderen - een nieuwe definitie te formuleren. Ook al is hier op zich geen bezwaar tegen, ons gevoel is dat aan een nieuwe definitie - de 39e in successie (cf. Vaughan & Hodges, 1973) - geen enkele behoefte bestaat zolang het theoretisch fundament nog gelegd moet worden. Daar komt bij dat een term als leerstoornissen, in zijn praktische uitwerking bezien, minder gelukkig gekozen lijkt. Zo wijst Kessler (1972) op het verschijnsel dat een term als leerstoornissen tot zwart-wit uitspraken verleidt: er zijn kinderen met en kinderen zonder leerstoornissen. Toch is een dergelijk duidelijk onderscheid in de onderwijspraktijk zelden te maken, evenmin als in de praktijk van het onderzoek zoals wij aan de hand van de diverse definities konden constateren. Daarnaast (ver)leidt het gebruik van de term stoornis tot 'etikettering' met alle gevolgen van dien. Volgens Kessler:

"As a result of the work of Rosenthal (1966), educators have become acutely aware of the power of a label in determining how a child will be treated and the attitude taken towards him. These labels directly affect what a child will learn and how he will feel about himself" (p. 195).

Ofschoon op het werk van Rosenthal, met name vanuit de methodologische hoek, wel enig kritisch commentaar gevolgd is (cf. Davie, 1976), zien wij - gegeven het plausibele karakter van bovenstaande argumenten - af van de keuze voor een van de bestaande definities, ook al zullen wij de termen leerstoornissen of leermoeilijkheden in het vervolg nog wel eens gemakshalve hanteren. Evenmin zullen wij een poging doen tot het formuleren van een nieuwe, meer bevredigende definitie. Hiervoor in de plaats zullen wij naar een andere oplossing van ons criteriumprobleem zoeken: een oplossing die in paragraaf 4.1 nader uiteengezet wordt.

1.3 Uitgangspunten en eerste doel van het onderzoek

Heeft onze speurtocht langs een aantal definities geen bevredigende definitie opgeleverd, in ieder geval verschaft zij enkele uitgangspunten voor de ontwikkeling van instrumenten en technieken voor de vroegtijdige signalering en diagnostiek van kinderen (4-8 jaar) met potentiële of feitelijke 'leermoeilijkheden'. In deze studie zullen wij ons echter alleen met die uitgangspunten bezighouden die van belang zijn voor de ontwikkeling van de LDT en zijn experimentele voorloper de LDT-E; bovendien zullen wij ons hier tot het psychologisch-cognitieve onderzoeksgebied beperken. Voor de overige twee gebieden waarop het onderzoek naar leermoeilijkheden zich dient te richten, namelijk het medisch-biologische en het sociaal-educatieve, verwijzen wij naar Groenendaal et al. (1978).

Het eerste uitgangspunt betreft het tijdstip waarop tot opsporing moet worden overgegaan, vóór of ná de entrée op de lagere school. Om redenen van verschillende aard besloten wij het aksent in ons onderzoek te leggen op de oudste kleuters (5- en 6-jarigen), ook wel schoolverlatende kleuters genaamd. In het kort komen onze argumenten hierop neer²:

- zolang de strikte scheiding tussen kleuter- en lager onderwijs nog niet bij de wet is opgeheven, zullen kinderen niet veel eerder dan in de eerste klas lagere school systematisch met lezen, rekenen, taal en schrijven te maken krijgen. Op dit tijdstip kunnen de eerste van een serie faalervaringen opgedaan worden; uit een oogpunt van preventie een ongewenste zaak (cf. par. 1.1).

- van de kleuters is vooral de oudste leeftijdsgroep (5- en 6-jarigen) van belang om redenen van ontwikkelingspsychologische aard. Uit onderzoek (cf. Stott & Ball, 1965; White, 1965) is gebleken dat de onderliggende psychologische processen en functies met betrekking tot de kennis- en vaardigheidsverwerving op school, aan tamelijk snelle veranderingen onderhevig zijn. Een van de manieren waarop men dit heeft kunnen constateren is via het faktoranalytisch onderzoek van tests, afgenomen op verschillende leeftijdsgroepen. Waar wij juist in de onderliggende psychologische processen en functies enig inzicht proberen te krijgen, ligt het voor de hand het tijdstip van onderzoek zo laat mogelijk te kiezen - i.c. op 5- à 6-jarige leeftijd - vlak vóór de entrée op de lagere school³.

Ons tweede uitgangspunt heeft betrekking op de vraag welke psychologisch-cognitieve kenmerken voor onderzoek in aanmerking komen. Uit de verschillende definities, met name de eerste definitie van Haring c.s. onder punt a), blijkt dat de bepaling van een aantal - voor het leren op school relevante - discrepanties tussen cognitieve functies belangrijk geacht wordt. Daarnaast wordt in iedere definitie gewezen op het onderscheid tussen debiele en imbeciele kinderen (mental retardation) enerzijds en zwak tot normaal of meer dan normaal begaafde kinderen anderzijds. Ofschoon dit onderscheid minder hard is dan het lijkt (cf. par. 1.2), wijst het wel op de unaniem geaccepteerde noodzaak van een intelligentie-onderzoek. Falen in het lager onderwijs kan immers alle mogelijke oorzaken hebben. Een van de meest voor de hand liggende oorzaken betreft een te geringe intelligentie. In een individueel psychodiagnostisch onderzoek zal men daarom deze mogelijkheid nauwkeurig willen nagaan. Voor de LDT-E betekende dit dat deze in ieder geval een bepaling van de intelligentie mogelijk moest maken. In dit opzicht bieden de verschillende

definities een stevig houvast; ons enige, zoals we in het volgende zullen zien.

Aanzienlijk minder houvast biedt namelijk de opmerking dat de bepaling van een aantal - voor het leren op school relevante - discrepanties tussen cognitieve functies belangrijk wordt geacht. Zo rijst de vraag welke functies en welke discrepanties daartussen dan wel relevant zijn voor het leren lezen, rekenen, schrijven etc. Uiterst vage termen als 'basic psychological processes' en 'sensory-motor, perceptual, cognitive, academic or related developmental levels' geven hooguit een idee van wat bedoeld wordt. Waar de verschillende functies al zo onduidelijk omschreven worden, hoeveel te meer onduidelijkheid mag men dan niet verwachten over de relatie van deze functies met de vaardigheden die op school bijgebracht worden? In dit licht moet dan ook onze eerder gemaakte opmerking bezien worden dat de bestaande definities theoretisch weinig houvast bieden (cf. par. 1.2).

Vandaar dat wij op zoek gegaan zijn naar een theorie over de ontwikkeling van leer- en denkprocessen, welke niet alleen de mogelijkheid biedt tot een voor het leren op school relevante beschrijving van het cognitief functioneren bij kinderen, maar ook gereede kansen verschaft op een operationalisatie van dit functioneren door middel van psychologische tests. Een dergelijke theorie - of liever model - meenden wij gevonden te hebben in het werk van de Amerikaanse psycholoog Henry J. Mark (1962; 1969), waarvan in hoofdstuk 2 verslag wordt gedaan.

Het derde uitgangspunt tenslotte is praktisch van aard. Individueel testonderzoek van kinderen kan geruime tijd in beslag nemen, afhankelijk van de problemen die zich voordoen en de snelheid waarmee deze diagnostisch in kaart te brengen zijn. In de praktijk beperkt men meestal de duur van het onderzoek tot één à anderhalf uur. Een langere onderzoeksduur is te vermoeiend voor 5- à 6-jarigen: de aandacht verslapt en de bereidheid tot medewerking neemt zienderogen af. De LDT-E diende dan ook zo samengesteld te worden dat het hele testonderzoek niet veel langer dan één uur zou duren.

Gegeven deze drie uitgangspunten kan nu het eerste doel van ons onderzoek omschreven worden als *de ontwikkeling van een meetinstrument ter bepaling van zowel de intelligentie - het algemeen cognitief ontwikkelingsniveau - als het niveau van een aantal afzonderlijke cognitieve functies, die mogelijk van belang zijn voor het leren op school*. Volledigheidshalve moeten wij hieraan toevoegen dat *dit meetinstrument, de*

LDT-E, gericht is op een individueel psychologisch onderzoek van schoolverlatende kleuters (5- en 6-jarigen) binnen het tijdsbestek van plm. één uur.

In deel A 'Cognitieve ontwikkeling en schoolprestaties' wordt verslag gedaan van dit onderzoek. Hiertoe starten wij met een uiteenzetting van Mark's theorie over de ontwikkeling van leer- en denkprocessen, gevolgd door een beschrijving van het model dat wij aan deze theorie ontleenden (hoofdstuk 2). Dit model ligt ten grondslag aan de konstruktie van de LDT-E, zoals beschreven in hoofdstuk 3. In hetzelfde hoofdstuk wordt tevens een paragraaf gewijd aan de plaats van de LDT-E binnen een onderzoek dat gericht is op de vroegtijdige opsporing van leermoeilijkheden. In hoofdstuk 4 wordt de opzet en uitvoering besproken van een deel van het empirisch onderzoek dat met de LDT-E verricht is. In essentie komt dit onderzoek neer op een betekenisanalyse, inclusief prediktieve validering van de LDT-E. In de hoofdstukken 5, 6 en 7 - achtereenvolgens aangeduid met Betrouwbaarheid, Structuur en Prediktieve validiteit - worden vervolgens de resultaten van het empirisch onderzoek weergegeven en bediscussieerd. Hoofdstuk 8 vormt de afsluiting van deel A 'Cognitieve ontwikkeling en schoolprestaties'. Een samenvatting van de belangrijkste onderzoeksresultaten, evenals een algemene discussie en konklusie, bepaalt de inhoud van dit slothoofdstuk.

In deel B 'Leervermogen en schoolprestaties' wordt het vervolg van ons onderzoek beschreven. Dit vervolg vormt - althans in de tijd gezien - een onlosmakelijke eenheid met het onderzoek zoals beschreven in deel A; inhoudelijk komen echter in deel B enigszins anders getinte onderwerpen aan de orde. Dit laatste is dan ook de reden waarom wij de beschrijving van ons onderzoek in twee delen gesplitst hebben en met de inleiding tot het vervolg wachten tot hoofdstuk 9.

A. COGNITIEVE ONTWIKKELING EN SCHOOLPRESTATIES

2. MARK-MODEL

2.1 Inleiding tot de theorie van Mark

Alvorens Mark's theorie over leer- en denkprocessen en het daaruit voortvloeiende model te bespreken, lijkt het nuttig eerst onze opvatting over wat leren op school nu globaal inhoudt, vanuit een psychologisch-cognitieve visie weer te geven. De relatie tussen een meetinstrument als de psychologische test en dit 'schoolse' leren zal daardoor ook duidelijker worden. In deel I schreven wij hierover het volgende:

"Lezen, rekenen, taal en schrijven zijn ingewikkelde vaardigheden. Algemeen wordt aangenomen dat een verscheidenheid aan psychologische functies - perceptuele, cognitieve en motorische - een rol speelt bij het leren van deze vaardigheden. De opvatting dat een stoornis in een of meer functies tot moeilijkheden bij het leren kan leiden, ligt in het verlengde hiervan. Met behulp van psychologische tests die deze functies meten, is het mogelijk stoornissen of dysfuncties op te sporen. De resulterende diagnose dient vervolgens als uitgangspunt voor een remediaal advies" (p. 12).

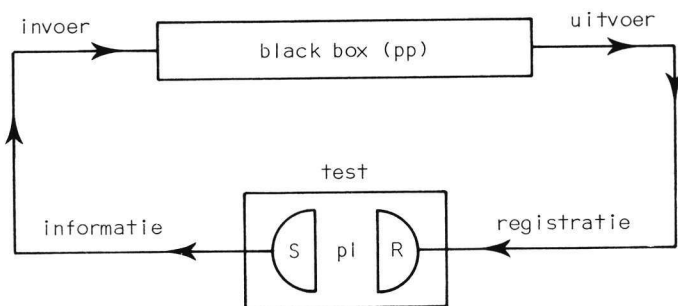
Deze visie stoelt op een benadering in de psychologie waarin alle vormen van leren - of dit nu lezen, rekenen en schrijven dan wel het aanleren van sociale gedragspatronen betreft - in laatste instantie herleid worden tot een informatieverwerkende activiteit van het zenuwstelsel. Lezen bijvoorbeeld, zou men kunnen beschouwen als een uitermate complexe, elektrofysiologische en biochemische activiteit van miljoenen zenuwcellen.

Het zal duidelijk zijn dat een directe meting van deze activiteit met behulp van een of ander meetinstrument bij voorbaat een hopeloze zaak is, zowel door het invasieve karakter van de meetprocedure als door het aantal zenuwcellen dat in het leesproces betrokken is. Toch is het wenselijk iets van de aard van deze activiteit te achterhalen, met name als het leesproces niet ongestoord verloopt. Aangezien directe meting niet mogelijk is, zal men dit op indirecte wijze moeten trachten te doen. Hiervoor is het echter niet strikt noodzakelijk dat men van tevoren reeds een nauwkeurig idee heeft hoe het zenuwstelsel in elkaar zit en funktioneert. Men kan zich voorlopig ook tevreden stellen met een model van dit zenuwstelsel waarin de activiteit van miljoenen zenuwcellen niet zozeer op biochemisch of elektrofysiologisch niveau als wel op psychologisch niveau beschreven wordt.

Dit model is het bekende 'input - black box - output' model, waarbij de input betrekking heeft op de invoer van informatie in het zenuwstelsel via de zintuigen (receptoren), de black box op het zenuwstelsel zelf en

de output op de uitvoer van informatie via de spieren (effectoren). Door via een proefopzet de invoer van informatie systematisch te variëren en de uitvoer van informatie nauwkeurig te registreren, is het mogelijk een aantal kenmerken van de informatieverwerking in de black box te achterhalen. Deze kenmerken betitelen wij - in navolging van Carroll (1976) en De Coster et al. (1977) - met de term 'cognitieve functies' als verzamelnaam voor alle processen waardoor binnenkomende informatie getransformeerd, opgeslagen of op andere wijze verwerkt wordt. Hiertoe rekenen wij ook zulk soort processen, c.q. functies als perceptie, onmiddellijk geheugen, denken, probleem-oplossen, etc.

Een psychologische test nu is in feite niets anders dan een instrument waarmee de proefleider (pl) in een gestandaardiseerde proefopzet en informatie ter verwerking aanbiedt aan het zenuwstelsel (black box) van de proefpersoon (pp) en verwerkte informatie registreert om zodoende via de samenhang tussen testinvoer (stimuli / items) en testuitslag (responses / scores) een indruk te krijgen van het cognitief functioneren van de proefpersoon. In schema (figuur 1):



Figuur 1. Schema van de psychologische test als meetinstrument in een 'input-black box-output' model
S = stimulus (item); R = geregistreeerde response (score)

In de loop der tijd is het simpele 'input-black box-output' model door functiepsychologen aanzienlijk verfijnd. Via experimenten in laboratorium-situaties slaagde men erin steeds nauwkeuriger afzonderlijke cognitieve functies (in relatie tot hun omgeving) te beschrijven, en wel zódanig dat sommigen al van een 'white box' spreken (cf. Apostel, 1977). Van ontwikkelingspsychologische zijde werd daarnaast getracht om het verloop en de opbouw van cognitieve functies in de tijd in kaart te brengen.

In schrille tegenstelling tot de vorderingen op functiepsychologisch en ontwikkelingspsychologisch terrein staan de resultaten van het testpsychologisch onderzoek. Als men al in termen van het simpele 'input-black box-output' model denkt, dan nog wordt daar bij de konstruktie, betekenisanalyse en praktische toepassing van tests geen rekening mee gehouden. Zo volstaat men nog steeds met tamelijk slordige testbenamingen als 'perceptueel', 'performaal' of 'verbaal', daarmee in het midden latend of met 'perceptueel' nu de auditieve invoer aangeduid wordt dan wel de visuele invoer. Hieruit blijkt dat men de complexiteit van informatiekkanalen (in- en uitvoerkombinaties) en cognitieve functies (binnen deze kanalen), welke te zamen het informatieverwerkend systeem vormen, sterk onderschat. Accepteert men eenmaal dat het (centrale) zenuwstelsel het meest complexe orgaan vormt en dientengevolge een informatieverwerkend systeem als model hiervan eveneens zeer complex van aard moet zijn, dan zal men op zoek gaan naar een theorie of, zo men wil, model van dit systeem, welke niet alleen recht doet aan deze complexiteit maar waarvan de theoretische componenten ook een redelijk houvast bieden voor de konstruktie, analyse en uitbreiding van een testbatterij.

In de nu volgende paragraaf wordt de aanzet tot een dergelijke theorie - waarin zowel het neuro- of functiepsychologische gezichtspunt als het ontwikkelingspsychologische gezichtspunt met elkaar verenigd zijn - beknoot weergegeven.

2.2 Mark's theorie over de ontwikkeling van leer- en denkprocessen

Voor een verklaring van het gedrag van organismen, in het bijzonder leer- gedrag, postuleert Mark (1962) een aantal logisch noodzakelijke 'operaties', welke het organisme:

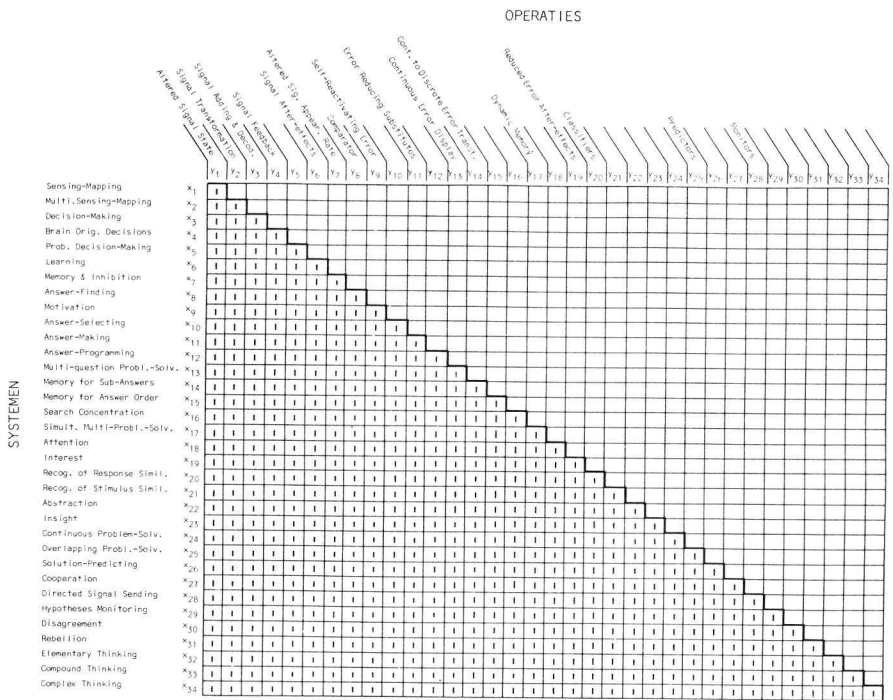
- motiveren of doen bewegen;
- in staat stellen om nieuwe informatie te verzamelen, zowel over de structuur van zijn omgeving als over de eigen structuur en operaties;
- in staat stellen dergelijke informatie te gebruiken bij het zoeken naar antwoorden op stimuli, vragen of problemen welke door de omgeving gesteld worden of binnen het eigen organisme rijzen;
- in staat stellen om informatie te verzamelen over 'within-brain answer finding operations' (dat wil zeggen in staat stellen zich te herinneren hoe het bepaalde antwoorden vond);
- in staat stellen om alle beschikbare informatie te gebruiken zowel

over zijn omgeving als over de eigen operaties gedurende een proces waarbij nieuwe kombinaties beschikbare informatie gegenereerd worden; deze nieuwe kombinaties vormen niet alleen een antwoord op oude, maar ook op nieuwe vragen, waarmee het organisme zich dan zou kunnen bezighouden.

Het begrip *operatie* verdient een nadere toelichting. Bedoeld wordt een informatieverwerkende aktiviteit van de hersenen, c.q. zenuwcellen of zenuwcelverzamelingen, die in haar simpelste vorm bestaat uit het met elkaar in verband brengen van afzonderlijke informatie-eenheden. Het leervermogen is in een dergelijk geval uitsluitend een functie van de mate waarin deze relaties tot stand gebracht kunnen worden. Afhankelijk van de plaats van het organisme op de phylogenetische ladder, zijn meer complexe operaties, bijvoorbeeld probleem-oplossen, mogelijk. Een centrale rol speelt hierbij de capaciteit van de hersenen tot het vormen van relaties en de capaciteit van het geheugen voor deze relaties. [Onder geheugen verstaat Mark (op.cit.) "the altered probability of within-brain pattern appearance" (p. 75).] Zo gauw een organisme meer complexe operaties dan de allereenvoudigste kan uitvoeren, wordt het leervermogen gedefinieerd als de inverse functie van de kans dat het organisme gokte bij het vinden van een antwoord op de vraag dan wel volgens een bepaald programma de vraag beantwoordde⁴.

In totaal worden 34 verschillende operaties onderscheiden, uitgaande van de neurofysiologische en neuropsychologische bevindingen van Amerikaanse en Russische onderzoekers (cf. Mark, op.cit.; p. 87). Kan Mark voor het eerste tweetal operaties - 'altered signal state' en 'signal transformation' (cf. fig. 2) - voldoende neurofysiologische evidentie aandragen, voor de daaropvolgende operaties is dit minder het geval; de gehanteerde begrippen en verklaringen worden steeds hypothetischer van aard. Gezien de huidige kennis van zaken hoeft dit echter geen verwondering te wekken.

Kenmerk van een goede theorie is dat zij onder andere de mogelijkheid schept tot toetsing van de gepostuleerde hypothetische begrippen. Voor Mark's theorie betekent dit dat de verschillende operaties, c.q. neurofysiologische processen en aktiviteiten, vertaalbaar moeten zijn in gedragstermen welke observatie en toetsing mogelijk maken. Deze gedragstermen worden door Mark aangeduid met de term *systeem*. In totaal worden 34 systemen onderscheiden, korresponderend met evenzovele operaties. Operaties en systemen vormen een matrix, aangeduid met *kontrolesysteem* (figuur 2).



Figuur 2. Kontrolesysteem
Uit Mark (1962; p. 78)

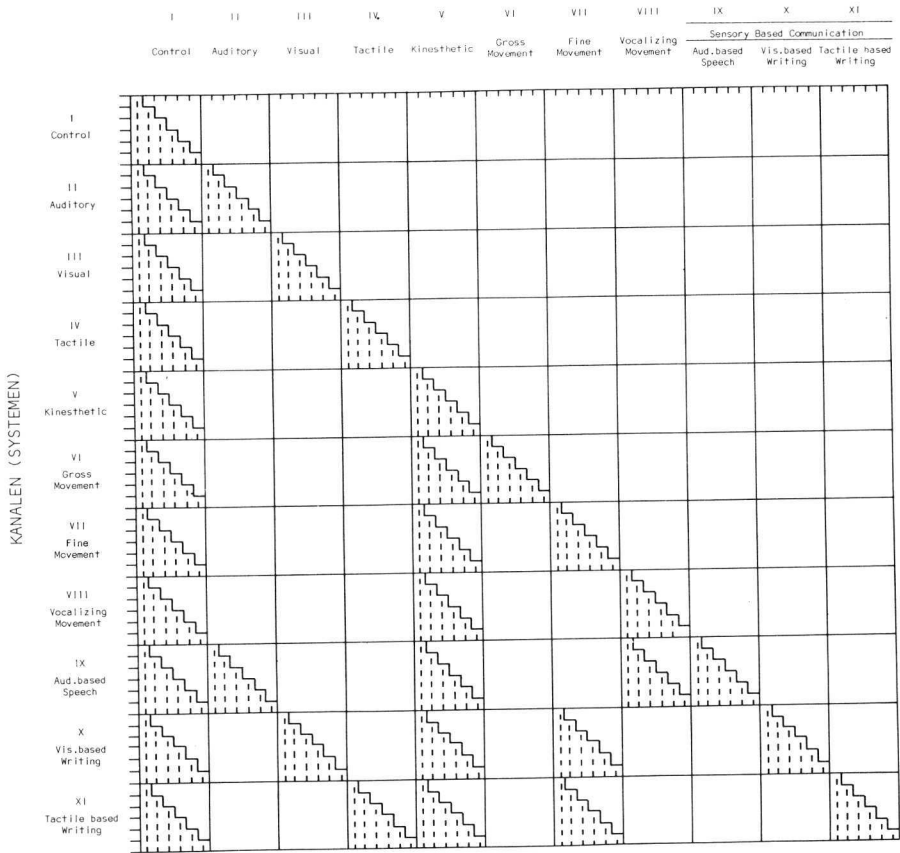
Met de benaming kontrolesysteem wordt, overeenkomstig de terminologie van Newell et al. (1958), aangegeven dat geen specifieke input- of outputmodaliteit vereist is. In de woorden van Mark (op.cit.):

"The control system, therefore, consists of those within-brain operations which act on transformed inputs into the brain and serve the receptor-effector systems by carrying on all the within-brain operations described so far, including the activation of outputs or effectors" (p. 79).

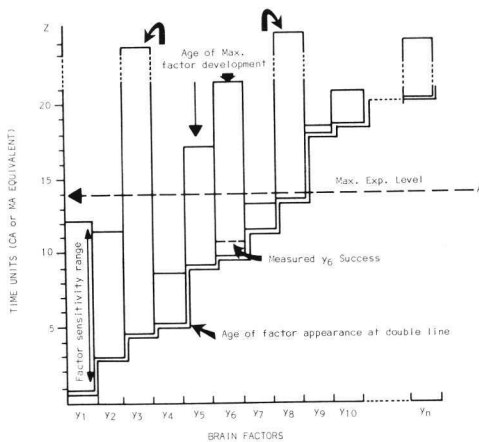
Binnen ieder receptor-effector systeem treft men dus dit kontrolesysteem aan. In zijn 1962-artikel geeft Mark een gedetailleerde beschrijving van de verschillende informatieverwerkende systemen - te samen het kontrolesysteem vormend - die ieder afzonderlijk, door middel van gedrags-tests, geïdentificeerd kunnen worden.

Een vooral uit psychodiagnostisch oogpunt belangrijke eigenschap van het kontrolesysteem is de hiërarchische opbouw. Korrespondeert systeem x_1 alleen met operatie y_1 , systeem x_2 korrespondeert zowel met operatie y_1 als operatie y_2 . In feite vormt ieder niveau (rij) van de hiërarchie een logische vergelijking, waarvan de algemene formule luidt:

KANALEN (OPERATIES)



Figuur 3. Gedragsklassifikatieschema. Uit Mark (1962; p. 79)



Figuur 4. Dwarsdoorsnee van een kanaal. Uit Mark (1962; p. 85)

$$x_n \rightarrow y_n \wedge y_{n-1} \cdots y_1$$

waarbij operationeel sukses op niveau x_{n+1} altijd afhankelijk is van de output gegenereerd door het laatste y -as mechanisme - ook wel 'brain-factor' genoemd - van niveau x_n . [Onder een mechanisme verstaat Mark een verzameling operaties, waarbij de output van de eerste operatie dient als input voor de tweede operatie, de output van de tweede als input voor de derde, etc.]

Het voordeel van een dergelijke hiërarchie is dat zeer snel gecontroleerd kan worden of een bepaald systeem naar behoren funktioneert. Men toetst daartoe, door middel van een (gedrags)test, het systeem op dat niveau, waarvan men op grond van externe criteria - leeftijd bijvoorbeeld - mag aannemen dat het naar behoren funktioneert. Indien dysfunktionen gekonstateerd wordt kiest men een lager niveau. Als het gekozen systeem goed funktioneert, moeten logischerwijze alle hiërarchisch voorafgaande systemen eveneens goed funktioniëren. Controle op de validiteit van de hiërarchie gebeurt automatisch; indien met de theorie strijdige testresultaten gevonden worden, is herziening van de hiërarchie noodzakelijk⁵.

Voor de toetsing van een systeem is in ieder geval één input (receptor) en één output (effector) noodzakelijk. In het menselijk organisme kan men een groot aantal receptor-effector systemen, of combinaties daarvan, onderscheiden. Binnen ieder receptor-effector systeem, door Mark aangeduid met de term *kanaal*, funktioneert het controlesysteem.

In figuur 3 wordt de relatie tussen dit controlesysteem (rij I, kolom I) en enkele belangrijke kanalen (II t/m VIII) en hun combinaties (IX t/m XI)anschouwelijk gemaakt. Voor de overzichtelijkheid is niet in iedere daarvoor in aanmerking komende cel van figuur 3 het controlesysteem volledig weergegeven. Minimaal worden echter 34 (= het aantal systemen in figuur 2) x 11 (= het aantal kanalen en hun combinaties) = 374 verschillende te toetsen elementen onderscheiden.

De verzameling kanalen en systemen binnen deze kanalen wordt door Mark aangeduid met de term *gedragsklassifikatieschema*. Als belangrijkste eigenschap van dit schema kan worden aangemerkt dat het in principe een zeer gedetailleerde operationalisatie mogelijk maakt van een theorie over leergedrag, welke

- een klassifikatie verschaft van een eindige verzameling dysfuncties, en
- zodanig is opgesteld dat iedere dysfunctie uit deze verzameling, inclusief het daaraan ten grondslag liggende mechanisme, geïdentificeerd

en beschreven kan worden in termen van de theorie zelf.

Om het gedragsklassifikatieschema bruikbaar te maken voor de identificatie van dysfuncties, voegt Mark een derde dimensie aan het tweedimensionale schema toe, namelijk een tijdschaal (figuur 4).

Met deze schaal wordt het *ontogenetische* of ontwikkelingstijdverschil aangeduid tussen de verschijning van een 'brainfactor' (mechanisme) en de verschijning van de hiërarchisch daaropvolgende brainfactor (y-as). De trapvormige, dubbele lijn in de figuur geeft aan wanneer zo'n brainfactor voor het eerst verschijnt. De bovenste horizontale lijn van iedere faktor laat zien op welk tijdstip een faktor - naar verwachting - tot volledige ontwikkeling is gekomen. Door als tijdseenheid het chronologische leeftijdequivalent (CAE) of, in het geval van psychologische tests, het mentale leeftijdequivalent (MAE) te kiezen, is het schema klaar voor klinisch gebruik.

Bijvoorbeeld: stel dat iemand een MAE-score van 14 punten haalt op een test die faktor y_3 meet en dat dezelfde persoon op geen enkele andere test een duidelijk hogere score haalt. In dat geval is de MAE-score van 14 punten het *maximale verwachtingsniveau* van de onderzochte (aangegeven door de gestippelde lijn A in figuur 4); dat wil zeggen, dat tests voor de overige factoren eveneens een MAE-score van 14 punten zouden moeten laten zien bij een normale ontwikkeling. In dit voorbeeld haalt de onderzochte echter een MAE-score van 11 punten op faktor y_6 (aangegeven met het kleine stippellijntje). In zo'n geval spreekt Mark van een dysfunctie.

2.3 Model

De sprong van theorie naar praktijk is meestal groot; vaak ook wordt deze in het geheel niet gemaakt. Het is de verdienste van Mark dat hij in een later artikel (1969), en nog later (1972) via een seminar voor het NIPG-TNO, onze aandacht gericht heeft op de praktische mogelijkheden en beperkingen van zijn theorie aan de hand van een aantal voorbeelden. Met behulp van deze voorbeelden konden wij het model konstrueren dat aan de LDT-E, en dus ook aan de LDT, ten grondslag ligt. Alvorens dit model wat uitgebreider te bespreken dan in de handleiding voor de LDT (deel I) mogelijk was, zullen we eerst de belangrijkste elementen daarvan - zoals beschreven in deze handleiding - de revue laten passeren.

"In de eerste plaats wordt het begrip *kanaal* onderscheiden. Andere termen hiervoor zijn 'receptor-effector systeem' of 'modaliteit'. Via een kanaal wordt informatie verwerkt, waardoor een leerproces kan plaats-

vinden. Een groot aantal kanalen staat voor de informatieverwerking ter beschikking. Deze kanalen worden gekarakteriseerd aan de hand van input-output kenmerken. De input kan van verschillende aard zijn: visueel, auditief, taktiel of kinesthetisch. Ook de output varieert: fijnmotorisch, grofmotorisch of spraakmotorisch. Enkele bekende leerkanalen zijn: auditief-spraakmotorisch (luisteren-praten) en visueel-fijnmotorisch (kijken-schrijven). Uiteindelijk kunnen evenveel kanalen onderscheiden worden als er input-output combinaties zijn.

Binnen ieder kanaal verricht het centrale zenuwstelsel (CZS) een aantal *funkties*, het tweede element in dit model. Onder functie verstaat Mark een informatieverwerkende activiteit van zenuwcellen. In totaal worden 34 functies onderscheiden. De allereenvoudigste is de oriëntatiereactie, welke men reeds bij een baby kan constateren: op een plotselinge geluidsprikkel bijvoorbeeld zal de baby met een lichte schrikbeweging reageren. De meest complexe functies van het CZS zijn die waar ingewikkelde probleemoplossingsstrategieën aan te pas komen.

Het derde element betreft de *hiërarchische organisatie* van de verzameling functies. Een bepaalde functie kan pas dan in werking treden indien alle andere, minder complexe functies successievelijk geactiveerd zijn.

Het vierde en laatste element heeft betrekking op de *ontogenie* of ontwikkeling van de verzameling functies binnen het CZS. Aan deze ontwikkeling liggen bepaalde wetmatigheden ten grondslag. De allereenvoudigste functies, zoals de oriëntatiereactie en de gewennings- en waakzaamheidsfuncties, kunnen vrijwel direct bij de geboorte gekonstateerd worden. Voor andere, meer complexe functies is een bepaalde ontwikkeling van het CZS noodzakelijk. Tot op zekere hoogte is deze ontwikkeling en daarmee het voorkomen van functie x op tijdstip y voorspelbaar. De vorming van een twee-woordszin bijvoorbeeld geschiedt meestal niet veel eerder dan op de leeftijd van anderhalf jaar. Met behulp van leeftijdsnormen voor psychologische tests is het verder mogelijk om redelijk nauwkeurig het tijdstip te bepalen waarop bepaalde functies, zoals gemeten door deze tests, tot ontwikkeling komen." (p. 13 en 14).

Bovenstaande vier elementen van wat wij in het vervolg het Mark-model zullen noemen, vormen met elkaar een wel zeer vereenvoudigde weergave van zijn theorie. Zo maakt Mark (1969) aan de invoerzijde van een leerkanaal niet alleen onderscheid tussen de vier bekende sensorische modaliteiten 'visueel', 'auditief', 'taktiel' en 'kinesthetisch', maar differentieert hij daarnaast ook nog binnen de auditieve en visuele invoeren aanzien van het *soort* informatie dat ingevoerd wordt. Bijvoorbeeld: naast een 'auditive input' als zodanig wordt ook een 'auditive-spoken language input' onderscheiden of een 'auditive-spoken mathematical input', afgekort tot Aud-SPL Math. In combinatie met de drie verschillende soorten output 'fijnmotorisch', 'grofmotorisch' en 'spraakmotorisch' komt Mark dan tot niet minder dan 24 verschillende leerkanalen, waarvan alleen de laatste drie (22, 23 en 24) gewoonlijk in testbatterijen aange troffen worden (tabel 1). Lang niet alle mogelijke leerkanalen zijn echter in deze tabel vertegenwoordigd, aangezien bijvoorbeeld de taktiele en kinesthetische modaliteiten met hun varianten duidelijk ontbreken.

Tabel 1. Verkorte lijst van 24 kanalen^a

input		output
1	audio (aud)	gross motor (G)
2	aud	fine motor (F)
3	aud	speech motor (S)
4	aud-spoken language (SPL)	G
5	aud-SPL	F
6	aud-SPL	S
7	aud-SPL Math	G
8	aud-SPL Math	F
9	aud-SPL Math	S
10	visual (vis)	G
11	vis	F
12	vis	S
13	vis pantomime language (PL)	G
14	vis PL	F
15	vis PL	S
16	vis recorded language (reading & writing, R & W)	G
17	vis R & W	F
18	vis R & W	S
19	vis R & W and Math	G
20	vis R & W and Math	F
21	vis R & W and Math	S
22	aud-SPL + vis PL	G
23	aud-SPL + vis PL	F
24	aud-SPL + vis PL	S

^a uit Mark (1969; p. 79)

Tabel 2. Verkorte lijst van 14 systemen^a

(1)	(2)	(3)	(4)	(5)
1	arousal threshold changes	birth	audio	
2	habituation-vigilance	birth	A	
3	learned (L) localization	2 mos	visual	
4	L between contour discrimination	2½ "	A	
5	L within contour resolution	3 "	V	
6	L input-output feedback monitoring	3½ "	V	
7	between input channel conditioning	4½ "	AV	
8	between and within channel conditioning	7 "	A	
9	directed secondary signal sending	8 "	AV	
10	concept formation	14 "	AV	x ₂₆
11	3-unit concept integration or sentence creation	36 "	AV	
12	2-sentence message creation	54 "	AV	
13	mathematical word problem solving	72 "	AV	x ₂₉
14	use of recorded language(s)	72 "	AV	

^a uit Mark (1969; p. 79)

Bij een systematisch diagnostisch onderzoek naar het cognitief functioneren dient men volgens Mark in ieder geval met het bestaan van deze 24 leerkanalen rekening te houden. Uit een oogpunt van eenvoud en overzichtelijkheid hebben wij ons echter bij de samenstelling van het model tot genoemde vier input- en drie outputcategorieën beperkt.

Het tweede element van het Mark-model wordt gevormd door de verzameling van 34 functies. Het moge duidelijk zijn dat wij met de term functie de door Mark gehanteerde begrippen 'systeem', 'brainfactor' of 'mechanisme' aanduiden, in navolging van het moderne spraakgebruik op dit terrein (cf. Michon et al., 1976). Over dit tweede element het volgende. In 1969 geeft Mark een nieuwe (verkorte) lijst van 14 systemen (functies) (tabel 2). Vergelijking van het volledige controlesysteem in het 1962-artikel met dat van de verkorte lijst in tabel 2 leert dat Mark in de loop van zeven jaren tussen de publikatie van zijn 1962- en 1969-artikel, de eerste 25 systemen gereduceerd heeft tot een negental 'key'-systemen welke klinisch relevant zijn en bovendien dit negental systemen geëtiketteerd heeft met namen welke aan de psychologie ontleend zijn in plaats van aan een systeemtheorie van informatieverwerking⁶. Aangezien Mark de relatie tussen zijn controlesysteem en de verkorte lijst niet duidelijk heeft uitgewerkt, kan niet achterhaald worden welke 'key'-systemen corresponderen met de systemen uit het (uitgebreide) controlesysteem. Een uitzondering moet echter gemaakt worden voor de 'key'-systemen 10 en 13, aangezien door voorbeelden in de tekst van het 1962-artikel duidelijk blijkt dat deze corresponderen met de oorspronkelijke systemen x_{26} en x_{29} (tabel 2, kolom 5).

In kolom 3 van tabel 2 geeft Mark schattingen van het tijdstip waarop een systeem behoort te functioneren. Deze informatie, gekombineerd met het feit dat systeem 10 uit de verkorte lijst correspondeert met systeem x_{26} uit het (uitgebreide) controlesysteem, laat zien dat de eerste 26 systemen (x_1 t/m x_{26}) reeds op een geschatte leeftijd van 14 maanden aanwezig zijn. Combineren wij op dezelfde manier systeem 13 uit de verkorte lijst met systeem x_{29} uit het controlesysteem, dan blijkt dat slechts drie nieuwe systemen (11, 12 en 13) in de periode van ± 14 maanden tot 72 maanden gepostuleerd worden. Trekken wij deze lijn nog verder door dan blijken er vanaf de leeftijd van ± 72 maanden nog slechts vijf systemen volgens het controlesysteem (x_{30} t/m x_{34}) over te schieten. De konklusie die wij hieruit mogen trekken is dat Mark een zeer gedetailleerde indeling geeft in systemen voor de laagste cognitieve ontwikke-

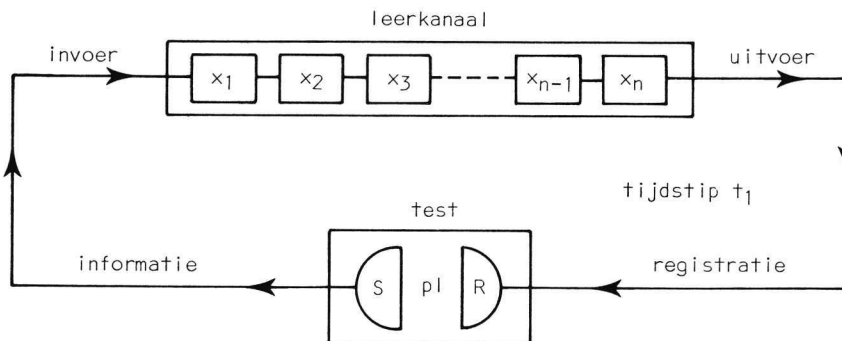
lingsniveau's - hiertoe in staat gesteld door uitgebreid neurofysiologisch en -psychologisch onderzoek op dit terrein - echter een zeer grove systeemindeling geeft voor de meer complexe ontwikkelingsniveau's vanaf de leeftijd van ± 14 maanden. Zowel de verkorte als de uitgebreide lijst van systemen heeft dus een beperkte gevoeligheid met betrekking tot het cognitief functioneren van kinderen ouder dan 14 maanden.

In de handleiding voor de LDT (deel I) merkten wij hierover het volgende op:

"De zwakke plek in dit model wordt gevormd door de verzameling functies welke nauwelijks een empirische basis heeft (Mark, 1972). Dit betekent in feite dat de gemaakte functie-indeling nog slechts heuristische waarde heeft en men vrij is om ook andere functies op hun waarde voor het model te toetsen; dit overeenkomstig het theoretisch inzicht van de onderzoeker en elders verkregen empirische onderzoeksresultaten.

Literatuuronderzoek brengt aan het licht dat een grote verscheidenheid aan functies belangrijk geacht wordt voor de leeftijdsgroep van 4 tot 8 jaar. Globaal zou men deze functies als volgt kunnen rubriceren: aandacht en concentratie (Dykman et al., 1971), sensoriek, perceptie en motoriek (Bannatyne, 1973; Delacato, 1963; Frostig, 1964; Keogh, 1971; Kephart, 1960), geheugen (Kluever, 1971), taal en cognitie (Bereiter & Engelmann, 1966; Blank & Frank, 1971; Kirk, 1966). De door Mark onderscheiden functies kunnen als een verbijzondering beschouwd worden van de hierboven genoemde. Bij elkaar bieden zij een overvloed aan keuzemogelijkheden voor de samenstelling van een testbatterij. Deze keuze wordt echter beperkt - zoals het volgende hoofdstuk laat zien - indien ook enkele praktische overwegingen in aanmerking worden genomen." (p. 14 en 15).

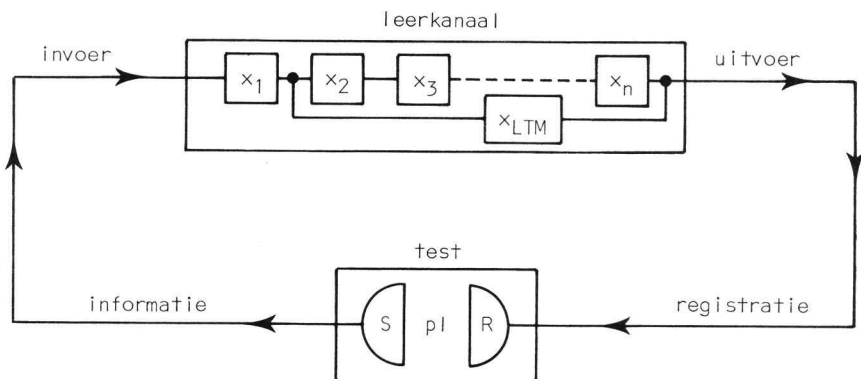
In schema zou men een vereenvoudigd Mark-model, waarin alleen de eerste twee elementen van dit model zijn opgenomen, nu als volgt kunnen weer-geven (figuur 5):



Figuur 5. Schema van één enkel leerkanaal waarin n functies onderscheiden worden op ontwikkelingstijdstip t_1
S = stimulus (item); R = geregistreeerde response (score)

Twee dingen vallen in dit schema op. In de eerste plaats wordt een seriële of sekventiële schakeling van functies gesuggereerd; in de tweede plaats wordt de indruk gewekt dat de verzameling van n functies kanaalgebonden is, dat wil zeggen per leerkanaal verschillend. Het laatste is echter in tegenspraak met Mark's theorie, waarin een modaliteits- of kanaalonafhankelijk geheel van functies gepostuleerd wordt, het zgn. controlesysteem. Waarschijnlijk ligt de waarheid in het midden: sommige functies, met name de meer perifere, zintuigelijke en onmiddellijk geheugenfuncties zullen kanaalgebonden zijn, terwijl meer centrale functies als 'coding', 'retrieval', etc. waarschijnlijk kanaalonafhankelijk zijn (cf. Boekaerts, 1977).

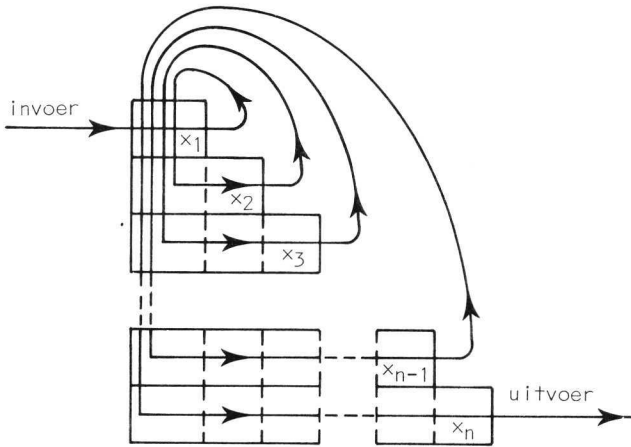
Wat betreft de suggestie dat de verzameling functies serieel geschakeld zou zijn, deze is op zijn minst misleidend. In geen enkel functiepyschologisch model van het menselijk informatieverwerkingssysteem komt men een dergelijke voorstelling van zaken tegen: zo ziet men altijd naast serieel geschakelde ook parallel geschakelde functies, waarvan het meest sprekende voorbeeld wel het LTM of 'Long Term Memory' is. In schema (figuur 6):



Figuur 6. Schema van één enkel leerkanaal waarin de functies $x_1, x_2, x_3 \dots x_n$ serieel en de functie x_{LTM} parallel geschakeld zijn
S = stimulus (item); R = geregistreerde response (score)

Gegeven de laatste twee opmerkingen over respectievelijk het kanaal(on)-afhankelijke en seriële, c.q. parallelle karakter van de verzameling functies, zal het duidelijk zijn dat het tweede element van het Mark-model door ons uiterst voorzichtig geformuleerd is. De redenen hiervoor zullen we aan het slot van deze paragraaf nader uiteenzetten.

In figuur 5 wordt, evenmin als in figuur 6, recht gedaan aan het derde element van het Mark-model: de hiërarchische organisatie van functies; dat wil zeggen de werking van de meest complexe functie, x_n , is afhankelijk van de werking van minder complexe functies: x_{n-1} tot en met x_1 . In figuur 7 wordt schematisch weergegeven hoe men zich de informatieverwerking volgens het hiërarchische principe moet indenken:



Figuur 7. Schema van informatieverwerking via n hiërarchisch gerangschikte functies

Uit deze figuur kan men aflezen dat de meest complexe functie x_n , opgebouwd is uit de hiërarchisch lager geordende (minder complexe) functies $x_{n-1} \dots x_1$. Het resultaat van een dergelijke opbouw is dat als één willekeurige functie niet naar behoren werkt alle hiërarchisch hoger geordende (meer complexe) functies evenmin naar behoren werken.

Ook bij het derde element van het Mark-model past weer enig commentaar. Wij beginnen daartoe met een citaat uit de handleiding voor de LDT (deel I), waarin tevens het vierde element - de ontogenie van functies - besproken wordt.

"Een tweetal belangrijke eigenschappen van de verzameling functies, namelijk de hiërarchische organisatie en de ontogenie, zou met het wegvallen van het empirisch fundament eveneens slechts heuristische waarde behouden⁷. Guilford (1967) trekt zelfs de heuristische waarde van deze elementen uit het Mark-model in twijfel. Uitgangspunt voor zijn stellingname is de Garrett-hypothese (Garrett, 1946). Deze houdt in dat er een hiërarchische rangschikking zou bestaan van intellectuele factoren welke in de loop van de ontwikkeling van het kind tot stand komt. Toetsing van deze hypothese aan de hand van 27 faktoranalytische studies

leidt Guilford tot de bevinding dat "there is ... no need for assuming a general ability from which differentiation can proceed in development" (p. 416). Deze konklusie wordt echter niet door iedereen gedeeld. Cronbach (1970; p. 309 e.v.) bijvoorbeeld komt tot de slotsom dat het hiërarchisch model de beste perspectieven biedt⁸. Daarnaast vraagt hij zich af of faktoranalyse wel de meest geschikte methode is om een bepaalde theorie te onderzoeken: "Useful as factor analysis is in reducing sets of correlations so that their information can be digested, it is not by itself an adequate basis for theory because it neglects non-correlational findings" (p. 345). Met andere woorden, informatie die niet opgesloten zit in een korrelatiematrix zal ook na faktoranalyse van de matrix niet tevoorschijn komen.

Voor zover er enige aarzeling bestaat omtrent de heuristische waarde van de hiërarchische organisatie en ontwikkeling van functies, wordt deze weggelaten door de resultaten van experimenteel ontwikkelingspsychologisch onderzoek. In een uitgebreid literatuuroverzicht, waarin onder andere het werk van Kendler & Kendler (1962), Luria (1961) en Piaget (1960) besproken wordt, toont White (1965) aan dat bij 5- tot 7-jarigen zich fundamentele ontwikkelingen voordoen in het cognitief functioneren. Deze ontwikkelingen dragen een duidelijk hiërarchisch karakter, in die zin dat het cognitief functioneren op een bepaald niveau voorwaarde is voor het functioneren op een hoger niveau. De typering van de niveauverschillen door White in termen van concreet - abstrakt, associatief - cognitief, intuïtief - concreet, minder en meer complex, is daarbij van ondergeschikt belang. Hoofdzaak is dat een hiërarchische ordening in de tijd duidelijk onderkend wordt." (p. 15 en 16).

Onze konstatering dat de hiërarchische organisatie en ontwikkeling van functies alleen een heuristische waarde heeft, is in de handleiding voor het Stroomdiagram van de LDT (deel II; Schroot's et al., 1978a; p. 84 e.v.) duidelijk gedemonstreerd; althans indien men uitgaat van psychologische tests, haasten wij ons hieraan toe te voegen. In de testpraktijk blijkt namelijk een valide hiërarchische organisatie van functies moeilijk terug te vinden. Als één van de redenen gaven wij hiervoor op dat tests vrijwel altijd meer dan één functie tegelijk meten.

Er is echter nog een, misschien wel meer fundamentele reden, namelijk dat functies niet alleen hiërarchisch maar ook heterarchisch georganiseerd zijn; dat wil zeggen er is niet zoiets als één vaste hiërarchie van functies maar een verschuivende hiërarchie, waarbij eventuele verschuivingen afhankelijk zijn van het type informatie dat verwerkt wordt (De Mey, 1977). Uiteraard rijst dan de vraag welke functie(s) deze verschuivingen weer controleren, etc.⁹. In de cognitieve psychologie is men over dit punt - evenals over het kanaal(on)afhankelijke karakter van de gepostuleerde functies - nog druk in discussie.

Het element 'ontogenie' behoeft na hetgeen hiervoor reeds gezegd is, eigenlijk geen commentaar. In navolging van Mark (cf. p. 22) hebben ook wij het begrip maximaal verwachtingsniveau (MVN) hierop gebaseerd. Voor een nadere toelichting raadplege men paragraaf 3.4

Zoals wij reeds op verschillende plaatsen aanstipten, valt er vanuit een cognitief psychologisch standpunt bezien nogal wat commentaar te leveren op het Mark-model. De vraag rijst dan ook waarom wij met dit commentaar geen rekening gehouden hebben bij de konstruktie van dit model. Elders (deel I; p. 18 e.v.) hebben wij deze vraag reeds indirekt beantwoord door te stellen dat men bij het psychodiagnostisch onderzoek door middel van tests vrijwel altijd verschillende funkties tegelijkertijd meet, zij het dat bij de ene test meer het aksent valt op één bepaalde funktie en bij een andere test meer het aksent op een andere funktie als onderdeel van de mogelijke verzameling funkties. Als zodanig is het bijzonder lastig - en met behulp van de huidige psychologische tests misschien ook wel onmogelijk - om één afzonderlijke cognitieve funktie, welke experimenteel-psychologisch gedefinieerd en geoperationaliseerd is, redelijk nauwkeurig te meten. Dit geldt eens te meer indien de verzameling funkties zowel serieel als parallel geschakeld is, gedeeltelijk hiërarchische en gedeeltelijk heterarchische trekken vertoont en daarnaast nog zowel kanaalafhankelijk als -onafhankelijk kan zijn.

Het relatief complexe karakter van de psychologische test is echter niet het volledige antwoord op de vraag. Ook het min of meer hypothetische karakter van de verzameling funkties heeft ons doen besluiten het Mark-model zo simpel mogelijk te houden. Het leek ons namelijk weinig zinvol een ingewikkeld model van de informatieverwerking te ontwerpen, waar over een belangrijk element van dat model - de verzameling funkties - nog zo weinig harde empirische gegevens bekend zijn, althans vanuit een testpsychologisch standpunt bezien. Naar analogie van de 'Law of Parsimony' voor empirisch-wetenschappelijk onderzoek zouden wij hieraan willen toevoegen dat het ontwerp van een zo simpel mogelijk model bovendien de voorkeur verdient boven het ontwerp van een tamelijk ingewikkeld model.

Waar bij voorbaat zo veel commentaar mogelijk is, kan men zich afvragen wat dan wél de verdienste is van het Mark-model. Onzes inziens zijn de kracht en de betekenis van dit model terug te voeren op zijn heuristisch karakter; dat wil zeggen het Mark-model vormt een leidraad voor de psychodiagnosticus bij de konstruktie, analyse en uitbreiding van een testbatterij waarmee het cognitief funktioneren van een kind zo goed mogelijk in kaart wordt gebracht. Het nu volgende hoofdstuk is de gedeeltelijke illustratie hiervan.

3. LDT-E: SAMENSTELLING, BESCHRIJVING EN PLAATSBEPALING

3.1 Samenstelling

In paragraaf 1.3 formuleerden wij het eerste doel van ons onderzoek als volgt: "de ontwikkeling van een meetinstrument ter bepaling van zowel de intelligentie - het algemeen cognitief ontwikkelingsniveau - als het niveau van een aantal afzonderlijke cognitieve functies, die mogelijk van belang zijn voor het leren op school" (p. 11). Deze doelstelling vereist een nadere toelichting.

Voor een betrouwbare bepaling van de intelligentie is de individueel af te nemen intelligentietest het meest geëigende instrument. Hierbij doet zich direkt het probleem voor wat precies onder intelligentie verstaan moet worden. Reeds een vluchtig onderzoek van de literatuur leert dat de meningen verdeeld zijn. Wechsler (1958) bijvoorbeeld geeft de volgende, vrij brede definitie:

"... intelligence, operationally defined, is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (p. 7).

In feite is men echter volgens Guilford (1967; p. 13) nooit veel verder gekomen dan de opvatting van Boring (1923) dat:

"... intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test".

Met andere woorden: intelligentie is datgene wat een intelligentietest meet. Ook Drenth et al. (1968) hebben zich bij de ontwikkeling van de Amsterdamse Kinder Intelligentie Test (AKIT) door deze gedachtengang laten leiden. In laatste instantie geeft een dergelijke operationele definitie aanleiding om te spreken van bijvoorbeeld AKIT-intelligentie, WISC-intelligentie (Wechsler, 1949), Stanford-Binet-intelligentie (Terman & Merrill, 1960) en LDT-E-intelligentie. Het is duidelijk dat, bij gebrek aan een unaniem geaccepteerd intelligentieconcept, de mogelijkheid bestaat om vele psychologische tests als intelligentietest te betitelen, titelen.

logische begrip intelligentie, bestaande uit die elementen of factoren welke intelligentietests met elkaar gemeen hebben. In het verleden zijn dan ook herhaalde pogingen gedaan om door middel van begripsvalideringsonderzoek het intelligentieconcept te verhelderen. Volgens Drenth (1975; p. 33) blijken nog tot vandaag de dag de door Thurstone (1938) geïdentificeerde factoren Verbaal (verbal), Numeriek (numerical), Ruimtelijk

(space), Woordenrijkdom (word fluency), Geheugen (memory) en Abstrakt redeneren (reasoning) belangrijke onderscheidingen te zijn in de intelligentietheorie.

Bij de konstruktie van de LDT-E als intelligentietest is met deze onderscheidingen zoveel mogelijk rekening gehouden; zoveel mogelijk, omdat de konstruktie van een intelligentietest op gespannen voet staat met de konstruktie van een diagnostische test; dat wil zeggen een test bestemd voor het meten van een aantal afzonderlijke cognitieve funkties (of factoren). Probeert men in een intelligentietest een zo breed mogelijk terrein van het complexe begrip intelligentie te bestrijken, bij een diagnostische test is men veeleer geïnteresseerd in bepaalde deelaspekten. De betrouwbaarheid waarmee een deelaspekt (functie of faktor) gemeten wordt dient verder bij voorkeur van dezelfde orde te zijn als de betrouwbaarheid waarmee intelligentie doorgaans gemeten wordt. Een dergelijke voorwaarde stelt bijzonder hoge eisen aan de betrouwbaarheid van tests voor individueel psychodiagnostisch gebruik; dit in tegenstelling tot de (sub)tests van een intelligentietest, welke minder betrouwbaar hoeven te zijn om toch, in lineaire combinatie, een zeer betrouwbare schatting van de intelligentie mogelijk te maken.

De gangbare manier om de betrouwbaarheid van een test te verhogen is testverlenging door uitbreiding van het aantal items. Testverlenging betekent echter een langere afnameduur. Uitgaande van een beperkte hoeveelheid onderzoekstijd komt men nu voor het probleem te staan een groot aantal deelaspekten (zoals vereist voor een intelligentietest) te combineren met een hoge mate van betrouwbaarheid van de verschillende deelaspekten (zoals vereist voor een diagnostische test). Cronbach en Gleser (1965) karakteriseren dit probleem als het 'Bandwijdte-Zuiverheid'-dilemma, waarbij bandwijdte staat voor veelsoortigheid van informatie en zuiverheid voor de betrouwbaarheid van de verschillende soorten informatie. Bandwijdte en zuiverheid staan in komplementaire relatie tot elkaar: een test met een brede band en dus een groot bereik aan allerlei soorten informatie, zal wat de afzonderlijke informatie-aspekten betreft, minder zuiver (betrouwbaar) zijn en omgekeerd. Zo bestrijkt de LDT-E als intelligentietest een minder breed terrein van het complexe begrip intelligentie dan voor deze leeftijd wenselijk zou zijn. Hiertegenover staat een grotere lengte en dus hogere betrouwbaarheid van de afzonderlijke tests uit de LDT-E, welke diagnostisch gezien belangrijk is.

Tenslotte: evenals in de voorafgaande delen worden ook in dit deel de begrippen 'intelligentie' en 'algemeen cognitieve ontwikkeling' naast

elkaar gebruikt alsof deze equivalent zijn. Gelet op de maat die wij tot nog toe voor deze begrippen hanteerden - namelijk het deviatie IQ als de getransformeerde som van de genormaliseerde standaardscores die een kind op de tests van de LDT haalt - mag men zelfs van volstrekt identieke begrippen spreken¹⁰. Historisch gezien hebben beide begrippen echter weinig met elkaar te maken. Zo stoelt het intelligentiebegrip hoofdzakelijk op faktoranalytische studies van tests die in de loop der tijd het predikaat 'intelligentietest' verworven hebben, terwijl aan het begrip 'algemeen cognitieve ontwikkeling' - als noemer en samenvatting van het cognitief functioneren - de gedachtengang van het Mark-model ten grondslag ligt. Nu hebben wij de faktoranalytische traditie bij de konstruktie van de LDT-E niet geheel en al verwaarloosd, al was het maar om de bevindingen met de LDT-E in een ruim, historisch kader te kunnen plaatsen en de potentiële gebruiker van de test niet bij voorbaat in verwarring te brengen omtrent zijn betekenis. Om deze reden hebben wij dan ook besloten om de begrippen 'intelligentie' en 'algemeen cognitieve ontwikkeling', voortvloeiend uit geheel verschillende tradities maar volstrekt identiek in de wijze waarop ze geoperationaliseerd zijn, naast elkaar te gebruiken.

In de hoofdstukken 1 en 2, evenals in de inleidende beschouwing van deze paragraaf zijn enkele overwegingen genoemd welke een rol speelden bij de samenstelling van de LDT-E. In het nu volgende overzicht zullen deze overwegingen puntsgewijs aangevuld met een aantal resterende opmerkingen opnieuw aan de orde komen als verantwoording voor de uiteindelijke samenstelling.

Ten eerste: in de LDT-E dienen de belangrijkste leerkanalen zo zuiver mogelijk vertegenwoordigd te zijn. Voor een grondig psychodiagnostisch onderzoek van potentiële of feitelijke leermoeilijkheden acht Mark een systematische toetsing op dysfuncties binnen tenminste 24 kanalen noodzakelijk. De beperkte opzet van de LDT-E laat een dergelijk uitvoerige toetsing niet toe: slechts de allerbelangrijkste leerkanalen kunnen bij het onderzoek naar potentiële of feitelijke leermoeilijkheden gecontroleerd worden. Algemeen wordt aangenomen dat in het onderwijs-leerproces de auditief-spraakmotorische en visueel-fijnmotorische kanalen een overheersende rol spelen bij de informatieverwerking. Deze veronderstelling wordt weerspiegeld in de keuze van input-output kenmerken van tests als de AKIT, ITPA (Kirk et al., 1968), SON (Snijders & Snijders-Oomen, 1958), UTANT (Kohnstamm et al., 1971), WISC en WPPSI, die bij het onderzoek naar

leermoeilijkheden een veelvuldig gebruik vinden. Uit onderzoek van Bakker (1972) blijkt daarnaast dat intermodale informatieverwerking, dat wil zeggen informatieverwerking waarbij van twee verschillende modaliteiten of kanalen binnen een taakopdracht gebruik wordt gemaakt, een minstens zo belangrijke rol speelt in het onderwijs-leerproces. Met name het auditief-spraakmotorisch/visueel-fijnmotorisch kanaal lijkt van de vele mogelijke combinaties veelbelovend voor het onderzoek naar leesmoeilijkheden.

Ten tweede: in de LDT-E dienen de belangrijkste functies zo zuiver mogelijk vertegenwoordigd te zijn. In het vorige hoofdstuk zijn functies als aandacht, concentratie, sensoriek, perceptie en motoriek, geheugen, taal en cognitie aangeduid als de belangrijkste functies die voor een onderzoek naar leermoeilijkheden in aanmerking komen. Ondanks de ogenschijnlijke eenvoud van deze benamingen zijn genoemde functies in feite uitermate grove aanduidingen voor wat in de experimentele (neuro)psychologie aanzienlijk gedetailleerder bestudeerd wordt. Een eenvoudig voorbeeld vormt de geheugenfunctie: in onderzoek maakt men een duidelijk onderscheid tussen het onmiddellijk geheugen en een geheugen voor informatie die meer blijvend wordt vastgelegd. Binnen het onmiddellijk geheugen wordt vervolgens weer een aantal andere functies onderscheiden (Thomasen & Kempen, 1976). Het experimentele onderzoek van het geheugen vereist tamelijk geavanceerde proefopzetten en meetapparatuur. In vergelijking met de laboratoriumuitrusting die de experimenteel psycholoog tot zijn beschikking heeft, maakt het functie-onderzoek met behulp van psychologische tests een povere indruk. Door de aard van het testmateriaal en de opzet van het psychodiagnostisch onderzoek worden vrijwel altijd verschillende functies tegelijkertijd gemeten. In tegenstelling tot de eis van kanaalzuiverheid valt de eis van functiezuiverheid in de praktijk van het psychologisch testonderzoek dan ook moeilijk te handhaven. Ofschoon het ideaal van functiezuiverheid misschien met behulp van nieuw te konstrueren tests bereikbaar zou zijn, stemt de geschiedenis van de psychologische test tot pessimisme: sinds de konstruktie van de eerste psychologische test door Binet en Simon in 1905 zijn geen fundamentele wijzigingen tot stand gebracht. Wel zijn de psychometrische kenmerken van tests aanzienlijk verbeterd. De keuze voor een selectie uit bestaande min of meer functie-onzuivere tests lijkt daarom voor de hand liggend. Bij de ontwikkeling van een volledig nieuwe test moet men bovendien maar afwachten of deze valide is en zich ook in de praktijk goed laat afnemen. Objectieve criteria voor een selectie uit bestaande tests zijn, zoals uit

het voorafgaande blijkt, moeilijk vast te stellen. Daarentegen kunnen wel enkele richtlijnen gegeven worden. Vanuit het Mark-model biedt de hiërarchische organisatie van functies een duidelijke vingerwijzing. Per kanaal zijn tests geselecteerd die meer en minder complexe functies meten. Om een voorbeeld te geven: tests die hoofdzakelijk een beroep doen op de imitatieve vaardigheid van een kind (onder andere onmiddellijk geheugentestjes) zijn van een minder complex funktieniveau dan tests die ook eenvoudige redeneeraspekten bevatten. Om verder een vergelijking mogelijk te maken tussen het auditief-spraakmotorische (AS) en het visueel-fijnmotorische (VF) kanaal, is ernaar gestreefd dat de functie-opbouw in beide kanalen zo parallel mogelijk verloopt. In zowel het AS- als het VF-kanaal zijn daarom zo identiek mogelijke functies opgenomen. Vanuit de onderzoeksliteratuur is reeds een aantal functies genoemd. Aandacht en onmiddellijk geheugen, evenals eenvoudige begrips- en redeneerfuncties lijken zowel binnen het AS- als het VF-kanaal vanuit het belangrijkste te zijn.

Ten derde: de LDT-E dient een schatting mogelijk te maken van het algemeen cognitief ontwikkelingsniveau, c.q. de intelligentie¹⁰. Hiertoe behoren de belangrijkste, door Thurstone (1938) geïdentificeerde factoren al dan niet in combinatie met elkaar aanwezig te zijn in de LDT-E. Een uitzondering willen wij echter maken voor de factoren 'Numeriek' en 'Woordenrijkdom'; voor de faktor 'Numeriek' omdat de leeftijd van de te onderzoeken kinderen (5- en 6-jarigen) de realisatie van een dergelijke faktor in de LDT-E op voorhand vrijwel onmogelijk maakt, en voor de faktor 'Woordenrijkdom' omdat deze - in vergelijking tot de overige factoren - minder relevant lijkt voor het diagnostisch onderzoek naar leer-moeilijkheden.

Ten vierde: de LDT-E dient gemakkelijk afgenomen te kunnen worden. Deze eis valt voor een deel samen met de overweging dat de totale afnameduur van de LDT-E niet veel meer dan één uur mag bedragen.

In tabel 3 wordt een overzicht gegeven van de uiteindelijke samenstelling van de LDT-E. In kolom (1) zijn de acht tests gerangschikt overeenkomstig hun kanaal- en funktiekenmerken met Plaatjes aanwijzen als typische overgangstest. Door het ontbreken van een systematische naamgeving binnen de psychologie zijn de tests tamelijk willekeurig benoemd. Soms hebben de namen betrekking op het materiaal (bijvoorbeeld Blokpatronen), dan weer op de handeling die verricht moet worden (Natieken) of op de inhoud van de test (Begrip en inzicht). Zoveel mogelijk is echter aansluiting gezocht.

Tabel 3. Kanaal- en functiekenmerken van de tests

test (1)	kanaal		
	input ^a (2)	funkties (3)	output ^b (4)
1. Blokpatronen (BP)	V	- patroonwaarneming - matching - analyse - synthese - redeneren - oog - hand koördinatie	F
2. Vouwblaadjes (VB)	V	- aandacht - geheugen voor ruimtelijke transformaties - redeneren - oog - hand koördinatie	F
3. Natikken (NT)	V	- aandacht - onmiddellijk geheugen voor ruimtelijke sekwenties - oog - hand koördinatie	F
4. Woordenspan (WS)	A	- aandacht - onmiddellijk geheugen voor woordsekwenties	S
5. Plaatjes aanwijzen (PA)	AV	- aandacht - geheugen voor woordsekwenties - matching (woorden-plaatjes)	F
6. Zinnen nazeggen (ZN)	A	- aandacht - onmiddellijk geheugen voor zinnen	S
7. Verhaaltje vragen (VV)	A	- begrijpen van vragen - semantisch geheugen - (re)productieve taalvaardigheid	S
8. Begrip en inzicht (BI)	A	- begrijpen van vragen - redeneren - productieve taalvaardigheid	S

^a V = visueel, A = auditief, AV = auditief-visueel

^b F = fijnmotorisch, S = spraakmotorisch

bij de konventie dienaangaande¹¹.

Kolom (2) en kolom (4) geven de input-output kenmerken aan van het kanaal waarop de betrokken test zich richt. Kolom (3) omvat een korte beschrijving van de belangrijkste functies binnen een kanaal per test.

Bij het inventariseren van de funkties die door één test gemeten worden, doet zich het probleem voor van de keuze van de belangrijkste funkties die door een test gemeten worden. Het lijkt immers geen twijfel dat nog andere funkties dan de hier genoemde onderscheiden kunnen worden. De meest bevredigende oplossing is gevonden door bij de keuze van de funkties uit te gaan van de grootste gemene deler zoals deze wordt aangetroffen in de bestaande literatuur (cf. par. 3.2).

Uit het overzicht in kolom (3) blijkt dat de combinatie van funkties per test telkens verschillend is. In een profielanalyse wordt zodoende de mogelijkheid geschapen om naar afzonderlijke funkties binnen een kanaal of vergelijkenderwijs tussen kanalen onderling een onderzoek te doen. In paragraaf 3.4 wordt in het kort aangegeven hoe een dergelijke analyse in zijn werk gaat.

3.2 Beschrijving van de tests

Blokpatronen

Taak van de proefpersoon (pp): binnen een bepaalde tijd een voorbeeldpatroon reproduceren met behulp van platte rood- en witgekleurde blokjes. De eerste zeven voorbeeldpatronen worden door de proefleider (pl) gevormd. De resterende drie patronen staan ieder afzonderlijk op een kaart afgebeeld.

Blokpatronen is identiek met de subtest Block Design uit de WPPSI (Wechsler, 1967). Materiaal, instructie, tijdgrenzen, afbreekregel en scoring zijn zonder essentiële wijzigingen overgenomen uit de handleiding van de WPPSI (cf. Schroots, 1973).

Over Block Design is onder andere gepubliceerd door Austin en Carpenter (1970), Herman (1968), Massey (1968) en Wechsler (1967). Mede aan de hand van deze gegevens zou men Blokpatronen kunnen typeren als een test voor het visueel-fijnmotorische kanaal, waarin de volgende funkties onderscheiden worden: patroonwaarneming, matching, analyse van een geheel (patroon) in delen (rood- en witgekleurde blokjes), synthese van delen tot een geheel, redeneren (planning, organisatie, begripsvorming) en oog-hand coördinatie.

Vouwblaadjes

Taak van pp: vouwen van een vierkant blaadje papier (vouwblaadje) volgens hetzelfde patroon als door pl wordt voorgedaan. Het gevouwen

voorbeeld (eindpatroon) blijft als model voor pp duidelijk zichtbaar op tafel liggen.

Vouwblaadjes is een verlengde versie van de subtest Paper Folding uit de Hiskey-Nebraska Test of Learning Aptitude (HNTLA) van Hiskey (1966). Voor 5- en 6-jarigen bevat Paper Folding een te gering aantal items. Ter verhoging van het differentiërend vermogen en de betrouwbaarheid van Vouwblaadjes zijn daarom vijf nieuwe items toegevoegd, i.c. item 6 t/m 8, 10 en 11. Bij de konstruktie van de nieuwe items is rekening gehouden met de principes welke aan Paper Folding ten grondslag liggen. Gelet is op het alterneren van de beginpatronen (rechthoekig of driehoekig), op het aantal deelpatronen per item en de complexiteit van de afzonderlijke deelpatronen, zowel naar richting als beweging. Daarnaast is de kleur van de vouwblaadjes gewijzigd in meigroen ter verhoging van de attractiviteit. De instructie, afbreekregel en scoring van Vouwblaadjes is ten slotte op sommige punten verbeterd, zonder dat hiermee de essentie van de test werd aangetast (cf. Schroots, 1973).

Ofschoon de HNTLA, primair bedoeld voor het dove en slechthorende kind, al sinds 1941 vrij frekvent gebruikt wordt, is naar verhouding weinig bekend over de betekenis van de afzonderlijke subtests. Buros' Seventh Mental Measurements Yearbook (1972) verstrekt slechts informatie over de korrelatie tussen de totaalscore op de HNTLA en andere intelligentietests, zoals de Stanford-Binet (Form L-M) en de WISC. Mede op basis van analyses van Anastasi (1961) en Sattler (1972) zou men Vouwblaadjes kunnen typeren als een test voor het visueel-fijnmotorische kanaal, waarin de volgende functies onderscheiden worden: continue aandacht, geheugen voor ruimtelijke transformaties, redeneren en oog-hand koördinatie.

Natikken

Taak van pp: aantikken van een serie blokjes in dezelfde volgorde als door pl wordt voorgedaan. Het aantikken vindt plaats met behulp van een hamertje.

Natikken is een licht gewijzigde versie van de Revised Knox Cube Test uit A Point Scale of Performance Tests, Revised Form II van Arthur (1947). De wijzigingen hebben betrekking op items, testmateriaal, instructie en scoring. De items 1 en 2 uit de Arthur zijn in Natikken proefitems a en b en tellen niet bij de beoordeling. Items 1 en 13 uit Natikken zijn nieuwe items, toegevoegd aan de oorspronkelijke versie om een grotere differentiatie onder respektievelijk de jongste en oudste kleuters te

bewerkstelligen.

Ter verhoging van de attractiviteit van Natikken is het testmateriaal, bestaande uit een plankje met vier houten blokjes, uitgevoerd in een helderrode kleur. De potloden waarmee de blokjes oorspronkelijk aange-tikt moesten worden zijn vervangen door twee hamertjes (xylofoonstokjes): een grote voor pl en een kleine voor pp.

De oorspronkelijke instructie voor de Revised Knox Cube Test, bedoeld voor het dove en slechthorende kind, wordt in gebarentaal gegeven. Voor Natikken is daarom een geschikte mondelinge versie opgesteld. Natikken hoeft verder slechts éénmaal afgenomen te worden, in tegenstelling tot de Revised Knox Cube Test. De scoring is dienovereenkomstig aangepast (cf. Schroots, 1973).

Ondanks het feit dat Knox al in 1927 deze test samenstelde, is er betrekkelijk weinig onderzoek naar zijn betekenis verricht. Studies van enig belang hebben alle betrekking op de verdienste van de 'Arthur' als intelligentietest (zie ook Buros' Sixth Mental Measurements Yearbook, 1965). Een uitzondering vormt het onderzoek van Orpet en Meyers (1966) waarin een duidelijke relatie aangetoond wordt tussen Natikken (c.q. de Knox Cube Tapping Test) en testjes die het Piagetiaanse begrip 'Conser-vatie' meten. Mede op basis van een analyse van Anastasi (1961) zou men Natikken kunnen typeren als een test voor het visueel-fijnmotorische kanaal, waarin de volgende functies onderscheiden worden: continue aandacht, onmiddellijk geheugen voor ruimtelijke sekwenties en oog-hand koördinatie.

Woordenspan

Taak van pp: herhaling van een serie woorden in dezelfde volgorde als door pl wordt voorgezegd.

Woordenspan is een nieuwe test, gebaseerd op de experimentele subtest Temporele Ordening Auditief-Visueel (TOAV) uit de Leesvaardigheid Pre-diktie Testbatterij van Bakker (1967; 1972) en Groenendaal en Bakker (1971).

In de TOAV is een aantal fundamentele wijzigingen aangebracht met betrekking tot het testmateriaal, de samenstelling van de items en de af-nameprocedure. De belangrijkste wijziging betreft een ingreep in de af-nameprocedure, waardoor twee tests verkregen werden. In de eerste test, Woordenspan, moet pp een serie woorden nazeggen. In de tweede test, Plaatjes aanwijzen, moet pp de bij deze serie woorden behorende plaatjes

aanwijzen op een testkaart. Plaatjes aanwijzen volgt direkt op het na-zeggen van een serie woorden en kan, in tegenstelling tot Woordenspan, niet als een zelfstandig af te nemen test worden beschouwd.

Items en testmateriaal zijn verder volledig nieuw gekonstrueerd. Een uitvoerige beschrijving van het konstruktieproces en de gedachtengang achter beide tests wordt gegeven door Schroots et al. (1975). In de LDT-E heette Woordenspan nog LPT-a (Lees Prediktie Test-a), respektievelijk LPT-na-zeggen (cf. Schroots, 1973).

In een publikatie van Rispens (1974) is reeds een aantal voorlopige onderzoeksbevindingen over Woordenspan en Plaatjes aanwijzen vermeld. Mede aan de hand van deze en bovenstaande gegevens zou men Woordenspan kunnen typeren als een test voor het auditief-spraakmotorische kanaal, waarin de volgende funkties onderscheiden worden: kontinue aandacht en onmiddellijk geheugen voor woordsekwenties.

Plaatjes aanwijzen

Taak van pp: de bij een serie woorden horende plaatjes in de juiste volgorde aanwijzen op een testkaart. Het aanwijzen vindt direkt plaats nadat een serie woorden door pp is opgenoemd.

Bij de bespreking van Woordenspan is reeds gewezen op de herkomst van Plaatjes aanwijzen en de nauwe samenhang met deze test¹². Voor een uitvoerige bespreking van de konstruktie raadplege men Schroots et al. (1975). In de LDT-E wordt Plaatjes aanwijzen aangeduid met LPT-b (Lees Prediktie Test-b), respektievelijk LPT-aanwijzen (cf. Schroots, 1973).

werking verlangd wordt. Als zodanig onderscheidt hij zich van alle overige intramodale tests van de LDT-E, welke slechts op één leerkanaal voor de informatieverwerking een beroep doen. In tabel 3, kolom (2), wordt door middel van het symbool AV aangegeven dat het intermodale karakter van Plaatjes aanwijzen betrekking heeft op de input. Pp dient eerst auditieve informatie (woorden) te verwerken via het auditief-spraakmotorische kanaal, alvorens deze te combineren (matchen) met visuele informatie (plaatjes).

Op basis van de bij Woordenspan vermelde onderzoeksbevindingen zou men Plaatjes aanwijzen kunnen typeren als een test voor het auditief/visueel-fijnmotorische kanaal, waarin de volgende funkties onderscheiden worden: aandacht, geheugen voor woordsekwenties en matching (woorden-plaatjes).

Zinnen nazeggen

Taak van pp: letterlijke herhaling van iedere zin uit een verhaaltje dat door pl wordt voorgelezen. Het verhaaltje heeft betrekking op een jongen die met zijn moeder boodschappen doet.

Zinnen nazeggen is een door Reesink ontwikkelde experimentele taalttest, waarover geen officiële publikatie verschenen is. Volgens Groenendaal en Barritt (1975a; b) ligt het werk van Blank en Frank (1971) en Arlman-Rupp (1971) ten grondslag aan de konstruktie van deze test. De kern van Reesink's taalttest wordt gevormd door een verhaaltje, opgebouwd uit een aantal zinnen waarin doelbewust verschillende grammatikale structuren verwerkt zijn.

Evenals bij Woordenspan en Plaatjes aanwijzen is de test gesplitst in twee afzonderlijke gedeelten: Zinnen nazeggen en Verhaaltje vragen. In Zinnen nazeggen moet pp iedere zin letterlijk herhalen. In Verhaaltje vragen moet pp een aantal vragen over het verhaaltje beantwoorden. Verhaaltje vragen volgt direkt op Zinnen nazeggen en kan, in tegenstelling tot Zinnen nazeggen, niet als een zelfstandig af te nemen test beschouwd worden. Groenendaal en Barritt (1975a; b) beschrijven voorts de ingewikkelde psycholinguïstische scoring die voor Zinnen nazeggen is gehanteerd. In de LDT-E wordt Zinnen nazeggen tenslotte aangeduid met Verhaaltje nazeggen (cf. Schroots, 1973).

Door Groenendaal en Barritt (1975a; b) zijn reeds enige onderzoeksbevindingen over Zinnen nazeggen en Verhaaltje vragen gepubliceerd. Mede op basis van deze gegevens zou men Zinnen nazeggen kunnen typeren als een test voor het auditief-spraakmotorische kanaal, waarin de volgende functies onderscheiden worden: continue aandacht en onmiddellijk geheugen voor zinnen.

Verhaaltje vragen

Taak van pp: beantwoording van een aantal vragen over een verhaaltje dat door pl wordt voorgelezen. De vragen worden direkt gesteld nadat de zinnen uit het verhaaltje door pp letterlijk zijn herhaald.

Bij de bespreking van Zinnen nazeggen is reeds gewezen op de herkomst van Verhaaltje vragen en de nauwe relatie met deze test. Ook Verhaaltje vragen is qua items, instructie en scoring letterlijk overgenomen van Reesink. Voor een uitvoerige toelichting raadplege men Groenendaal en Barritt (1975a; b), cf. Schroots (1973).

Op basis van de bij Zinnen nazeggen vermelde onderzoeksgegevens zou men

Verhaaltje vragen kunnen typeren als een test voor het auditief-spraak-motorische kanaal, waarin de volgende functies onderscheiden worden: begrijpen van vragen, semantisch geheugen en (re)productieve taalvaardigheid.

Begrip en inzicht

Taak van pp: beantwoording van een aantal vragen, deels normatief, deels praktisch en sociaal van aard. De vragen hebben betrekking op uiteenlopende onderwerpen zoals gezondheid, hygiëne, omgeving en tussenmenselijke relaties.

Begrip en inzicht is de letterlijke vertaling van de subtest Comprehension uit de WPPSI (Wechsler, 1967). Deze laatste omvat zes items van de subtest Comprehension uit de WISC (Wechsler, 1949). Instructie, afbreekregel en scoringscriteria zijn zonder essentiële wijzigingen overgenomen uit de handleiding van de WPPSI (cf. Schroots, 1973).

Over de Wechsler-tests zijn vrij veel onderzoeksgegevens gepubliceerd (onder andere Austin & Carpenter, 1970; Herman, 1968; Massey, 1968; Wechsler, 1967). Mede aan de hand van deze gegevens zou men Begrip en inzicht kunnen typeren als een test voor het auditief-spraakmotorische kanaal, waarin de volgende functies onderscheiden worden: begrijpen van vragen, logisch redeneren (gebaseerd op ervaringskennis en inzicht in sociale verhoudingen) en productieve taalvaardigheid.

3.3 Overeenkomst en verschil tussen LDT-E en LDT

Zoals wij in paragraaf 1.1 reeds opmerkten, is de LDT-E niet volledig identiek met de LDT. In verband met de vraag naar de geldigheid van onze bevindingen met de LDT-E voor de LDT, lijkt het dan ook nuttig de overeenkomsten en verschillen tussen de tests uit de experimentele versie (par. 3.2) en de definitieve versie (deel I; p. 23-28) beknopt weer te geven. Vergelijking van beide versies leert het volgende:

Blokpatronen uit de LDT-E omvat in totaal 10 items (patronen); dit in tegenstelling tot Blokpatronen uit de LDT, welke in totaal 15 items omvat. Daarnaast komen de items 5 en 10 van Blokpatronen uit de LDT niet voor in de LDT-E. Voor het overige zijn er geen verschillen tussen de experimentele en de definitieve versie;

Vouwblaadjes uit de LDT-E omvat in totaal 12 items, welke gelijk zijn aan de eerste 12 items van Vouwblaadjes uit de LDT. Het enige verschil tussen beide versies is dus dat de items 13 t/m 18 van Vouwblaadjes uit

de LDT niet in de LDT-E voorkomen;

Natikken uit de LDT-E omvat 5 items méér dan Natikken uit de LDT. Daarnaast zijn de items 7 en 8, evenals de items 10 en 11 in de experimentele versie van plaats verwisseld. Voor het overige zijn er geen verschillen tussen Natikken uit de LDT en de LDT-E;

Woordenspan en *Plaatjes aanwijzen* uit de LDT-E zijn volledig identiek met de gelijknamige versies uit de LDT;

Zinnen nazeggen uit de LDT-E omvat 2 items (zinnen) minder dan Zinnen nazeggen uit de LDT (zin 3 en 7). Verder zijn er geen verschillen tussen beide versies, uitgezonderd de ingewikkelde psycholinguïstische scoringsmethode die voor de experimentele versie is gehanteerd. Uit later onderzoek (Van Alphen de Veer & Plomp, 1976) is gebleken dat de ingewikkelde methode .95 korreleert met de simpele methode welke voor Zinnen nazeggen uit de LDT is gebruikt. Op basis van dit gegeven mogen wij dan ook alvast konkluderen dat de betekenis van Zinnen nazeggen uit de LDT-E niet noemenswaard verschilt van die van Zinnen nazeggen uit de LDT;

Verhaaltje vragen uit de LDT-E wijkt op verschillende punten af van Verhaaltje vragen uit de LDT. Zo zijn de items (vragen) uit de experimentele versie iets abstrakter geformuleerd dan in de definitieve versie en komt de eerste vraag uit de LDT - Hoe heet de jongen uit het verhaaltje? - nog niet voor in de LDT-E. Ook de scoringscriteria en antwoordelementen van Verhaaltje vragen uit de LDT-E zijn minder gepolijst dan die van Verhaaltje vragen uit de LDT. Voor een uitvoerig onderzoeksverslag van genoemde wijzigingen verwijzen wij verder naar Zaal (1976);

Begrip en inzicht uit de LDT-E omvat in totaal 15 vragen, een vraag méér dan in de LDT. Daarnaast komen de items 7 en 15 uit de experimentele versie niet voor in de definitieve versie welke, zoals bekend, één item uit de Vlaamse versie van de WISC (Lagae & Coetsier, 1963) omvat. Tenslotte zijn ook de scoringscriteria en antwoordvoorbeelden van Begrip en inzicht op een aantal punten verbeterd..

3.4 Plaatsbepaling

Zoals bekend is de LDT-E bestemd voor het individueel psychodiagnostisch onderzoek van kinderen. Een dergelijke individuele benaderingswijze laat zich echter moeilijk rijmen met het onderzoek van grote aantallen kinderen ter vroegtijdige opsporing van leermoeilijkheden. Om het aantal kinderen dat voor een individueel onderzoek in aanmerking komt te beperken, last men daarom meestal een signaleringsonderzoek in als eerste fase van

een meertraps sekventiële testprocedure ('multistage sequential testing'; cf. Cronbach & Gleser, 1965). Voor dit signaleringsonderzoek maakt men dan gebruik van groepsgewijs af te nemen psychologische tests, waarover wij in paragraaf 4.2 meer informatie zullen verstrekken.

De LDT-E nu is mede bestemd voor de tweede fase van een sekventiële testprocedure, welke in principe met nog een aantal andere fasen uitgebreid kan worden. Uitdrukkelijk schrijven wij 'mede' omdat de LDT-E ook als afzonderlijk instrument of binnen een combinatie van andere instrumenten en procedures gehanteerd kan worden, afhankelijk van de oogmerken van de psychodiagnosticus.

Of de LDT-E gebruikt wordt in de eerste dan wel tweede fase van een sekventiële testprocedure, in beide gevallen dient men na afloop van het testonderzoek het LDT-E profiel te analyseren op mogelijk sterke en zwakke punten in de cognitieve ontwikkeling van een kind.

In principe kan dit profiel op talloze wijzen geanalyseerd worden. Sattler (1974; p. 201 e.v.) laat bijvoorbeeld zien hoe het WISC profiel in de loop der tijd op tenminste negen manieren geanalyseerd is. Wij zullen de voor- en nadelen van de verschillende methoden hier niet bespreken; men raadplege daarvoor Sattler. Wel willen wij opmerken dat de meest gebruikelijke analysemethoden uitgaan van een standaard waartegen de testprestaties van een kind systematisch afgezet worden.

Zo kiest Mark bijvoorbeeld - volgens de redenering beschreven in paragraaf 2.2 - als standaard de beste testprestatie, c.q. hoogste testscore van een kind; een standaard die hij met de term 'maximaal verwachtingsniveau' betitelt en baseert op de ontogenie van functies zoals die weerspiegeld wordt in de leeftijdsnormen van psychologische tests (cf. par. 2.3). In navolging van Mark hebben ook wij gekozen voor een dergelijke standaard (cf. deel II; p. 85 e.v.), zij het dat onze standaard gevormd wordt door het gemiddelde van de hoogste drie testprestaties (scores) van een kind op de LDT¹³. De reden hiervoor is dat wij één enkele testscore te onbetrouwbaar achten.

Nadat aldus de standaard, i.c. het maximaal verwachtingsniveau (MVN) is bepaald, verloopt de rest van de analyse in principe als volgt. Systematisch wordt nagegaan welke scores significant verschillen van het MVN. Indien een significant verschil gevonden wordt, dan betekent dit dat er sprake is van een zwak punt in de cognitieve ontwikkeling van het kind. Indien het gevonden verschil niet significant is, dan kan men dit weer als een sterk punt opvatten. Aangezien van iedere test(score) bekend is

welke functies en welk kanaal hij representeert (cf. tabel 3), is het daarna betrekkelijk eenvoudig om de sterke en zwakke punten in het LDT-profiel te vertalen of te beschrijven in termen van deze functies onder vermelding van het betrokken leerkanaal. Ter illustratie van dit laatste een voorbeeld:

Binnen het visueel-fijnmotorische leerkanaal blijkt de score voor Vouwblaadjes duidelijk onder het vastgestelde maximaal verwachtingsniveau te liggen. Voor Blokpatronen en Natikken werd daarentegen geen duidelijk verschil gevonden.

Dit kan betekenen dat de volgende cognitieve functie binnen dit kanaal relatief zwak ontwikkeld is: geheugen voor ruimtelijke transformaties. Relatief sterk ontwikkeld zijn: patroonwaarneming, matching (het in overeenstemming brengen van patronen), analyse van een geheel (patroon) in delen (rood- en witgekleurde blokjes), synthese van delen tot een geheel, en onmiddellijk geheugen voor ruimtelijke sekwenties.

Deze beschrijving vormt een willekeurig gekozen fragment uit de verzameling uitspraken die op basis van een systematische analyse van het LDT-profiel opgesteld kan worden. De regels voor deze analyse worden 'en détail' weergegeven in het Stroomdiagram voor de LDT (deel III; Schroots et al., 1978b), terwijl de uitspraken apart gebundeld zijn in deel IV (Schroots et al., 1978c).

Zoals gesteld is de LDT-E bedoeld voor de eerste dan wel tweede fase van een psychodiagnostisch onderzoek dat (na analyse van het testprofiel) nu eens antwoord geeft op een bepaalde vraag dan weer indicaties oplevert voor uitgebreider psychodiagnostisch onderzoek. Gehanteerd op de laatste manier zou men de LDT-E kunnen beschouwen als een instrument dat hypothesen omtrent de oorzaak van (potentiële) leermoeilijkheden oproept¹⁴ welke nadere toetsing behoeven door vervolgonderzoek (tweede of derde fase, etc.).

Evenals bij de samenstelling van de LDT-E en de analyse van het testprofiel kan ook nu het Mark-model weer nuttige diensten bewijzen voor eventueel vervolgonderzoek. In principe kan men hierbij op twee manieren te werk gaan:

a) men selecteert tests met variërende functiecomplexiteit voor die leerkanalen welke nog niet onderzocht zijn. Bijvoorbeeld: het taktiel-fijnmotorische kanaal of het visueel-spraakmotorische kanaal. Uiteraard zijn ook ingewikkelder input-output combinaties te onderscheiden, zoals in het geval van een samengestelde input (auditief en visueel tegelijkertijd) met een enkelvoudige output (alleen fijnmotorisch of grofmotorisch) en omgekeerd een enkelvoudige input en samengestelde output. Een dergelijk vervolgonderzoek vindt plaats indien de resultaten van het onderzoek

met de LDT-E geen dysfunctie (zwak punt) opleveren in de reeds onderzochte leerkanalen, terwijl desondanks het vermoeden bestaat dat de oorzaak van de (potentiële) leermoeilijkheden gezocht moet worden in een vertraagde of gestoorde cognitieve ontwikkeling van het kind;

b) worden daarentegen wel een of meer dysfuncties aangetroffen binnen een bepaald kanaal dan selecteert men tests voor onderliggende of tussenliggende funktieniveau's, afhankelijk van het niveau waarop dysfuncties geconstateerd wordt. Ter illustratie een voorbeeld uit deel I:

"In de LDT zijn twee tests opgenomen waarin primair het auditieve geheugen een belangrijke rol speelt, namelijk Woordenspan (onmiddellijk geheugen voor woordsekwenties) en Zinnen nazeggen (onmiddellijk geheugen voor zinnen). In een derde test, Plaatjes aanwijzen, wordt naast het auditieve geheugen voor woordsekwenties ook de auditief-visuele matchings- of integratievaardigheid gemeten. Volgens Flower (1968) liggen de volgende hiërarchisch gerangschikte auditieve functies ten grondslag aan de leesvaardigheid: auditieve sensibiliteit, aandacht, discriminatie, geheugen, analyse, synthese en auditief-visuele integratie. Indien nu uit het LDT profiel blijkt dat op het gebied van de auditieve aandacht en/of geheugen en/of auditief-visuele integratie van een dysfunctie gesproken mag worden, dan is de volgende stap om in een vervolgonderzoek door middel van aanvullende (sub)tests de resterende functies te onderzoeken, namelijk auditieve synthese, analyse, discriminatie en sensibiliteit (audiogram). Afhankelijk van het niveau waarop dysfuncties geconstateerd wordt, selecteert men tests voor onderliggende of tussenliggende funktieniveau's. Voor dit doel zijn inmiddels verschillende tests ontwikkeld (Rispen, 1974)" (p. 21 en 22).

De keuze van tests voor vervolgonderzoek vereist een grondige kennis van de testliteratuur op het gebied van leermoeilijkheden. Vaak zijn echter geschikte tests voor een bepaalde leeftijdsgroep niet voorhanden. In een dergelijk geval is men dan gedwongen om improvisatorisch te werk te gaan en eventueel zelf het testmateriaal en de opgaven samen te stellen. Bestaat geen eenvoudige opgave, maar gegeven de stand van zaken de moeite van het proberen waard.

4. TYPERING VAN HET ONDERZOEK

4.1 Inleiding

Onder de validiteit van een test verstaat Drenth (1975; p. 236) de mate waarin deze aan zijn doel beantwoordt. In deel I schreven wij reeds welk doel wij met de LDT voor ogen hadden:

"Hij beoogt een bijdrage te leveren aan het individueel psychologisch onderzoek naar de cognitieve ontwikkeling van kinderen in de leeftijd van 4 tot 8 jaar. Deze bijdrage kan variëren van een maat voor de intelligentie tot meer gedetailleerde informatie over specifieke cognitieve vaardigheden van het kind. (...) Op basis van deze informatie kan dan bijvoorbeeld een onderwijsprogramma gekozen of ontwikkeld worden dat beter aansluit bij de mogelijkheden van het kind." (p. 11).

In deel I en II hebben wij vervolgens een begin gemaakt met het uitgebreide valideringsproces dat zo kenmerkend is voor de ontwikkeling van een psychologische test. In dit deel willen wij deze lijn nu verder doortrekken door middel van een betekenisanalyse, inclusief prediktieve validering van de LDT-E.

In de *betekenisanalyse* van een test gaat het volgens Drenth (1975; p. 282) om het vinden en konfirmeren¹⁵ van een theorie of theoretisch begrip als verklaring voor het testgedrag. Op basis van het voorafgaande zouden wij kunnen konkluderen dat de theorie, i.c. het Mark-model als summier afspiegeling van Mark's theorie, reeds gevonden is en dat slechts de toetsing van de hypothetische elementen uit dit model nog rest. Afgezien van de vraag of sommige elementen zich wel laten toetsen - men denke aan de toetsing van één afzonderlijke cognitieve functie in een test die meer dan een functie tegelijk meet - lijkt dit een wat voorbarige konklusie.

In de eerste plaats omdat wij het Mark-model als een heuristisch model hebben gehanteerd dat weliswaar redelijke aanwijzingen verschaft voor de konstruktie van een test, maar onvoldoende suggesties oplevert omtrent zijn betekenis; althans vanuit een traditioneel, sterk faktoranalytisch gekleurd, testpsychologisch standpunt bezien.

In de tweede plaats omdat de eerste fase van het betekenisanalytisch proces - het vinden van een theorie of theoretisch begrip als verklaring voor het LDT-E testgedrag - nog niet is afgesloten, ondanks het reeds verrichte onderzoek. Op een aantal punten zouden wij graag wat méér willen weten over de betekenis van de LDT-E, alvorens tot gerichte toetsing van de theorie (of onderdelen daarvan) over te gaan.

In de derde plaats omdat het de onderzoeker vrij staat beide fasen van het betekenisanalytisch proces - het vinden en konfirmeren van een theorie etc. - gedeeltelijk náást elkaar af te wikkelen in plaats van ná elkaar.

Om genoemde redenen besluiten wij dan ook de betekenis van de LDT-E, zowel formeel als inhoudelijk, verder te exploreren en het toetsingsonderzoek tot een latere datum te verschuiven. Ten behoeve van dit exploratieve onderzoek zullen wij onder andere gebruik maken van gegevens welke verkregen zijn met behulp van de Nijmeegse Schoolbekwaamheidstest (NST) van Mönks et al. (1969) en de Frankforter Schoolrijpheidstest (FST) van Gerritsen en Hulsman (1974). Beide tests worden in paragraaf 4.2 kort beschreven.

Een van de belangrijkste aanknopingspunten voor de vraag wat een test meet, ligt in het onderzoek van wat de test kan voorspellen (Drenth, 1975; p. 293). Dit onderzoek, aangeduid met de term *prediktieve validering*, betreft in ons geval het aanvoeren van enig bewijs voor de relevantie van de door de LDT-E gemeten begrippen voor de vroegtijdige opsporing van leermoeilijkheden.

Zoals we in paragraaf 1.2 gezien hebben, leidde een speurtocht langs verschillende definities niet tot een bevredigende definitie van de term 'leermoeilijkheden' en daardoor evenmin tot een oplossing van ons criteriumprobleem. Om die reden besloten wij een andere oplossing te zoeken, welke simpelweg hierop neerkomt dat wij de lees-, reken-, taal- en schrijfprestaties van een kind op de gewone lagere school (GLO) als maatstaf zullen nemen. De prestaties van een schoolverlatende kleuter op de LDT-E, op basis waarvan de kans groot geacht mag worden dat zij tot onvoldoende prestaties met betrekking tot lezen, rekenen, taal en/of schrijven op het GLO aanleiding geven, vormen dan een indikatie voor het optreden van (latere) leermoeilijkheden.

Twee dingen vallen in deze formulering op. In de eerste plaats het kansaspect. Met behulp van geen enkele test, hoe betrouwbaar ook, zal men met 100% zekerheid een voorspelling kunnen doen over de toekomstige schoolprestaties van een kind, al is het maar omdat zich tussen het tijdstip van de testafname en de feitelijke schoolprestaties nog allerlei ontwikkelingen kunnen voordoen in het cognitief functioneren van het kind. Op basis van een eenmaal vastgestelde empirische relatie tussen testgedrag en latere schoolprestaties, is het echter wel mogelijk een kansuitspraak te doen. De hoogte van de kans op sukses of falen bij het leren

lezen, rekenen, schrijven etc. is dan afhankelijk van de eerder vastgestelde relatie.

Het tweede aspekt dat opvalt is de zinsnede 'onvoldoende prestaties etc.'. Wat als onvoldoende beoordeeld wordt is sterk afhankelijk van de persoonlijke opvattingen van de leerkracht en/of de onderzoeker, evenals van het onderwijsprogramma waarmee het kind geconfronteerd wordt. Het laatste wil nogal eens van school tot school verschillen. Ook in de beoordeling van wat voldoende of onvoldoende schoolprestaties zijn, zit dus een kansaspect verborgen, zoals wij al eerder (cf. par. 1.2) hebben mogen konstateren aan de hand van enkele definities. De meest redelijke benadering van het probleem 'leermoeilijkheden' is onzes inziens dan ook die waarbij men uitgaat van de schoolprestaties van het kind en het vervolgens aan de onderzoeker en/of leerkracht overlaat om te bepalen wat 'onvoldoende schoolprestaties' zijn.

Op basis van het voorafgaande kunnen wij nu de prediktieve validering van de LDT-E nader preciseren als het onderzoek naar de vraag hoe goed de tests van de LDT-E - afzonderlijk of in wisselende combinaties met elkaar - de latere schoolprestaties van een schoolverlatende kleuter voorspellen en daarmee ook het eventuele optreden van leermoeilijkheden.

Zowel op de vraag wat de LDT-E nu precies meet als op de vraag hoe goed de LDT-E latere schoolprestaties kan voorspellen, hopen wij in het verdere verslag van dit onderzoek een antwoord te kunnen geven door de gegevens te presenteren van een groep van 411 schoolverlatende kleuters uit 's-Hertogenbosch en Rosmalen. Deze groep is in februari/maart 1973 onderzocht met de LDT-E, terwijl ongeveer tegelijkertijd ook de reeds eerdergenoemde NST en FST zijn afgenomen. Genoemde groep kleuters is daarna gevolgd tot en met het eind van de eerste klas lagere school, waarbij tussentijds de prestaties gemeten werden door middel van schoolvorderingentoetsen en een vragenlijst. In paragraaf 4.3 en 4.4 wordt van de verdere opzet en uitwerking van dit onderzoek verslag gedaan.

Tot slot van deze paragraaf willen wij nog de vraag stellen naar de geldigheid van onze bevindingen met de LDT-E voor de LDT. De sleutel voor het antwoord hierop wordt gegeven in paragraaf 3.3 waarin de overeenkomsten en verschillen tussen de LDT-E en LDT besproken werden. Zuiver en alleen gelet op de verschillen tussen beide versies mogen wij stellen dat deze niet zó ingrijpend van aard zijn dat zij een voortzetting van de betekenisanalyse van de LDT op basis van LDT-E gegevens in de weg

staan. Helaas verhinderen de verschillen tussen beide versies wèl de toepassing van het LDT Stroomdiagram op de LDT-E gegevens en de opstelling van verwachtingstabellen. Prediktieve validiteitsgegevens zullen daarom voorsnog beperkt moeten blijven tot een weergave van de gevonden samenhangen tussen het testgedrag op de LDT-E en latere schoolprestaties. Ofschoon deze samenhangen zeker indikatief zijn voor de samenhangen die men op basis van LDT gegevens mag verwachten, willen wij nog wel enige terughoudendheid betrachten waar het de generalisatie van LDT-E naar LDT gegevens betreft. Als zodanig is het kenmerk 'exploratief' volledig van toepassing op deze studie. Als enige uitzondering hierop mogen echter de resultaten van Woordenspan en Plaatjes aanwijzen beschouwd worden, aangezien beide tests ongewijzigd opgenomen zijn in de LDT.

4.2 Frankforter Schoolrijpheidstest en Nijmeegse Schoolbekwaamheidstest

In paragraaf 3.4 merkten wij op dat onderzoek gericht op de vroegtijdige opsporing van leermoeilijkheden in verschillende fasen wordt verricht, waarbij de eerste fase aangeduid wordt met signalering. Gezien de grote aantallen kinderen die onderzocht moeten worden, maakt men voor de signaleringsfase bij voorkeur gebruik van groepsgewijs af te nemen psychologische tests, welke de belangrijkste aspecten meten van het gedrag dat men op het oog heeft, i.c. die aspecten van de cognitieve ontwikkeling welke van belang zijn voor het leersukses in de eerste klas lagere school. Dergelijke tests duidt men ook wel aan met schoolrijpheids- of schoolbekwaamheidstests, daarmee tot uitdrukking brengend dat de kleuter over een bepaalde mate van rijpheid, c.q. bekwaamheid moet beschikken alvorens hij met sukses het onderwijs op de lagere school kan volgen. Ofschoon over de termen 'schoolrijpheid' en 'schoolbekwaamheid' nogal wat discussie mogelijk is - onder andere veroorzaakt door het feit dat niet zozeer de 'rijpheid' of 'bekwaamheid' van het kind gemeten wordt als wel zijn intelligentie (Krapp & Mandl, 1971) - voldoen deze tests redelijk in een signaleringsonderzoek dat volgens bepaalde regels verricht wordt (cf. Van Alphen de Veer et al., 1978). Mede met het oog op het opstellen van deze regels zijn door ons de Frankforter Schoolrijpheidstest (FST) van Gerritsen en Hulsman (1974) en de Nijmeegse Schoolbekwaamheidstest (NST) van Mönks et al. (1969) afgenomen op dezelfde groep van 411 schoolverlatende kleuters als waarop de LDT-E is afgenomen. Uitdrukkelijk schrijven wij 'mede', omdat hierdoor tevens de kans geboden werd om de

betekenis van de LDT-E te onderzoeken in het licht van de gegevens die over de FST en NST bekend zijn. Een aantal van deze gegevens wordt vermeld in een eerdere publikatie (Van Alphen de Veer et al., op.cit.). Wij zullen ons daarom hier beperken tot een beschrijving van die kenmerken van beide tests welke van belang zijn voor een goed begrip van de resultaten in hoofdstuk 6. In tabel 4 wordt een overzicht gegeven van de subtests en de bijbehorende taken in de FST en de NST.

Tabel 4. Overzicht van subtests en bijbehorende taken in de Frankforter Schoolrijpheidstest en de Nijmeegse Schoolbekwaamheidstest

subtests	taken
<i>Frankforter Schoolrijpheidstest</i>	
1. Figuren natekenen	- kopiëren van letters, figuren en onderdelen van plaatjes
2. Hekpatroon voortzetten	- kopiëren en verder afmaken van figuren
3. Ramen tekenen	- tekenen of aangeven van een hoeveelheid, welke kort wordt getoond
<i>Nijmeegse Schoolbekwaamheidstest</i>	
1. Vormwaarneming en onderscheidingsvermogen	- identieke plaatjes aankruisen
2. Fijne motoriek	- kopiëren van onderdelen van plaatjes
3. Begrip voor grootte, hoeveelheid en verhoudingen	- kleinste, middelste, vierde etc. plaatje of figuur aankruisen
4. Scherp waarnemen	- verborgen figuren kleuren
5. Kritische waarneming	- plaatje voltooien
6. Koncentratie	- herkennen en doorstrepen van één bepaalde figuur temidden van andere figuren
7. Geheugen	- herkennen en doorstrepen van plaatjes die in vorige subtests voorkwamen
8. Objektbegrip en situatiebeoordeling	- plaatjes aankruisen die aangeven dat men bepaalde situaties begrijpt
9. Weergeven van een verhaaltje	- plaatjes aankruisen die bij een verhaaltje van pl horen
10. Menstekening	- tekenen van een poppetje

Uit deze tabel kunnen we aflezen dat de FST uit drie subtests is opgebouwd en de NST uit tien. Alle subtests doen in wisselende mate een beroep op het auditief/visueel-fijnmotorische leerkanaal, gezien de aard

van de testtaken en de testinstructie (pl geeft mondelinge, en in geval van de FST ook visuele aanwijzingen aan een groep kinderen om met een potlood iets te tekenen, door te strepen of aan te kruisen in een testboekje). Als zodanig lenen de FST en de NST zich minder goed voor een betekenisanalyse van de LDT-E, waarin het aksent juist ligt op de meting van een aantal cognitieve functies binnen het visueel-fijnmotorische leerkanaal enerzijds en het auditief-spraakmotorische leerkanaal anderzijds. Individueel af te nemen psychologische tests die met zorg op hun kanaal- en functie-eigenschappen geselecteerd zijn, zouden dan ook de voorkeur verdiend hebben, ware het niet dat praktische omstandigheden beletten hieraan uitvoering te geven. Als doorslaggevend argument voor de introductie van de FST en de NST pleit echter dat hun betekenis - vanuit een traditioneel testpsychologisch standpunt bezien - voldoende vaststaat om het inzicht in de betekenis van de LDT-E verder te helpen verduidelijken.

Uit eerder onderzoek (Van Alphen de Veer, op.cit.; p. 72 e.v.) is ons namelijk gebleken dat de NST subtests Vormwaarneming en onderscheidingsvermogen (1), Fijne motoriek (2), Begrip voor grootte, hoeveelheid en verhoudingen (3), Kritische waarneming (5) en Menstekening (10) op een-en-dezelfde faktor laden, aangeduid met *Visueel-motorische differentiatie*. Op dezelfde faktor laden ook de drie subtests van de FST, respectievelijk Figuren natekenen (1), Hekpatroon voortzetten (2) en Ramen tekenen (3). Op een tweede faktor, aangeduid met *Waarneming en verbaal begrip*, laden de resterende NST subtests, te weten Begrip voor grootte, hoeveelheid en verhoudingen (3), Scherp waarnemen (4), Concentratie (6), Geheugen (7), Objektbegrip en situatiebeoordeling (8) en Weergeven van een verhaaltje (9). Gegeven de faktorstructuur van de FST en NST zullen wij dan ook in hoofdstuk 6 nagaan in hoeverre de nog te rapporteren faktorstructuur van de LDT-E verder opgehelderd kan worden na faktoranalyse van de FST, NST en LDT-E gezamenlijk.

4.3 Steekproef

Om gegevens te verzamelen ten behoeve van de betekenisanalyse en predictieve validering van de LDT-E, is deze test - ongeveer tegelijkertijd met de FST en NST - afgenomen op een groep van 411 kinderen. Deze groep vormt een steekproef uit de populatie kinderen van 5 en 6 jaar die gedurende de maanden februari en maart 1973 een kleuterschool bezochten in 's-Hertogenbosch en Rosmalen. Zowel opzet als trekking van de steekproef zijn

tot stand gekomen in samenwerking met de Vrije Universiteit (VU) en het Gewestelijk Onderwijs Advies Centrum 's-Hertogenbosch e.o. (GOAC)¹. Aan de steekproefopzet liggen de volgende overwegingen ten grondslag.

De eerste overweging heeft betrekking op de leeftijd. Alleen schoolverlatende kleuters zouden, gelet op onze bedoelingen (cf. par. 1.3), in de steekproef vertegenwoordigd mogen zijn; dat wil zeggen alleen die kleuters die vóór 1 oktober van het jaar waarin ze voor het eerst leerplichtig zijn, de leeftijd van 6 jaar en 0 maanden (6;00) bereikt hebben. Indien we als peildatum februari 1973 hanteren dan betekent dit dat de te trekken kinderen minimaal 5 jaar en 5 maanden (5;05) moeten zijn. Als maximum-leeftijdsgrens besloten wij voorts 6 jaar en 4 maanden (6;04) te hanteren, aangezien nog oudere kleuters - in het merendeel der gevallen kleuters die het advies 'nog een jaar extra kleuterschool' hadden gekregen - niet meer als representatief voor de schoolverlatende kleuter beschouwd kunnen worden.

De tweede overweging betreft het aantal te trekken schoolverlatende kleuters. In feite wordt hiermee de vraag naar de gewenste nauwkeurigheid van de onderzoeksresultaten gesteld. Een voor ieder bevredigend antwoord valt hierop niet te geven, aangezien overwegingen van theoretische, financiële en organisatorische aard, evenals het subjectieve belang dat men hecht aan een bepaalde mate van nauwkeurigheid, een rol spelen bij het vaststellen van het aantal (Angoff, 1971; p. 558 e.v.). Vanuit een exploratief, psychometrisch standpunt bezien en gelet op de mogelijkheid van een stratifikatie naar leeftijd, geslacht en sociaal-ekonomisch niveau, lijkt echter een aantal van ± 400 schoolverlatende kleuters ruimschoots voldoende.

Een derde overweging bij de steekproefopzet heeft betrekking op de trekingsprocedure. Nadat eenmaal het aantal schoolverlatende kleuters is vastgesteld, lijkt trekking van een aselekte steekproef uit de populatie schoolverlatende kleuters van 's-Hertogenbosch en Rosmalen voor de hand te liggen, indien een zekere representativiteit zou worden nagestreefd. Voor het doel van ons onderzoek is representativiteit echter geen eerste vereiste. Daar komt bij dat een dergelijke trekingsprocedure, uit te voeren door het GOAC, op enkele onoverkomelijke financiële en organisatorische moeilijkheden zou stuiten en bovendien de voordelen van een gezamenlijk onderzoek op een-en-dezelfde groep kleuters door het NIPG en de VU (cf. Groenendaal, 1978) volledig verloren zouden gaan. Om deze redenen besloten wij althans een deel van de benodigde kleuters te rekruteren uit

een eerder getrokken steekproef van de VU (Groenendaal, op.cit.; p. 66 e.v.) en vervolgens het resterende aantal kleuters te trekken volgens onderstaande procedure.

Als uitgangspunt voor de trekking diende een eerder (oktober 1972) door de VU getrokken steekproef van 275 5- en 6-jarige kleuters, afkomstig van een negental kleuterscholen in 's-Hertogenbosch¹⁶. Deze scholen waren getypeerd door het GOAC als volgt:

- scholen met een groot aantal leergedepriiveerde kinderen; dat wil zeggen kinderen die een gebrekkige cognitieve of denkgerichte attitude bezitten,
- gesitueerd in zwak-sociale milieu's (arbeidersbuurt).

Na de keuze van deze scholen door het GOAC is voor de bepaling van het sociaal-ekonomisch niveau van de kleuters een vragenlijst opgesteld met betrekking tot het beroepsniveau van vader en moeder, ontleend aan een klassifikatieschema van beroepsniveau's van het Instituut voor Toegepaste Sociologie (ITS) te Nijmegen (Van Westerlaak et al., 1975)¹⁷. Deze vragenlijst is ter beantwoording voorgelegd aan de leerkrachten van de betrokken kleuters. Met behulp van de bij het schema verstrekte lijst met beroepen konden vervolgens alle beroepen in één van de zes categorieën ondergebracht worden.

Naast gegevens met betrekking tot het beroepsniveau zijn, eveneens met behulp van dezelfde vragenlijst, ten overvloede gegevens verzameld over het opleidingsniveau van vader en moeder volgens de indeling van Rupp (1969). Deze varieert van (1) alleen LO of VGLO of onvoltooide voortgezette opleiding tot (6) volledige universitaire opleiding.

Op basis van de gekombineerde beroeps- en opleidingsgegevens zijn ten slotte de kleuters ingedeeld in één van de drie volgende sociaal-ekonomische niveau (SEN) - groepen:

- SEN-I: hoog; hogere beroepen en middelbare employé's, respectievelijk universitaire opleiding, hogere beroepsopleiding en VWO;
[kleuters van ouders van wie beroeps- en opleidingsniveau in de categorieën 5 en 6 vallen]
- SEN-II: midden; kleine zelfstandigen en lagere employé's, respectievelijk HAVO, middelbare beroepsopleiding en MAVO;
[kleuters van ouders van wie beroeps- en opleidingsniveau in de categorieën 3 en 4 vallen]
- SEN-III: laag; geschoolde en ongeschoolde arbeiders, respectievelijk

lagere beroepsopleiding, lagere school en BLO;
[kleuters van ouders van wie beroeps- en opleidings-
niveau in de categorieën 1 en 2 vallen]

Naar schatting loopt deze indeling parallel aan de bij het CBS gebruikelijke driedeling van aanmerkelijk oudere datum (CBS, 1962). In die gevallen waarbij de oorspronkelijke indeling in zes categorieën verschillen van twee of meer punten opleverde tussen het beroeps- en opleidingsniveau van vader, respectievelijk moeder, is als criterium het opleidingsniveau gekozen, tenzij van een hogere opleiding, gezien het beroep, geen gebruik gemaakt werd; bijvoorbeeld HAVO en als beroep vrachtrijder. Bij verschillen tussen het beroeps- en opleidingsniveau van vader en moeder - eveneens twee punten verschil - zijn de gegevens van moeder als indelingscriterium gehanteerd.

Uit deze door de VU getrokken steekproef van 275 5- en 6-jarigen zijn voor het onderhavige onderzoek 205 schoolverlatende kleuters getrokken, ongeveer evenveel jongens (102) als meisjes (103) en allen behorend tot SEN-III.

Vervolgens is deze steekproef van 205 kinderen aangevuld met 112 schoolverlatende kleuters uit SEN-II en 94 uit SEN-I, in totaal 411, waarbij gelet werd op een zo evenredig mogelijke vertegenwoordiging van jongens en meisjes binnen de stratifikatievariabele sociaal-ekonomisch niveau. Deze trekking verliep als volgt. Eerst is een aantal kleuterscholen in 's-Hertogenbosch en Rosmalen getypeerd door het GOAC als behorend tot SEN-II en SEN-I. Hieruit zijn, aan de hand van lijsten met beroeps- en opleidingsgegevens over de ouders, de benodigde schoolverlatende kleuters getrokken binnen de strata geslacht en sociaal-ekonomisch niveau. Van tevoren was reeds de steekproefomvang voor SEN-I en -II gelijkelijk bepaald op ca. 100 kinderen per SEN-groep, dit in verband met de toepassing van enkele statistische analysemethoden en het doel van ons onderzoek. Volgens een schatting van het GOAC¹⁸ omvat de steekproef ca. 20% van de populatie schoolverlatende kleuters in 's-Hertogenbosch en Rosmalen.

In februari en maart 1973 is de LDT-E afgenomen bij deze groep schoolverlatende kleuters, aangewezen voor onderzoek volgens de hierboven beschreven procedure. Afnamē vond plaats op de scholen zelf in een ruimte welke zoveel mogelijk voldeed aan de eisen van een gestandaardiseerd testonderzoek. Per dag zijn maximaal drie kinderen onderzocht door een proefleid(st)er. In totaal 20 proefleid(st)ers namen aan het onderzoek deel, allen doktoraal-studenten psychologie of pedagogiek van de Vrije

Universiteit.

Ter verzekering van een zo zorgvuldig mogelijke afname zijn alle proefleid(st)ers in het kader van een VU-seminar over 'Neuropsychologie en leerstoornissen' getraind in het verrichten van psychodiagnostisch onderzoek met de LDT-E. Het trainingsprogramma bestond onder andere uit een LDT-E demonstratie met behulp van video, evenals twee verplichte proefonderzoeken op kinderen in de leeftijd van 5 tot 6 jaar, gevolgd door een nabespreking.

De NST is in februari 1973 klassikaal afgenomen op de gehele populatie schoolverlatende kleuters uit de regio 's-Hertogenbosch e.o., als onderdeel van het jaarlijks terugkerende routine-testonderzoek door het GOAC. Aangezien de hier beschreven steekproef deel uitmaakt van deze populatie, konden NST gegevens op eenvoudige wijze verzameld worden.

De FST is in maart 1973 klassikaal afgenomen op de hier beschreven steekproef. Als proefleidsters voor de NST en FST fungeerden de kleuterleidsters van de bij het onderzoek betrokken kleuterscholen. De testinstructie voor de NST werd verzorgd door het GOAC; de instructie voor de FST werd gegeven door een psychologe van de Rijksuniversiteit Leiden¹⁹.

In tabel 5 wordt een overzicht gegeven van de frekwentieverdeling naar leeftijd, geslacht en sociaal-ekonomisch niveau voor het onderzoek met de LDT-E.

Tabel 5. Steekproefverdeling naar leeftijd, sociaal-ekonomisch niveau en geslacht^a

N = 411 leerlingen van kleuterscholen in 's-Hertogenbosch en Rosmalen; februari en maart 1973

leeftijd	SEN - I			SEN - II			SEN - III			totaal		
	j	m	t	j	m	t	j	m	t	j	m	t
5;05 - 5;10	31	19	50	27	36	63	59	52	111	117	107	224
5;11 - 6;04	22	22	44	27	22	49	43	51	94	92	95	187
totaal	53	41	94	54	58	112	102	103	205	209	202	411

^a leeftijd in jaren en maanden; SEN = sociaal-ekonomisch niveau (SEN-I = hoog, SEN-II = midden, SEN-III = laag); j = jongen, m = meisje, t = totaal

Uit deze tabel kan men aflezen dat ongeveer evenveel jongens (209) als meisjes (202) in de steekproef vertegenwoordigd zijn en dat de jongste

leeftijdsgroep (5;05 - 5;10) van 224 kinderen, vergeleken met de 187 kinderen uit de oudste leeftijdsgroep (5;11 - 6;04), naar verhouding iets oververtegenwoordigd is. Het laatste zou men echter ook als een artefakt kunnen beschouwen van de methode om de gehele groep van 411 kinderen in twee leeftijdsgroepen van een halfjaar te verdelen in plaats van bijvoorbeeld de gehele groep te verdelen in twee groepen met de mediaan van de leeftijd in maanden als snijpunt.

Ten overvloede vermelden wij hier nog de landelijke percentages van de verdeling naar sociaal-ekonomisch niveau, afkomstig uit een aselekt steekproefonderzoek van Prinssen en Kropman (1975): SEN-I: 15 (23), SEN-II: 39 (27), SEN-III: 46 (50). Tussen haakjes zijn de percentages voor de verschillende sociaal-ekonomische niveau's uit het onderzoek met de LDT-E vermeld. Uit de verschillende percentages blijkt dat SEN-I oververtegenwoordigd, SEN-II ondervertegenwoordigd en SEN-III enigszins oververtegenwoordigd is in onze steekproef.

Tot slot van deze paragraaf willen wij opmerken dat het niet mogelijk bleek alle kleuters uit de steekproef te onderzoeken met de FST en NST. De redenen hiervoor waren ziekte of verhuizing enerzijds en organisatorische problemen anderzijds. Hieruit volgt dat het totaal aantal onderzochte kinderen met de FST en NST kleiner is dan in tabel 5 wordt aangegeven, respektievelijk $N = 356$ voor de FST en $N = 371$ voor de NST²⁰.

4.4 Schoolprestaties

Met het oog op de prediktieve validering van de LDT-E hebben wij de in paragraaf 4.3 beschreven groep kleuters gevolgd tot en met het eind van de eerste klas gewoon lager onderwijs (GLO). Uitval door ziekte en verhuizing bleek onvermijdelijk. Ook het 'restriction of range'-effect deed zich gelden, aangezien 25 van de 411 schoolverlatende kleuters op basis van hun NST resultaten als niet-schoolrijp beoordeeld werden door het GOAC. Deze groep van 25 kleuters is dan ook na overleg met ouders en leerkrachten niet toegelaten tot de eerste klas GLO. Daarnaast stapte een aantal kinderen over naar een vorm van buitengewoon onderwijs. Het lijkt aannemelijk dat het 'restriction of range'-effect een onderschatting oplevert van de prediktieve validiteit van de LDT-E.

Gedurende het vervolgonderzoek zijn verschillende toetsen en een vragenlijst afgenomen. Te zamen vormen zij het criterium 'schoolprestaties' - opgesplitst in twee groepen van ieder vier variabelen - waartegen de LDT-E gevalideerd wordt.

Voor de meting van de vaardigheden die in de eerste klas GLO bijgebracht worden, zijn wij op zoek gegaan naar zo objectief mogelijke meetinstrumenten, waarvan in ieder geval de kanaaleigenschappen goed zouden aansluiten bij die van de LDT-E. Onzes inziens verdienen dan individueel af te nemen schoolvorderingentoetsen de voorkeur. Om praktische en organisatorische redenen enerzijds en een gebrek aan individuele toetsen voor sommige vaardigheden anderzijds, waren wij echter gedwongen om vrijwel alleen klassikale toetsen af te nemen. Voor zulke belangrijke vaardigheden als taal en schrijven bleken zelfs in het geheel geen toetsen ontwikkeld te zijn, althans volgens het desbetreffende overzicht van het Centraal Instituut voor Toetsontwikkeling (Van Lijf, 1971). Over deze laatste twee onderdelen hebben wij dan ook noodgedwongen het oordeel van de leerkracht moeten vragen; noodgedwongen omdat dit oordeel tamelijk subjectief is (cf. De Groot, 1967). In totaal zijn vier toetsen afgenomen²¹. Iedere toets wordt hieronder kort toegelicht onder vermelding van afnametijdstip en aantal onderzochte kinderen²².

Toetsen

Begrijpend lezen I (Van Calcar et al., 1974a): een klassikale toets welke de vaardigheid van de leerling om de betekenis van een gedrukte tekst te begrijpen, meet. Afnametijdstip: eind april 1974. N = 337.

Beelddiktee I (Van Calcar et al., 1974b): een klassikale toets die de spellingsvaardigheid van de leerling meet. Afnametijdstip: eind april 1974. N = 336.

Brus-Eén-Minuut-Test, vorm B (Brus & Voeten, 1973): een individuele toets welke de snelheid meet waarmee de leerling afzonderlijke, niet samenhangende woorden kan lezen (technische leesvaardigheid). Afnametijdstip: juni 1974. N = 327.

Schiedamse rekentest (Heesen et al., 1969): een klassikale toets welke de rekenvaardigheid meet van de leerling in het maken van eenvoudige optel- en aftreksommen. Afnametijdstip: juni 1974. N = 311.

Om een zo volledig mogelijke indruk te krijgen van de schoolprestaties in de eerste klas GLO is naast toetsinformatie ook nog andere informatie verzameld door middel van een vragenlijst. In deze lijst wordt het oordeel van de leerkracht gevraagd over de lees-, reken-, taal- en schrijfprestaties van de leerling in vergelijking met klasgenootjes²³. Dit oordeel moest worden aangegeven op een vijfpuntsschaal met als schaalwaarden: slecht/zeer slecht (1) - onvoldoende (2) - matig (3) - voldoende (4)

en goed/zeer goed (5). Hieronder worden deze vier oordeelsvariabelen weergegeven, onder vermelding van het aantal beoordeelde kinderen²².

Oordeel leerkracht

Lezen N = 322

Rekenen N = 322

Taal N = 322

Schrijven N = 322

Tabel 6. Moeilijkheidsgraad (P) en discriminatie-index (r_{ir})^a van de items, per test
N = 411

item	Blok- patronen		Vouw- blaadjes		Natikken		Woorden- span		Plaatjes aanwijzen		Zinnen nazeggen		Verhaaltje vragen		Begrip en inzicht	
	P	r_{ir}	P	r_{ir}	P	r_{ir}	P	r_{ir}	P	r_{ir}	P	r_{ir}	P	r_{ir}	P	r_{ir}
1	100	-. ^a	100	-	95	.08	100	-	90	.16	67	.65	22	.16	68	.20
2	100	-	100	-	93	.01	93	.27	74	.40	90	.41	44	.43	82	.10
3	95	.23	99	-	85	.17	95	.24	70	.41	79	.56	48	.51	96	.19
4	87	.20	96	.14	64	.40	95	.33	58	.56	88	.59	65	.49	87	.22
5	96	.38	88	.28	56	.47	93	.40	54	.51	77	.41	74	.31	44	.13
6	85	.33	72	.27	53	.44	63	.59	38	.47	82	.61	13	.28	85	.30
7	95	.43	81	.31	33	.46	53	.61	21	.53	85	.56	63	.49	34	.27
8	84	.36	56	.37	45	.58	52	.57	22	.53	84	.58	36	.30	45	.39
9	32	.45	41	.36	30	.54	44	.60	14	.47	79	.69	83	.34	35	.33
10	45	.44	44	.40	13	.40	3	-	1	-	76	.57	34	.37	46	.51
11			34	.42	10	.35	6	.30	1	-					36	.50
12			17	.36	14	.44	2	-	1	-					60	.37
13					1	-									38	.38
14					1	-									44	.52
15					0	-									18	.28
16					0	-										
17					0	-										

^a discriminatie-indices van items met een P >96 of <4 zijn weggelaten vanwege hun onbetrouwbaar karakter

5. BETROUWBAARHEID

5.1 Itemkenmerken

Voor de interpretatie van de onderzoeksbevindingen, met name in paragraaf 5.2 en deel B, worden in tabel 6 twee 'klassieke' typen itemgegevens verstrekt. Het eerste type heeft betrekking op de moeilijkheidsgraad (P) of P-waarde van de items per test. De berekening van de P-waarden vond plaats aan de hand van een formule waarin de gemiddelde score per item genomen wordt²⁴. Het tweede type gegevens betreft de discriminatie-index (r_{ip}) of item-restkorrelatie van de items per test, berekend volgens Nunnally (1967; p. 262). Mede op basis van beide typen gegevens is de definitieve versie van de LDT tot stand gekomen.

Een evaluatie van de verdeling en hoogte van de P-waarden en discriminatie-indices dient te gebeuren in relatie tot het doel van een test. Aangezien de LDT bestemd is voor het onderzoek naar zowel de specifieke als de algemeen cognitieve ontwikkeling binnen verschillende leeftijdsgroepen (4- tot 8-jarigen), moeten de verschillende items per test variëren in moeilijkheidsgraad. Zo dienen er voldoende 'makkelijke' items voor de jongste leeftijdsgroep en voldoende 'moeilijke' items voor de oudste leeftijdsgroep te zijn. In de konstruktiefase van de LDT hebben wij dan ook primair gestreefd naar de ontwikkeling van zogenaamde 'equidiscriminating' tests voor de leeftijd van vier tot acht jaar, in plaats van 'peaked' tests (Nunnally, op.cit.; p. 237 e.v.), waarin alle P-waarden bij voorkeur rondom de waarde van 50 schommelen.

Onze opzet wordt weerspiegeld in de tamelijk regelmatig aflopende reeks P-waarden van Blokpatronen, Vouwblaadjes, Natikken, Woordenspan, Plaatjes aanwijzen en in mindere mate van Begrip en inzicht (tabel 6). Een uitzondering hierop vormen Zinnen nazeggen en Verhaaltje vragen. Deze 'peaked' tests waren oorspronkelijk bestemd voor de middelste leeftijdsgroepen (5- en 6-jarigen), de groepen waaraan de hier beschreven experimentele versie van de LDT in eerste instantie is afgenomen. Zowel Zinnen nazeggen als Verhaaltje vragen laten daarom wisselende, maar niet al te hoge of lage P-waarden zien.

Ofschoon P-waarden de betrouwbaarheid van een test aanzienlijk kunnen beïnvloeden - met name de extreme waarden oefenen een ongunstige invloed uit - zijn discriminatie-indices aanzienlijk belangrijker dan de moeilijkheidsgraad van de items. Via deze indices wordt namelijk de mate van samenhang uitgedrukt tussen item en testscore, of anders geformuleerd,

de mate waarin item en test hetzelfde psychologische kenmerk meten. Hieruit volgt dat een inspectie van de hoogte van de discriminatie-indices een voorlopige indruk kan verschaffen van de homogeniteit of interne consistentie van een test (Ebel, 1965; p. 346 e.v.).

Voor lang niet alle tests zijn echter bevredigende discriminatie-indices verkregen, uitgaande van een minimumwaarde voor de item-restkorrelatie van .30. Ook de verdeling van de indices per test voldoet niet in alle opzichten. Hier wreekt zich het feit dat hoofdzakelijk 'equidiscriminatie' tests, bestemd voor de leeftijdsgroep van vier tot acht jaar, afgenomen zijn bij 5- en 6-jarigen, waarvoor de eerste en de laatste items achtereenvolgens te makkelijk en te moeilijk waren. Hiermee wordt overigens fraai geïllustreerd wat de invloed van extreme P-waarden is op de hoogte van de item-restkorrelaties en daarmee ook op de betrouwbaarheid van een test (Lienert, 1961; p. 122), zoals we in paragraaf 5.2 nog zullen zien.

Na deze inleidende opmerkingen kunnen wij tenslotte de moeilijkheidsgraad en de discriminatie-index van de items per test bespreken.

Blokpatronen. Deze test is in zijn geheel wat aan de makkelijke kant voor 5- en 6-jarigen. Alleen de laatste twee items (9 en 10) laten een vrij sterke stijging in moeilijkheidsgraad zien. De lage discriminatie-indices van het derde en vierde item vinden waarschijnlijk mede hun oorzaak in de hoge P-waarden.

Vouwblaadjes. Voor de eerste zes items worden lage of zelfs helemaal geen discriminatie-indices gevonden, grotendeels te wijten aan de zeer hoge P-waarden van deze items. De door ons toegevoegde items 6 t/m 8, 10 en 11 aan de oorspronkelijke subtest Paper folding uit de Hiskey-Nebraska Test of Learning Aptitude (HNTLA) van Hiskey (1966) zijn blijkens hun P-waarden en discriminatie-indices een gunstige keuze geweest. Desondanks vertonen slechts de laatste zes items discriminatie-indices $>.30$.

Natikken. P-waarden en discriminatie-indices laten over het geheel genomen een bevredigend beeld zien.

Woordenspan. De verdeling van de P-waarden mag gunstig genoemd worden: makkelijke items aan het begin en moeilijke items aan het eind. De discriminatie-indices zijn zeker voor de middelste groep items ruimschoots bevredigend.

Plaatjes aanwijzen. Uit de hoogte van de P-waarden - in hun geheel wat lager dan bij Woordenspan - valt af te lezen dat de antwoorden op deze test afhankelijk zijn van een goede of foute beantwoording op Woorden-

span¹². De lagere P-waarden oefenen echter een gunstige invloed uit op de discriminatie-indices, met name voor de eerste vier items, zodat Plaatjes aanwijzen in vergelijking tot Woordenspan alleen maar aan homogeniteit kan winnen.

Zinnen nazeggen. Vrij konstante P-waarden over de gehele linie. Mogelijk zijn de items wat aan de makkelijke kant. De discriminatie-indices zijn zonder uitzondering zeer bevredigend.

Verhaaltje vragen. P-waarden en discriminatie-indices laten een iets ongunstiger beeld zien dan bij Zinnen nazeggen, alhoewel slechts twee indices (item 1 en 6) onder de .30 dalen.

Begrip en inzicht. Voor een 'equidiscriminating' test laat de verdeling van de P-waarden een tamelijk onregelmatig verloop zien. Daarnaast blijven de discriminatie-indices bij niet minder dan zeven items onder de .30.

5.2 Interne consistentie en homogeniteit

Om een indruk te krijgen van de betrouwbaarheid van de LDT-E zijn per test twee typen koëfficiënten berekend, i.c. de koëfficiënt van interne consistentie en de homogeniteitskoëfficiënt. Op basis van de verkregen koëfficiënten is vervolgens per type koëfficiënt de betrouwbaarheid van de som van alle acht tests te zamen geschat, aangeduid met 'totaal'. De hiervoor benodigde formule, afkomstig van Mosier, wordt gegeven door Lienert (1961; p. 371). In tabel 7 worden de resultaten weergegeven, evenals de maximumscore, het gemiddelde en de standaarddeviatie per test.

Zowel de koëfficiënt van interne consistentie als de homogeniteitskoëfficiënt zijn klassieke betrouwbaarheidsmaten, ieder met hun eigen specifieke toepassingsgebied en theoretische achtergrond, welke onder anderen door Drenth (1975; p. 186 e.v.) en Nunnally (1967; p. 172 e.v.) uitvoerig beschreven worden. Wij volstaan daarom met een korte toelichting op beide koëfficiënten per test in het licht van de reeds beschreven P-waarden en discriminatie-indices en onder verwijzing naar de overige gegevens in tabel 7. Een uitvoerige toelichting op deze maten lijkt eens te meer overbodig aangezien de moderne generaliseerbaarheidstheorie van Cronbach et al. (1972) langzamerhand de plaats inneemt van de klassieke betrouwbaarheidstheorie.

Uitgaande van Drenth's indeling (op.cit.; p. 195) in acht factoren die de testvariantie bepalen, willen wij echter nog wèl opmerken dat in

Tabel 7. Maximumscore^a, gemiddelde (m), standaarddeviatie (s), koëfficiënt van interne consistentie (r_{11})^b en homogeniteitskoëfficiënt (r_{kk}) van de tests

tests	max.- ^a score	m	s	r_{11} ^b	r_{kk}
Blokpatronen	20	16.38	2.42	.70	.63
Vouwblaadjes	12	8.27	1.95	.66	.63
Natikken	17	5.94	2.52	.83	.73
Woordenspan	12	6.99	2.04	.80	.75
Plaatjes aanwijzen	12	4.44	2.34	.81	.75
Zinnen nazeggen	55	44.39	8.79	.87	.85
Verhaaltje vragen	35	15.54	6.88	.74	.69
Begrip en inzicht	30	16.35	4.43	.73	.71
totaal ^c	-	-	-	.93	.92

^a theoretische spreidingsbreedte per test: 0 - max.score

^b gekorrigeerd voor testhalvering met behulp van de Spearman-Brown formule

^c geschat volgens Mosier, in Lienert (1961; p. 371)

beide koëfficiënten de niet-specifieke variantie de bedoelde variantie vormt en al het specifieke de niet-bedoelde variantie. Met een variant op de woorden van Drenth (op.cit.) betekent dit dat deze koëfficiënten: "een *overschatting* geven van de betrouwbaarheid voor zover de algemene, maar toevallige variantie bijdraagt tot de korrelatiehoogte, evenals dat het geval is voor de mate waarin beide testhelften (koëfficiënt van interne consistentie) of de afzonderlijke items (homogeniteitskoëfficiënt) toch niet geheel onafhankelijk van elkaar zijn en testhelften of items aan dezelfde 'specifieke' fouten mank gaan. Aan de andere kant geven deze koëfficiënten in de eliminatie van alle specifieke factoren weer een *onderschatting* van de betrouwbaarheid voor zover kenmerken of eigenschappen van de persoon zelf, die typisch aan het specifieke instrument gebonden zijn, wel worden geëlimineerd" (p. 207).

Dit citaat - dat overigens niet geheel letterlijk is - heeft in feite betrekking op de koëfficiënt van α -specificiteit, ontwikkeld voor de parallelvormmethode²⁵. Bij het konkluderen tot betrouwbaarheid op basis van de koëfficiënt van interne consistentie zal men dan ook voorzichtig moeten zijn. Dit geldt eens te meer voor de homogeniteitskoëfficiënt. Zo wil een lage homogeniteitskoëfficiënt niet per se zeggen dat de test onbetrouwbaar is; het is immers best mogelijk dat de items op betrouwbare wijze niet één, maar verschillende functies (of factoren) meten,

zodat de test alleen als minder homogeen beschouwd moet worden. Door dit wisselende perspectief op de betrouwbaarheid van een test is dan ook onze keuze voor twee betrouwbaarheidscoëfficiënten bepaald in plaats van één.

Blokpatronen. In het licht van de hoge P-waarden en het naar verhouding geringe aantal items met bevredigende item-restkorrelaties (tabel 6) zou men de betrouwbaarheid van Blokpatronen, gemeten naar de coëfficiënt van interne consistentie (.70), nog tamelijk hoog kunnen noemen (tabel 7). Vanuit een diagnostisch standpunt gezien is deze echter aan de lage kant, zeker indien men op basis van deze test alléén een indruk wil krijgen van het specifieke funktieniveau²⁶. Uit het tamelijk grote verschil van zeven punten tussen de coëfficiënt van interne consistentie (.70) en de homogeniteitscoëfficiënt (.63) zou men kunnen afleiden dat Blokpatronen niet zozeer onbetrouwbaar is als wel een aantal verschillende functies in principe op betrouwbare wijze meet. Een dergelijke voorstelling van zaken lijkt in overeenstemming met onze eerder verwoorde inhoudsanalyse van de functies die door Blokpatronen gemeten zouden worden binnen het visueel-fijnmotorische kanaal (cf. par. 3.2). Het hoge gemiddelde ($m = 16.38$; $\text{max.score} = 20$) en de geringe spreiding ($s = 2.42$) bevestigen en verklaren voorts de gepresenteerde bevindingen.

In het algemeen dient men bij deze versie van Blokpatronen, welke identiek is met de subtest Block Design uit de WPPSI (Wechsler, 1967), ernstig rekening te houden met het zgn. 'plafond'-effekt; dat wil zeggen de test differentieert onvoldoende binnen de betrokken leeftijdsgroep, aangezien de items kennelijk te makkelijk zijn. Dit is dan ook een van de redenen waarom aan de definitieve versie van Blokpatronen uit de LDT moeilijker items zijn toegevoegd (cf. par. 3.3). Hiermee werd tevens de betrouwbaarheid (interne consistentie) van Blokpatronen tot een zeer acceptabel niveau opgevoerd, namelijk .80 voor leeftijdsgroep IV (5;06 - 5;11) en .90 voor leeftijdsgroep V (6;00 - 6;05)²⁷, cf. deel I; p. 42.

Vouwblaadjes. Zowel de coëfficiënt van interne consistentie als de homogeniteitscoëfficiënt laten lage waarden zien, respectievelijk .66 en .63. Niet alleen is deze test minder betrouwbaar, maar ook weinig homogeen, zoals de P-waarden en discriminatie-indices (tabel 6) reeds deden vermoeden. De geringe spreiding van de items ($s = 1.95$) lijkt ons hieraan mede debet.

Gelet op het hoge groepsgemiddelde van de betrokken leeftijdsgroep ($m = 8.27$; $\text{max.score} = 12$) is uitbreiding van Vouwblaadjes met een aantal

moeilijker items noodzakelijk, zeker met het oog op 7-jarigen. In de definitieve versie van Vouwblaadjes uit de LDT is deze uitbreiding dan ook gerealiseerd (cf. par. 3.3). Desondanks is de betrouwbaarheid (interne consistentie) van Vouwblaadjes niet sterk toegenomen, namelijk .71 voor leeftijdsgroep IV (5;06 - 5;11) en .72 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

Natikken. De koëfficiënt van interne consistentie (.83) mag bevredigend genoemd worden. Minder bevredigend, ofschoon nog ruimschoots voldoende, is de homogeniteitskoëfficiënt met een waarde van .73. Het grote verschil van tien punten tussen beide koëfficiënten lijkt erop te wijzen dat niet alle items dezelfde functie(s) meten binnen het visueel-fijn-motorische leerkanaal; dit in tegenstelling tot onze eerder geuite verwachtingen omtrent het geringe aantal functies van deze test (cf. par. 3.2). Mogelijk doen de wat moeilijker items een beroep op nieuwe of iets anderssoortige functies dan de makkelijker items.

Gezien het relatief lage gemiddelde op Natikken ($m = 5.94$; $\text{max.score} = 17$) lijkt uitbreiding van deze test met wat makkelijker items voor de jongste leeftijdsgroep (4-jarigen) aan te bevelen. De spreiding ($s = 2.52$) is verder redelijk. Nauwelijks gemodificeerd - alleen de moeilijkste vijf items zijn weggelaten (cf. par. 3.3) - laat de betrouwbaarheid (interne consistentie) van Natikken in zijn definitieve versie iets lagere waarden zien, namelijk .75 voor leeftijdsgroep IV (5;06 - 5;11) en .80 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

Woordenspan. Deze test laat qua betrouwbaarheid ongeveer hetzelfde beeld zien als Natikken, zij het dat de koëfficiënt van interne consistentie (.80) drie punten lager en de homogeniteitskoëfficiënt (.75) twee punten hoger is. Gelet op de P-waarden, item-restkorrelaties, het gemiddelde ($m = 6.99$; $\text{max.score} = 12$) en de spreiding ($s = 2.04$) enerzijds en de resultaten bij jongere en oudere leeftijdsgroepen²⁸ anderzijds, is de hier beschreven versie van Woordenspan ongewijzigd opgenomen in de definitieve versie van de LDT.

De betrouwbaarheid (interne consistentie) van Woordenspan in de definitieve versie is .82 voor leeftijdsgroep IV (5;06 - 5;11) en .76 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

Plaatjes aanwijzen. Zowel de koëfficiënt van interne consistentie (.81) als de homogeniteitskoëfficiënt (.75) zijn bevredigend te noemen. Zoals aan het verloop en de hoogte van de P-waarden reeds te zien viel (cf. tabel 6), is het gemiddelde van Plaatjes aanwijzen ($m = 4.44$; $\text{max.score} = 12$) wat lager dan bij Woordenspan. De spreiding van Plaatjes aanwijzen

($s = 2.34$) is daarentegen iets hoger. Hetzelfde geldt voor de discriminatie-indices van deze test, welke over het geheel genomen iets hoger uitvallen dan bij Woordenspan. Het 'drukkende' effect van het wat lagere gemiddelde op de homogeniteit van Plaatjes aanwijzen enerzijds en het 'verhogende' effect van de iets grotere spreiding en de gemiddeld iets hogere discriminatie-indices anderzijds, kan mogelijk een verklaring bieden voor de bevinding dat de homogeniteit van Woordenspan en Plaatjes aanwijzen identiek is.

Evenals Woordenspan is Plaatjes aanwijzen ongewijzigd opgenomen in de definitieve versie van de LDT. De betrouwbaarheid (interne consistentie) van Plaatjes aanwijzen in de definitieve versie is .76 voor leeftijdsgroep IV (5;06 - 5;11) en .86 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

Zinnen nazeggen. De waarden van de koëfficiënt van interne consistentie (.87) en de homogeniteitskoëfficiënt (.85) zijn volledig in overeenstemming met de P-waarden, de discriminatie-indices, het gemiddelde ($m = 44.39$; max.score = 55) en de spreiding ($s = 8.79$) van deze test. Zinnen nazeggen in deze versie beschouwen wij dan ook als een zeer betrouwbare en homogene test voor de betrokken leeftijdsgroep.

Voornamelijk met het oog op de 4-jarigen is een tweetal items (zinnen) aan de definitieve versie van Zinnen nazeggen toegevoegd. Daarnaast is ook de scoring aanzienlijk vereenvoudigd (cf. par. 3.3). Het resultaat hiervan is een nog hogere betrouwbaarheid (interne consistentie) van Zinnen nazeggen in zijn definitieve versie, namelijk .92 voor leeftijdsgroep IV (5;06 - 5;11) en .91 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

Verhaaltje vragen. Gelet op de P-waarden en discriminatie-indices (tabel 6) enerzijds en het gemiddelde ($m = 15.54$; max.score = 35) en de spreiding ($s = 6.88$) anderzijds, zijn de koëfficiënt van interne consistentie (.74) en de homogeniteitskoëfficiënt (.69) lager dan verwacht. Vanuit een diagnostisch standpunt gezien mag de betrouwbaarheid van Verhaaltje vragen in deze versie als minder bevredigend gekenschetst worden. In de definitieve versie van Verhaaltje vragen zijn dan ook enkele wijzigingen doorgevoerd met betrekking tot de items, de formulering van de vragen en de scoring van de antwoorden (cf. par. 3.3).

Een en ander heeft geresulteerd in een verhoging van de betrouwbaarheid (interne consistentie) van Verhaaltje vragen tot .79 voor leeftijdsgroep IV (5;06 - 5;11) en .80 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

Begrip en inzicht. De koëfficiënt van interne consistentie (.73) is in redelijke overeenstemming met de homogeniteitskoëfficiënt (.71). Beide koëfficiënten zijn echter aan de lage kant, zoals de geringe hoogte van een aantal discriminatie-indices (tabel 6) reeds deed vermoeden. Het gemiddelde ($m = 16.35$; $\text{max.score} = 30$) geeft verder geen aanleiding tot opmerkingen, in tegenstelling tot de spreiding ($s = 4.43$), welke aan de krappe kant is.

Ook bij Begrip en inzicht zijn in de definitieve versie enkele wijzigingen doorgevoerd met betrekking tot de items en de scoring van de antwoorden (cf. par. 3.3). Het resultaat hiervan is een verhoging van de betrouwbaarheid (interne consistentie) tot .82 voor leeftijdsgroep IV (5;06 - 5;11) en .83 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

*Totaal*¹⁰. Voor de berekening van het 'totaal' zijn eerst de ruwe scores van de tests uit de LDT-E omgezet in standaardscores (z-scores). De som van deze standaardscores vormt dan het LDT-E totaal. De betrouwbaarheid van dit totaal, tot uitdrukking komend in de koëfficiënt van interne consistentie (.93) en de homogeniteitskoëfficiënt (.92) mogen wij als zeer bevredigend kenschetsen.

Dit geldt eens te meer omdat een met het 'totaal' redelijk vergelijkbare maat als het LDT deviatie IQ nauwelijks hogere waarden voor zijn betrouwbaarheid (interne consistentie) laat zien, namelijk .94 voor leeftijdsgroep IV (5;06 - 5;11) en .95 voor leeftijdsgroep V (6;00 - 6;05), cf. deel I; p. 42.

6. STRUKTUUR

6.1 Invloed van leeftijd, geslacht en sociaal-ekonomisch niveau

Om uiteenlopende redenen is het van belang om te weten hoe testprestaties variëren in relatie tot de leeftijd, het geslacht en het sociaal-ekonomisch niveau van de onderzochte. Bij een variabele als leeftijd is dit belang vrij duidelijk: een enigszins betekenisvolle samenhang tussen leeftijd en testprestatie duidt erop dat men met een ontwikkelingsgevoelig aspekt van het cognitief functioneren te maken heeft en dat is op zich weer belangrijk in verband met de konstatering of er al dan niet sprake is van een ontwikkelingsachterstand. Om over dat laatste een uitspraak te kunnen doen maakt men gebruik van leeftijdsnormen, waarbij de ruwe scores in zgn. 'leeftijdequivalenten' worden omgezet, overigens met alle nadelen van dien (Angoff, 1971; p. 521 e.v.; Drenth, 1975; p. 174 e.v.)²⁹.

Ten aanzien van de variabelen geslacht en sociaal-ekonomisch niveau liggen de zaken minder duidelijk. Als bekend is dat meisjes van een bepaalde leeftijd gemiddeld betere testprestaties leveren dan jongens van dezelfde leeftijd, moeten er dan leeftijdsnormen ontwikkeld worden voor jongens en meisjes apart? Men zou denken van wel indien men de handleiding van de Nederlandse WAIS (Stinissen et al., 1969) doorneemt en tot de ontdekking komt dat aparte leeftijdsnormen voor mannen en vrouwen ontwikkeld zijn.

Hetzelfde, maar dan voor het sociaal-ekonomisch niveau, kan men constateren bij inspectie van de normtabellen in de UTANT (Kohnstamm et al., 1971). Een verantwoording van deze handelwijze zal men in beide handleidingen tevergeefs zoeken, tenzij men afgaat op het argument dat geslacht en met name sociaal-ekonomisch niveau (SEN) van invloed zijn op de testprestaties (cf. Kohnstamm, 1968). Een gelijkloidend argument kan men overigens voor nog een aantal andere variabelen aanvoeren, variërend van urbanisatiegraad tot het lidmaatschap van culturele, religieuze of ethnische groeperingen.

Voor de LDT hebben wij besloten geen leeftijdsnormen naar geslacht of sociaal-ekonomisch niveau te ontwikkelen; niet omdat wij geen verschillen in gemiddelde testprestaties naar geslacht of sociaal-ekonomisch niveau konden constateren, maar omdat onze bedoelingen bij de konstruktie van normen onder meer betrekking hadden op de (prediktieve) relatie

cognitief ontwikkelingsniveau - schoolprestaties. Wij gingen er hierbij vanuit dat de leerstof in het Nederlandse schoolsysteem niet verschilt naar geslacht of sociaal-ekonomisch niveau; meisjes krijgen bijvoorbeeld geen andere sommen dan jongens. Dit wil echter nog niet zeggen dat er geen verschillen in gemiddelde leerprestaties zouden bestaan tussen jongens en meisjes of tussen hogere en lagere sociaal-ekonomische niveau-groepen.

Juist dit laatste fenomeen van verschillen in gemiddelde leerprestaties tussen onderscheiden groepen wordt nogal eens impliciet als argument gehanteerd om dan ook maar de verschillen in testprestaties - die zoals bekend samenhangen met leerprestaties - als het ware weg te werken door gedifferentieerde normen op te stellen. Het resultaat hiervan is desastreus indien men niet over per groep berekende voorspellingsformules beschikt, bijvoorbeeld in de vorm van regressievergelijkingen. Voor de groep met de hoogste testprestaties vindt dan namelijk een onderschatting plaats van hun schoolprestaties en voor de groep met de laagste testprestaties vindt omgekeerd een overschatting plaats. Bedoelde regressie-vergelijkingen worden echter zelden of nooit opgesteld, gezien het vele werk dat daaraan verbonden is. Onze konklusie moet dan ook luiden dat het konstrueren van normen anders dan voor leeftijdsgroepen in de praktijk veelal onjuiste verwachtingen wekt omtrent de leer- of schoolprestaties van de onderzochte kinderen³⁰.

Om andere redenen dan bovengenoemde is het onderzoek naar de invloed van leeftijd, geslacht en sociaal-ekonomisch niveau weer wèl van belang. De eerste reden heeft betrekking op de verzameling van indirecte evidentie voor de betrouwbaarheid van een test. Variërend op Kalverboer en Sanders-Allersma (1973; p. 76): als er inzichtelijke en herhaalbare samenhangen gevonden worden tot een aantal criteria als leeftijd, geslacht en sociaal-ekonomisch niveau dan suggereren dergelijke bevindingen iets over de stabiliteit van een test. De tweede reden betreft het mogelijke leeftijds-, geslachts- en/of milieugebonden karakter van de verzameling cognitieve functies - zoals gemeten door de LDT-E - waardoor een nader licht geworpen kan worden op de betekenis van de tests afzonderlijk en in hun totaliteit voor de onderscheiden groepen.

Met andere woorden: de vraag is in hoeverre met de leeftijd, het geslacht of het sociaal-ekonomisch niveau van de onderzochte ook de betekenis van de tests uit de LDT-E verandert. Voor de beantwoording van deze vraag staan verschillende wegen open, waarvan de variantie-analytische en faktoranalytische weg wel het meest in het oog springend zijn. Wij beginnen

met de variantie-analytische benadering van het gestelde probleem om daarna - in paragraaf 6.4 - ook de faktoranalytische benadering te volgen.. Om de invloed na te gaan van leeftijd, geslacht en sociaal-ekonomisch niveau op de testprestaties van schoolverlatende kleuters, is per test en over het totaal van de LDT-E een drieweg variantie-analyse uitgevoerd volgens een $2 \times 2 \times 3$ faktoriële opzet (Kirk, 1968; p. 171 e.v.)³¹. Hier toe zijn de kleuters verdeeld in de volgende groepen. Faktor 1: Jongste leeftijdsgroep (5;05 - 5;10) vs. Oudste leeftijdsgroep (5;11 - 6;04), fixed; Faktor 2: Jongens vs. Meisjes, fixed; Faktor 3: Sociaal-ekonomisch niveau, drie niveau's (SEN-I, SEN-II, SEN-III), fixed.

Gezien het grote aantal variabelen en F-toetsen dat in deze analyse betrokken is, is de kans op minimaal één significant resultaat bij afzonderlijke toetsing niet onaanzienlijk, althans vergeleken bij het gekozen significantieniveau per afzonderlijke toets. Ter oplossing van dit probleem dat in de waarschijnlijkheidsleer de naam van 'Bonferroni-ongelijkheid' heeft meegekregen, beveelt Mellenbergh (1976; p. 15 e.v.) eerst de specificatie van het multivariate significantieniveau aan, waarna vervolgens de bepaling van de (univariate) significantieniveau's kan plaatsvinden via de formule:

$$p_u = \frac{p_m}{n}$$

waarin:

p_u = univariate significantieniveau

p_m = multivariate significantieniveau

n = aantal toetsen op n variabelen, i.c. het aantal tests.

Uitgaande van een $p_m = .05$ en $n = 8$ (aantal toetsen op acht tests per variantiebron) is het significantieniveau per F-toets (p_u) dan gelijk aan .63% ($p = .0063$). Inspektie van de verkregen F-ratio's leverde noch multivariaat noch univariaat bezien, significante interacties op. Wel werd een aantal significante hoofdeffekten gevonden. De resultaten van de uitgevoerde drieweg variantie-analyse treft men aan in tabel 8. Kursief gedrukte F-ratio's zijn significant op 1%-niveau ($p < .0063$, afgerond tot $p < .01$).

Uit deze tabel kan men aflezen dat de faktor *leeftijd* voor Blokpatronen, Natikken, Plaatjes aanwijzen, Begrip en inzicht en het totaal een significant effect oplevert. Dit betekent dat oudere kleuters betere prestaties leveren dan jongere kleuters op deze tests. Voor Vouwblaadjes, Woordenspan, Zinnen nazeggen en Verhaaltje vragen laat leeftijd daaren-

Tabel 8. Drieweg variantie-analyse^a over de tests en het totaal met leeftijd, geslacht en sociaal-ekonomisch niveau als onafhankelijke variabelen

bron	df	F-ratio's ^b								totaal ^c
		Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span	Plaatjes aanwijzen	Zinnen nazeggen	Verhaaltje vragen	Begrip en inzicht	
A: leeftijd	1	<i>22.90</i>	5.86	<i>17.08</i>	1.31	<i>10.77</i>	1.41	0.12	<i>9.35</i>	<i>17.57</i>
B: geslacht	1	0.17	<i>19.20</i>	0.01	1.40	2.90	3.69	0.76	4.03	1.47
C: soc.-ekon. niveau	2	1.32	1.58	<i>11.89</i>	<i>6.84</i>	<i>7.72</i>	<i>20.62</i>	3.16	<i>13.52</i>	<i>14.23</i>
A x B	1	2.20	0.12	2.33	1.22	0.81	0.22	0.14	0.02	0.54
A x C	2	0.66	2.26	1.81	0.12	1.09	1.50	0.29	0.43	1.22
B x C	2	0.16	1.46	1.33	3.59	2.87	0.24	2.41	2.16	2.81
A x B x C	2	1.61	0.14	0.64	0.09	1.54	0.24	0.17	2.46	0.56

^a drieweg variantie-analyse volgens Kirk (1968; p. 171 e.v.)

^b kursief gedrukte F-ratio's zijn op 1%-niveau significant ($p < .01$)

^c voor de berekening van het totaal is de som van de z-scores op alle acht tests genomen

tegen geen significant effect zien; met andere woorden jongere en oudere schoolverlatende kleuters doen het gemiddeld ongeveer even goed.

De faktor *geslacht* beïnvloedt alleen de prestaties op Vouwblaadjes in die zin dat meisjes het beter doen dan jongens. Voor alle overige tests en het totaal valt geen effect van betekenis te konstaten.

De derde en laatste faktor, *sociaal-ekonomisch niveau* (SEN) levert een significant effect op voor Natikken, Woordenspan, Plaatjes aanwijzen, Zinnen nazeggen, Begrip en inzicht en het totaal. Voor Blokpatronen, Vouwblaadjes en Verhaaltje vragen werd daarentegen geen significant SEN-effect gevonden.

Uit tabel 8 kan men tenslotte ook aflezen dat geen enkele eerste en tweede orde interactie significant is.

Om na te gaan in hoeverre de gevonden SEN-effecten op Natikken etc. verklaard kunnen worden door verschillen in gemiddelde testprestaties tussen de drie onderscheiden SEN-groepen, respectievelijk SEN-I, SEN-II en SEN-III, is vervolgens een a posteriori-analyse uitgevoerd, waarbij de verschillen in gemiddelden paarsgewijs getoetst werden met een variant van de t-toets (Nie et al., 1975; p. 270). Het resultaat van deze analyse wordt weergegeven in tabel 9.

Tabel 9. A posteriori-analyse^a van SEN-effecten^b op enkele tests en het totaal

tests	t-waarden ^c		
	SEN I - II	SEN II - III	SEN I - III
Natikken	-2.45*	-2.81	-5.11
Woordenspan	n.s.	-2.96	-3.52
Plaatjes aanwijzen	n.s.	-2.63	-3.95
Zinnen nazeggen	-2.70	-4.68	-8.09
Begrip en inzicht	-3.09	-2.65	-6.62
totaal	-2.54*	-3.30	-6.20

^a volgens Nie et al. (1975; p. 270)

^b SEN = sociaal-ekonomisch niveau (SEN-I = hoog; SEN-II = midden; SEN-III = laag)

^c n.s. = niet significant ($p > .05$); * = asterisk, significant op 5%-niveau ($p < .05$); zonder asterisk significant op 1%-niveau ($p < .01$)

Tabel 10. Korrelaties^a tussen de tests en het totaal^b

N = 411

tests	Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span	Plaatjes aanwijzen	Zinnen nazeggen	Verhaaltje vragen	Begrip en inzicht
Vouwblaadjes	.30							
Natikken	.38	.31						
Woordenspan	.20	.15	.35					
Plaatjes aanwijzen	.29	.30	.45	.63				
Zinnen nazeggen	.24	.29	.39	.55	.55			
Verhaaltje vragen	.22	.25	.32	.27	.34	.55		
Begrip en inzicht	.16	.16	.29	.28	.34	.54	.46	
totaal ^b	.53	.54	.67	.66	.75	.79	.65	.61
totaal ^c	.37	.38	.54	.53	.64	.70	.51	.46

^a korrelaties $\geq .12$ zijn op 1%-niveau significant ($p \leq .01$)

^b voor de berekening van het totaal is de som van de z-scores op alle acht tests genomen, *zonder* correctie voor kontaminatie

^c idem, *met* correctie voor kontaminatie (Lienert, 1961; p. 404)

Aangezien voor Blokpatronen, Vouwblaadjes en Verhaaltje vragen geen significant SEN-effekt gevonden is, kunnen we volstaan met de opmerking dat de prestaties op deze tests nauwelijks of niet beïnvloed worden door het sociaal-ekonomisch niveau van de kleuter. Dit in tegenstelling tot de gevonden effecten op de overige tests en het totaal. Afgaande op de hoogte van de t-waarden voor de verschillende SEN-paren in tabel 9, mag men stellen dat het effect van het sociaal-ekonomisch niveau zich het sterkst manifesteert bij Zinnen nazeggen, iets minder sterk bij Natikken, Begrip en inzicht en het totaal, gevolgd door Woordenspan en Plaatjes aanwijzen als de naar verhouding minst gevoelige tests voor sociaal-ekonomisch niveau. Eveneens gelet op de hoogte van de t-waarden, valt tenslotte op dat de verschillen in gemiddelde testprestaties tussen SEN-I en SEN-II beduidend kleiner zijn dan die tussen SEN-II en SEN-III.

6.2 Korrelaties

In tabel 10 worden de korrelaties tussen de tests en het totaal van de LDT-E weergegeven. Voor de berekening van het totaal is de som van de standaardscores (z-scores) op alle acht tests genomen. Zoals men uit de tabel kan aflezen zijn voor de korrelaties van het totaal met de afzonderlijke tests twee typen koëfficiënten berekend: met en zonder correctie voor 'kontaminatie' (Lienert, 1961; p. 404). De gekorrigeerde korrelatiekoëfficiënten zijn kursief weergegeven.

De korrelaties tussen de tests variëren van .15 (Woordenspan - Vouwblaadjes) tot .63 (Woordenspan - Plaatjes aanwijzen); met andere woorden alle tests uit de LDT-E vertonen een duidelijke samenhang, welke significant is op 1%-niveau ($p < .01$). Een dergelijke samenhang kan men verwachten indien de functies - zoals gemeten door de verschillende tests - elkaar gedeeltelijk overlappen (cf. tabel 3). Als zodanig biedt deze samenhang een eerste, zij het zeer summiere vingerwijzing voor de steekhoudendheid van het Mark-model bij de samenstelling van een test (cf. hfdst. 3). Ofschoon een zekere mate van samenhang gewenst en ook noodzakelijk is, zijn al te hoge korrelaties tussen de tests weer minder begerenswaardig, aangezien hierdoor te weinig informatie over het cognitief functioneren van de onderzochte verkregen wordt binnen het tijdsbestek van een plm. één uur durende testafname. Ondanks de tamelijk hoge samenhang tussen Woordenspan en Plaatjes aanwijzen (.63) - welke althans gedeeltelijk verklaard kan worden uit de afhankelijkheid tussen beide tests¹² - is onze indruk dat wij nog redelijk tussen het Scylla en Charybdis van

verwaarloosbare en al te hoge interkorrelaties zijn doorgevaaren. Vrij hoge verbanden met het voor kontaminatie gekorrigeerde totaal treffen wij aan bij Plaatjes aanwijzen en Zinnen nazeggen, respektievelijk .64 en .70. In percentages variantie uitgedrukt betekent dit dat Plaatjes aanwijzen en Zinnen nazeggen respektievelijk 41% en 49% van de variantie van het LDT-E totaal voor hun rekening nemen, waarbij het LDT-E totaal opgevat mag worden als een maat voor de intelligentie¹⁰. Vanuit dit perspektief gezien zou Zinnen nazeggen en in mindere mate ook Plaatjes aanwijzen geen gek figuur slaan als simpel 'intelligentie'-testje.

In tabel 11 worden de korrelaties tussen de tests nogmaals weergegeven, zij het dat nu alle korrelaties gekorrigeerd zijn voor 'attenuatie', uitgaande van de koëfficiënt van interne consistentie van iedere test afzonderlijk (tabel 7). Via deze korrektie voor 'attenuatie' wordt een schatting verkregen van de relatie tussen twee tests indien beide maximaal betrouwbaar zouden zijn (Drenth, 1975; p. 228 e.v.). De reden voor deze korrektie is dat wij niet zozeer geïnteresseerd zijn in de feitelijke waargenomen samenhang van twee tests uit een (toevallige) experimentele versie van de LDT als wel in de werkelijke samenhang van twee tests indien de metingen niet gestoord zouden zijn door waarnemingsfouten.

Een ander opmerkelijk verschil met tabel 10 is dat de tests nu gegroepeerd zijn per leerkanaal en bovendien de volgorde der tests lichtelijk gewijzigd is om redenen die zo meteen duidelijk zullen worden. Om dezelfde redenen zullen wij voorts de geattenueerde korrelaties van Plaatjes aanwijzen met alle overige tests uit de LDT-E niet in de bespreking van de resultaten betrekken.

Om te onderzoeken in hoeverre het eerste element van het Mark-model - de indeling naar kanalen - een nuttige leidraad vormde voor de samenstelling van de LDT-E, hebben wij de tests gegroepeerd in drie kanalen, respektievelijk het visueel-fijnmotorische (V-F) kanaal, het auditief-spraakmotorische (A-S) kanaal en het auditief/visueel-fijnmotorische (AV-F) kanaal (cf. tabel 3). Vergelijking van de korrelaties per kanaal leert dat de korrelaties tussen tests die ieder voor zich betrekking hebben op een ander leerkanaal in het algemeen lager zijn dan de korrelaties tussen tests die betrekking hebben op één-en-hetzelfde leerkanaal³². Dit betekent dat de functies die door de verschillende tests gemeten worden, een grotere samenhang vertonen binnen respektievelijk het V-F en het A-S kanaal dan tussen het V-F en het A-S kanaal; met

Tabel 11. Korrelaties tussen de tests, gekorrigeerd voor 'attenuatie'^a
N = 411

tests		V - F ^b			A - S ^c			
		NT	BP	VB	WS	ZN	VV	BI
V - F ^b	Natikken (NT)							
	Blokpatronen (BP)	.50						
	Vouwblaadjes (VB)	.42	.44					
A - S ^c	Woordenspan (WS)	.43	.27	.21				
	Zinnen nazeggen (ZN)	.46	.31	.38	.66			
	Verhaaltje vragen (VV)	.41	.31	.36	.35	.69		
	Begrip en inzicht (BI)	.37	.22	.23	.37	.68	.63	
AV - F ^d	Plaatjes aanwijzen (PA)	.55	.39	.41	.78	.66	.44	.44

^a geschat volgens Drenth (1975; p. 229), uitgaande van de koëfficiënten van interne consistentie voor iedere test afzonderlijk (tabel 7)

^b V - F = visueel-fijnmotorisch kanaal

^c A - S = auditief-spraakmotorisch kanaal

^d AV - F = auditief/visueel-fijnmotorisch kanaal

andere woorden er is sprake van een zekere kanaalafhankelijkheid van de verzameling functies, zoals gemeten door de LDT-E. Als zodanig vormt deze kanaalafhankelijkheid een tweede aanwijzing voor de steekhoudendheid van het Mark-model.

Het geconstateerde verschil in korrelaties tussen het V - F en A - S kanaal is echter niet groot, beïnvloed als het wordt door de naar verhouding hoge korrelaties tussen Natikken enerzijds en Woordenspan (.43), Zinnen nazeggen (.46), Verhaaltje vragen (.41) en Begrip en inzicht (.37) anderzijds. De vraag rijst naar het waarom van deze hoge korrelaties, te meer daar wij eerder de veronderstelling opperden dat onmiddellijk geheugenfuncties - en Natikken meet althans volgens onze inhoudsanalyse primair het onmiddellijk geheugen voor ruimtelijke sekwenties (tabel 3) - meer kanaalgebonden zijn (cf. par. 2.3). Ofschoon wij het antwoord op deze vraag (nog) niet kunnen geven, suggereren de korrelaties met de tests uit het A - S kanaal - en zeker de korrelaties met Plaatjes aanwijzen (.55) uit het AV - F kanaal - dat Natikken naast de onmiddellijk geheugenfunctie óók nog op een relatief centrale, kanaal-onafhankelijke mediërende functie als bijvoorbeeld 'coderen' een beroep doet. Als deze suggestie enige waarheid bevat, dan wordt deze in ieder geval niet tegengesproken door de eerder gevonden betrouwbaarheids-

koëfficiënten van Natikken. In paragraaf 5.2 constateerden wij immers dat Natikken over een weliswaar goede interne consistentie (.83) beschikt, maar tevens minder homogeen (.73) is. Met andere woorden: de items van Natikken meten op betrouwbare wijze niet één, maar verschillende functies tegelijkertijd. In een latere analyse van de gegevens zullen wij nagaan in hoeverre deze voorstelling van zaken door andere bevindingen ondersteund wordt.

Uitgaande van het derde element van het Mark-model - de hiërarchische organisatie van functies - kunnen de tests, c.q. interkorrelaties van tabel 11 nog op een andere manier dan in tabel 10 gegroepeerd worden. Hiertoe plaatst men de korrelaties in een zodanige rangorde dat de uiteindelijke korrelatiematrix - of onderdelen daarvan - een zgn. 'simplexstructuur' vertoont, zoals beschreven door Guttman (1954). Deze structuur houdt in dat korrelaties aan weerszijden van de diagonaal (van linksboven naar rechtsonder) het hoogst zijn en vervolgens langzamerhand afnemen naarmate zij verder van de diagonaal verwijderd zijn. In de rechterbovenhoek en de linkerbenedenhoek treft men dan de laagste korrelaties aan. De rationale van dit korrelatiepatroon voor de LDT-E is dat tests van aangrenzende functiecomplexiteit, dat wil zeggen tests die volgens het postulaat van de hiërarchische organisatie op één na dezelfde functie meten, hoger met elkaar correleren dan tests van niet-aangrenzende functiecomplexiteit. Bovendien zullen de korrelaties dalen naarmate de tests sterker verschillen in complexiteit, dat wil zeggen een steeds geringer aantal functies gemeenschappelijk hebben.

Bekijken we nu de interkorrelaties van tabel 11 per kanaal, uitgaande van het kanaalgebonden karakter der verzameling functies, dan treffen we voor het V-F kanaal een perfecte simplexstructuur aan, met Natikken als naar alle waarschijnlijkheid minst complexe en Vouwblaadjes als meest complexe test. Om een dergelijke structuur te bereiken moesten we echter wel de volgorde der tests wijzigen, aangezien Blokpatronen - volgens een eerdere inhoudsanalyse de meest complexe test (cf. tabel 3) - in zijn oorspronkelijke positie de simplexstructuur te niet deed³³.

Ondanks systematische pogingen om ook voor het A-S kanaal een perfecte simplexstructuur te bereiken, zijn we hierin niet geslaagd. Wel leverde herhaalde rangschikking van de tests een tweetal zgn. 'quasi-simplex' structuren op, waarin de perfecte simplex redelijk benaderd wordt. Eén van deze twee structuren laat qua rangschikking van de tests, hetzelfde beeld zien als tabel 3, met Woordenspan als minst complexe en Begrip en

inzicht als meest (?) complexe test. Dit is dan ook de reden waarom wij deze rangschikking in tabel 11 verkozen boven de tweede, rivaliserende quasi-simplex, waarvan de hiërarchische rangschikking als volgt luidt: Woordenspan, Zinnen nazeggen, Begrip en inzicht en Verhaaltje vragen. Zowel de perfecte simplex voor het V-F kanaal als de quasi-simplex voor het A-S kanaal vormen onzes inziens een derde aanwijzing voor de steekhoudendheid van het Mark-model.

6.3 Faktorstructuur

Gegeven de keuze van een aantal tests voor de LDT-E welke in meer of mindere mate dezelfde (hypothetische) functies meten, is het niet verwonderlijk dat deze tests een bepaalde, zij het wisselende onderlinge samenhang vertonen, zoals gedemonstreerd via de korrelaties in tabel 10. Langs verschillende wegen zijn wij tot nog toe op zoek geweest naar een verklaring voor de aangetroffen verbanden in de hoop daarmee een nader licht te kunnen werpen op de betekenis van de LDT-E en - impliciet - daarmee ook op het model dat aan de LDT-E ten grondslag ligt.

Een vaak bewandelde weg in het exploratieve onderzoek naar de betekenis van tests is die van de faktoranalyse. In het kort komt deze techniek hierop neer dat men door middel van een kleiner aantal niet-waargenomen variabelen of factoren een verklaring probeert te geven voor de empirisch aangetroffen verbanden tussen de tests. Als zodanig leidt faktoranalyse altijd tot informatiereductie. In een exploratief onderzoek lijkt dit laatste een weinig aantrekkelijk vooruitzicht, aangezien het in deze onderzoeksfase nu juist gaat om een zo grondig mogelijke verkenning van alle informatie die de LDT-E mogelijkerwijs kan opleveren. Dat deze informatie in principe tamelijk omvangrijk is, bewijzen de overwegingen die aan de konstruktie van de LDT-E ten grondslag liggen en de daarop volgende inhoudsanalyse van de tests. Alleen al vanwege het feit dat een faktoroplossing van de LDT-E minder factoren zal tellen dan er tests zijn, kunnen wij bij voorbaat stellen dat faktoranalyse slechts ten dele recht doet aan de in aanleg rijke verscheidenheid van kanaal- en funktiekenmerken in de LDT-E.

Ondanks bovenstaande bedenkingen kan faktoranalyse toch een nuttige rol vervullen in ons onderzoek. In de eerste plaats omdat de LDT-E naast diagnostische test - gericht op de meting van een aantal afzonderlijke functies - ook de rol van intelligentietest vervult en als zodanig voor een faktoranalytische exploratie en interpretatie in aanmerking komt

(cf. par. 3.1). In de tweede plaats omdat faktoranalyse de techniek bij uitstek is om een gegeven interkorrelatiematrix zó samen te vatten dat van iedere test(variabele) duidelijk wordt wat zij gemeenschappelijk heeft met andere test(variabelen). Indien daarnaast de betekenis van die andere testvariabelen al min of meer vaststaat, zoals in het geval van de FST- en NST-subtests, dan kan de betekenis van de tests uit de LDT-E alleen nog maar in duidelijkheid winnen. Argumenten pro en contra tegen elkaar afwegend hebben wij vervolgens besloten de LDT-E aan een faktor-analytisch onderzoek te onderwerpen als aanvulling op voorafgaande en komende analyses.

In tabel 12 (Bijlage; p. 256) wordt de korrelatiematrix van de acht tests uit de LDT-E gegeven, op basis waarvan een faktoranalyse is uitgevoerd volgens Jöreskog's maximum likelihood methode (Bethlehem et al., 1978; p. 30)³⁴. Uitgaande van een restriktiviteitsgrens van 4.47 voor het gekozen faktormodel volgens de vuistregel $m < p + \frac{1}{2} - \frac{1}{2} \sqrt{3p+1}$, waarin m = aantal factoren en p = aantal tests, is eerst het aantal factoren bepaald dat benodigd is om de samenhang tussen de tests te verklaren. Hier-toe zijn toetsen voor achtereenvolgens 0-, 1-, 2-, 3- en 4-faktoroplos-singen uitgevoerd, waarbij de 3-faktoroplossing nog juist significant bleek ($\chi^2 = 24.29$, $df = 14$, $p = .04$) en de 4-faktoroplossing net niet ($\chi^2 = 14.03$, $df = 9$, $p = .12$). Gegeven het feit dat de hypothese met betrekking tot de 3-faktoroplossing nog juist verworpen moest worden voordat de re-striktiviteitsgrens werd bereikt³⁵, hebben wij vervolgens gekozen voor een oplossing naar vier factoren. Het resultaat van deze analyse is als volgt. Het totaal percentage verklaarde gemeenschappelijke variantie (som uitgevoerde kommunaliteiten) bedraagt 53%. Hiervan neemt de eerste faktor 38% in beslag en de drie daaropvolgende factoren respectievelijk 7%, 6% en 2%. Opvallend - maar niet geheel onverwacht gezien de samen-stelling van de LDT-E en de bevindingen tot nog toe - is de dominante positie van de eerste (algemene) faktor, welke wij in aansluiting op de faktoranalytische traditie aanduiden met 'g', als afkorting voor 'gen-eral (intelligence)'. Op basis van deze duidelijke g-faktor mogen wij konkluderen dat de waargenomen variantie van de tests uit de LDT-E in ieder geval in één betekenisdimensie beschreven kan worden. Uitdrukke-lijk schrijven wij 'kan', aangezien orthogonale rotatie (varimaxrotatie) van de verkregen 4-faktoroplossing ons inzicht in de structurele bete-kenis van de LDT-E aanzienlijk kan verdiepen. In tabel 13 wordt de des-betreffende faktormatrix, geroteerd naar een eenvoudige structuur vol-

gens het varimaxcriterium van Kaiser, weergegeven.

Tabel 13. Faktorladingen van de tests na varimaxrotatie over vier factoren

N = 392

tests	faktorladingen			
	I	II	III	IV
Blokpatronen	.09	.11	<u>.54</u>	.21
Vouwblaadjes	.15	.10	.32	<u>.52</u>
Natikken	.23	<u>.30</u>	<u>.57</u>	.11
Woordenspan	.20	<u>.76</u>	.14	.04
Plaatjes aanwijzen	.24	<u>.68</u>	.32	.15
Zinnen nazeggen	<u>.63</u>	<u>.49</u>	.14	.22
Verhaaltje vragen	<u>.64</u>	.14	.17	.21
Begrip en inzicht	<u>.68</u>	.18	.13	-.01

Wanneer we ons beperken tot tests met een faktorlading $\geq .40$ (in de tabel onderstreept)³⁶ dan wordt de eerste faktor (I) gevormd door Zinnen nazeggen, Verhaaltje vragen en Begrip en inzicht. Een juiste benaming voor deze faktor lijkt *Taalbegrip* te zijn aangezien in de tests die op deze faktor laden het vermogen om verbale betekenisinhouden te begrijpen een belangrijke rol speelt.

Van de drie tests laadt Zinnen nazeggen ook duidelijk op de tweede faktor (II), welke voornamelijk de variantie van Woordenspan en Plaatjes aanwijzen beschrijft. Dit valt te verklaren uit het feit dat Zinnen nazeggen niet alleen een beroep doet op het vermogen verbale betekenisinhouden te begrijpen maar ook op de vaardigheid om datgene wat begrepen is te reproduceren. De veronderstelling hierbij is dat een zin beter gereproduceerd wordt indien deze ook begrepen wordt. Als zodanig bevestigt de combinatie Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen onze eerder geuite verwachting (cf. par. 3.2) omtrent de geheugenfunctie van deze tests. Opvallend is echter dat Vouwblaadjes en Natikken, eveneens tests die volgens onze inhoudsanalyse de geheugenfunctie zouden moeten meten, nauwelijks of niet laden op deze tweede faktor (II), welke wij voorlopig met *Verbaal geheugen* zullen aanduiden.

De derde faktor (III) wordt gevormd door Blokpatronen en Natikken. In-

dien we deze faktor zouden moeten interpreteren dan lijkt de waarneming van visueel-ruimtelijke patronen, eerder dan een of andere geheugenfunctie, het bindend element tussen deze twee tests. In afwachting van verdere analyses zullen wij deze derde faktor (III) dan ook korthedshalve aanduiden met *Ruimtelijk-analytisch inzicht*.

De vierde faktor (IV) - waarop alleen Vouwblaadjes substantieel laadt - lijkt op het eerste gezicht een restfaktor en is als zodanig moeilijk te interpreteren. Vanuit het Mark-model gezien wordt echter de interpretatie eenvoudiger indien men het volgende bedenkt.

In de eerste plaats wordt het onderscheid naar kanalen weerspiegeld in de oplossing naar vier factoren. Zo hebben de eerste twee factoren (I en II) betrekking op het auditief-spraakmotorische kanaal³⁷ en de laatste twee factoren (III en IV) op het visueel-fijnmotorische kanaal. Gegeven deze 'kanaalindeling' van de factoren is het vervolgens interessant om na te gaan hoe de veronderstelde hiërarchische organisatie van functies tot uiting komt in de faktorladingen van ieder paar factoren. Met de voorkennis die wij reeds bezitten op basis van onze eerder uitgevoerde simplexanalyse, mogen we verwachten dat op iedere faktor per leerkanaal òf minder complexe òf meer complexe tests zullen laden. Inspectie van de faktorladingen per leerkanaal bevestigt globaal onze verwachtingen: zo laden de naar verhouding meest complexe tests binnen het auditief-spraakmotorische leerkanaal op één en dezelfde faktor (I) en de minst complexe tests binnen hetzelfde leerkanaal alleen op de tweede faktor. Voor het visueel-fijnmotorische leerkanaal vinden we hetzelfde patroon terug: de minst complexe tests - Blokpatronen en Natikken - laden op dezelfde faktor (III) en Vouwblaadjes als meest complexe test op de vierde faktor. Met andere woorden: ook al valt de vierde faktor (IV) binnen de gegeven samenhang van acht tests niet direkt te benoemen, toch kan men haar interpreteren als een specifieke faktor welke primair de variantie van de meest complexe test binnen het visueel-fijnmotorische kanaal beschrijft. Welke verdere psychologische betekenis wij aan deze complexiteit moeten hechten kan mogelijk in relatie tot de subtests van de FST en NST bepaald worden.

Ofschoon de samenhang van de tests uit de LDT-E grotendeels verklaard kan worden door middel van vier factoren, is de vraag naar het praktisch nut en de theoretische houdbaarheid van dit aantal niet zonder enige betekenis. Indien bijvoorbeeld voorspelling door middel van een zo klein mogelijk aantal factoren het doel is van de analyse, dan lijkt een

3-faktoroplossing te prefereren boven een 4-faktoroplossing: het percentage te verklaren variantie van de laatste twee factoren is immers tamelijk gering (respektievelijk 6% en 2%). Nu kan men aanvoeren dat een techniek als principale componentenanalyse zich beter leent voor een dergelijke datareduktie met het oog op voorspelling. Wij kunnen dit slechts beamen. Dat wij hierna desondanks een 3-faktoroplossing zullen presenteren vindt dan ook haar grond in het volgende.

In de eerste plaats zijn wij primair geïnteresseerd in voorspellingen waarbij de oorspronkelijke betekenisstructuur van de tests uit de LDT-E zoveel mogelijk behouden blijft. Principale componentenanalyse doet, gezien haar aard als pure datareduktietechniek, geen recht aan dit laatste aspect, ook al lijken de numerieke resultaten van beide technieken vaak sterk op elkaar. Deze uitspraak van Bethlehem et al. (1978; p. 16) wordt bevestigd door de resultaten van een eerder uitgevoerde principale componentenanalyse (eenheden op de diagonaal, varimaxrotatie van vier componenten) door Schroots en Van Alphen de Veer (1974), zij het dat de ladingen van de tests op de vier componenten, evenals het totaalpercentage verklaarde variantie (77%) naar verwachting hoger zijn dan bij de hier gepresenteerde faktoranalyse.

Ten tweede: alle criteria ter bepaling van een optimaal aantal factoren, zoals hoogte en verloop van eigenwaarden, hoogte van de ladingen, interpreteerbaarheid en toetsing van het aantal factoren etc., zijn erop gericht om het risico van de interpretatie van foutenvariantie te voorkomen. Ofschoon een 4-faktoroplossing volgens de hier gehanteerde criteria nog juist toelaatbaar is, kunnen wij ons niet aan de indruk onttrekken dat wij bij een 4-faktoroplossing gevaarlijk dicht in de buurt van deze foutenvariantie balanceren, getuige onder andere het feit dat het percentage te verklaren variantie van de vierde faktor slechts 2% bedraagt. Deze indruk wordt versterkt door de hoogte en het verloop van de vier eigenwaarden in de principale componentenoplossing, welke respektievelijk 3.45, 1.08, .95 en .72 bedragen. Volgens het vaak gehanteerde criterium: eigenwaarde >1.0, zou strikt genomen alleen een 2-komponentenoplossing toegestaan zijn.

In de derde plaats achten wij de theoretische en praktische betekenisvraag van belang in hoeverre een 3-faktoroplossing tot wezenlijk andere resultaten zal leiden dan een 4-faktoroplossing, en zo niet, in welke factoren de vierde faktor zich zal oplossen.

Een vierde reden tenslotte waarom wij geïnteresseerd zijn in een 3-faktoroplossing betreft de vergelijkbaarheid van verschillende faktor-

oplossingen indien de gehele onderzoeksgroep opgesplitst wordt naar leeftijd, geslacht en sociaal-ekonomisch niveau. Aangezien een dergelijke opsplitsing onvermijdelijk tot geringere aantallen per subgroep leidt en schatting van het aantal te extraheren factoren per subgroep na toetsing van respectievelijk een 2-, 3- en 4-faktoroplossing een 3-faktoroplossing indiceert, verdient laatstgenoemde oplossing duidelijk de voorkeur boven een 4-faktoroplossing. Het risico van de interpretatie van foutenvariantie nemen wij liever niet.

In tabel 14 wordt de faktormatrix van een oplossing naar drie factoren weergegeven, geroteerd naar een eenvoudige structuur volgens het varimaxcriterium van Kaiser. Tussen haakjes zij vermeld dat het percentage verklaarde gemeenschappelijke variantie 50% bedraagt. Hiervan neemt de eerste ongeroteerde faktor 38% in beslag en de twee daaropvolgende factoren respectievelijk 7% en 6%. Evenals bij de 4-faktoroplossing kunnen wij dus ook bij de 3-faktoroplossing konkluderen dat de waargenomen variantie van de tests uit de LDT-E in één betekenisdimensie ('g') beschreven kan worden.

Tabel 14. Faktorladingen van de tests na varimaxrotatie over drie factoren

N = 392

tests	faktorladingen		
	I	II	III
Blokpatronen	.08	.13	<u>.56</u>
Vouwblaadjes	.19	.08	<u>.53</u>
Natikken	.20	.33	<u>.53</u>
Woordenspan	.22	<u>.75</u>	.13
Plaatjes aanwijzen	.25	<u>.67</u>	.34
Zinnen nazeggen	<u>.64</u>	<u>.48</u>	.23
Verhaaltje vragen	<u>.64</u>	.14	.25
Begrip en inzicht	<u>.66</u>	.19	.11

Zoals uit tabel 14 af te lezen valt, leidt een 3-faktoroplossing geenszins tot dramatische verschuivingen in de faktorstructuur van de LDT-E, vergeleken bij een 4-faktoroplossing. De vierde faktor is opgelost in de derde faktor welke wij eerder benoemden met de term Ruimtelijk-analy-

tisch inzicht. Gegeven de ladingen van Blokpatronen, Vouwblaadjes en Natikken op deze faktor en gelet op hun gemeenschappelijke betekenis met betrekking tot de waarneming van visueel-ruimtelijke patronen, zien wij geen aanleiding om de derde faktor (III) anders te benoemen dan als *Ruimtelijk-analytisch inzicht*.

De eerste en tweede faktor zijn vrijwel identiek met de eerste twee factoren van de 4-faktoroplossing. Gezien onze eerdere interpretatie en benoeming van deze factoren, duiden wij de eerste faktor (I) dan ook weer aan met *Taalbegrip* en de tweede faktor (II) met *Verbaal geheugen*. Vanuit het Mark-model geredeneerd beschouwen wij voorts Taalbegrip en Verbaal geheugen als de noemers van de functies die door de betrokken tests binnen het auditief-spraakmotorische kanaal gemeten worden³⁷ en Ruimtelijk-analytisch inzicht op overeenkomstige wijze als de noemer van de functies etc. binnen het visueel-fijnmotorische kanaal.

6.4 Faktorstructuur naar leeftijd, geslacht en sociaal-ekonomisch niveau

In paragraaf 6.1 hebben wij uiteengezet waarom het ons nuttig leek om de samenhang tussen de testprestaties op de LDT-E en de variabelen leeftijd, geslacht en sociaal-ekonomisch niveau te bestuderen. In dezelfde paragraaf hebben wij vervolgens deze samenhang onderzocht. Uit dit onderzoek bleek onder andere dat oudere kinderen (5;11-6;04) betere prestaties leveren op respectievelijk Blokpatronen, Natikken, Plaatjes aanwijzen en Begrip en inzicht dan jongere kinderen (5;05-5;10); leeftijd speelde verder geen rol van betekenis bij Vouwblaadjes, Woordenspan, Zinnen nazeggen en Verhaaltje vragen.

De vraag is nu in hoeverre met de leeftijd ook de psychologische betekenis van eerstgenoemde tests is veranderd, aangezien het vierde element van het Mark-model - de ontogenie van functies - inhoudt dat op latere leeftijd méér (en complexere) functies tot ontwikkeling gekomen zijn dan op jongere leeftijd. Indien dit zo is dan zullen psychologische tests die deze functies meten door hun veranderde psychologische betekenis tevens de veronderstelde ontogenie van de functies reflektieren.

Een van de manieren om dit te onderzoeken is door middel van faktoranalytisch onderzoek van tests, afgenomen op verschillende leeftijdsgroepen (Stott & Ball, 1965). Indien eenzelfde faktorstructuur gevonden wordt voor verschillende leeftijdsgroepen, dan mag men konkluderen dat ondanks het vorderen van de leeftijd de psychologische betekenis van de tests niet is veranderd en dat dientengevolge dezelfde functies gemeten worden.

Een en ander sluit uiteraard niet uit dat de functies als zodanig niet verder zouden kunnen 'rijpen'. In feite is 'rijping' in geval van een stabiele faktorstructuur dan ook datgene wat men konstateert indien een significant leeftijdseffekt gevonden wordt (cf. par. 2.2; fig. 4).

In tabel 15 (Bijlage; p. 257) worden de interkorrelaties van de tests uit de LDT-E weergegeven voor respectievelijk de jongste (5;05 - 5;10) en de oudste (5;11 - 6;04) leeftijdsgroep. Op basis van deze twee korrelatiematrices is vervolgens per matrix een faktoranalyse uitgevoerd volgens Jöreskog's maximum likelihood methode (Bethlehem et al., op.cit.). Gekozen is voor een oplossing naar drie factoren op grond van eerder genoemde overwegingen (par. 6.3). Het resultaat van deze analyse is als volgt.

Voor de jongste leeftijdsgroep bedraagt het totaal percentage verklaarde gemeenschappelijke variantie 48. Hiervan neemt de eerste faktor 34% in beslag en de twee daaropvolgende factoren respectievelijk 7% en 7%. Voor de oudste leeftijdsgroep zijn deze waarden iets verschillend, namelijk 51% (totaal percentage verklaarde gemeenschappelijke variantie), 39% (eerste faktor), 7% (tweede faktor) en 5% (derde faktor). Op basis van deze resultaten mogen wij konkluderen dat de waargenomen variantie van de tests uit de LDT-E - zowel voor de jongste als voor de oudste leeftijdsgroep - in één betekenisdimensie beschreven kan worden.

De twee faktormatrices zijn vervolgens geroteerd naar een eenvoudige structuur volgens het varimaxcriterium van Kaiser. In tabel 16 worden de geroteerde faktormatrices voor respectievelijk de jongste en oudste leeftijdsgroepen weergegeven.

Uit de tabel kan men aflezen dat binnen de leeftijdsgroep van 5 jaar en 5 maanden tot en met 6 jaar en 4 maanden zich geen verschuivingen van betekenis voordoen in de faktorstructuur van de LDT-E. Met andere woorden: de psychologische betekenis van de tests is vrijwel dezelfde voor zowel de jongste als de oudste leeftijdsgroep. Vertaald in het Mark-model betekent dit dat ongeacht het leeftijdsverschil van een half jaar praktisch dezelfde functies gemeten worden door de tests uit de LDT-E. Kennelijk verloopt de functie-ontwikkeling binnen de onderzochte leeftijdsgroep weer niet zó snel dat zich fundamentele verschuivingen in de testbetekenis voordoen. Uiteraard sluiten wij hierbij de mogelijkheid niet uit dat de LDT-E of de gebruikte analysetechniek weinig gevoelig is voor een zich feitelijk voltrekkende ontwikkeling.

Tabel 16. Faktorladingen van de tests na varimaxrotatie over drie factoren, naar leeftijd^a

N₁ = 219; N₂ = 173

tests	faktorladingen					
	I		II		III	
	1	2	1	2	1	2
Blokpatronen	-.01	.18	.12	.13	<u>.54</u>	<u>.55</u>
Vouwblaadjes	.28	.05	.04	.12	<u>.51</u>	<u>.54</u>
Natikken	.16	.22	.33	.32	<u>.46</u>	<u>.56</u>
Woordenspan	.19	.22	<u>.75</u>	<u>.75</u>	.08	.18
Plaatjes aanwijzen	.22	.25	<u>.65</u>	<u>.71</u>	.29	.35
Zinnen nazeggen	<u>.66</u>	<u>.60</u>	<u>.48</u>	<u>.47</u>	.17	.33
Verhaaltje vragen	<u>.60</u>	<u>.68</u>	.18	.12	.36	.16
Begrip en inzicht	<u>.64</u>	<u>.70</u>	.13	.24	.03	.12

^a 1 = jongste leeftijdsgroep (5;05 - 5;10); 2 = oudste leeftijdsgroep (5;11 - 6;04)

Blijkt leeftijd binnen de beperkingen van dit onderzoek nauwelijks van invloed te zijn op de betekenis van de LDT-E, voor geslacht liggen de zaken misschien anders. Zo leveren meisjes bijvoorbeeld betere testprestaties op Vouwblaadjes dan jongens, ongeacht hun leeftijd of sociaal-ekonomisch niveau (cf. par. 6.1). Dit wettigt de veronderstelling dat meisjes over gedeeltelijk andere cognitieve functies beschikken - eerder dan over meer 'gerijpte' of complexe functies - waardoor verschillen in testbetekenis tussen jongens en meisjes verwacht kunnen worden. Gezien het exploratieve karakter van dit onderzoek zullen wij echter met de verklaring van mogelijk optredende verschillen wachten tot de resultaten van de nu volgende factoranalyse gepresenteerd zijn.

In tabel 17 (Bijlage; p. 258) worden de interkorrelaties van de tests uit de LDT-E weergegeven voor jongens en meisjes apart. Op basis van deze twee korrelatiematrices is vervolgens per matrix een factoranalyse uitgevoerd volgens Jøreskog's maximum likelihood methode met een oplossing naar drie factoren.

Voor jongens bedraagt het totaal percentage verklaarde gemeenschappelijke variantie 51; voor meisjes eveneens 51. De percentages verklaarde variantie per faktor voor jongens zijn respectievelijk 37, 9 en 5; voor meisjes 39, 7 en 5. Evenals voor de twee onderscheiden leeftijdsgroepen

mogen wij ook nu weer konkluderen dat de waargenomen variantie van de tests uit de LDT-E zowel voor jongens als voor meisjes in één betekenisdimensie beschreven kan worden.

De resultaten van een hiernavolgende varimaxrotatie naar geslacht worden in tabel 18 weergegeven.

Tegen de verwachting in doet zich in de faktoriële betekenis van Vouwblaadjes geen opmerkelijke verschuiving voor. Met andere woorden: de veronderstelling dat Vouwblaadjes bij meisjes op gedeeltelijk andere cognitieve functies een beroep doet, wordt door de resultaten van deze analyse niet bevestigd. Een verklaring voor dit resultaat kunnen wij vooralsnog niet geven. Wèl willen wij de mogelijkheid openhouden dat een faktoranalyse op Vouwblaadjes in combinatie met andere tests tot het verwachte resultaat zal leiden. Bekend is immers dat de resultaten van een faktoranalyse sterk afhankelijk zijn van de geselecteerde variabelen, i.c. de tests van de LDT-E.

Letten wij in tabel 18 voorts alleen op tests met een faktorlading $\geq .40$ dan valt meteen op dat van de drie tests die op faktor III 'Ruimtelijk-analytisch inzicht' laden, Natikken ook op faktor II 'Verbaal geheugen' laadt, althans voor meisjes. Voor jongens is dit niet het geval. De vraag rijst welke betekenis wij hieraan moeten hechten, mede gelet op het gegeven dat jongens en meisjes gemiddeld even goede testprestaties op Natikken leveren ($m_j = 5.99$ en $m_m = 6.01$; cf. tabel 17).

Ofschoon deze vraag niet met zekerheid te beantwoorden valt, lijkt het minder waarschijnlijk - gezien de vrijwel identieke testprestaties - dat dezelfde functies van Natikken meer bij meisjes dan bij jongens gerijpt zouden zijn. Ook de veronderstelling dat bij meisjes meer complexe functies tot ontwikkeling gekomen zouden zijn op grond van het duidelijke leeftijdseffekt voor Natikken, klinkt weinig aannemelijk omdat de hiervoor gepresenteerde 3-faktoroplossing per leeftijdsgroep geen opmerkelijke verschuiving in de ladingen van Natikken heeft opgeleverd. Wèl opmerkelijk, maar op basis van onze gegevens onvoldoende bewijsbaar, is de veronderstelling dat bij meisjes gedeeltelijk verschillende (andere) functies tot ontwikkeling gekomen zijn.

Vrij vertaald komt deze veronderstelling hierop neer dat meisjes en jongens de testopgaven van Natikken ieder op hun eigen wijze oplossen met overigens hetzelfde eindresultaat: de jongens door voornamelijk een beroep te doen op de (kanaalafhankelijke) onmiddellijke geheugenfunctie voor ruimtelijke sekwenties (cf. par. 3.2) en de meisjes door tévens

Tabel 18. Faktorkladingen van de tests na varimaxrotatie over drie faktoren, naar geslacht^a

N_j = 199; N_m = 193

tests	faktorkladingen					
	I		II		III	
	j	m	j	m	j	m
Blokpatronen	.08	.07	.11	.19	<u>.63</u>	<u>.47</u>
Vouwblaadjes	.19	.22	.06	.07	<u>.57</u>	<u>.51</u>
Natikken	.10	.28	.25	<u>.42</u>	<u>.58</u>	<u>.52</u>
Woordenspan	.23	.19	<u>.75</u>	<u>.75</u>	.11	.14
Plaatjes aanwijzen	.28	.23	<u>.67</u>	<u>.69</u>	.29	.38
Zinnen nazeggen	<u>.63</u>	<u>.65</u>	<u>.50</u>	<u>.46</u>	.25	.18
Verhaaltje vragen	<u>.58</u>	<u>.68</u>	.18	.13	.33	.16
Begrip en inzicht	<u>.66</u>	<u>.68</u>	.19	.17	.04	.25

^a j = jongen; m = meisje

een beroep te doen op een of andere (kanaalonafhankelijke) centrale mediërende funktie als 'coderen' (cf. par. 6.2); dat wil zeggen visuele informatie over de positie van de aan te tikken blokjes wordt - na het passeren van de onmiddellijk geheugenfunktie - stilzwijgend benoemd met bijvoorbeeld links, rechts of één, twee, etc., alvorens een fijnmotorisch antwoord tot stand komt. Onzes inziens vormt deze extra 'codeer'-funktie bij meisjes de band tussen Natikken als onmiddellijk geheugentest voor het V-F kanaal enerzijds en Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen als onmiddellijk geheugentest voor het A-S en AV-F kanaal anderzijds.

Nog een tweetal opmerkingen over tabel 18 tot slot. In de eerste plaats kunnen wij - met uitzondering van Natikken - geen verschuivingen van betekenis konstateren in de faktorstruktuur van de LDT-E voor jongens en meisjes afzonderlijk. In de tweede plaats achten wij de benoeming van de tweede faktor met 'Verbaal' geheugen bij nader inzien minder korrekt; niet alleen omdat een substantiële lading voor Natikken (een test voor het V-F kanaal) op de tweede faktor gevonden is, maar ook omdat de kanaalkenmerken van Plaatjes aanwijzen (AV-F) - eveneens ladend op de tweede faktor - minder goed harmoniëren met het veronderstelde 'verbale' karakter van deze geheugenfaktor. Alvorens faktor II opnieuw te benoemen,

zullen wij echter eerst de resultaten van een faktoranalyse per sociaal-ekonomische niveaugroep afwachten.

Naast leeftijd en geslacht kan ook het sociaal-ekonomische niveau (SEN) van het kind van invloed zijn op de betekenis van de LDT-E. In paragraaf 6.1 vonden wij bijvoorbeeld duidelijke SEN-invloeden terug op de gemiddelde testprestaties bij Natikken, Woordenspan, Plaatjes aanwijzen, Zinnen nazeggen en Begrip en inzicht. De vraag is nu ook hier weer in hoeverre zich verschuivingen zullen voordoen in het patroon van faktorladingen per SEN-groep en in hoeverre deze verschuivingen - in het licht van de verschillen in gemiddelde testprestaties - iets zeggen over de functies die door de tests uit de LDT-E gemeten worden. De redenering die wij hierbij volgen is om en nabij dezelfde als die voor leeftijd en geslacht.

In tabel 19 (Bijlage; p. 259) worden de interkorrelaties van de tests uit de LDT-E weergegeven voor de drie SEN-groepen afzonderlijk (SEN-I = hoog; SEN-II = midden; SEN-III = laag). Op basis van deze drie korrelatiematrices is vervolgens een faktoranalyse uitgevoerd volgens Jöreskog's maximum likelihood methode met een oplossing naar drie faktoren.

Voor SEN-I bedraagt het totaal percentage verklaarde gemeenschappelijke variantie 46; voor SEN-II 46; voor SEN-III 52. De percentages verklaarde variantie per faktor zijn voor SEN-I respectievelijk 32, 8 en 6; voor SEN-II 32, 9 en 5; voor SEN-III 39, 7 en 6. Ook nu weer mogen wij konkluderen dat de waargenomen variantie van de tests uit de LDT-E zowel voor SEN-I en SEN-II als voor SEN-III in één betekenisdimensie beschreven kan worden.

In tabel 20 worden de geroteerde faktormatrices voor de drie SEN-groepen weergegeven.

Uit deze tabel kan men aflezen dat de faktorstructuur van de LDT-E voor de onderscheiden groepen grotendeels dezelfde is. Vergeleken bij de 3-faktoroplossing voor de gehele onderzoeksgroep (par. 6.3; tabel 14), kunnen wij echter een tweetal nieuwe, betekenisvolle faktorladingen ($\geq .40$) onderscheiden voor respectievelijk Natikken (SEN-III) op de tweede faktor en Plaatjes aanwijzen op de derde faktor.

Te beginnen met Natikken: deze test laadt voor SEN-III zowel op de derde faktor Ruimtelijk-analytisch inzicht als op de tweede faktor Verbaal geheugen. Volgen wij dezelfde redenering als voor geslacht dan mogen wij in eerste instantie konstaten dat Natikken niet alleen de (kanaalafhankelijke) onmiddellijk geheugenfunctie voor ruimtelijke sekwenties

Tabel 20. Faktorieladingen van de tests na varimaxrotatie over drie factoren, naar sociaal-ekonomisch niveau^a

N₁ = 94; N₂ = 112; N₃ = 205

tests	faktorieladingen								
	I			II			III		
	1	2	3	1	2	3	1	2	3
Blokpatronen	.03	.00	.14	.13	.10	.16	<u>.54</u>	<u>.62</u>	<u>.59</u>
Vouwblaadjes	.05	.13	.22	.09	.17	.04	<u>.59</u>	<u>.48</u>	<u>.51</u>
Natikken	.32	.17	.14	.20	.11	<u>.41</u>	<u>.51</u>	<u>.56</u>	<u>.54</u>
Woordenspan	.11	.18	.24	<u>.81</u>	<u>.69</u>	<u>.76</u>	.11	.18	.09
Plaatjes aanwijzen	.19	.26	.23	<u>.68</u>	<u>.67</u>	<u>.69</u>	<u>.41</u>	.19	.36
Zinnen nazeggen	<u>.52</u>	<u>.54</u>	<u>.68</u>	<u>.58</u>	<u>.47</u>	<u>.47</u>	.18	.21	.19
Verhaaltje vragen	<u>.58</u>	<u>.67</u>	<u>.68</u>	.06	.14	.17	.30	.08	.31
Begrip en inzicht	<u>.46</u>	<u>.68</u>	<u>.63</u>	.13	.22	.15	-.04	.14	.14

^a 1 = SEN-I (hoog); 2 = SEN-II (midden); 3 = SEN-III (laag)

meet maar ook een (kanaalafhankelijke) centrale, mediërende functie als coderen. In het licht van de gemiddelde testprestaties op Natikken voor respectievelijk SEN-I $m = 6.96$, SEN-II $m = 6.16$ en SEN-III $m = 5.35$ (tabel 19) lijkt echter een mediërende functie als coderen in tweede instantie weinig waarschijnlijk. Voor zover op deze functie al een beroep wordt gedaan, leidt dit niet tot gelijke of betere testprestaties. Eerder vermoeden wij dat alleen het onmiddellijk geheugen als zodanig een rol speelt bij het oplossen van de testopgaven van Natikken voor SEN-III; een tamelijk controversieel vermoeden aangezien wij eerder veronderstelden dat onmiddellijk geheugenfuncties relatief kanaalafhankelijk zijn. Als zodanig 'hoort' Natikken - een test voor het V-F kanaal - dan ook niet op de tweede faktor te laden, tenzij men ervan uitgaat dat niet zozeer het onmiddellijk geheugen of zijn veronderstelde kanaalafhankelijkheid het bindend element vormt tussen Natikken, Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen als wel hun gemeenschappelijke aandachtsfunctie of het temporele orde-karakter van de items uit deze tests (cf. par. 3.2). Meer speculatie dan exploratie dus, te meer daar noch de aandachtsfunctie noch het temporele orde-karakter van Natikken goed kan verklaren waarom deze test - met uitzondering van de groep meisjes en SEN-III - alleen op de derde faktor laadt.

Ten aanzien van Plaatjes aanwijzen vinden wij voor SEN-I niet alleen betekenisvolle ladingen op de tweede faktor, maar ook op de derde faktor, aangeduid met Ruimteliijk-analytisch inzicht. Een verklaring hiervoor is niet moeilijk te vinden indien men bedenkt dat Plaatjes aanwijzen 'gemengde' kanaaleigenschappen heeft; dat wil zeggen er wordt zowel auditieve als visuele informatieverwerking verlangd alvorens een fijnmotorisch antwoord gegeven kan worden. Kennelijk speelt het visueel-fijnmotorische aspekt in SEN-I een belangrijker rol bij het oplossen van de testopgaven van Plaatjes aanwijzen dan in SEN-II en SEN-III. Een aanduiding voor het feit dat een meer complexe functie in de hoogste SEN-groep tot ontwikkeling gekomen is? Of, danweg een aanduiding voor het feit dat deze groep met gedeeltelijk verschillende (andere) cognitieve functies toegerust is? Ofschoon wij het verlossende antwoord op basis van onze gegevens niet kunnen geven, zijn de gemiddelde testprestaties van SEN-I op Plaatjes aanwijzen wel hoger dan van SEN-II en SEN-III (respektievelijk $m = 5.11$, $m = 4.68$ en $m = 3.97$; tabel 19), zij het dat het verschil tussen SEN-I en SEN-II niet significant is (par. 6.1; tabel 9).

Tot slot van deze paragraaf nog een paar algemene opmerkingen. In de eerste plaats: ondanks de verschuivingen in faktoriële betekenis van Natikken en Plaatjes aanwijzen, en ondanks onze tamelijk spekulatieve veronderstellingen omtrent de betekenis van deze tests, mogen wij wel stellen dat de verkregen faktorstructuur van de LDT-E vrij stabiel is; het patroon van faktorladingen per onderscheiden subgroep wijkt immers niet noemenswaard af van het eerder gevonden patroon voor de gehele onderzoeksgroep (3-faktoroplossing; par. 6.3). De vraag rijst dan ook in hoeverre de opgetreden verschuivingen niet eerder te wijten zijn aan artefakten ten gevolge van bijvoorbeeld de gekozen rotatietechniek dan aan werkelijke betekenisverschillen van de tests per subgroep. Ofschoon wij deze vraag niet afdoende kunnen beantwoorden, spelen mogelijke artefakten - zoals ook bij nader onderzoek bleek - een ondergeschikte rol³⁸. Een eerder uitgevoerde principale componentenanalyse leverde namelijk zowel voor de gehele onderzoeksgroep als per SEN-groep (varimaxrotatie van drie componenten) praktisch hetzelfde beeld op als bij de hier gepresenteerde faktoranalyses (Schroofs & Van Alphen de Veer, 1974). De tweede opmerking betreft de benoeming van de tweede faktor met Verbaal geheugen. Eerder voerden wij reeds argumenten aan voor het weglaten van het woord 'verbaal'. In deze argumenten worden wij nu gesterkt door de resultaten van de faktoranalyse per SEN-groep, waarbij Natikken (V-F kanaal), evenals Plaatjes aanwijzen (AV-F kanaal), het meer 'algemene'

geheugen-karakter van de tweede faktor benadrukt.

Om mogelijk misverstand omtrent het puur verbale karakter van deze faktor te vermijden, benoemen wij daarom faktor II van de 3-faktoroplossingen alsnog en alleen met de term *Geheugen*.

6.5 Faktorstructuur in het licht van de FST en NST

Een van de manieren waarop men de betekenis van de LDT-E verder kan verduidelijken is door de toevoeging van nieuwe tests met een vaststaande betekenis aan de oorspronkelijke korrelatiematrix, met het doel deze aan een hernieuwd faktoranalytisch onderzoek te onderwerpen. Als 'nieuwe' tests hebben wij de FST en de NST gekozen op grond van enkele, in paragraaf 4.2 genoemde overwegingen. Indien nu uit dit onderzoek blijkt dat de tests uit de LDT-E en de subtests van de FST en NST met een afzonderlijk vastgestelde, maar ongeveer gelijksoortige faktoriële betekenis als de LDT-E op dezelfde factoren laden, dan vormt dit gegeven extra evidentie voor de eerder gegeven interpretatie van de faktorstructuur van de LDT-E.

Dat deze evidentie beperkt wordt door de faktorstructuur van de FST en NST spreekt voor zich. Zo kan de samenhang van de subtests uit de FST en NST grotendeels verklaard worden door middel van twee factoren, respectievelijk 'visueel-motorische differentiatie' en 'waarneming en verbaal begrip', terwijl de samenhang van de tests uit de LDT-E daarentegen ten minste door drie factoren te verklaren valt.

In tabel 21 (Bijlage; p. 260 en 261) worden de interkorrelaties van de tests uit de LDT-E en de subtests van de FST en NST weergegeven, evenals de interkorrelaties van het totaal van de LDT-E, FST en NST. Ofschoon wij deze interkorrelaties hier niet verder zullen analyseren, valt wel op dat het NST-totaal een grotere samenhang vertoont met het LDT-E totaal (.74) dan het FST-totaal (.67); niet verwonderlijk indien men de grotere verscheidenheid aan subtests en taken van de NST in aanmerking neemt. Afgaande op een eerder geuite opmerking dat de FST en NST voornamelijk intelligentie meten (par. 4.2), vormt het relatief hoge percentage gemeenschappelijke variantie van de NST en LDT-E enerzijds (55%) en de FST en LDT-E anderzijds (45%) enige ondersteuning voor onze opvatting dat de LDT-E ook als intelligentietest te gebruiken valt.

Op basis van de interkorrelaties in tabel 21 - de verschillende totalen niet meegerekend - is vervolgens een faktoranalyse uitgevoerd volgens

de maximum likelihood methode van Jöreskog. De keuze viel hierbij op een oplossing naar drie factoren; oplossingen naar respectievelijk vier, vijf en zes factoren bleken volgens de in paragraaf 6.3 gestelde criteria heel goed mogelijk, maar wierpen bij nader inzien meer licht op de faktorstructuur van de FST en NST dan op die van de LDT-E.

Het totaal percentage verklaarde gemeenschappelijke variantie bedraagt 40. Hiervan neemt de eerste faktor 33% in beslag, gevolgd door de tweede en derde faktor met respectievelijk 4 en 3%. De verkregen 3-faktoroplossing is hierna geroteerd naar een eenvoudige structuur volgens het varimaxcriterium van Kaiser. In tabel 22 worden de resultaten weergegeven.

Tabel 22. Faktorladingen van de (sub)tests uit de LDT-E, FST en NST na varimaxrotatie over drie factoren
N = 340

(sub)tests	faktorladingen		
	I	II	III
<i>LDT-E</i>			
Blokpatronen	.17	.09	.60
Vouwblaadjes	.22	.10	.38
Natikken	.30	.27	.43
Woordenspan	.16	.70	.20
Plaatjes aanwijzen	.26	.63	.35
Zinnen nazeggen	.53	.55	.27
Verhaaltje vragen	.51	.28	.20
Begrip en inzicht	.60	.26	.15
<i>FST</i>			
Figuren natekenen (1)	.28	.15	.74
Hekpatroon voortzetten (2)	.11	.18	.54
Ramen tekenen (3)	.23	.29	.56
<i>NST</i>			
Vormwaarneming etc. (1)	.32	.18	.49
Fijne motoriek (2)	.29	.19	.68
Begrip voor grootte etc. (3)	.49	.29	.47
Scherp waarnemen (4)	.45	.00	.31
Kritische waarneming (5)	.32	.09	.38
Koncentratie (6)	.40	.17	.24
Geheugen (7)	.55	.08	.17
Objektbegrip etc. (8)	.57	.14	.29
Weergeven verhaaltje (9)	.55	.16	.18
Menstekening (10)	.13	.14	.34

Letten we vooralsnog op ladingen $\geq .40$ dan blijkt dat alle subtests uit de NST die de faktor 'Waarneming en verbaal begrip' representeren (subtests 3, 4, 6, 7, 8, en 9), evenals de drie tests uit de LDT-E die de faktor 'Verbaal begrip' vormen, op de eerste faktor (I) laden. In zekere zin is dit resultaat verrassend aangezien de subtests van de NST geen spraakmotorisch antwoord vergen - zoals bij Zinnen nazeggen, Verhaaltje vragen en Begrip en inzicht - maar een fijnmotorisch antwoord. Daarnaast doen de betrokken subtests van de NST blijkens eerder betekenisonderzoek (Van Alphen de Veer et al., 1978) een beroep op functies die met de waarneming van een kind te maken hebben (plaatjes aankruisen, kleuren of herkennen; cf. tabel 4), waardoor men in ieder geval ook substantiële ladingen op de derde faktor (III) zou verwachten. Met uitzondering van subtest 3 van de NST is dit echter niet het geval. Kennelijk speelt het waarnemingsaspect in deze subtest een minder belangrijke rol dan wij aanvankelijk dachten. Des te belangrijker wordt daardoor het verbale begripstige karakter van de subtests 3, 4, 6, 7, 8 en 9 van de NST, waarbij men het woord 'verbaal' niet zozeer als (uit)gesproken taal moet opvatten alswel als onuitgesproken, interne verwoording van bepaalde begrippen. Zo moet het kind, althans volgens de resultaten van het reeds genoemde betekenisonderzoek, de testopdrachten en -taken niet alleen begrijpen, maar de gehanteerde begrippen ook kunnen toepassen of onthouden. Met andere woorden: het manipuleren en onthouden van verbale betekenisinhouden of semantische representaties, staat in deze subtests centraal. Dit laatste vormt teven het bindend element met Zinnen nazeggen, Verhaaltje vragen en Begrip en inzicht, zodat het ladingenpatroon op de eerste faktor (I) een redelijke ondersteuning vormt voor de faktoranalytische betekenis van deze drie tests.

Op de tweede faktor (II) laden alleen Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen. Dieper inzicht in de betekenis van deze faktor, welke wij aanduiden met 'Geheugen', kan helaas niet via de FST of NST verkregen worden, aangezien geen enkele subtest een lading van enige betekenis vertoont. Dit geldt met name voor de subtest Geheugen (7) uit de NST, waarbij het kind plaatjes moet herkennen en doorstrepen die in de eerste zes subtests voorkwamen. Gelet op de testtaak - en niet in de laatste plaats op de naam van deze test - zou men een redelijke lading op faktor II mogen verwachten. De relatief hoge lading op de eerste faktor en de zeer bescheiden lading op de derde faktor, welke altijd nog hoger is dan die op de tweede faktor, wijzen echter totaal niet op een geheugenfunctie. Dit laatste is trouwens conform eerdere bevindingen met deze subtest

(Van Alphen de Veer et al., op.cit.).

De derde faktor (III) levert nauwelijks enige moeilijkheden op bij de interpretatie en vormt als zodanig weer een redelijke ondersteuning voor de verkregen faktorstructuur van de LDT-E. Zo laden de subtests 1, 2 en 3 van de FST en de eerste drie subtests van de NST - te zamen met de subtests 5 en 10 van de NST de faktor 'visueel-motorische differentiatie' representerend - duidelijk op een en dezelfde faktor, evenals Blokpatronen en Natikken, welke deel uitmaken van de faktor 'Ruimtelijk-analytisch inzicht'. De visuele waarneming van (ruimtelijke) patronen (figuren, plaatjes) is duidelijk het bindend element tussen deze (sub)tests. Iets lagere ladingen dan .40 op de derde faktor laten Vouwblaadjes, Plaatjes aanwijzen en de subtests 5 en 10 van de NST zien. Gezien onze eerdere opmerkingen over de faktoriële betekenis van Vouwblaadjes (par. 6.3) en Plaatjes aanwijzen (par. 6.4), en gelet op eerdere bevindingen omtrent het tamelijk complexe karakter van de subtests 5 en 10 uit de NST (Van Alphen de Veer et al., op.cit.), bevestigen deze ladingen slechts onze vermoedens omtrent de specifieke betekenis van Vouwblaadjes als meest complexe test binnen het visueel-fijnmotorische kanaal en, in mindere mate ook van Plaatjes aanwijzen als een test met 'gemengde' kanaaleigenschappen.

7. PREDIKTIEVE VALIDITEIT

7.1 Korrelaties tussen LDT-E en criteriumvariabelen

Met het oog op de bepaling van de prediktieve validiteit van de LDT-E zijn de schoolprestaties in de eerste klas van het gewoon lager onderwijs (GLO) gemeten aan de hand van acht criteriumvariabelen. In paragraaf 4.4 zijn deze variabelen - bestaande uit een viertal toetsen enerzijds en het oordeel van de leerkracht met betrekking tot Lezen, Rekenen, Taal en Schrijven anderzijds - nader toegelicht.

Tabel 23 geeft de korrelaties tussen de criteriumvariabelen weer volgens een indeling naar drie groepen, respectievelijk Lezen/taal, Rekenen en Schrijven. Deze indeling vond plaats op basis van de (veronderstelde) inhoud van de variabelen en niet naar meetmethode (toetsen vs. oordeel leerkracht) of chronologische volgorde. Op een dergelijke manier hoopten wij het inzicht in de onderlinge (on)afhankelijkheid van de criteriumvariabelen te vergroten en daarmee ook het inzicht in het criterium schoolprestaties. Gezien de hoogte van de korrelaties tussen Taal en Lezen (volgens het oordeel van de leerkracht) enerzijds en de verschillende leestoetsen anderzijds, hebben wij de variabele Oordeel Taal in dezelfde groep als de leesvariabelen ondergebracht.

Zoals uit tabel 23 valt af te lezen, hangen de variabelen in de Lees/taalgroep nauw met elkaar samen: de verkregen korrelaties variëren van .67 (Oordeel Taal - Brus) tot .85 (Oordeel Lezen - Oordeel Taal). Met elkaar geven de variabelen uit deze groep een goede indicatie van de lees- en taalvaardigheid van het kind, te meer daar uit onderzoek van Zwarts (1976) gebleken is dat de toetsen Begrijpend lezen en de Brus niet alleen zeer betrouwbaar zijn (homogeniteitscoëfficiënten respectievelijk .90 en .95) maar ook valide; dat wil zeggen de toets Begrijpend lezen meet inderdaad de vaardigheid in het begrijpen van een tekst, terwijl de Brus primair de technische leesvaardigheid van het kind meet.

De twee variabelen uit de Rekengroep hangen aanzienlijk minder nauw met elkaar samen dan de variabelen uit de Lees/taalgroep. Een verklaring hiervoor is niet moeilijk te geven indien men op het verschil in hoogte let tussen de korrelaties van Rekenen (volgens het oordeel van de leerkracht) en de Schiedamse rekentest met alle overige variabelen. Zo variëren de korrelaties met Oordeel Rekenen van .51 voor de Brus en Oordeel Schrijven tot .63 voor Oordeel Taal. De korrelaties met de Schie-

Tabel 23. Korrelaties tussen acht criteriumvariabelen

N = 322

criteriumvariabelen	Lezen / taal				Rekenen		Schrijven	
	Begrijpend lezen I	Beeld- diktee I	Brus-Eén- Minuut-Test vorm B	Oordeel Lezen	Oordeel Taal	Schiedamse rekentest	Oordeel Rekenen	Oordeel Schrijven
<i>Lezen / taal</i>								
Begrijpend lezen I								
Beelddiktee I	.82							
Brus-Eén-Minuut-Test B	.71	.69						
Oordeel Lezen	.69	.72	.71					
Oordeel Taal	.69	.75	.67	.85				
<i>Rekenen</i>								
Schiedamse rekentest	.49	.37	.40	.30	.32			
Oordeel Rekenen	.54	.52	.51	.59	.63	.53		
<i>Schrijven</i>								
Oordeel Schrijven	.45	.44	.38	.49	.54	.21	.51	

damse rekentest variëren daarentegen van .21 voor Oordeel Schrijven tot .49 voor Begrijpend lezen. Kennelijk is de variantie van de Schiedamse rekentest specifiekere dan van Rekenen volgens het oordeel van de leerkracht; dit verklaart verder meteen hun relatief lage korrelatie van .53. Aangenomen dat de Schiedamse rekentest een betrouwbaar instrument is - helaas zijn ons hierover geen gegevens bekend - achten wij deze variabele dan ook een duidelijk betere maat voor de rekenvaardigheid van het kind dan Rekenen volgens het oordeel van de leerkracht³⁹.

Over de variabele Schrijven valt verder weinig te vertellen, aangezien de Schrijfgroep slechts door deze ene 'oordeels'variabele gerepresenteerd wordt. Opvallend zijn echter wel de tamelijk hoge korrelaties met Oordeel Taal (.54) en Oordeel Rekenen (.51). Mede gelet op de overige 'oordeels'korrelaties kunnen wij ons niet aan de indruk onttrekken dat (vragenlijst)methodevariantie een niet onaanzienlijk deel van de gevonden gemeenschappelijke variantie van de oordeelsvariabelen in beslag neemt. Zo korreleren de oordeelsvariabelen Lezen, Rekenen, Taal en Schrijven aanzienlijk hoger met elkaar dan men op basis van hun veronderstelde, geheel verschillende inhoud zou vermoeden. Daarnaast achten wij het niet onwaarschijnlijk dat de hoge korrelaties tussen de oordeelsvariabelen het bekende 'halo'effekt weerspiegelen.

Tabel 24 geeft een overzicht van de korrelaties tussen de tests en het totaal van de LDT-E enerzijds en de acht criteriumvariabelen anderzijds. Opvallend is dat de meest complexe tests binnen het V-F en A-S kanaal - respectievelijk Blokpatronen, Vouwblaadjes (V-F), Verhaaltje vragen en Begrip en inzicht (A-S) - de laagste korrelaties met de criteriumvariabelen vertonen, alhoewel Begrip en inzicht er nog het best van afkomt. De minst complexe tests daarentegen, welke wij eerder als 'geheugen'tests bestempelden (Natikken, Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen) laten de hoogste korrelaties zien. Uitgaande van de redenering dat de meest complexe tests ook de grootste variantie gemeen zouden hebben met tamelijk ingewikkelde (en ondoorzichtige) vaardigheden als lezen, rekenen, schrijven, etc., hadden wij eerder een omgekeerd korrelatiepatroon verwacht. Een mogelijke verklaring hiervoor biedt de interne consistentie van de tests: deze ligt voor de simpele 'geheugen'-tests gemiddeld zo'n 12 punten hoger dan voor de meer complexe tests (cf. par. 5.2).

Het totaal van de LDT-E - op te vatten als een maat voor de intelligentie of algemeen cognitieve ontwikkeling - laat verreweg de hoogste

Tabel 24. Korrelaties^a van de tests en het totaal met acht criteriumvariabelen

N = 322

criteriumvariabelen	Lezen / taal					Rekenen		Schrijven
	Begrijpend lezen	Beeld- diktee	Brus-Eén- Minuut-Test vorm B	Oordeel Lezen	Oordeel Taal	Schiedamse rekentest	Oordeel Rekenen	Oordeel Schrijven
<i>V - F^b</i>								
Natikken	.45	.35	.35	.32	.37	.30	.41	.27
Blokpatronen	.24	.22	.16	.13	.14	.23	.28	.29
Vouwblaadjes	.27	.22	.23	.27	.29	.22	.30	.27
<i>AV - F^c</i>								
Plaatjes aanwijzen	.45	.39	.44	.42	.42	.25	.41	.30
<i>A - S^d</i>								
Woordenspan	.38	.36	.32	.31	.29	.18	.31	.28
Zinnen nazeggen	.49	.48	.40	.40	.44	.25	.48	.37
Verhaaltje vragen	.29	.26	.25	.24	.29	.18	.28	.15
Begrip en inzicht	.35	.23	.24	.20	.24	.29	.35	.19
totaal	.59	.51	.48	.46	.52	.40	.58	.42

^a alle korrelaties zijn significant op 1%-niveau ($p < .01$)

^b V - F = visueel-fijnmotorische kanaal

^c AV - F = auditief/visueel-fijnmotorische kanaal

^d A - S = auditief-spraakmotorische kanaal

korrelaties zien met alle criteriumvariabelen. De korrelaties, i.c. predictieve validiteitscoëfficiënten, variëren van .40 voor de Schiedamse rekentest tot .59 voor Begrijpend lezen. Als zodanig is het 'totaal' de beste voorspeller van schoolprestaties in de eerste klas GLO, op de voet gevolgd door Zinnen nazeggen met korrelaties variërend van .25 voor de Schiedamse rekentest tot .49 voor Begrijpend lezen.

In dit verband is het opvallend te noemen dat Natikken van alle tests uit de LDT-E - inclusief Zinnen nazeggen - de hoogste korrelatie met de Schiedamse rekentest laat zien, namelijk .30. Voor het overige zijn de korrelaties tussen Natikken en de verschillende criteriumvariabelen in het algemeen iets lager dan de korrelaties tussen Plaatjes aanwijzen en de criteriumvariabelen. Als afzonderlijke voorspeller van schoolprestaties in de eerste klas GLO komt Plaatjes aanwijzen dan ook ná Zinnen nazeggen op de tweede plaats, zij het dat de korrelatie tussen Plaatjes aanwijzen en technisch lezen volgens de Brus iets hoger is (.44) dan de korrelatie tussen Zinnen nazeggen en de Brus (.40).

Proberen wij tenslotte de tests uit de LDT-E - als afzonderlijke voorspellers van schoolprestaties in de eerste klas GLO - globaal te rangschikken naar volgorde van belangrijkheid, dan ziet deze rangorde er voor de beste vier voorspellers per criteriumgroep als volgt uit:

- *Lezen / taal* 1. Zinnen nazeggen, 2. Plaatjes aanwijzen, 3. Natikken, 4. Woordenspan;
- *Rekenen* 1. Zinnen nazeggen en Natikken (ex equo), 2. Plaatjes aanwijzen en Begrip en inzicht (ex equo);
- *Schrijven* 1. Zinnen nazeggen, 2. Blokpatronen, 3. Woordenspan, 4. Vouwblaadjes en Natikken (ex equo).

Om te onderzoeken tot hoever de verkregen korrelaties kunnen stijgen indien alle tests uit de LDT-E - maar vooral de meest complexe - maximaal betrouwbaar zouden zijn, hebben wij alle korrelaties van tabel 24 éénzijdig gekorrigeerd voor 'attenuatie' (Drenth, 1975; p. 229); éénzijdig wil in dit verband zeggen dat wij alleen een correctie voor de (on)betrouwbaarheid (interne consistentie) van de tests uitgevoerd hebben en niet voor die van de criteriumvariabelen.

Na deze correctie stijgen alle korrelaties met zo'n 2 tot 7 punten, afhankelijk van de hoogte van de coëfficiënt voor interne consistentie van een test (r_{11}) en zijn korrelatie met een criteriumvariabele. Zo stijgt bijvoorbeeld de korrelatie van Vouwblaadjes ($r_{11} = .66$) met Oordeel Rekenen van .30 tot .37, terwijl de korrelatie van Zinnen nazeggen ($r_{11} = .87$)

Tabel 25. Kumulatieve multipele korrelaties^a (kum.R) van de tests met acht criteriumvariabelen

N = 322

Lezen / taal									
Begrijpend lezen I		Beeld-diktee I		Brus-Eén-Minuut-Test vorm B		Oordeel Lezen		Oordeel Taal	
kum.R		kum.R		kum.R		kum.R		kum.R	
Zinnen nazeggen	.49	Zinnen nazeggen	.48	Plaatjes aanwijzen	.44	Plaatjes aanwijzen	.42	Zinnen nazeggen	.44
Natikken	.59	Natikken	.52	Zinnen nazeggen	.49	Zinnen nazeggen	.47	Natikken	.52
Plaatjes aanwijzen	.61	Plaatjes aanwijzen	n.s. ^b	Natikken	.52	Natikken	.49	Plaatjes aanwijzen	.54
Rekenen					Schrijven				
Schiedamse rekentest		Oordeel Rekenen				Oordeel Schrijven			
kum.R		kum.R				kum.R			
Natikken	.30	Zinnen nazeggen	.48			Zinnen nazeggen	.37		
Begrip en inzicht	.37	Natikken	.56			Blokpatronen	.42		
Vouwblaadjes	n.s. ^b	Plaatjes aanwijzen	.57			Vouwblaadjes	.44		

met Oordeel Rekenen daarentegen slechts drie punten stijgt (van .48 tot .51).

Opvallend is verder dat het patroon van de korrelaties niet duidelijk door deze korrektie wordt beïnvloed: alle korrelaties 'groeien' weliswaar iets naar elkaar toe en zijn gemiddeld iets hoger, maar hun rangorde blijft vrijwel dezelfde. Konkluderend mogen wij dan ook stellen dat de prediktieve validiteitscoëfficiënten van de LDT-E bij maximale verhoging van de betrouwbaarheid der tests een geringe tot matige stijging zullen laten zien, zonder dat het patroon van deze coëfficiënten noemenswaard verandert.

7.2 Stapsgewijze multipele regressie

In tabel 25 vindt men de kumulatieve multipele korrelaties van de in rangorde geplaatste tests (prediktoren) met iedere criteriumvariabele afzonderlijk⁴⁰. Voor de overzichtelijkheid hebben wij ook hier weer de acht criteriumvariabelen ingedeeld in drie groepen, namelijk: Lezen/taal, Rekenen en Schrijven. Op enkele uitzonderingen na worden slechts die tests weergegeven die tot een statistisch significante verhoging van de multipele korrelatie op 1%-niveau leidden ($p < .01$).

Zoals men uit de tabel kan aflezen zijn in vrijwel alle gevallen drie tests voldoende om een nauwelijks voor verbetering vatbare multipele korrelatie met de desbetreffende criteriumvariabele te verkrijgen. Bij twee criteriumvariabelen, i.c. Beelddiktee en Schiedamse rekentest, zijn twee tests reeds voldoende.

Uit de kumulatieve multipele korrelaties met de criteriumvariabelen in de Lees/taalgroep blijkt dat de testkombinatie Zinnen nazeggen, Natikken en Plaatjes aanwijzen de lees-taalprestaties in de eerste klas GLO het best voorspelt (multipele korrelaties variërend van .49 voor Oordeel Lezen tot .61 voor Begrijpend lezen). Genoemde combinatie heeft zelfs een iets hogere voorspellende waarde dan het totaal van de LDT-E (cf. tabel 24), waarvoor korrelaties variërend van .46 (Oordeel Lezen) tot .59 (Begrijpend lezen) met dezelfde criteriumvariabelen werden gevonden. Dit laatste wil echter niet zeggen dat men nu de scores van een kind op Zinnen nazeggen, Natikken en Plaatjes aanwijzen maar gewoon bij elkaar moet optellen om de lees-taalprestaties beter te kunnen voorspellen dan met behulp van het LDT-E totaal of - ceteris paribus - het LDT-IQ. Voor de praktische toepassing van dergelijke testkombinaties dient men namelijk eerst de nodige regressievergelijkingen op te stellen. Aange-

zien de gepresenteerde bevindingen betrekking hebben op een experimentele versie van de LDT, heeft het weinig zin om deze vergelijkingen hier te geven. Belangrijker dan de praktische toepassing is daarom de konstatering dat Zinnen nazeggen, Natikken en Plaatjes aanwijzen als combinatie de beste voorspellers vormen van de lees-taalprestaties.

Ten aanzien van de rekenprestaties ligt de zaak minder eenvoudig; één enkele combinatie van tests als beste voorspeller van rekenprestaties in de eerste klas GLO valt niet te geven. Zo zijn Natikken en Begrip en inzicht de beste voorspellers van rekenprestaties indien de Schiedamse rekentest het criterium vormt voor deze prestaties (multiële korrelatie .37). Indien daarentegen het Oordeel Rekenen het criterium is dan zijn Zinnen nazeggen, Natikken en Plaatjes aanwijzen weer de beste voorspellers (multiële korrelatie .57).

Al eerder drukten wij onze scepsis uit over de validiteit van het oordeel van de leerkracht met betrekking tot de verschillende leerprestaties van een kind (cf. par. 7.1). In deze opinie worden wij nu gesterkt door de testkombinatie Zinnen nazeggen, Natikken en Plaatjes aanwijzen als beste voorspeller van de rekenprestaties (Oordeel Rekenen); een combinatie die identiek is aan de eerder gevonden combinatie met betrekking tot de lees-taalprestaties van het kind. Aangezien het weinig aannemelijk is dat lezen en rekenen op precies dezelfde functies een beroep doen en diensgevolge door precies dezelfde tests het best voorspeld zouden worden, hechten wij dan ook aan de combinatie Zinnen nazeggen, Natikken en Plaatjes aanwijzen als zogenaamde beste voorspeller van de feitelijke rekenprestaties van een kind weinig waarde. Een uitzondering willen wij echter maken voor Natikken, aangezien deze test - in combinatie met Begrip en inzicht - de rekenprestaties op de Schiedamse rekentest het best voorspelt; zij het dat de verkregen multiële korrelatie van .37 bepaald niet hoog genoemd kan worden, zeker niet in vergelijking tot de korrelatie van het LDT-E totaal met de Schiedamse rekentest (.40), cf. tabel 24.

De kumulatieve multiële korrelaties van respektievelijk Zinnen nazeggen, Blokpatronen en Vouwblaadjes met het Oordeel Schrijven spreken verder voor zich. Wel willen wij nog opmerken dat de multiële korrelatie van de combinatie van deze drie tests iets hoger (.44) is dan de korrelatie van het LDT-E totaal met Oordeel Schrijven (.42).

Eenzelfde stapsgewijze multiële regressie-analyse hebben wij tenslotte ook uitgevoerd voor jongere en oudere kleuters afzonderlijk, evenals

voor jongens en meisjes en de verschillende SEN-groepen. Het resultaat hiervan is in kort bestek als volgt⁴¹.

In bijna alle gevallen bleken twee tests voldoende om een vrijwel maximale multiële korrelatie met de criteriumvariabelen te verkrijgen. Voor de Lees/taalgroep waren dit Zinnen nazeggen en Natikken, weinig verrassend dus. Voor de Rekengroep daarentegen werd een afwijkend resultaat gevonden, met name in SEN-III: zo bleek Blokpatronen een belangrijke voorspeller te zijn van zowel de Schiedamse rekentest als van Oordeel Rekenen [kumulative R van Blokpatronen en Natikken met Schiedamse rekentest respectievelijk .30 en .36; kumulative R van Zinnen nazeggen en Blokpatronen met Oordeel Rekenen respectievelijk .42 en .51.] In alle overige subgroepen (jongens, meisjes, etc.) konden wij verder geen duidelijke verschillen konstateren, althans vergeleken met de bevindingen in tabel 25 voor de gehele onderzoeksgroep. Voor de Schrijfgroep werd vergelijkenderwijs evenmin een duidelijk afwijkend resultaat verkregen.

7.3 Canonische korrelaties tussen LDT-E en criteriumvariabelen

Een techniek waarmee men de onderliggende samenhang en betekenis van een verzameling prediktorvariabelen enerzijds en een verzameling criteriumvariabelen anderzijds kan exploreren, is canonische korrelatie-analyse. In het kort komt deze techniek hierop neer dat men zo veel mogelijk variantie probeert te beschrijven van de relatie tussen de verzamelingen prediktor- en criteriumvariabelen door zodanige lineaire combinaties van variabelen te vormen dat de samenhang tussen de twee lineaire combinaties zo hoog mogelijk wordt. Deze samenhang wordt uitgedrukt in de canonische korrelatiecoëfficiënt R_C , terwijl het kwadraat van de canonische korrelatie (R_C^2) de hoeveelheid gemeenschappelijke variantie weergeeft. Sommige auteurs - onder anderen Nie et al. (1975; p. 515 e.v.) - vergelijken canonische korrelatie-analyse wel met principale componentenanalyse om daarmee het begrip 'canonische variaat' te kunnen uitleggen, de tweede vorm van informatie die een canonische korrelatie-analyse ons verschaft (de eerste is R_C). Volgens deze uitleg mag men een canonische variaat als een soort 'faktor' opvatten, opgebouwd uit de ladingen van de variabelen op deze faktor voor een specifieke lineaire combinatie van prediktor- en criteriumvariabelen. Korrekter is echter de zienswijze van Kerlinger & Pedhazur (1973; p. 341 e.v.) die canonische korrelatie-analyse beschouwen als een gegeneraliseerde vorm van multiële regressie-

Tabel 26. Canonische korrelaties (R_c)^a tussen prediktor- en criteriumvariabelen en de gewichten van deze variabelen in drie combinaties (I, II en III)^b voor twee canonische variaten (can. var.) per combinatie

variabelen	canonische gewichten					
	I ^b		II ^b		III ^b	
	can. var. 1	can. var. 2	can. var. 1	can. var. 2	can. var. 1	can. var. 2
CRITERIUMVARIABLEN						
<i>Toetsen</i>						
Begrijpend lezen I	-.57	.05	-.75	-.93		
Beelddiktee I	-.01	.47	.08	1.19		
Brus-Eén-Minuut-Test, vorm B	-.13	-.42	-.25	.44		
Schiedamse rekentest	.03	.37	-.21	-.60		
<i>Oordeel leerkracht</i>						
Oordeel Lezen	.30	-.49			-.02	-.02
Oordeel Rekenen	-.52	.20			.61	.21
Oordeel Taal	-.17	-.73			.39	-1.06
Oordeel Schrijven	-.08	.73			.17	1.06
PREDIKTORVARIABLEN						
<i>LDT-E</i>						
Blokpatronen	-.09	.81	-.07	-.11	.09	.74
Vouwblaadjes	-.17	-.02	-.12	-.12	.22	.13
Natikken	-.40	-.18	-.40	-.19	.32	-.38
Woordenspan	.00	.19	-.06	.04	.00	.28
Plaatjes aanwijzen	-.24	-.67	-.30	.20	.25	-.38
Zinnen nazeggen	-.37	.21	-.30	.83	.46	.26
Verhaaltje vragen	-.04	-.41	.00	.02	.01	-.69
Begrip en inzicht	-.21	.44	-.24	-1.00	.14	.23
canonische korrelatie (R_c) ^a	.69	.30	.61	.24	.64	.27
χ^2	232.80	64.36	164.86	31.09	197.04	41.04
vrijheidsgraden (df)	64	49	32	21	32	21
kans (p)	.00	.07	.00	.07	.00	.01
aantal (N)	275	275	292	292	307	307

^a gebaseerd op de korrelatiematrices, zoals weergegeven in tabel 12, 23 en 24

^b I = LDT-E en Toetsen plus Oordeel leerkracht; II = LDT-E en Toetsen; III = LDT-E en Oordeel leerkracht

analyse met k onafhankelijke en m afhankelijke variabelen. Een canonische variataat bestaat dan uit de verzameling regressiecoëfficiënten of -gewichten van een specifieke lineaire combinatie van prediktor- en criteriumvariabelen. Hierbij is de grootte van de gewichten indicatief voor het belang dat men aan de variabelen mag hechten bij de interpretatie van het onderliggende verband tussen de twee verzamelingen variabelen. Canonische korrelatie-analyse kan meer dan één R_c plus canonische variataat opleveren, aangezien verschillende lineaire combinaties van de verzamelingen prediktor- en criteriumvariabelen te vormen zijn. Of een tweede dan wel derde, vierde etc. canonische variataat meer dan toevallige betekenis heeft en daarom weergegeven wordt in een tabel, hangt echter af van de significantie van de desbetreffende R_c , te toetsen met behulp van de χ^2 -toets.

Een laatste opmerking betreft de relatie tussen de verschillende canonische variaten. De eerste canonische variataat beschrijft altijd de grootste variantiebron van de variabelen. De variantie die na extractie van deze eerste variantiebron overblijft wordt vervolgens onafhankelijk daarvan beschreven door de tweede canonische variataat, etc. Hieruit volgt dat de verschillende canonische korrelaties en hun kwadraten steeds kleiner worden.

Uitgaande van de korrelatiematrices, zoals weergegeven in tabel 12, 23 en 24, zijn de canonische korrelaties berekend tussen de acht prediktorvariabelen (tests) van de LDT-E enerzijds en drie verschillende combinaties van criteriumvariabelen anderzijds, i.c. Toetsen plus Oordeel leerkracht (I), Toetsen (II) en Oordeel leerkracht (III)⁴². In tabel 26 worden deze canonische korrelaties weergegeven, evenals de gewichten van de variabelen in de drie combinaties voor telkens twee canonische variaten per combinatie.

Alle eerste canonische korrelaties per combinatie zijn significant op 1%-niveau ($p < .01$). Alle tweede canonische korrelaties per combinatie zijn daarentegen niet significant op 5%-niveau ($p > .05$), met uitzondering van de derde combinatie (III), waarvoor een kans p van .006 gevonden werd (in tabel 26 afgerond tot .01). Aan de tweede canonische korrelaties (2) van de combinaties I en II zullen wij dan ook in eerste instantie geen aandacht besteden; hetzelfde geldt voor de gewichten van de variabelen welke de tweede canonische variataat vormen.

Voor combinatie I bedraagt de waarde van de eerste canonische korrelatie .69. Te oordelen aan de grootte van de gewichten van de eerste canonische

variaat (I-1) dragen Natikken en Zinnen nazeggen van de prediktorvariabelen en Begrijpend lezen, Oordeel Rekenen en Oordeel Lezen van de criteriumvariabelen, het meest bij aan deze korrelatie. Anders geformuleerd: Natikken, Zinnen nazeggen, Begrijpend lezen, Oordeel Rekenen en Oordeel Lezen nemen het grootste deel van de 48% gemeenschappelijke variantie van de LDT-E en schoolprestaties (Toetsen plus Oordeel leerkracht) in beslag.

In combinatie II is de waarde van de eerste canonische korrelatie gedaald tot .61, aangezien de criteriumvariabelen 'Oordeel leerkracht' in de analyse zijn weggelaten. Desondanks hebben de acht tests van de LDT-E en de vier overblijvende toetsen nog 37% van de variantie gemeen. De gewichten van de eerste canonische variaat (II-1) laten ongeveer hetzelfde beeld zien als bij de eerste combinatie (I-1). Zo dragen Natikken, Zinnen nazeggen en nu ook Plaatjes aanwijzen, evenals Begrijpend lezen, relatief het meest bij aan de verkregen canonische korrelatie van .61.

De derde combinatie (III) van variabelen (LDT-E en Oordeel leerkracht) levert twee significante canonische korrelaties op, respectievelijk .64 (III-1) en .27 (III-2). Op de eerste canonische variaat (III-1) krijgen Natikken en Zinnen nazeggen, evenals Oordeel Rekenen en Oordeel Taal de hoogste gewichten. Ook hier vinden wij dus - met uitzondering van Oordeel Taal - ongeveer hetzelfde beeld terug als bij de eerste combinatie (I-1), zij het dat de gemeenschappelijke variantie van de prediktor- en criteriumvariabelen lager is, namelijk 41%. Op de tweede canonische variaat (III-2) - die zoals gezegd onafhankelijk is van de eerste en slechts 7% variantie beschrijft - laten Blokpatronen, Natikken, Plaatjes aanwijzen, Verhaaltje vragen, Oordeel Taal en Oordeel Schrijven de hoogste gewichten zien.

Hoe moet men nu bovenstaande uitkomsten interpreteren? Kerlinger en Pedhazur (op.cit.) merken hierover het volgende op:

"When studying the underlying relations between two sets of variables, there has to be some reasonable source of common variance in the two sets of variables. One can easily see that if one set of measures presumably reflects an underlying phenomenon or construct and the second set, similarly, reflects another related phenomenon, then canonical correlation is appropriate and valuable." (p. 345).

Letten wij alleen op die variabelen van iedere eerste canonische variaat die stelselmatig een hoog gewicht krijgen, dan valt meteen op dat Natikken en Zinnen nazeggen van de prediktorvariabelen, en Begrijpend lezen en Oordeel Rekenen van de criteriumvariabelen de hoofdmoot van de gemeenschappelijke variantie voor zich opeisen. In het licht van onze

vorige analyses (par. 7.1 en 7.2), mag dit resultaat weinig verrassend genoemd worden. Uitgaande van de gedachtengang dat deze vier variabelen ongeveer een-en-hetzelfde onderliggende 'konstrukt' of psychologisch begrip reflekteren, is dit resultaat echter weer wel opmerkelijk. Zo blijken tamelijk simpele 'geheugen'tests als Zinnen nazeggen en Natikken een zeer nauwe samenhang te hebben met relatief complexe vaardigheden als het begrijpen van een tekst en rekenen volgens het oordeel van de leerkracht. Gegeven de validiteit van de toets Begrijpend lezen (cf. Zwarts, 1976) - het hoogstwaarschijnlijk minder valide Oordeel Rekenen laten wij hier buiten beschouwing - lijkt deze tegenstrijdigheid alleen maar te verklaren indien we aannemen dat qua functies simpele tests als Zinnen nazeggen en zeker Natikken minder simpel zijn dan wij aanvankelijk veronderstelden. Met andere woorden: Zinnen nazeggen en Natikken meten naast de veronderstelde onmiddellijk geheugenfunctie voor zinnen en ruimtelijke sekwenties (cf. tabel 3) ook nog een meer complexe functie. Het ligt in de lijn van eerdere analyses en redeneringen (cf. par. 6.2 en 6.4) om deze meer complexe functie aan te duiden met de naam 'coderen'. Onmiddellijk geheugen en coderen vormen dan het onderliggende psychologische begrip van de variabelen Begrijpend lezen, Oordeel Rekenen, Zinnen nazeggen en Natikken.

Interpretatie van de tweede canonische variaten is aanzienlijk lastiger aangezien alleen de derde combinatie (III-2) een significant resultaat heeft opgeleverd. Indien wij echter het net niet significante karakter van de tweede canonische variaten in de eerste twee combinaties buiten beschouwing laten - tenslotte heeft de keuze van een kans $p < .05$ altijd iets arbitrairs - dan zijn er desondanks wel enkele trends te bespeuren. Zo leidt vergelijking van de tweede canonische variaten uit de eerste en derde combinatie (I-2 vs. III-2) tot de bevinding dat met name Oordeel Taal en Oordeel Schrijven van de criteriumvariabelen en Blokpatronen, Plaatjes aanwijzen en Verhaaltje vragen van de prediktorvariabelen het merendeel van de gemeenschappelijke variantie bepalen. Eenzelfde vergelijking, maar dan voor de eerste en tweede combinatie (I-2 vs. II-2), suggereert weer dat de toets Beelddiktee - en in mindere mate ook de Schiedamse rekentest en de Brus-Eén-Minut-Test - primair verwantschap vertonen met Begrip en inzicht van de LDT-E. Bij beide vergelijkingen is het echter tamelijk onduidelijk welk onderliggend psychologisch begrip hierdoor gereflekteerd wordt.

Eenzelfde canonische korrelatie-analyse hebben wij tenslotte ook uitgevoerd voor jongere en oudere kleuters afzonderlijk, evenals voor jongens

en meisjes en de verschillende SEN-groepen. In alle subgroepen bleken Begrijpend lezen en Natikken van alle prediktor- en criteriumvariabelen het zwaarste gewicht te krijgen op de eerste canonische variaat. Met betrekking tot de overige variabelen werd geen consistent patroon van gewichten gevonden over de verschillende subgroepen, nòch op de eerste nòch op de tweede canonische variaat voor zover deze laatste al significant was⁴¹.

8. DISKUSSIE

Een vruchtbaar uitgangspunt voor de discussie is de vraag in hoeverre de resultaten van het exploratieve onderzoek naar de betekenis en voorspellende waarde van de LDT-E een bevestiging of ontkenning vormen van de pretenties die wij met dit meetinstrument voor ogen hebben, namelijk: de bepaling van zowel het algemeen cognitief ontwikkelingsniveau, c.q. intelligentieniveau van schoolverlatende kleuters, als het niveau van enkele cognitieve functies die van belang zijn voor het leren op school. Bij voorbaat willen wij stellen dat een definitief antwoord op deze vraag niet te geven valt, niet alleen omdat de bevindingen van het empirisch onderzoek in een exploratief in plaats van hypothese-toetsend kader zijn gepresenteerd, maar ook omdat valideringsonderzoek - toetsend dan wel exploratief - in principe oneindig lang voortgezet kan worden. Daarnaast vormt het door praktische omstandigheden afgedwongen gebrek aan adequate referentievariabelen evenzeer een belemmering voor het geven van een meer definitief antwoord.

Een en ander heeft tot gevolg dat wij in dit afsluitende hoofdstuk meer onderstellingen dan stellingen zullen opperen omtrent de betekenis en voorspellende waarde van de LDT-E. Aangezien het hier om een experimentele versie van de LDT gaat, is het bovendien maar de vraag of duidelijker uitspraken wel zin hebben. Veeleer beschouwen wij de resultaten van dit onderzoek als hopelijk nuttige informatie- en inspiratiebron voor de praktiserend psycholoog die met de LDT werkt enerzijds en voor de onderzoeker-psycholoog die tot begripsvalidering van de LDT overgaat anderzijds. Wat dit laatste betreft wachten wij op de resultaten van een onderzoek naar de relatie tussen de LDT en enkele belangrijke intelligentietests, i.c. AKIT (Drenth et al., 1968), SON 2½-7 (Snijders & Snijders-Oomen, 1975), WPPSI en WISC-R (Wechsler, 1967; 1974), evenals op de bevindingen van een LDT-onderzoek bij leerlingen van de LOM-school (Van Alphen de Veer, in voorbereiding).

Terugkerend naar de vraag in de aanhef van dit hoofdstuk zullen wij deze in drie paragrafen proberen te beantwoorden. Iedere paragraaf gaat in op één hoofdaspekt van de LDT-E, te beginnen met het model dat aan deze testen grondslag ligt (par. 8.1). In de daaropvolgende paragraaf (8.2) wordt de LDT-E als intelligentietest onder de loep genomen. De laatste paragraaf (8.3) vormt de afsluiting van dit hoofdstuk en behandelt de LDT-E als diagnostische test. Paragraaf 8.3 vormt tevens het slot van deel A 'Cognitieve ontwikkeling en schoolprestaties'.

8.1 Mark-model: terugblik en vooruitzicht

Bestudering van de testliteratuur rondom het thema 'intelligentie en cognitie' is een ontmoedigende aangelegenheid. Niet alleen is de literatuur uiterst omvangrijk - men denke slechts aan het werk van Cattell (1971) en Guilford (1967) - maar ook is overeenstemming omtrent de aard en structuur van de intelligentie nog ver te zoeken. Dit gebrek aan overeenstemming is onder andere te danken aan het feit dat men op basis van faktor-analytische onderzoeksresultaten, verkregen met behulp van vaak zeer uiteenlopende technieken (cf. Harman, 1967), een model van de intelligentie probeerde te ontwikkelen, terwijl juist het omgekeerde voor de hand ligt, namelijk: begin met een of ander intelligentiemodel of theorie over cognitieve processen en interpreteer op basis daarvan de tests en de al dan niet via factoranalyse verkregen onderzoeksresultaten (cf. Carroll, 1976). In feite hebben wij deze laatste, weinig orthodoxe werkwijze gevolgd door uit te gaan van het Mark-model, zowel bij de beschrijving en samenstelling van de LDT-E als bij het daaropvolgende exploratieve onderzoek naar de betekenis van deze test. De vraag is nu in hoeverre het Mark-model inderdaad zo'n nuttige leidraad is geweest als wij in hoofdstuk 2 en 3 suggereerden.

Te beginnen met het 'kanaal'-begrip zullen wij opnieuw de verschillende elementen van het Mark-model ter discussie stellen, mede in het licht van enkele literatuurgegevens en de door ons verkregen onderzoeksresultaten.

Hoe zinnig is een onderscheid naar kanalen in een model - en dientengevolge ook naar kanaalkenmerken van een test - waarbij niet alleen de invoer (modaliteit) maar ook de uitvoer (motorisch antwoord) nauwkeurig gespecificeerd is? Een tamelijk overbodige vraag zal men wellicht denken, aangezien informatieverwerking altijd een in- en uitvoer impliceert. Toch zijn wij in de literatuur ettelijke voorbeelden tegengekomen van modellen waarbij de uitvoer van informatie 'domweg' vergeten was. Een sprekend voorbeeld hiervan is het bekende 'distributive memory model' van Hunt (1971), waarin op geen enkele wijze gerefereerd wordt aan een motorisch antwoord als laatste schakel in een keten van cognitieve functies en processen. Een onbetekenend detail misschien, maar geen enkele neuroloog zal de aard en de sterkte van de reacties van een patient op de toediening van prikkels over het hoofd zien, juist omdat hij op de hoogte is van het neuronale verband tussen sensoren en effectoren, althans voor de

lagere neuronale niveau's.

Wat voor de neuroloog geldt, is van zo mogelijk nog groter belang voor de psycholoog. Niet alleen heeft hij minder inzicht in de aard en structuur van zijn studie-objekt, maar ook beschikt hij over geen enkel ander tastbaar aanknopingspunt voor zijn testonderzoek dan de in- en uitvoer van het menselijk organisme (cf. par. 2.1). Het is dan ook enigszins wonderlijk om bij herhaling te moeten konstateren dat alleen de modaliteit als variabele in de opzet en uitwerking van een onderzoek betrokken wordt met veronachtzaming van de motorische antwoord-variabele.

Een typisch voorbeeld hiervan is het onderzoek van Torgeson et al. (1978) waarin het effect van twee taakvariabelen bij een Cijferspan-test werd bestudeerd. Deze twee variabelen waren respectievelijk taakstructuur (gelijktijdige vs. successieve presentatie van cijfers) en modaliteit (visuele vs. auditieve presentatie van cijfers). Uit dit onderzoek bleek onder andere dat de variabele 'taakstructuur' tot aanzienlijk grotere verschillen in de testprestaties op Cijferspan leidde dan de variabele 'modaliteit'. De vraag is echter of dit onderzoeksresultaat ook verkregen zou zijn indien het motorisch antwoord (bijvoorbeeld fijnmotorisch vs. spraakmotorisch) als variabele in het onderzoek was opgenomen. Nu konkludeerden Torgeson et al. terecht dat taakstructuur een veel belangrijker variabele was dan modaliteit, daarmee echter wel de suggestie wekkend dat de modaliteit waarop een test zich richt er niet veel toe doet. Psychologen die ooit slechtziende of slechthorende kinderen hebben moeten onderzoeken, weten wel beter. Uit persoonlijke ervaring is ons voorts bekend dat een ogenschijnlijk zwakbegaafde kleuter niet minder dan 30 punten in IQ steeg nadat een ernstige visusafwijking was gecorrigeerd. Wij kunnen ons voorstellen dat spraakmotorische stoornissen evenzeer de prestaties op een test als Zinnen nazeggen kunnen beïnvloeden omdat een spraakmotorisch antwoord wordt verlangd.

Uit het feit dat men tot nog toe tamelijk nonchalant met de specificatie van in- en uitvoerkenmerken van een test is omgesprongen, zou men kunnen afleiden dat deze kenmerken bij niet duidelijk zintuigelijk of motorisch gestoorde kinderen er weinig toe doen, althans in verhouding tot de meer of minder complexe functies die door deze tests gemeten worden. In ons onderzoek hebben wij dit helaas voor de afzonderlijke in- en uitvoerkenmerken van de tests uit de LDT-E niet kunnen nagaan, aangezien de mogelijkheid tot vergelijking ontbrak. Zo zijn bijvoorbeeld in de LDT-E alleen tests voor het auditief-spraakmotorische kanaal vertegenwoordigd en niet voor het auditief-fijnmotorische kanaal. Wel konden wij konstateren

dat de kanaalvariabele, waarin de in- en uitvoervariabelen een vaste combinatie vormen (auditief-spraakmotorisch vs. visueel-fijnmotorisch), een bron van individuele verschillen in testprestaties vertegenwoordigt, zij het geen overtuigende. Men vergelijkte hiertoe onze opmerkingen in paragraaf 6.2 over de kanaalafhankelijkheid van de verzameling functies, zoals gemeten door de LDT-E.

Zoals gezegd speelt de kanaalvariabele, vergeleken bij alle andere variabelen (elementen) van het Mark-model en de tests die volgens dit model beschreven zijn, geen overtuigende rol als verklarende variabele in ons onderzoek naar verschillen in testprestaties. Met name geldt dit voor een test als Natikken, die volgens eerdere analyses niet alleen een veronderstelde kanaalafhankelijke functie als onmiddellijk geheugen voor ruimtelijke sekwenties zou meten, maar ook een kanaalafhankelijke, centrale mediërende functie als 'coderen'. Ofschoon dit laatste onzes inziens goed te verdedigen valt, vormt de veronderstelling dat meer complexe functies als bijvoorbeeld 'coderen' kanaalafhankelijk zijn slechts een gedeeltelijke verklaring voor de minder overtuigende rol van de kanaalvariabele. Als namelijk meer complexe functies kanaalafhankelijk zouden zijn dan valt niet goed uit te leggen waarom qua functies tamelijk complexe tests als bijvoorbeeld Blokpatronen en Vouwblaadjes uit het visueel-fijnmotorische kanaal hoger met elkáár korreleren dan met tamelijk complexe tests als Verhaaltje vragen en Begrip en inzicht uit het auditief-spraakmotorische kanaal. Enkele aanvullende onderstellingen zijn daarom noodzakelijk.

In de eerste plaats: het belang van een indeling naar kanalen neemt af naarmate de in- en uitvoer van een organisme beter functioneert en wel ten gunste van andere variabelen (cognitieve functies) die door een test gemeten worden.

Ten tweede: gegeven een qua in- en uitvoer redelijk goed functionerend organisme, zal de samenhang tussen tests met verschillende of gelijke kanaalkenmerken hoofdzakelijk bepaald worden door de functies die deze tests gemeenschappelijk meten. In dit licht bezien hoeven de korrelaties van Blokpatronen en Vouwblaadjes met bijvoorbeeld Verhaaltje vragen en Begrip en inzicht geen verwondering te wekken, althans in vergelijking tot de korrelaties van Natikken met laatst genoemde tests.

Kort samengevat is het belang van een onderscheid naar kanalen dan ook dit, dat de kanaalvariabele een potentiële bron van individuele verschillen in testprestaties vormt welke diagnostisch gezien niet zonder meer

genegeerd mag worden.

Al eerder schreven wij dat het tweede element van het Mark-model - de verzameling van 34 functies - de zwakke plek vormt in dit model. Niet alleen omdat psychologische tests zich moeilijk laten beschrijven in termen van deze functies, maar ook omdat de onderscheiden functies te gering in aantal zijn voor een adequate beschrijving van tests die bij 5- en 6-jarigen worden afgenomen (cf. par. 2.3). Ter oplossing van dit probleem hebben wij daarom onze toevlucht genomen tot de bestaande testliteratuur (cf. par. 3.2), waarin de verschillende (deel)betekenissen van de tests uit de LDT-E zo goed en zo kwaad als het gaat beschreven staan. Helaas zijn deze beschrijvingen in vrijwel alle gevallen traditioneel van karakter - dat wil zeggen vanuit de psychometrie ontstaan - en bepaald niet gefundeerd op een of andere cognitieve theorie waarin allerlei interne processen en functies met betrekking tot de menselijke informatieverwerking uiteengezet worden. Dit laatste was voor ons nu juist een van de aantrekkelijke kanten van Mark's theorie over de ontwikkeling van leer- en denkprocessen.

Het resultaat van onze 'vlucht' in de bestaande testliteratuur is dat het Mark-model een hybridisch karakter heeft: enerzijds bevat het elementen of begrippen (kanalen, functies etc.) die duidelijk aan de cognitieve psychologie ontleend zijn, anderzijds bevat het elementen of begrippen die een meer traditionele, psychometrische benaderingswijze verdragen, vooral tot uiting komend in de (functie)beschrijvingen van de tests. Het is dan ook niet verwonderlijk dat in de beschrijving en analyse van onze onderzoeksgegevens (hfdst. 5-7) eveneens dit tweeslachtige karakter tot uitdrukking komt. Vooral in de paragrafen over de factorstructuur van de LDT-E moesten wij herhaaldelijk terugvallen op het oude vertrouwde, tamelijk subjectieve interpretatiekader van een psychometrisch georiënteerd betekenisonderzoek.

Sinds 1972 - het jaar waarin de LDT-E en daarmee ook de LDT gekoncipieerd werd - heeft de tijd echter niet stilgestaan. Zo liet John B. Carroll in 1976 een publikatie het licht zien onder de titel "Psychometric tests as cognitive tasks: a new 'Structure of Intellect'". In zekere zin is het jammer dat wij deze publikatie pas zo laat onder ogen kregen. Uitgaande van het 'distributive memory model' van Hunt (1971), aangevuld met het door Newell (1973) geïntroduceerde begrip 'production system', geeft Carroll namelijk een aanzienlijk gedetailleerder model van het cognitief functioneren dan waarover wij beschikken in de vorm van het Mark-model,

dat uiteindelijk zijn ontstaansgrond heeft in het begin van de jaren zestig. Bovendien is Carroll in staat om de resultaten van faktoranalytisch onderzoek te 'vertalen' in termen van zijn model, te oordelen naar de manier waarop hij 24 factoren uit de 'Kit of Reference Tests for Cognitive Factors' van French et al. (1963) zeer nauwkeurig analyseert en beschrijft. Hij maakt daartoe gebruik van een codeerschema dat niet minder dan vijf pagina's taak- c.q. testvariabelen omvat en helaas te uitgebreid is om hier weer te geven. Wèl kunnen wij enkele voorbeelden geven van de manier waarop Carroll zijn factoren (en de tests die daarop betrekking hebben) omschrijft.

"Factor CS (Speed of Closure) requires a search of a LTM (Long Term Memory) visual-representational memory store for a match for a partially degraded stimulus cue. Individual differences appear primarily in the rate of this search, but the probability of certain special strategies may also be involved: (1) Searching and utilizing hypotheses drawn from associations in LTM, (2) (consciously) searching in different portions of LTM, and (3) restructuring the perception of the stimulus (an operation involving the addressing of a sensory buffer, and similar to the alternation of the perception of ambiguous figures such as the Necker cube).

Factor MS (Memory Span) involves storage and retrieval of information (nonspecific as to modality) in STM (Short Term Memory). The capacity of STM for this operation is the primary individual difference determiner. Strategies of chunking or grouping stimulus elements may be helpful to some subjects" (p. 49).

Het door Carroll gehanteerde model, evenals het daarop gebaseerde codeerschema voor factoren en tests, biedt onzes inziens uitstekende perspectieven voor een versteviging van de 'zwakke plek' in het Mark-model. Uitbreiding en/of herformulering van het Mark-model in de richting zoals deze door Carroll geschetst wordt, lijkt ons dan ook alleszins de moeite van het nastreven waard.

Hoe gedegen Carroll's model ook in elkaar zit en hoe nauwkeurig hij ook allerlei tests in termen van zijn model kan beschrijven, toch zal men in zijn publikatie tevergeefs zoeken naar een taxonomie waarin alle door hem onderscheiden variabelen op een of andere wijze systematisch aan elkaar gerelateerd zijn. Zowel het eerste, het derde als het vierde element van het Mark-model - respectievelijk de indeling naar kanalen en de hiërarchische organisatie en ontogenie van functies - geven in ieder geval de principes aan van de manier waarop men zich een bepaalde, niet noodzakelijkerwijs korrekte taxonomie van functies kan voorstellen. Voor het ontbreken van een taxonomie geeft Carroll de volgende verklaring:

"The multifaceted nature of psychometric tasks also further supports the conclusion that it is impossible, in principle, to construct a 'structure

of intellect' model containing an n -way classification of tasks such that a 'factor' can be found for each cell in the classification" (p. 52).

Met andere woorden: het feit dat een test meer dan één functie tegelijkertijd meet (cf. par. 3.2) verhindert de opstelling van een n -weg klassifikatiesysteem zoals bijvoorbeeld door Guilford (1967) is voorgesteld. Het gevolg hiervan is echter wel dat de samenstelling van een testbatterij, geënt op Carroll's model, een tamelijk arbitraire aangelegenheid wordt, tenzij men op puur empirische gronden - bijvoorbeeld via test-criterium relaties - tot samenstelling overgaat.

Zoals wij in hoofdstuk 2 en 3 beargumenteerd hebben, is een meer systematische benadering - uitgaande van onder andere de hiërarchische organisatie en ontogenie van functies - niet volstrekt uitgesloten. In hoofdstuk 6 hebben wij vervolgens via een simplexanalyse onderzocht in hoeverre de veronderstelde hiërarchische rangschikking naar functiecomplexiteit van de tests uit de LDT-E ook empirisch gedemonstreerd kon worden. De resultaten van deze analyse suggereren dat het inderdaad mogelijk is om een redelijk valide hiërarchische organisatie van functies in de testpraktijk te realiseren ongeacht het feit dat tests meer dan een functie tegelijkertijd meten. Zo werd een perfecte simplex bereikt voor de tests in het V-F kanaal [hiërarchische rangschikking: Natikken (minst complexe test), Blokpatronen en Vouwblaadjes (meest complexe test)], terwijl voor het A-S kanaal een quasi-simplex gekonstrueerd kon worden [hiërarchische rangschikking: Woordenspan (minst complex), Zinnen nazeggen en (ex equo) Verhaaltje vragen en Begrip en inzicht (meest complexe tests)]. In paragraaf 3.4 hebben wij reeds uiteengezet wat de voordelen zijn van een dergelijke hiërarchie indien men de LDT-E wil uitbreiden met andere tests. Ofschoon in de experimentele ontwikkelingspsychologie voldoende evidentie aangedragen is voor een hiërarchische rangschikking van functies welke in de loop van de ontwikkeling van het kind tot stand komt - men denke slechts aan Bruner (1968) en Piaget (1960) - verlenen de resultaten van ons onderzoek met de LDT-E geen steun aan deze evidentie. Weliswaar werden voor vier tests significante leeftijdseffekten verkregen in die zin dat oudere kinderen (5;11-6;04) betere testprestaties leverden op Blokpatronen, Natikken, Plaatjes aanwijzen en Begrip en inzicht dan jongere kinderen (5;05-5;10), maar een daaropvolgende faktoranalyse op beide leeftijdsgroepen afzonderlijk leverde geen verschil van betekenis op in de faktorstructuur van de LDT-E (cf. par. 6.4). Uit de combinatie van deze gegevens werd vervolgens de konklusie getrokken dat er eerder sprake was van 'rijping' van sommige functies, zoals gemeten door genoemde

vier tests, dan dat er méér (complexere) functies tot ontwikkeling gekomen zouden zijn in de oudste leeftijdsgroep als bewijs voor het ontogenetisch aspekt in de gevonden hiërarchie. Voorzichtigheidshalve voegden wij hieraan toe dat de LDT-E of de gebruikte analysetechniek mogelijk weinig gevoelig is voor een zich feitelijk voltrekkende ontwikkeling. Daar komt bij dat ook de spreidingsbreedte van de leeftijd (5;05-6;04) niet al te groot is. Ter relativering van het verkregen resultaat willen wij bovendien nog in herinnering brengen dat het onderhavige onderzoek 'slechts' exploratief van aard is en zeker niet gericht is op toetsing van het Mark-model.

Het laatste neemt echter niet weg dat faktoranalyse - evenals trouwens simplexanalyse - inderdaad niet de meest geëigende techniek lijkt voor een valideringsonderzoek naar de hiërarchische organisatie en ontwikkeling van functies zoals gemeten door tests. De Corte en Van Bowel (1978) komen althans tot een vergelijkbare konklusie in een (literatuur)onderzoek naar de hiërarchisch-kumulatieve structuur van Bloom's taxonomie voor onderwijsdoelen, 'cognitief domein' (Bloom, 1956).

In dit onderzoek vergeleken zij verschillende analysetechnieken waarmee de hiërarchisch-kumulatieve hypothese in de loop der tijd is onderzocht, namelijk door middel van een bepaling der moeilijkheidsgraad van de categorieën der taxonomie, simplexanalyse, causale modelanalyse, faktoranalyse, Guttman-Lingoes smallest space analysis en hiërarchische syndroomanalyse. Het resultaat van dit vergelijkend onderzoek laat zich, mede gezien het ingewikkelde karakter van de verschillende analysetechnieken, niet makkelijk samenvatten.

In ieder geval stemmen zij in met de opvatting van Geisinger (1973) dat 'Guttman-Lingoes smallest space analysis' (radexanalyse) in vergelijking met faktoranalyse een meer geperfectioneerde en meer geëigende methode is voor het valideren van hiërarchische klassifikatieschema's. Hun voorkeur gaat echter uit naar causale modelanalyse (Madaus et al., 1973) als sterk verbeterde opvolger van een eenvoudige simplexanalyse, aangezien Madaus et al. door middel van causale modelanalyse niet alleen tot een interessante nieuwe hypothese omtrent de relaties tussen de taxonomische categorieën zijn gekomen, maar ook in staat bleken tot het "leggen van de brug naar de 'information-processing approach', die de laatste jaren snel opgang heeft gemaakt in het kader van de cognitieve psychologie" (De Corte & Van Bowel, 1978; p. 234).

Samenvattend willen wij stellen dat de opvattingen van Carroll, zoals

neergelegd in zijn model, in combinatie met de door Madaus et al. gehanteerde causale modelanalyse, waarin onder andere gebruik wordt gemaakt van multiple regressieprocedures, voor de toekomst het beste perspectief lijken te bieden op de uitbreiding en/of herformulering van het Mark-model als uitgangspunt bij de samenstelling en beschrijving van een testbatterij. In het licht van het voorafgaande lijkt dan ook de term 'model' voor het Mark-model te zwaar beladen. Eerder zouden wij nu willen spreken van een proto-model, als aanzet voor het hierboven geschetste perspectief. Hiermee is tevens ons antwoord gegeven op de vraag in hoeverre het Mark-model een nuttige leidraad is geweest bij de samenstelling en beschrijving van de LDT-E: in eerste aanzet een nuttige leidraad, zeker, maar minder adequaat voor een inspirerend vervolg.

8.2 LDT-E als intelligentietest

Of een test 'intelligentie' meet, is een probleem dat wel nooit opgelost zal worden, tenzij men uitgaat van een extreem operationele definitie. In paragraaf 3.1 hebben wij laten zien dat een dergelijke vorm van probleemoplossen door de tijd achterhaald is door heel summier te verwijzen naar de faktoranalytische traditie waarop het historisch gegroeide, psychologische begrip intelligentie gestoeld is. In dit verband ook merkten wij op dat de belangrijkste door Thurstone (1938) geïdentificeerde factoren - al dan niet in combinatie met elkaar - aanwezig behoorden te zijn in de LDT-E, met uitzondering overigens van de factoren 'Numeriek' en 'Woordenrijkdom'. Door het ontbreken van korrelationele gegevens met 'erkende' intelligentietests - tenzij men de FST of NST als zodanig wil beschouwen - konden wij in deze studie helaas niet nagaan in hoeverre de bedoelde factoren door de LDT-E gemeten worden. Maar zelfs indien wij wèl de beschikking hadden gehad over intelligentietests als referentievariabelen voor de LDT-E, dan nog zou het gestelde probleem niet direkt opgelost kunnen worden. Starren (1973) illustreert dit in zijn proefschrift over een nieuwe versie van de SON als volgt:

"Omdat geen enkele empirische beschrijving (variabele) van "intelligentie" volledige representativiteit kan opeisen voor het begrip in kwestie, is het antwoord op de vraag naar de begripsvaliditeit van de S.S.O.N. noodzakelijkerwijs ineksakt en indirekt. Dat wil zeggen, dat we aannemen dat de S.S.O.N. "intelligentie" meet, indien de test die mate van samenhang met andere variabelen vertoont, die we op grond van "common sense" of anderszins verwachten (konvergente resp. divergente validiteitsindicaties; Campbell & Fiske, 1959)" (p. 84).

In navolging van deze gedachtengang en aan de hand van de onderzoeksresultaten zullen wij nu eveneens de (indirekte) aanwijzingen bespreken dat

de LDT-E zoiets als 'intelligentie' meet. Omdat wij van tevoren geen duidelijke hypothesen opgesteld hebben omtrent de aard en mate van samenhang met allerlei variabelen zoals in een begripsvaliderend onderzoek te doen gebruikelijk, zullen wij onze aanwijzingen voor de betekenis van de LDT-E als intelligentietest ondersteunen met enkele bevindingen uit de literatuur.

Ter wille van de overzichtelijkheid groeperen wij aanwijzingen en literatuurbevindingen in vier categorieën, aangeduid met 'Betrouwbaarheid', 'Leeftijd, geslacht en sociaal-ekonomisch niveau', 'Faktorstructuur' en 'Schoolprestaties'.

Betrouwbaarheid. Niet zozeer met het oog op de betekenis van de LDT-E als wel met het oog op de stelregel dat intelligentietests aan zekere minimumeisen dienen te beantwoorden inzake hun betrouwbaarheid, vermelden wij hier nogmaals de twee betrouwbaarheidswaarden voor het LDT-E totaal, namelijk .93 (interne consistentie) en .92 (homogeniteit), cf. par. 5.2. Eerder typeerden wij deze waarden als zeer bevredigend, ofschoon het LDT deviatie IQ voor de betrokken leeftijdsgroepen IV (5;06-5;11) en V (6;00-6;05) nog iets hogere interne consistentiewaarden te zien geeft, respectievelijk .94 en .95 (deel I; p. 42). Vergeleken met de interne consistentie van het AKIT totaal (.92) en het WPPSI Full Scale IQ (.97), evenals met de homogeniteit van de SON 2½-7 (.91)⁴³ voldoet de LDT-E in ieder geval aan de minimumeisen welke aan de betrouwbaarheid van erkende intelligentietests gesteld worden.

Leeftijd, geslacht en sociaal-ekonomisch niveau. Belangrijker dan de vraag of de LDT-E een bepaald begrip op betrouwbare wijze meet, is de vraag of dit begrip ook met intelligentie aangeduid mag worden. Kenmerkend voor intelligentietests - zeker voor jongere leeftijdsgroepen - is het gegeven dat de prestaties met de leeftijd variëren. Voor het totaal van de LDT-E werd overeenkomstig dit gegeven een significant leeftijdseffekt verkregen: oudere kinderen (5;11-6;04) leveren betere testprestaties dan jongere kinderen (5;05-5;10). Een welhaast triviale bevinding, indien niet tevens een mogelijke verschuiving in de betekenis van het gemetene onderzocht wordt. Algemeen bekend is immers dat met het vorderen van de leeftijd ook de psychologische betekenis van het begrip 'intelligentie' verandert. De resultaten van ons onderzoek naar de faktorstructuur van de LDT-E voor jongere en oudere kinderen kunnen dit laatste echter niet bevestigen: binnen de onderzochte leeftijdsgroep (5;05-6;04) vertoont de

LDT-E een opmerkelijk stabiele structuur, zodat het onaannemelijk geacht moet worden dat de betekenis van het LDT-E totaal ingrijpend verandert met de leeftijd. De geringe variatiebreedte van slechts één jaar lijkt ons mede debet aan dit resultaat, naast de mogelijk geringe gevoeligheid van de gebruikte analysetechniek (cf. par. 8.1).

In tegenstelling tot de leeftijd oefent de variabele geslacht nauwelijks enige invloed uit op het totaal van de LDT-E: meisjes leveren weliswaar iets betere testprestaties dan jongens, maar de verschillen zijn te gering om daaraan enige waarde te hechten. Deze resultaten zijn konform de gegevens die over intelligentietests als bijvoorbeeld de Stanford-Binet (McNemar, 1942), de WPPSI (Ruschival & Way, 1971) en de SON (Starren, 1973) bekend zijn. Ook Anastasi (1958) en Maccoby en Jacklin (1975) wijzen op de geringe betekenis van de variabele geslacht, althans voor zover het om een totaalscore gaat bij kinderen van deze leeftijd (5;05-6;04). Anastasi (op.cit.; p. 460) verklaart dit uit het feit dat sommige testkonstrukteurs, zoals in het geval van de Stanford-Binet, doelbewust de items of subtests verwijderen die nauw samenhangen met geslacht. Als tweede reden voert zij aan dat mogelijke geslachtsverschillen op bepaalde subtests elkaar in de totaalscore van een intelligentietest opheffen. Deze laatste verklaring lijkt ook voor de LDT-E niet onaannemelijk, aangezien jongens op Blokpatronen en Begrip en inzicht gemiddeld iets betere testprestaties leveren dan meisjes, en omgekeerd meisjes weer iets betere testprestaties leveren dan jongens op de overige tests van de LDT-E (cf. Bijlage; tabel 17). Het zeer significante verschil in gemiddelde testprestaties op Vouwblaadjes (cf. par. 6.1) is kennelijk onvoldoende om een significant geslachtsverschil in het LDT-E totaal te bewerkstelligen. Evenals voor leeftijd is tenslotte ook voor de variabele geslacht nagegaan of er verschillen zijn in de faktorstructuur van de LDT-E en - daarmee samenhangend - in het begrip intelligentie zoals gerepresenteerd door het LDT-E totaal. Op het gevonden geringe verschil met betrekking tot Natikken (cf. par. 6.3) zullen wij in paragraaf 8.3 nog terugkomen. Een van de best gedocumenteerde bevindingen uit de onderzoeksliteratuur is de relatie intelligentie en sociaal-ekonomisch niveau (SEN). In een samenvattend overzicht rapporteert Guilford (1967; p. 387 e.v.) zeer significante waarden die tussen de .25 en .50 variëren, afhankelijk van het meer of minder verbale karakter van de gebruikte intelligentietest en het beroeps- dan wel opleidingsniveau van de ouders als indicatie voor SEN. In overeenstemming met deze literatuurbevindingen is het significante effect van SEN op het totaal van de LDT-E: zo leveren kinderen uit het

hoogste milieu (SEN-I) betere totaalprestaties dan kinderen uit lagere milieus (SEN-II en SEN-III), en kinderen uit het midden-milieu (SEN-II) weer betere prestaties dan kinderen uit het laagste milieu (SEN-III). Tenslotte is ook voor de variabele SEN nagegaan in hoeverre deze aanleiding geeft tot betekenisverschillen in het intelligentiebegrup zoals gemeten door de LDT-E. Enkele kleine verschillen werden gevonden tussen de verschillende SEN-groepen, met name ten aanzien van de faktoriële betekenis van Natikken en Plaatjes aanwijzen. Ook hierop zullen wij in paragraaf 8.3 nog terugkomen.

Faktorstruktuur. Zoals wij in de vorige paragraaf reeds aanstipten biedt het Mark-model nauwelijks aanknopingspunten voor een beschrijving en analyse van onze faktoranalytische onderzoeksresultaten. Dit betekent dat een theoretisch kader voor de discussie van deze resultaten nagenoeg ontbreekt en wij dientengevolge tot tamelijk subjektieve, vaak arbitrair aandoende interpretaties van de LDT-E als intelligentietest veroordeeld zijn, geheel in overeenstemming overigens met de traditie op dit terrein. Eén van deze tradities houdt in dat men nagaat in hoeverre het totaal percentage verklaarde gemeenschappelijke variantie bepaald wordt door een of meer (ongeroteerde) factoren. Neemt één faktor het grootste deel van deze variantie in beslag dan konkludeert men tot een algemene (g) faktor, als indicatie voor een zeer gemeenschappelijke betekenis van de tests. De sprong van dit zeer gemeenschappelijke naar de term intelligentie blijkt dan in de praktijk erg klein te zijn. Alle door ons uitgevoerde faktoranalyses van de LDT-E leverden, zowel voor de 3- en 4-faktoroplossingen als voor de 3-faktoroplossingen naar leeftijd, geslacht en sociaal-ekonomisch niveau, een dominante positie op voor de eerste ongeroteerde faktor; dat wil zeggen de eerste faktor neemt het grootste gedeelte (32-39%) van het totaal percentage verklaarde gemeenschappelijke variantie (46-53%) in beslag. Op basis hiervan konkludeerden wij dat de waargenomen variantie van de tests uit de LDT-E in ieder geval in één betekenisdimensie beschreven kan worden. Deze betekenisdimensie duiden wij nu aan met de term intelligentie, in navolging van de traditie. Het ligt in de lijn van dezelfde traditie om door middel van rotatie, i.c. varimaxrotatie van de verkregen faktoroplossingen, het inzicht in de strukturele betekenis van het intelligentiebegrup te verdiepen. Uitgaande van een 4-faktoroplossing voor de gehele onderzoeksgroep konden wij drie van de vier geroteerde factoren interpreteren en benoemen als respektievelijk Taalbegrup (Zinnen nazeggen, Verhaaltje vragen en Begrip en inzicht),

Verbaal geheugen (Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen) en Ruimtelijk-analytisch inzicht (Blokpatronen en Natikken). De vierde faktor - alleen gerepresenteerd door Vouwblaadjes - bleek vooralsnog moeilijk te benoemen, maar kon binnen het kader van het Mark-model wel tentatief geïnterpreteerd worden (cf. par. 6.3). Mede op grond van het risico dat de vierde faktor slechts foutenvariantie representeerde (verklaarde variantie: 2%) is ook een faktoranalyse naar drie factoren (varimaxrotatie) uitgevoerd, eveneens op de gehele onderzoeksgroep. Dit leverde geen dramatische verschuiving op in de faktorstructuur van de LDT-E: de moeilijk benoembare vierde faktor loste zich op in de derde faktor Ruimtelijk-analytisch inzicht (Blokpatronen, Vouwblaadjes en Natikken). De eerste en tweede faktor, respectievelijk Taalbegrip en Verbaal geheugen bleken tenslotte identiek te zijn met de gelijknamige eerste twee factoren van de 4-faktoroplossing.

Zoals hiervoor reeds werd vermeld, is ook nagegaan in hoeverre de factoriële betekenis van de tests uit de LDT-E - en daarmee tevens het door deze tests gemeten intelligentiebegrip - verandert met de leeftijd, het geslacht en het sociaal-ekonomisch niveau. Varimaxrotatie van de verkregen 3-faktoroplossing per leeftijdsgroep (5;05-5;10 en 5;11-6;04) leidde niet tot een verschuiving in de faktorstructuur van de LDT-E vergeleken met de eerder verkregen 3-faktoroplossing voor de gehele onderzoeksgroep. Eenzelfde analyse, maar dan op jongens en meisjes apart, leverde hetzelfde resultaat op, zij het dat Natikken (alléén bij meisjes) nu ook op de tweede faktor Verbaal geheugen een substantiële lading vertoonde. Varimaxrotatie van de verkregen 3-faktoroplossing per SEN-groep (SEN-I, SEN-II en SEN-III) liet eveneens een soortgelijk resultaat zien, ofschoon Natikken en Plaatjes aanwijzen nog extra ladingen vertoonden op respectievelijk de tweede faktor Verbaal geheugen (alléén voor SEN-III) en de derde faktor Ruimtelijk-analytisch inzicht (alléén voor SEN-I). Op grond van deze gegevens konkludeerden wij dat de faktorstructuur van de LDT-E, evenals het door deze test gemeten intelligentiebegrip, vrij stabiel genoemd mag worden, hoewel de verschuivingen in het ladingenpatroon van Natikken en Plaatjes aanwijzen voor ons aanleiding waren om de tweede faktor met 'Geheugen' te benoemen, in plaats van Verbaal geheugen.

Bleek het mogelijk om de LDT-E niet alleen zeer 'globaal' (één betekenisdimensie) maar ook zeer gedetailleerd (drie à vier factoren) te beschrijven, iets anders is of de verkregen faktorstructuur zich ook in combinatie met andere tests kan handhaven. De resultaten van een faktoranalyse op de LDT-E, FST en NST gezamenlijk, laten zien dat de drie factoren

Taalbegrip, Geheugen en Ruimtelijk-analytisch inzicht ook in een dergelijke combinatie naar voren komen. Onze interpretatie van de 3-faktorstructuur wordt bovendien versterkt door de bevinding dat de factoren 'visueel-motorische differentiatie' en 'waarneming en verbaal begrip' uit de FST en NST grotendeels blijken samen te vallen met de factoren Ruimtelijk-analytisch inzicht en Taalbegrip. Voor de faktor Geheugen werd - conform de verwachting - geen ondersteuning gevonden. Krapp en Mandl (1971) kwamen indertijd, na onderzoek van een groot aantal zgn. schoolrijpheidstests, tot de konklusie dat de FST en NST eerder intelligentie dan wat anders meten. De betrekkelijk hoge korrelaties tussen de LDT-E, FST en NST in aanmerking genomen (cf. par. 6.5), wijst dit erop dat de LDT-E althans een deel van het intelligentiedomein op deze leeftijd bestrijkt. Hoe groot dit deel is kunnen wij moeilijk aangeven, omdat de resultaten van een faktoranalyse in sterke mate afhankelijk zijn van de geselecteerde tests. Aan de uitkomsten van onze analyse te oordelen, bestrijkt de LDT-E een groter domein van de intelligentie dan de FST en NST, aangezien de LDT-E zich in ieder geval in drie factoren laat beschrijven, in tegenstelling tot de FST (één faktor) en NST (twee factoren). De vraag naar de omvang van het intelligentiedomein van de LDT-E laat zich echter pas goed beantwoorden, indien 'erkende' intelligentietests met een duidelijke faktoriële structuur en betekenis te zamen met de LDT-E geanalyseerd worden.

Schoolprestaties. In tal van onderzoeken is aangetoond dat intelligentie en schoolprestaties nauw met elkaar samenhangen. Het 'IQ' wordt dan ook als een van de beste voorspellers van schoolprestaties beschouwd. Dat een dergelijke nauwe samenhang voor de hand ligt, is op diverse plaatsen in deel A reeds duidelijk gemaakt. Aanwijzingen voor de validiteit van de LDT-E als intelligentietest zijn daarom direkt af te leiden uit de samenhang tussen het LDT-E totaal en de prestaties van de kinderen in de eerste klas GLO.

De resultaten van ons onderzoek naar de relaties tussen het LDT-E totaal enerzijds en een achttal criteriumvariabelen met betrekking tot de lees-, reken-, schrijf- en taalprestaties in de eerste klas GLO anderzijds, konfirmeren bovenstaande gedachtengang in grote lijnen. Zo variëren de verkregen korrelaties van .40 (Schiedamse rekentest) tot .59 (Begrijpend lezen), terwijl de mediane korrelatie met de criteriumvariabelen .50 bedraagt. Aangezien het verschil in afnametijdstip tussen de LDT-E en de diverse criteriumvariabelen ruim één jaar bedraagt, mogen deze korrelaties

tamelijk hoog genoemd worden, te meer daar het aannemelijk lijkt dat zonder 'restriction of range' effect (cf. par. 4.4) nog hogere waarden gevonden zouden zijn. Ter vergelijking vermelden wij hier enkele gegevens uit een prediktief valideringsonderzoek van Van der Linden (1970) met de PMA 5-7 (Thurstone & Thurstone, 1953 a en b), afgenomen in de eerste klas GLO. Uit dit onderzoek kwam naar voren dat het PMA totaal IQ tussen de .39 en .62 korreleerde met verschillende lees- en rekentoetsen die ongeveer drie maanden vóór of ná de PMA 5-7 waren afgenomen.

Een aanwijzing tenslotte dat het LDT-E totaal en het LDT deviatie IQ naar alle waarschijnlijkheid een maat met dezelfde betekenis voorstellen, wordt geleverd door de korrelaties tussen het LDT deviatie IQ en schoolprestaties, zoals weergegeven in tabel 14 (N= 164; leeftijdsgroep IV: 5;06-5;11) en tabel 15 (N= 150; leeftijdsgroep V: 6;00-6;05) van deel I (p. 52-53). De schoolprestaties zijn als volgt gemeten. Tijdens het landelijk normeringsonderzoek van de LDT is aan de leerkracht van het onderzochte kind een schriftelijk in te vullen vragenlijst voorgelegd waarin onder andere de volgende vraag voorkwam: "Hooft dit kind, naar Uw mening, qua schoolprestaties a. tot de beste 25 procent van zijn klas, b. de middelste 50 procent, c. de slechtste 25 procent?" Voor leeftijdsgroep IV vonden wij een korrelatie van .59 tussen het LDT deviatie IQ en schoolprestaties; voor leeftijdsgroep V een korrelatie van .52. Vergeleken met de mediane korrelatie van .50 voor het LDT-E totaal en het achttal criteriumvariabelen, ligt de korrelatie met schoolprestaties voor het LDT deviatie IQ dus iets hoger. Opgemerkt moet echter worden dat de variabele 'schoolprestaties' uit het landelijk normeringsonderzoek met de LDT een gelijktijdig criterium betreft.

Samenvattend willen wij stellen dat de resultaten van ons onderzoek naar de betekenis en voorspellende waarde van de LDT-E zonder uitzondering suggereren dat de LDT-E als intelligentietest beschouwd mag worden. Langs indirecte en directe weg werden hiertoe verschillende aanwijzingen verzameld. Zo bleek de invloed van leeftijd, geslacht en sociaal-ekonomisch niveau op het totaal van de LDT-E in redelijke overeenstemming te zijn met de invloed die deze variabelen op andere intelligentietests uitoefenen. Een andere aanwijzing dat de LDT-E intelligentie meet, werd verkregen via faktoranalytisch onderzoek. In principe kan de waargenomen variantie van de tests uit de LDT-E beschreven worden door middel van een algemene faktor; daarnaast leverde varimaxrotatie een drietal duidelijk te interpreteren factoren op: Verbaal begrip, Geheugen en Ruimtelijk-

analytisch inzicht. Ondersteuning van twee van de drie factoren werd verkregen via de FST en NST, welke in feite beschouwd mogen worden als (groeps)intelligentietests. De nauwe samenhang tussen het LDT-E totaal, de FST en NST, vormde een nieuwe aanwijzing voor de LDT-E als intelligentietest. Duidelijke aanwijzingen voor de voorspellende waarde van het LDT-E totaal werden tenslotte verkregen vanuit de positieve relatie tussen de prestaties van schoolverlatende kleuters op de LDT-E enerzijds en hun lees-, reken-, taal- en schrijfpresetaties in de eerste klas GL0 anderzijds (mediane $r = .50$).

8.3 LDT-E als diagnostische test

"It would be a waste of valuable time if the psychologist's only goal were to produce an IQ." Niet voor niets hebben wij dit citaat van Jerome M. Sattler (1974; p. viii) als motto voor de LDT-E gebruikt. In het verleden volstond men namelijk maar al te vaak met de rapportage van een simpel 'IQ' als enig gegeven dat een soms enkele uren durend intelligentie-onderzoek opleverde. Te weinig werd beseft dat intelligentietests aanzienlijk méér informatie over de onderzochte kunnen verschaffen, aangezien de meeste van deze tests opgebouwd zijn uit een aantal subtests die op deelaspecten van de intelligentie een beroep doen. Juist het gegeven dat de meeste intelligentietests meer informatie kunnen verschaffen dan een IQ, is een belangrijke reden geweest waarom wij de LDT-E als diagnostische test bestempeld hebben. In deel I schreven wij hierover het volgende:

"Afhankelijk van het doel waarvoor men de LDT wil gebruiken zou men de LDT als intelligentietest of als diagnostische test kunnen bestempelen. Zoals in de naam tot uitdrukking komt, wordt de voorkeur gegeven aan het gebruik van de LDT voor diagnostische doeleinden. Dit houdt in dat men naast een schatting van de intelligentie ook de sterke en zwakke punten in specifieke cognitieve vaardigheden poogt op te sporen"⁴⁴ (p. 11).

Zoals we bij de bespreking van het 'Bandwijdte - Zuiverheid'-dilemma gezien hebben (par. 3.1), worden aan een intelligentietest die tevens voor diagnostische doeleinden bestemd is, nogal wat eisen gesteld met betrekking tot de betrouwbaarheid van de afzonderlijke (sub)tests. Dat de tests van de LDT-E lang niet alle aan deze (hoge) betrouwbaarheidseisen voldoen, spreekt voor zich: tenslotte betreft het hier een experimentele versie van de LDT die - onderzocht op zijn betrouwbaarheid - juist een betrouwbaarder, definitieve versie van de LDT mogelijk heeft gemaakt.

Voor het onderzoek naar de betekenis en voorspellende waarde van de afzonderlijke tests uit de LDT-E, zijn de minder bevredigende betrouwbaar-

heidskoëfficiënten van enkele tests echter een handicap gebleken. Zo kunnen wij ons bijvoorbeeld niet aan de indruk onttrekken dat de betrekkelijk lage prediktieve validiteitskoëfficiënten van Blokpatronen, Vouwblaadjes, Verhaaltje vragen en Begrip en inzicht mede hun oorzaak vinden in de overwegend matige tot lage betrouwbaarheid van deze tests (cf. par. 7.1). Deze indruk wordt versterkt indien we de hier verkregen prediktieve validiteitskoëfficiënten vergelijken met de korrelaties tussen dezelfde tests uit de LDT en het criterium 'schoolprestaties' (deel I; p. 52 en 53), ondanks het feit dat het bij schoolprestaties om een gelijktijdig criterium gaat. Bij de bespreking van de afzonderlijke tests uit de LDT-E zullen wij hier nog uitvoeriger op ingaan.

Met het oog op de validiteit van de tests uit de LDT-E voor diagnostisch gebruik, zijn naast hoge betrouwbaarheidskoëfficiënten ook lage interkorrelaties van de tests gewenst, althans volgens Starren (1973):

"... subtests die ieder voor zich items bevatten die iets gemeenschappelijks meten en waarvan de somscores maximaal differentiëren tussen personen tegen zo laag mogelijke "skorekosten", kunnen scores opleveren die een dusdanige variantie-overlap vertonen dat daarmee onderzoek naar de partiële validiteit van mogelijke intelligentie-aspekten bij voorbaat tot een betrekkelijk zinloze bezigheid wordt. Naast hoge betrouwbaarheid geldt daarom de eis van lage interkorrelaties (met name tussen subtests die qua aard van de gevraagde verrichting duidelijk verschillen) als voorwaarde voor partiële validiteit" (p. 23).

Vanuit een traditioneel psychometrisch standpunt bezien heeft Starren met zijn eis van 'lage interkorrelaties' volkomen gelijk. Voor een intelligentietest die tevens diagnostische test is, gelden echter gedeeltelijk andere maatstaven, zoals wij in paragraaf 8.1 reeds hebben laten zien. Ter illustratie een voorbeeld.

Woordenspan en Plaatjes aanwijzen korreleren tamelijk hoog met elkaar (.63), te hoog volgens een 'pur sang' psychometricus naar wij mogen aannemen. Tot overmaat van ramp korreleert Woordenspan met vrijwel alle criteriumvariabelen nog beduidend lager dan Plaatjes aanwijzen (cf. tabel 24). De koëfficiënten van interne consistentie van beide tests, evenals hun homogeniteitskoëfficiënten zijn verder praktisch gelijk en in ieder geval bevredigend (cf. tabel 7). Alles bij elkaar genomen lijkt dit voldoende reden om Woordenspan uit de LDT-E te verwijderen en op zoek te gaan naar een andere test die minstens even betrouwbaar is als Plaatjes aanwijzen, daarmee laag korreleert en liefst ook nog hoog korreleert met dezelfde criteriumvariabelen als waarmee Plaatjes aanwijzen korreleert. Toch hebben wij dit niet gedaan en wel omdat een lage score op Plaatjes aanwijzen op verschillende manieren 'verklaard' kan worden. Eén van de

mogelijke verklaringsbronnen is een zwak onmiddellijk geheugen voor woordsekwenties, een functie die nu juist door Woordenspan gemeten wordt. Vindt men nu een hoge score op Woordenspan en een lage score op Plaatjes aanwijzen, dan betekent dit dat het (onmiddellijk) geheugen voor woordsekwenties - als één van de functies van Plaatjes aanwijzen - naar alle waarschijnlijkheid in takt is en dat, eveneens naar alle waarschijnlijkheid, een meer complexe functie als matching de minder goede prestatie op Plaatjes aanwijzen verklaart (cf. tabel 3). Voor het geval dat een lage score op Plaatjes aanwijzen gepaard gaat met een lage score op Woordenspan, verwijzen we naar het voorbeeld op pagina 46.

In het nu volgende zullen wij aan de hand van de onderzoeksresultaten en enkele aanvullende literatuurbevindingen de betekenis en voorspellende waarde van iedere test uit de LDT-E afzonderlijk bespreken.

Blokpatronen. Gezien het grote aantal functies dat door Blokpatronen, c.q. Block Design uit de WPPSI, gemeten zou worden, hoeft het geen verbazing te wekken indien deze test tamelijk hoog zou korreleren met de totaalscore (IQ) van allerlei intelligentietests. Wechsler (1967; p. 34) vermeldt althans korrelaties van .41 met de Stanford-Binet (Form L-M), van .35 met de Peabody Picture Vocabulary Test (Form A) en van .44 met de Pictorial Test of Intelligence. Blokpatronen wordt dan ook als een typische 'g'test' beschouwd (Milliren & Newland, 1969), die mede om deze reden wordt aanbevolen voor opname in verkorte versies van onder andere de WPPSI (King & Smith, 1972) en de LDT (deel I; p. 91). Daarnaast verschaft een zo veelzijdige en complexe test als Blokpatronen ruimschoots gelegenheid tot observatie. In de woorden van Wechsler (op.cit.):

"Persistence is often revealed by the way a child deals with the Block Design Test: does he give up when he encounters his first failure, or does he want to continue working after even the time limit has expired?" (p. 43).

Ook problemen met het visueel-ruimtelijk oriëntatievermogen (onder andere rotatiefouten) kunnen door middel van Blokpatronen geobserveerd en opgespoord worden (Matheney et al., 1972).

Over de betekenis en voorspellende waarde van Blokpatronen uit de LDT-E zijn wij in ons onderzoek maar betrekkelijk weinig te weten gekomen, en niet alleen door gebrek aan referentievariabelen. Zo liet Blokpatronen minder bevredigende betrouwbaarheidscoëfficiënten zien ($r_{11} = .70$; $r_{kk} = .63$), zeker in vergelijking met de coëfficiënten van interne consistentie die Wechsler (op.cit.; p. 22) voor vergelijkbare leeftijdsgroepen geeft: r_{11} varieert van .81 tot .88. Dit niet onaanzienlijke verschil in

betrouwbaarheid zou diegenen die Block Design uit de WPPSI in de Nederlandse versie afnemen op zijn minst aan het denken moeten zetten over het gebruik van deze test in de praktijk, te meer daar wij een duidelijk 'plafond'-effekt konden konstaten (cf. par. 5.2). Dit laatste is dan ook een van de redenen geweest waarom wij aan de definitieve versie van Blokpatronen uit de LDT moeilijker items hebben toegevoegd. Hiermee werd tevens de betrouwbaarheid (interne consistentie) van Blokpatronen tot een zeer acceptabel niveau opgevoerd, namelijk .80 voor leeftijdsgroep IV (5;06-5;11) en .90 voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42.

Ondanks het feit dat voor Blokpatronen een duidelijk 'plafond'-effekt werd gevonden, hangt de scorevariantie significant samen met verschillen in leeftijd (cf. par. 6.1). Dit betekent dat de door Blokpatronen gemeten functies nogal snel rijpen met de leeftijd van het kind. Weinig verrassend is verder het ontbreken van een duidelijk geslachtseffekt, ook al doen jongens het iets beter dan meisjes op Blokpatronen. Dit resultaat is in overeenstemming met de bevindingen van Maccoby en Jacklin (1975) die na uitvoerig literatuuronderzoek constateerden dat er nauwelijks enig verschil is in visueel-ruimtelijke vaardigheden tussen jongens en meisjes van deze leeftijd; als er al een verschil optreedt dan is dat ten gunste van de jongens. Het ontbreken van een duidelijk SEN-effekt is daarentegen wel verrassend, aangezien tests voor ruimtelijk-konstruktieve functies - waartoe ook Blokpatronen gerekend mag worden - niet echt 'culture-free' zijn (cf. Kalverboer & Sanders-Allersma, 1973).

De minder bevredigende betrouwbaarheid van Blokpatronen heeft er onzes inziens in niet geringe mate toe bijgedragen dat voor deze test teleurstellend lage prediktieve validiteitscoëfficiënten verkregen zijn, variërend van .13 (Oordeel Lezen) tot .29 (Oordeel Schrijven), cf. par. 7.1. Uitgesplitst naar sociaal-ekonomisch niveau werd echter voor SEN-III toch nog een korrelatie van .30 verkregen met de Schiedamse rekentest (cf. par. 7.2). Desondanks moeten wij konstaten dat Blokpatronen geen grote betekenis heeft als afzonderlijke voorspeller van de lees-, reken-, taal- en schrijfprestaties in de eerste klas GLO. Ook in combinatie met andere tests uit de LDT-E is zijn bijdrage aan de voorspelling gering. Een uitzondering willen wij echter maken voor de combinatie Zinnen nazeggen, Blokpatronen en Vouwblaadjes die - in genoemde volgorde - de schrijfprestaties van het kind het best voorspelt (R respectievelijk .37, .42 en .44), cf. par. 7.2.

Het vermoeden dat de prediktieve validiteit van Blokpatronen uit de LDT-E

een onderschatting vormt van zijn potentiële prediktieve validiteit, baseren wij niet alleen op de veel hogere betrouwbaarheid van Blokpatronen uit de LDT, maar ook op de tamelijk hoge korrelaties van de definitieve versie met een niet al te betrouwbaar (gelijktijdig) criterium als 'schoolprestaties', respectievelijk .49 voor leeftijdsgroep IV en .33 voor leeftijdsgroep V, cf. deel I; p. 52 en 53.

In zekere zin opmerkelijk zijn tenslotte de bevindingen van de canonische korrelatie-analyse. Hieruit blijkt dat Blokpatronen naar verhouding de meeste verwantschap vertoont met Oordeel Taal en Oordeel Schrijven (cf. par. 7.3). In combinatie met de hiervoor bediscussieerde gegevens leidt dit tot de veronderstelling dat de definitieve versie van Blokpatronen uit de LDT een tamelijk goede voorspeller zal zijn van de reken- en schrijfprestaties, en in mindere mate ook van de taalprestaties.

Vouwblaadjes. Evenals Blokpatronen leent ook deze test zich goed voor het verrichten van observaties zoals de nauwkeurigheid waarmee het kind vouwt, problemen met het materiaal, oog-hand coördinatie, fijne motoriek, persistentie, perseveratie en rotatieproblemen. In klinisch opzicht mag Vouwblaadjes dan ook een dankbare test genoemd worden. Qua betekenis en voorspellende waarde is Vouwblaadjes echter in meer dan een opzicht een zeer complexe test, alhoewel ondoorzichtig misschien een betere benaming lijkt.

Zo liet Vouwblaadjes uit de LDT-E, ondanks de toevoeging van een vijftal nieuwe items aan de oorspronkelijke subtest Paper Folding uit de HNTLA (cf. par. 3.2), teleurstellend lage betrouwbaarheidscoëfficiënten zien ($r_{11} = .66$; $r_{kk} = .63$), die voor de prediktieve validiteit van deze test weinig goeds voorspelden. Aan de definitieve versie van Vouwblaadjes uit de LDT is dan ook een aantal nieuwe items toegevoegd. Desondanks nam de betrouwbaarheid (interne consistentie) niet sterk toe, namelijk .71 voor leeftijdsgroep IV (5;06-5;11) en .72 voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42. Een duidelijke verklaring hiervoor kunnen wij niet geven. Het is echter mogelijk dat een verbetering in de scoring van Vouwblaadjes - onder andere door vergrote differentiatie in het aantal toe te kennen punten per item en door verscherping van de beoordelingscriteria - tot een verhoging van de betrouwbaarheid zal leiden.

In tegenstelling tot de bevindingen van Maccoby en Jacklin (1975) dat jongens van deze leeftijd in het algemeen iets betere prestaties dan meisjes leveren op tests voor visueel-ruimtelijke vaardigheden, staan de resultaten van ons onderzoek bij Vouwblaadjes: meisjes doen het duidelijk

beter dan jongens op deze test. Gelet op de functies die door Vouwblaadjes gemeten zouden worden (cf. par. 3.2), evenals op de bevindingen van McGuiness en Pribram (geciteerd door Goleman, 1978; p. 59) dat meisjes van jongs af aan over een betere handvaardigheid en fijne motoriek beschikken, lijkt het aannemelijk dat de functie 'oog - hand coördinatie' het verschil in testprestaties tussen jongens en meisjes verklaart. Men zou het verkregen resultaat ook nog anders kunnen interpreteren: van alle functies die door Vouwblaadjes gemeten worden, lijkt 'oog - hand coördinatie' niet het kleinste deel van de testvariantie in beslag te nemen (cf. tabel 3). Indien deze tamelijk spekulatieve interpretatie juist is, dan verklaart dit misschien ook waarom de functie 'geheugen voor ruimtelijke transformaties' minder belangrijk is dan wij aanvankelijk dachten, mede te oordelen naar het gegeven dat Vouwblaadjes niet op de faktor Geheugen laadt en in een 4-faktoroplossing op de vierde (rest)faktor terecht komt (cf. par. 6.3 en 6.4). Voor een discussie over het ontbreken van enigerlei leeftijds- en SEN-effekt op Vouwblaadjes verwijzen wij naar de desbetreffende passage bij Blokpatronen.

Ondanks het feit dat Vouwblaadjes iets lagere betrouwbaarheidscoëfficiënten laat zien dan Blokpatronen, zijn iets hogere prediktieve validiteitscoëfficiënten verkregen, variërend van .22 (Beelddiktee en Schiedamse rekentest) tot .30 (Oordeel Rekenen), cf. par. 7.1. Als afzonderlijke voorspeller van de lees-, reken-, taal- en schrijfprestaties in de eerste klas GLO, is Vouwblaadjes desondanks niet de meest belangrijke test uit de LDT-E. Ook in combinatie met andere tests is zijn bijdrage aan de voorspelling minimaal, zij het dat de schrijfprestaties nog het best voorspeld worden door de combinatie Zinnen nazeggen, Blokpatronen en Vouwblaadjes (R respectievelijk .37, .42 en .44), cf. par. 7.2.

De resultaten van de uitgevoerde canonische korrelatie-analyse verschaften verder geen enkele helderheid over de betekenis van Vouwblaadjes in relatie tot de diverse prediktor- en criteriumvariabelen.

Gegeven de matige betrouwbaarheid van de definitieve versie van Vouwblaadjes uit de LDT, zijn de korrelaties tussen deze test en het criterium 'schoolprestaties' op zijn minst verrassend, zeker in vergelijking met de veel betrouwbaarder definitieve versie van Blokpatronen. Zo werd voor leeftijdsgroep IV een korrelatie van .50 gevonden en voor leeftijdsgroep V een korrelatie van .41 (cf. deel I; p. 52 en 53). In combinatie met het feit dat Vouwblaadjes uit de LDT-E over de gehele linie van criteriumvariabelen ongeveer even goed voorspelt, leidt dit tot de veronderstelling dat de definitieve versie van Vouwblaadjes een redelijk goede

voorspeller zal zijn van de lees-, reken-, taal- en schrijfprestaties.

Natikken. Uit een oogpunt van observatie is Natikken minder interessant dan de twee voorafgaande tests, hoewel het natikken van blokjes met behulp van een hamertje wel eens 'agressieve' impulsen aan een kind kan ontlokken. Of het kind verbaal medieert bij het natikken valt te beoordelen aan spontane taaluitingen, variërend van hardop praten en mompelen tot het stilzwijgend bewegen van de lippen.

Natikken uit de LDT-E laat bevredigende betrouwbaarheidscoëfficiënten zien ($r_{11} = .83$; $r_{kk} = .73$), ofschoon het verschil tussen beide coëfficiënten indiceert dat verschillende functies door deze test gemeten worden, zij het op betrouwbare wijze (cf. par. 5.2). Nauwelijks gemodificeerd zijn voor de definitieve versie van Natikken uit de LDT iets lagere betrouwbaarheidscoëfficiënten verkregen, respectievelijk $r_{11} = .75$ voor leeftijdsgroep IV (5;06-5;11) en $r_{11} = .80$ voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42.

De scorevariantie van Natikken hangt significant samen met verschillen in leeftijd en sociaal-ekonomisch niveau. Tussen de prestaties van jongens en meisjes werd, ongeacht hun leeftijd of sociaal-ekonomisch niveau, geen duidelijk verschil aangetroffen. Het optreden van een duidelijk SEN-effekt suggereert - in combinatie met enkele andere bevindingen - dat in Natikken verbale mediatie een rol speelt; bekend is immers dat een test gevoeliger is voor SEN-invloeden naarmate er een groter beroep gedaan wordt op verbale functies (Guilford, 1967; p. 387 e.v.). Aan de samenhang tussen Natikken en de tests uit het AV-F en A-S kanaal te oordelen - geattenueerde korrelaties variërend van .37 (Begrip en inzicht) tot .55 (Plaatjes aanwijzen) - lijkt deze suggestie niet onredelijk (cf. par. 6.2).

Ondersteuning voor de veronderstelling dat Natikken naast een functie als 'onmiddellijk geheugen voor ruimtelijke sekwenties' ook nog een centrale, kanaalonaafhankelijke, verbaal mediërende functie als 'coderen' meet, werd verkregen via faktoranalyse. In een 3-faktoroplossing laadt Natikken - althans voor meisjes en SEN-III - niet alleen op de derde faktor Ruimtelijk-analytisch inzicht, maar ook op de tweede faktor Geheugen, waarin tests als Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen vertegenwoordigd zijn (cf. par. 6.4). De bevindingen van de canonische korrelatie-analyse, waarin Zinnen nazeggen en Natikken te zamen met Begrijpend lezen en Oordeel Rekenen de hoofdmoot van de gemeenschappelijke variantie voor zich opeisen, verlenen verdere steun aan deze veronderstelling (cf.

par. 7.3). Carroll (1973; p. 386) laat er voorts weinig twijfel over bestaan dat bij dit type tests inderdaad gebruik wordt gemaakt van verbale mediatie, zij het niet door alle kinderen: zo maakt hij onderscheid tussen 'mediators' en 'non mediators'. Dit onderscheid vormt mogelijk een verklaring voor de resultaten van de faktoranalyse bij jongens en de groepen SEN-I en -II, waarin Natikken alleen op de derde faktor Ruimtelijk-analytisch inzicht laadt.

Evenals de betrouwbaarheidscoëfficiënten zijn ook de prediktieve validiteitscoëfficiënten van Natikken bevredigend te noemen: de korrelaties met de diverse criteriumvariabelen variëren van .27 (Oordeel Schrijven) tot .45 (Begrijpend lezen), cf. par. 7.1. Als afzonderlijke voorspeller van met name de lees-, reken- en taalprestaties in de eerste klas GLO behoort Natikken dan ook tot de betere tests uit de LDT-E. De mediane korrelatie van Natikken met de leesprestaties (Begrijpend lezen en Brus) ligt zelfs nog iets hoger (.40) dan door Guthrie en Goldberg (1972) in een vergelijkbaar onderzoek gevonden is (.34). Indirekte steun voor de 'coding'functie van Natikken ontleen wij tenslotte aan een recent onderzoek van Swanson (1978), waarin hij met behulp van een redelijk vergelijkbare testtaak als Natikken aantoonde dat

"primary reading deficits in learning disabled children are related to verbal encoding deficiencies (visual-verbal integration) and not to deficiencies of visual memory, as suggested by the perceptual deficit hypothesis" (p. 539).

Gegeven de prediktieve validiteitscoëfficiënten van de overige tests uit de LDT-E, is het weinig verrassend om te konstaten dat Natikken - in combinatie met Zinnen nazeggen en Plaatjes aanwijzen - een substantiële bijdrage levert aan de voorspelling van de lees-, reken- en taalprestaties in de eerste klas GLO (cf. tabel 25). Des te verrassender is het dan ook om daarna te moeten konstaten dat de nauwelijks gemodificeerde, definitieve versie van Natikken uit de LDT tamelijk laag korreleert met het criterium 'schoolprestaties', respectievelijk .20 voor leeftijdsgroep IV en .27 voor leeftijdsgroep V (cf. deel I; p. 52 en 53). Een acceptabele verklaring kunnen wij hiervoor niet geven.

Woordenspan. Ook deze test biedt, evenals Natikken en Plaatjes aanwijzen, slechts beperkte mogelijkheden tot observatie, zij het dat de strategie die het kind volgt van belang kan zijn voor zijn uiteindelijke testprestatie: aarzelt het kind lang, alvorens het een antwoord geeft; doet het zijn best om eerst de woorden te 'leren' voordat deze nagezegd worden, of herhaalt het spontaan en zonder na te denken de voorgezegde woorden?

Daarnaast vormen omkeerfouten bij het nazeggen van de woorden een mogelijk interessant onderwerp voor nader onderzoek.

De betrouwbaarheidscoëfficiënten van Woordenspan uit de LDT-E zijn bevredigend: $r_{11} = .80$ en $r_{kk} = .75$. De definitieve versie van deze test uit de LDT, welke zoals bekend (cf. par. 3.3) identiek is met de experimentele versie, laat vergelijkbare waarden voor de betrouwbaarheid (interne consistentie) zien, namelijk: .82 voor leeftijdsgroep IV (5;06-5;11) en .76 voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42.

Voor Woordenspan is in de onderzochte leeftijdsgroep noch een duidelijk leeftijdseffekt noch een duidelijk geslachtseffekt gevonden. De variabele sociaal-ekonomisch niveau oefent daarentegen wel een zekere invloed uit op de testprestaties, zij het geen sterke: zo leveren kinderen uit SEN-I en -II significant betere testprestaties dan kinderen uit SEN-III. Tussen SEN-I en SEN-II hebben wij echter geen duidelijk verschil in testprestaties kunnen konstaten. Zowel het ontbreken van een leeftijdseffekt als van een geslachtseffekt gaat in tegen de verwachtingen die wij hadden op basis van de bevindingen van Bakker (1972) met een soortgelijke test als Woordenspan. Voor de variabele sociaal-ekonomisch niveau zijn de resultaten daarentegen weer wel conform de verwachtingen (Schroots et al., 1975). Een andere verklaring voor de gevonden discrepanties dan in termen van steekproeffluctuaties kan niet snel gegeven worden, te meer daar de steekproef uit Bakker's onderzoek representatief is voor de gehele Nederlandse populatie van 5- tot 7-jarigen.

De prediktieve validiteitscoëfficiënten van Woordenspan zijn over vrijwel de gehele linie van criteriumvariabelen iets lager dan die van Natikken, variërend van .18 (Schiedamse rekentest) tot .38 (Begrijpend lezen), cf. par. 7.1. Als afzonderlijke voorspeller van met name de leesprestaties in de eerste klas GL0 zijn de verkregen prediktieve validiteitscoëfficiënten echter niet onbevredigend; daarnaast zijn de coëfficiënten met betrekking tot de leesprestaties in overeenstemming met eerdere bevindingen van Bakker (op.cit.). Op het feit dat Natikken in het algemeen hoger korreleert met de diverse criteriumvariabelen dan Woordenspan, zullen wij bij Plaatjes aanwijzen nog terugkomen.

De resultaten van de stapsgewijze multiële regressie-analyse zijn op het eerste gezicht weinig bemoedigend: zo dragen van alle tests uit de LDT-E Zinnen nazeggen, Natikken en Plaatjes aanwijzen het meest bij aan de voorspelling van de lees- en taalprestaties in de eerste klas GL0, terwijl Woordenspan - als vierde test⁴⁵ - telkens net niet tot een significant verhoging van de kumulatieve korrelatie leidt. Mocht echter

één van de genoemde drie tests om een of andere reden komen te vervallen, dan suggereren onze bevindingen dat Woordenspan een goede remplaçant is (cf. par. 7.1). Aan het nut van Woordenspan voor diagnostisch gebruik hebben wij voorts aan het begin van deze paragraaf reeds enkele woorden gewijd. In het licht daarvan moeten dan ook de negatieve bevindingen van de multiële regressie- en canonische korrelatie-analyse geplaatst worden. Rest ons nog te melden dat de korrelatie van Woordenspan met het criterium 'schoolprestaties' .26 bedraagt voor zowel leeftijdsgroep IV als V (cf. deel I; p. 52 en 53).

Plaatjes aanwijzen. Voor deze test gelden dezelfde opmerkingen met betrekking tot de observatie als voor Woordenspan. Een fenomeen apart dat men bij Plaatjes aanwijzen nogal eens kan waarnemen, is het 'gokken' of 'willekeurig aanwijzen van plaatjes', vooral bij de wat moeilijker items en zeker indien het kind te veel tijd heeft laten passeren tussen het nazeggen en aanwijzen. In een doodenkel geval wil het bovendien wel voorkomen dat het kind de woorden fout nazegt, maar de plaatjes goed aanwijst. Men dient dan aan het desbetreffende item van Woordenspan 0 punten toe te kennen en één punt aan het bijbehorende item van Plaatjes aanwijzen.

De betrouwbaarheidskoefficienten van Plaatjes aanwijzen uit de LDT-E zijn, evenals van Woordenspan, bevredigend: $r_{11} = .81$ en $r_{kk} = .75$. De definitieve versie van Plaatjes aanwijzen uit de LDT, welke identiek is met de experimentele versie (cf. par. 3.3), laat vergelijkbare waarden voor de betrouwbaarheid (interne consistentie) zien, namelijk .76 voor leeftijdsgroep IV (5;06-5;11) en .86 voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42.

Evenals in het qua opzet vergelijkbare onderzoek van Bakker (1972) met een soortgelijke test als Plaatjes aanwijzen (cf. par. 3.2), variëren ook hier de testprestaties in sterke mate met de leeftijd. In tegenstelling tot Bakker (op.cit.) vinden wij daarentegen geen duidelijke prestatieverschillen tussen jongens en meisjes, ook al doen de meisjes het iets beter dan jongens. Hetzelfde geldt trouwens voor Woordenspan. Overeenkomstig eerder uitgesproken verwachtingen (Schroots et al., 1975) zijn verder de resultaten van ons onderzoek naar sociaal-ekonomisch niveau: zo leveren kinderen uit SEN-I en SEN-II signifikant betere prestaties op Plaatjes aanwijzen dan kinderen uit SEN-III. Tussen SEN-I en SEN-II konden wij echter geen duidelijk verschil in testprestaties konstaten (cf. par. 6.1), alhoewel de betekenis van deze prestaties voor SEN-I niet

hetzelfde is als voor SEN-II en SEN-III. In een 3-faktoroplossing van de LDT-E naar sociaal-ekonomisch niveau konden wij immers constateren dat Plaatjes aanwijzen niet alleen op de tweede faktor Geheugen, maar ook op de derde faktor Ruimtelijk-analytisch inzicht een substantiële lading vertoonde, althans voor SEN-I. Een verklaring hiervoor gaven wij reeds in paragraaf 6.4 door te wijzen op de 'gemengde' kanaaleigenschappen van Plaatjes aanwijzen. Erg bevredigend achten wij deze verklaring echter niet, aangezien hiermee niet uitgelegd wordt waarom SEN-II en SEN-III alleen op de tweede faktor Geheugen laden. Mogelijk speelt de 'coding'- of 'matching'functie van Plaatjes aanwijzen een rol in het geheel (cf. Bakker, op.cit.; Swanson, 1978), maar onduidelijk blijft hoe.

De prediktieve validiteitskoëfficiënten zijn in het algemeen bevredigend te noemen, variërend van .25 (Schiedamse rekentest) tot .45 (Begrijpend lezen), cf. par. 7.1. Als afzonderlijke voorspeller van de lees- en taalprestaties in de eerste klas GLO voldoet Plaatjes aanwijzen in ieder geval aan de verwachtingen (cf. par. 3.1); daarnaast zijn de verkregen koëfficiënten - met uitzondering van de reken- en schrijfvariabelen - in overeenstemming met eerdere bevindingen van Bakker (op.cit.).

Plaatjes aanwijzen laat qua korrelaties met de diverse criteriumvariabelen ongeveer hetzelfde patroon zien als Natikken, zij het dat de korrelaties van Natikken tussen die van Woordenspan en Plaatjes aanwijzen in liggen. Op het eerste gezicht lijken deze resultaten enigszins in strijd met de eerder verwoorde opvatting dat prestaties op tests waarbij intermodaal gematched moet worden - zoals op Plaatjes aanwijzen - hoger korreleren met leesprestaties dan tests waarbij uitsluitend intramodale matching plaatsvindt, zoals op Woordenspan en Natikken (cf. par. 3.1; Schroots et al., 1975). Indien wij echter aannemen dat Natikken, naast het onmiddellijk geheugen voor ruimtelijke sekwenties, ook nog een beroep doet op een functie als 'verbal (en) coding' - met het aksent op 'verbal' - dan valt de middenpositie van Natikken te verklaren vanuit het matching- of codingproces van visueel-ruimtelijke informatie in verbale informatie.

Gegeven de bevredigende prediktieve validiteitskoëfficiënten van Plaatjes aanwijzen, is het weinig verrassend dat deze test in combinatie met Zinnen nazeggen en Natikken de grootste bijdrage levert aan de voorspelling van de lees- en taalprestaties in de eerste klas GLO (cf. par. 7.2), met kumulatieve korrelaties variërend van .49 tot .61. De resultaten van de canonische korrelatie-analyse wijzen er verder op dat Plaatjes aanwijzen, getuige zijn gewichten op de tweede canonische variaten, nog wat

specifieke variantie beschrijft, welke onafhankelijk is van die van Zinnen nazeggen en Natikken. In overeenstemming met de verkregen validiteiten van Plaatjes aanwijzen zijn tenslotte de korrelaties van deze test met het criterium 'schoolprestaties', namelijk: .46 voor leeftijdsgroep IV en .37 voor leeftijdsgroep V (cf. deel I; p. 52 en 53).

Zinnen nazeggen. Deze test biedt verschillende mogelijkheden tot observatie. Zo kan de aandachtsspan van het kind afgemeten worden aan zijn luisterhouding tijdens het voorlezen van het verhaaltje en aan zijn taakgerichtheid tijdens het nazeggen van de zinnen. Vaak ook wekt het verhaaltje allerlei associaties op met persoonlijke belevenissen die dan prompt verteld worden. Wát precies teruggezegd wordt na het voorlezen per zin kan tenslotte ook van belang zijn voor een taalkundige foutenanalyse.

Zinnen nazeggen is in vrijwel ieder opzicht de beste test uit de LDT-E. De betrouwbaarheidskoëfficiënten bijvoorbeeld zijn zeer bevredigend, namelijk $r_{11} = .87$ en $r_{kk} = .85$. De betrouwbaarheidskoëfficiënten (interne consistentie) van de licht gemodificeerde, definitieve versie van deze test uit de LDT zijn zo mogelijk nog bevredigender, namelijk .92 voor leeftijdsgroep IV (5;06-5;11) en .91 voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42.

Gegeven het feit dat de experimentele versie van Zinnen nazeggen een zgn. 'peaked test' is (cf. par. 5.1), hoeft het geen verwondering te wekken dat de prestaties op deze test nauwelijks of niet variëren met de leeftijd. Voor geslacht werd evenmin een significant effect gevonden, geheel overeenkomstig de bevindingen van Maccoby en Jacklin (1975) met dit type tests. Naar verwachting (cf. Lawton, 1969) werd voor de variabele sociaal-ekonomisch niveau wel een heel duidelijk effect gevonden: zo leveren kinderen uit SEN-I en SEN-II significant betere testprestaties dan kinderen uit SEN-III, terwijl kinderen uit SEN-I weer significant betere prestaties leveren dan kinderen uit SEN-II (cf. par. 6.1). De verkregen verschillen in testprestaties gaan echter niet gepaard met verschillen in psychologische betekenis, afgaande op de stabiele faktoriële betekenis van deze test in de LDT-E (ladingen op zowel de eerste faktor Taalbegrip als op de tweede faktor Geheugen voor alle faktoroplossingen), cf. par. 6.3 en 6.4.

Evenals de betrouwbaarheidskoëfficiënten zijn ook de prediktieve validiteitskoëfficiënten van Zinnen nazeggen zeer bevredigend: de korrelaties met de diverse criteriumvariabelen variëren van .25 (Schiedamse

rekentest) tot .49 (Begrijpend lezen), cf. par. 7.1. Als afzonderlijke voorspeller van de lees-, taal- en schrijfprestaties neemt Zinnen nazeggen van alle tests uit de LDT-E dan ook de eerste plaats in; niet zo verwonderlijk overigens als men op de overeenkomst in testtaak en schooltaak let. Zo speelt het kunnen onthouden en uitvoeren van korte, mondelinge opdrachten in de eerste klas GLO een grote rol; een veel grotere rol wellicht dan men tot nog toe geneigd is aan te nemen. Niet voor niets levert een ogenschijnlijk 'simpel' geheugentestje als Zinnen nazeggen, te zamen met twee andere geheugentests als Plaatjes aanwijzen en Natikken, de belangrijkste bijdrage aan de voorspelling van de lees-, taal- en schrijfprestaties (kumulative korrelaties variërend van .44 tot .61), cf. par. 7.2. Dat naast het simpele onthouden en nazeggen van zinnen ook nog een beroep wordt gedaan op iets ingewikkelder funkties, suggereren niet alleen de bevindingen van de faktoranalyse (met een substantiële lading voor Zinnen nazeggen op de faktor Taalbegrip) maar ook die van de canonische korrelatie-analyse: van alle variabelen krijgen Zinnen nazeggen, Natikken, Begrijpend lezen en Oordeel rekenen stelselmatig het zwaarste gewicht op de eerste canonische variaten. De korrelaties van Zinnen nazeggen uit de LDT met het criterium 'schoolprestaties' beantwoorden tenslotte aan de verwachtingen die wij mochten koesteren op basis van de hier weergegeven prediktieve validiteitskoëfficiënten, namelijk: .44 voor leeftijdsgroep IV en .38 voor leeftijdsgroep V (cf. deel I; p. 52 en 53).

Verhaaltje vragen. Evenals Begrip en inzicht geeft deze test gelegenheid tot allerhande observaties: antwoordt het kind direkt op een gestelde vraag of denkt het eerst even na; bestaan de antwoorden uit één- of twee-lettergrepige woorden of worden al hele zinnen geformuleerd; is het taalgebruik nog erg kinderlijk voor zijn of haar leeftijd of doet dit juist erg ouwelijk aan? Ook de bereidheid tot het beantwoorden van de vragen wil nog wel eens verschillen van kind tot kind. Daarnaast blijkt soms uit de antwoorden dat het kind de vraag niet goed verstaan heeft. Als één test niet goed uit de verf gekomen is in ons onderzoek, dan is deze Verhaaltje vragen. Om met de betrouwbaarheidskoëfficiënten te beginnen, hiervoor zijn minder bevredigende waarden gevonden, namelijk $r_{11} = .74$ en $r_{kk} = .69$. Voor de definitieve versie uit de LDT zijn, dankzij enkele modifikaties (cf. par. 3.3), hogere waarden voor de betrouwbaarheid (interne konsistentie) verkregen, namelijk .79 voor leeftijdsgroep IV (5;06 - 5;11) en .80 voor leeftijdsgroep V (6;00 - 6;05), cf.

deel I; p. 42.

De scorevariantie van Verhaaltje vragen uit de LDT-E hangt nòch significant samen met verschillen in leeftijd, nòch met verschillen in geslacht of sociaal-ekonomisch niveau. Is dit voor leeftijd en geslacht een weinig opmerkelijk gegeven - Verhaaltje vragen is een zgn. 'peaked' test voor 5- tot 6-jarigen (cf. par. 5.1) en significante geslachtsverschillen zijn volgens Maccoby en Jacklin (1975) niet te verwachten - voor sociaal-ekonomisch niveau is het verkregen resultaat echter wèl opmerkelijk, aangezien Verhaaltje vragen een typische 'verbale' test is (cf. Lawton, 1969). Ofschoon er een tendens te bespeuren valt dat kinderen uit hogere SEN-groepen betere testprestaties leveren dan kinderen uit lagere SEN-groepen (zie tabel 19), kunnen wij een verklaring voor dit negatieve resultaat niet geven.

Evenmin als voor Blokpatronen en Vouwblaadjes zijn ook voor Verhaaltje vragen niet al te hoge prediktieve validiteitskoëfficiënten verkregen, variërend van .15 (Oordeel Schrijven) tot .29 (Begrijpend lezen en Oordeel Taal), cf. par. 7.1. In een vergelijkbaar onderzoek met een gelijksoortige test als Verhaaltje vragen (i.c. Verbal recall) vinden Stevenson et al. (1976) daarentegen hogere prediktieve validiteitskoëfficiënten, niet alleen voor de leesprestaties (mediane r voor jongens èn meisjes .33), maar ook voor de rekenprestaties (mediane r voor jongens èn meisjes .32). Daarnaast treffen zij op deze test, evenals op een groot aantal andere tests, een opmerkelijk aantal verschillen aan in prediktieve validiteit per geslacht, ook al laten de tests zelf weinig tot geen geslachtsverschillen zien. Zo korreleert bijvoorbeeld een test als 'Verbal recall' - afgenomen bij schoolverlatende kleuters - .20 met de rekenprestaties van meisjes in de eerste klas GLO, maar .54 (!) met de rekenprestaties van jongens. De konsekwenties van deze bevindingen voor toekomstig prediktief valideringsonderzoek met de LDT mogen duidelijk zijn: men dient zowel voor jongens als voor meisjes aparte regressievergelijkingen op te stellen of prediktieve validiteitskoëfficiënten te berekenen (cf. par. 6.1).

Ofschoon de prediktieve validiteit van Verhaaltje vragen uit de LDT-E - met name in relatie tot de lees- en taalprestaties - iets hoger is dan die van Blokpatronen en Vouwblaadjes, behoort deze test bepaald niet tot de beste afzonderlijke voorspellers van schoolprestaties in de eerste klas GLO. In combinatie met andere tests uit de LDT-E levert Verhaaltje vragen voorts geen substantiële bijdrage aan een verbetering van de voorspelling (cf. par. 7.2). Desondanks heeft deze test nog wat specifieke

variantie gemeen met Oordeel Taal en Oordeel Schrijven van de criteriumvariabelen en Blokpatronen en Plaatjes aanwijzen van de prediktorvariabelen, getuige zijn gewicht op de tweede canonische variaten (cf. par. 7.3). Erg betrouwbaar achten wij deze informatie echter niet. Het vermoeden dat de minder bevredigende betrouwbaarheid van Verhaaltje vragen uit de LDT-E niet alleen het onderzoek naar zijn betekenis maar ook zijn voorspellende waarde nadelig heeft beïnvloed baseren wij tenslotte op de hierboven vermelde bevindingen van Stevenson et al. (op.cit.) en de correlaties van de gemodificeerde, definitieve versie van deze test uit de LDT met het criterium 'schoolprestaties', namelijk .43 voor leeftijdsgroep IV en .35 voor leeftijdsgroep V (cf. deel I; p. 52 en 53).

Begrip en inzicht. Voor deze test gelden ongeveer dezelfde observatieopmerkingen als voor Verhaaltje vragen. Daarnaast kan Begrip en inzicht - identiek met Comprehension uit de WPPSI (Wechsler, 1967) - ook klinisch relevante informatie opleveren. Wechsler zelf zegt hierover:

"In response to one of the Comprehension questions, a child's statement that he must wash his face because his mother tells him to may reflect dependency. Similarly, aggression may be indicated if he says, "I'll beat him up", when asked what he should do if a smaller child starts to fight with him" (p. 43).

Evenals Blokpatronen, wordt ook Begrip en inzicht wel als een 'g'test' beschouwd. Wechsler (op.cit.; p. 34) vermeldt althans tamelijk hoge correlaties met de totaalscore (IQ) van verschillende intelligentietests, namelijk .58 (Stanford-Binet, Form L-M), .45 (Peabody Picture Vocabulary Test, Form A) en .45 (Pictorial Test of Intelligence). Mede om deze redenen wordt Begrip en inzicht vaak opgenomen in een verkorte versie van de WPPSI, te zamen met Blokpatronen (o.a. Silverstein, 1967; King & Smith, 1972).

Begrip en inzicht in de door ons gebruikte Nederlandse vertaling verdient - evenmin als Blokpatronen - aanbeveling voor afname in de praktijk. Zo vonden wij lagere waarden voor de betrouwbaarheid van deze test ($r_{11} = .73$ en $r_{kk} = .71$) dan Wechsler (op.cit.; p. 22) voor vergelijkbare leeftijdsgroepen rapporteert (r_{11} varieert van .78 tot .84). Bovendien laten de P-waarden van de items uit Begrip en inzicht allerminst een keurig afnemend verloop zien (cf. par. 5.1), waardoor de in de praktijk gehanteerde afbreekregel van nadelige invloed kan zijn op de uiteindelijke testscore. Ook de scoringscriteria en antwoordvoorbeelden van Begrip en inzicht leverden in Nederlandse vertaling nogal wat problemen op, zodat een revisie van deze test onvermijdelijk bleek (cf. par. 3.3).

Het uiteindelijk resultaat is dat de definitieve versie van Begrip en inzicht uit de LDT bevredigende waarden voor de betrouwbaarheid (interne consistentie) laat zien, namelijk .82 voor leeftijdsgroep IV (5;06-5;11) en .83 voor leeftijdsgroep V (6;00-6;05), cf. deel I; p. 42.

In tegenstelling tot de variabelen leeftijd en sociaal-ekonomisch niveau werd voor de variabele geslacht geen significant verschil geconstateerd. Dit laatste is in overeenstemming met de bevindingen van Maccoby en Jacklin (1975) dat er op deze leeftijd nauwelijks enig verschil in verbale vaardigheden tussen jongens en meisjes bestaat. Het geconstateerde significante verschil in testprestaties naar leeftijd klopt voorts met het gegeven dat Begrip en inzicht primair opgezet is als zgn. 'equidiscriminating'-test voor de leeftijd van 3;10-6;07 (Wechsler, op.cit.), cf. par. 5.1. De zeer significante verschillen in testprestaties tussen respectievelijk SEN-I, SEN-II en SEN-III demonstreren weer het puur verbaal-begripsmatige karakter van deze test (cf. Lawton, 1969).

Gezien de minder bevredigende betrouwbaarheid van Begrip en inzicht uit de LDT-E vormen de naar verhouding matige korrelaties met de diverse criteriumvariabelen geen verrassing (korrelaties variërend van .19 met Oordeel Schrijven tot .35 met Begrijpend lezen en Oordeel Rekenen; cf. par. 7.1). Desondanks behoort Begrip en inzicht nog tot de betere afzonderlijke voorspellers uit de LDT-E nà Zinnen nazeggen, Plaatjes aanwijzen, Natikken en Woordenspan. In combinatie met Natikken levert Begrip en inzicht zelfs nog een substantiële bijdrage aan de voorspelling van de rekenprestaties in de eerste klas GLO, gemeten met behulp van de Schiedamse rekentest ($R = .37$), cf. par. 7.2. Het ligt tenslotte in de lijn der verwachting dat een betrouwbaarder versie van Begrip en inzicht redelijk goed zal korreleren met de lees-, reken- en taalprestaties in de eerste klas GLO. Deze verwachting baseren wij niet alleen op de relatief hoge gewichten van Beelddiktee, Schiedamse rekentest, Brus en Begrip en inzicht op de tweede canonische variaten (cf. par. 7.3), maar ook op de korrelaties van Begrip en inzicht uit de LDT met het criterium 'schoolprestaties', namelijk: .47 voor leeftijdsgroep IV en .35 voor leeftijdsgroep V (cf. deel I; p. 52 en 53).

Samenvattend willen wij stellen dat ons onderzoek naar de betekenis en voorspellende waarde van de LDT-E als diagnostische test niet in alle opzichten tot bevredigende resultaten heeft geleid. De minder bevredigende betrouwbaarheid van een viertal tests - i.c. Blokpatronen, Vouwblaadjes, Verhaaltje vragen en Begrip en inzicht - evenals een gebrek

aan adequate referentievariabelen, is hier ongetwijfeld debet aan. Voor de betekenis van deze tests moeten wij dan ook grotendeels verwijzen naar de resultaten van de inhoudsanalyse (literatuuronderzoek) en faktoranalyses, zij het dat enkele bevindingen met betrekking tot Vouwblaadjes suggereerden dat de functie 'oog - hand-koördinatie' minstens zo belangrijk is als de functie 'geheugen voor ruimtelijke transformaties'. Gegeven een minder bevredigende betrouwbaarheid zijn naar verwachting ook de prediktieve validiteitscoëfficiënten van genoemde vier tests aan de lage kant. Door combinatie van de hier verkregen betrouwbaarheids- en validiteitsgegevens met die van dezelfde tests uit de LDT konden wij echter enkele veronderstellingen opperen omtrent de differentiële waarde van deze tests als voorspeller van de lees-, reken-, taal- en schrijfprestaties in de eerste klas GLO.

Van Natikken, Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen bleek de betrouwbaarheid aanzienlijk hoger te zijn. Ofschoon ook ten aanzien van deze tests zich een gebrek aan referentievariabelen deed voelen, leidde het onderzoek naar hun betekenis en voorspellende waarde toch tot bevredigende resultaten. Zo bleken Woordenspan, Natikken, Plaatjes aanwijzen en Zinnen nazeggen redelijk tot goede voorspellers te zijn van de lees- en taalprestaties in de eerste klas GLO, terwijl Zinnen nazeggen en Natikken daarnaast ook nog de rekenprestaties in voldoende mate bleken te kunnen voorspellen. De bevindingen met betrekking tot Natikken suggereerden voorts dat deze test naast een functie als 'onmiddellijk geheugen voor ruimtelijke sekwenties' ook nog een beroep doet op een functie als 'verbaal coderen'. Als zodanig behoeft de betekenisomschrijving van deze test een nadere aanvulling. Voor de overige drie tests - i.c. Woordenspan, Plaatjes aanwijzen en Zinnen nazeggen - bleek een wijziging of aanvulling van de betekenis, zoals gebaseerd op inhouds- en faktoranalytisch onderzoek, niet noodzakelijk.

Gelet op het 'restriction of range'effekt enerzijds en het relatief ruime tijdsinterval tussen de afname van de LDT-E en de diverse criteriumvariabelen anderzijds, lijkt het aannemelijk dat alle acht tests van de LDT-E - zowel afzonderlijk als in wisselende combinaties met elkaar - een grotere samenhang met gelijktijdige criteria als lees-, reken-, taal- en schrijfprestaties zullen vertonen dan in dit onderzoek gerapporteerd werd. Gegeven voorts de grote mate van overeenkomst tussen LDT-E en LDT mogen wij veronderstellen dat deze verwachte grotere samenhang ook geldt voor de tests uit de LDT en zeker voor de gelijknamige minder betrouwbare tests uit de LDT-E. Aangenomen dat deze veronder-

stelling juist is, dan lijkt het gebruik van de LDT als diagnostische test voor de bepaling van 'het niveau van enkele cognitieve functies die van belang zijn voor het leren op school', gerechtvaardigd. Hoe belangrijk deze functies precies zijn voor het verwerven van elementaire vaardigheden als lezen, rekenen, taal en schrijven, is echter een vraag die pas na voortgezette validering van de LDT met meer zekerheid beantwoord kan worden.

B. LEERVERMOGEN EN SCHOOLPRESTATIES

9. LEERVERMOGEN EN GRENZEN-TOETSEN

9.1 Psychodiagnostiek in beweging

In de afgelopen twintig jaar heeft het psychologisch testonderzoek zien-derogen aan populariteit ingeboet, en niet alleen onder het grote pu-bliek. Ook onder psychologen is reeds geruime tijd een kritische diskus-sie aan de gang over de technisch-wetenschappelijke, sociaal-ethische en politieke betekenis van het testonderzoek, evenals over de wijze waarop de praktizerend psycholoog de psychodiagnostiek beoefent (Cronbach, 1975a; Drenth, 1967; Van Strien, 1969). In de inleiding van deel II gaven wij een globaal overzicht van de technisch-wetenschappelijke kant van deze discussie onder de titels "Devaluatie van de psychodiagnostiek" en "Psychometrie erger dan de kwaal?". Aan het slot van dezelfde inleiding stipten wij vervolgens een tweetal mogelijkheden aan die tot een onzes inziens noodzakelijke revaluatie van de psychodiagnostiek zouden kunnen leiden.

De eerste mogelijkheid is die waarbij men te werk gaat volgens de *expe-rimentele* methode (cf. Shapiro, 1951). Dat wil zeggen: men voert een 'miniatuur' experimenteel-psychologisch onderzoek uit met als onderwerp het probleemgedrag of de vraag van een cliënt. In dit onderzoek worden drie fasen onderscheiden: (a) de nauwkeurige formulering van het probleem-gedrag of de vraagstelling in de vorm van een theorie, (b) de opstelling van een aantal hypothesen als verklaring voor het probleemgedrag of als antwoord op de vraag, (c) de toetsing van deze hypothesen.

De onderzoeker acht zich volgens deze methode niet gebonden aan een theo-rie, noch aan bepaalde meetinstrumenten zoals psychologische tests al-leen. Hij is daarentegen vrij die theorie te formuleren en die meetin-strumenten en -technieken te selekteren, welke zijns inziens het snelst een bevredigende oplossing bieden voor het gestelde probleem.

De tweede mogelijkheid wordt geboden door de *beslissingstheoretische* be-nadering. Deze houdt in dat men ter oplossing van een praktisch probleem de kans op succes bepaalt voor verschillende handelingsalternatieven, evenals de waarde die men aan dat succes hecht.

Een dergelijke benadering is niet gebonden aan een of andere psycholo-gische theorie over afwijkend of probleemgedrag, maar richt zich zuiver en alleen op het beslissingsproces zelf. Dit kan betekenen dat de psycho-diagnosticus gebruik maakt van psychologische tests, bijvoorbeeld ter oplossing van een selectieprobleem of ter beantwoording van de vraag

welk remediaal programma het meest in aanmerking komt voor een kind met leesstoornissen. Hij zal dit echter alleen doen, indien de afname van een test het effect van zijn beslissing of oordeel verbetert, daarbij rekening houdend met de waarde van de doelstelling voor zijn beslissing (cf. Drenth, 1975; p. 307 e.v.).

In tegenstelling tot de beslissingstheoretische benadering hebben wij de experimenteel-psychologische benadering volgens Shapiro (op.cit.) nauwelijks uitgewerkt (cf. deel II). De reden hiervoor is ongeveer dezelfde als die welke klinici wel aanvoeren tegen de experimentele methode: er moet voor ieder probleemgeval of voor iedere probleemsituatie telkens opnieuw een theorie bedacht worden, gevolgd door de hele machinerie van hypothese-afleiding en -toetsing. Een tamelijk arbeidsintensieve en tijdrovende methode, welke niet alleen in de drukbezette klinische praktijk nauwelijks te realiseren valt, maar in een handleiding als de onze (deel II) bij voorbaat moeilijk overdraagbaar lijkt aan praktizerend psychologen zonder te vervallen in de bespreking van een eindeloos aantal probleemgevallen. Ondanks het feit dat de strikt experimentele methode in de praktijk weinig weerklank gevonden heeft, zijn er met name van onderwijspsychologische zijde de nodige pogingen gedaan tot een meer experimenteel getinte benadering van de diagnostiek; experimenteel in die zin dat de klassieke, sterk psychometrisch georiënteerde onderzoeksinstrumenten en -procedures zoals de objektieve (intelligentie)test en de standaard-testprocedure, geheel of gedeeltelijk werden afgewezen om daar andere onderzoeksinstrumenten en -procedures voor in de plaats te stellen. Deze 'experimentele' diagnostiek zou men historisch gezien kunnen beschouwen als een reactie op de uitkomst van de kontroverse 'klinisch - aktuarisch', waarin de statistische prediktie van objektieve criteria uit objektieve variabelen superieur geacht werd aan de klinisch-intuïtieve manier van voorspellen, beslissen of oordelen (cf. deel II, p. 12 e.v.). De argumenten waarmee objektieve variabelen als het IQ of de scores van een test voor de cognitieve ontwikkeling door een aantal onderwijspsychologen werden afgewezen of ter discussie werden gesteld, luiden ruwweg als volgt:

- traditionele psychologische tests zoals de WPPSI en de ITPA geven een kwantitatieve karakterisering van de cognitieve ontwikkeling in de vorm van scores, welke het resultaat of het produkt zijn van cognitieve activiteiten. Deze zogenaamde 'produkt'scores geven geen inzicht in de cognitieve processen of funkties die tot het verkregen resultaat hebben geleid (Nelissen et al., 1978; Stevens, 1978; Venger, 1975);
- testcores reflekteren de som van ervaringen in allerlei leer- en

omgevingssituaties. De rijkdom aan ervaringen die een kind heeft kunnen opdoen is echter voor een groot deel afhankelijk van de cultureel-maatschappelijke of sociaal-ekonomische groepering waartoe hij behoort. Testscores geven daarom voornamelijk een indruk van de opgedane leerervaring en niet van het leerpotentieel (Cole & Bruner, 1971; Sewell & Severson, 1974; Whitely & Dawis, 1974);

- de informatie vervat in testcores is nauwelijks of niet relevant voor de hulpverlening aan kinderen met leermoeilijkheden of andere problemen. Anders geformuleerd: aan testcores zijn onvoldoende aanwijzingen of voorschriften te ontlelen voor handelingen die kunnen leiden tot een verandering van het probleemgedrag in het gewenste (leer)gedrag (Mann, 1971; Ysseldyke & Salvia, 1974).

Op basis van deze argumenten werd vervolgens gepleit voor procesmaten in plaats van produktscores, voor een dynamisch in plaats van statisch diagnostisch onderzoek, kortom voor een kwalitatieve in plaats van kwantitatieve diagnostiek⁴⁶.

Kratochwill (1977) geeft in navolging van Sewell & Severson (1974) een overzicht van enkele alternatieve vormen van diagnostiek, waarmee men de bezwaren tegen het traditionele psychologisch testonderzoek probeerde te ondervangen. Deze diagnostische alternatieven kan men in twee groepen verdelen: (a) een groep waarbij de psychologische test het uitgangspunt vormt, en (b) een groep waarbij andere instrumenten, methoden en technieken gebruikt worden. Voorbeelden van groep (b) zijn: 'paired-associate learning', 'diagnostic teaching' (nauw verwant aan 'task analysis' en 'criterion-referenced testing') en 'child development observation'. Onder groep (a) vallen experimenteel-diagnostische onderzoeksmethoden als 'learning potential assessment' en 'limit testing'. Laatstgenoemde methode vormt nu het onderzoeksobject van dit deel B "Leervermogen en schoolprestaties".

'Limit testing' of *grenzen-toetsen* is een methode die onder anderen door Mark (1969, 1972) en Mark & Mark (1971) werd toegepast als aanvulling op het traditionele testonderzoek. Met deze methode - in feite een klinische testprocedure - beoogt men de eerder genoemde gebreken die aan het standaard-testonderzoek kleven, althans gedeeltelijk te corrigeren. Aangezien een dergelijke korrektie veelbelovend klinkt en ook de toepassing van grenzen-toetsen niet veel extra tijd hoeft te kosten, besloten wij om deze methode praktisch te beproeven, in aansluiting op het standaard-testonderzoek met de LDT-E (cf. deel A).

Alvorens hiervan verslag te doen, zullen wij in paragraaf 9.2 kort aangeven welk kernprobleem aan een standaard diagnostisch onderzoek ten grondslag ligt. In de daarop volgende paragraaf (9.3) zullen wij dan het grenzen-toetsen zelf bespreken, met het aksent op de door Mark gehanteerde methode. In de laatste paragraaf (9.4) komen tenslotte de praktische en theoretische beperkingen, evenals het doel van het door ons verrichte onderzoek aan de orde.

9.2 Leervermogen

Een belangrijk probleem waarmee veel psychologen zich sinds het begin van de 20e eeuw hebben beziggehouden, betreft de "opvoedbaarheid der kindelijke intelligentie" (Waterink, 1951; p. 197 e.v.). Dat wil zeggen: is het IQ een onveranderlijke grootheid waar onderwijs en opvoeding geen greep op hebben of juist een veranderlijke grootheid waarin de individuele leerervaringen een meer of minder belangrijke rol spelen.

Ofschoon Binet - als konstrakteur van de eerste intelligentietest - een duidelijke tussenpositie innam, waren vooral Amerikaanse psychologen tot aan het eind van de tweede wereldoorlog geneigd om het IQ als een onveranderlijke grootheid te beschouwen. Sindsdien lopen de opvattingen over het mogelijk veranderlijke karakter van de intelligentie nogal uiteen, mede onder invloed van het onderzoek van Bloom (1964), Ferguson (1954, 1956) en Jensen (1969, 1973).

Op het ogenblik kunnen - zwart-wit geschetst - drie standpunten onderscheiden worden. Het eerste standpunt - dat van het primaat van de erfelijkheid - wordt onder anderen vertegenwoordigd door Jensen (op.cit.). Volgens zijn opvattingen bestaat er zoiets als een aangeboren mentale capaciteit (intelligentie), welke betrekkelijk onafhankelijk is van de individuele leerervaring. Het tweede standpunt, waarvoor geen typische vertegenwoordiger is aan te wijzen, wordt aangeduid met interaktionisme; dat wil zeggen men acht iemands mentale capaciteit veranderbaar onder invloed van (stimulerende) omgevingsfactoren en leerervaringen. Het derde en laatste standpunt wordt vertegenwoordigd door diegenen (onder anderen Skinner, 1968, 1972) die onze kennis over mentale capaciteiten en dergelijke nog erg gebrekkig achten en het daarom te vroeg vinden om hierover een uitspraak te doen. Zij stellen zich op het omgevingsstandpunt, aangezien de leeromgeving de enige variabele vormt die openstaat voor systematische manipulatie.

Verreweg de meeste psychologen houden er tegenwoordig een interaktionistische visie op na waarbij het IQ of een soortgelijke cognitieve maat beschouwd wordt als de resultante van een complexe interactie van erfe-

lijkheids- en omgevingsfactoren. Dit betekent dat het IQ een minder goede maat vormt voor iemands mentale capaciteit ('learning potential' of 'learning aptitude'), aangezien het IQ ook gedeeltelijk de individuele leerervaringen uit het verleden ('past learning') reflecteert. Bovendien - zo wordt gesteld - werkt het IQ discriminerend ten aanzien van minder bedeelde, dat wil zeggen in een minder stimulerende omgeving grootgebrachte groepen uit onze samenleving. Aan de hand van een voorbeeld kunnen wij dit laatste verduidelijken.

Stel dat wij de geschiktheid van twee schoolverlatende kleuters voor het volgen van gewoon lager onderwijs willen bepalen met behulp van een intelligentietest. De ene kleuter komt uit een in alle opzichten gedepriveerd milieu en de andere kleuter uit een in dezelfde opzichten bevoorrechte omgeving. Beiden halen hetzelfde lage IQ en (mede) op basis daarvan luidt het advies 'plaatsing op een school voor moeilijk lerende kinderen'. Het probleem is nu in hoeverre de twee IQ's met elkaar vergeleken mogen worden. Volgens de 'interaktionisten' is een dergelijke vergelijking niet volstrekt eerlijk, aangezien de potentiële leergeschiktheid van de kleuter uit het gedepriveerde milieu zich door een gebrek aan stimulerende omgevingsfactoren onvoldoende heeft kunnen ontwikkelen. Waren echter tijdig stimulerende opvoedkundige en onderwijskundige maatregelen genomen, dan zou het IQ misschien een stuk hoger zijn geweest en ook het advies anders geluid hebben, in tegenstelling tot dat van de kleuter uit het bevoorrechte milieu.

Tot nog toe zijn de termen mentale capaciteit, intelligentie en potentiële leergeschiktheid op een enigszins verwarrende manier door elkaar gebruikt. Deze terminologische verwarring weerspiegelt slechts de onduidelijkheid in de Engelstalige literatuur over begrippen als 'learning ability', 'learning capacity', 'learning aptitude' en 'learning potential'. Vooraanstaande auteurs als Cronbach & Snow (1977) hanteren echter consequent de termen 'learning ability' (als equivalent voor 'intelligence') en 'learning aptitude' (als equivalent voor 'learning capacity' en 'learning potential'). In navolging van deze auteurs zullen wij in het vervolg eveneens de termen 'ability' en 'aptitude' gebruiken, zij het in de vrije Nederlandse vertaling van respectievelijk intelligentie (cognitieve ontwikkeling) en leervermogen (leerzaamheid)⁴⁷. Intelligentie komt dan tot uitdrukking in de feitelijke prestatie (score) op een intelligentietest of test voor de cognitieve ontwikkeling zoals de LDT-E, terwijl leervermogen duidt op de prestatie die geleverd zou zijn indien

de omgevingskondities optimaal waren geweest voor de ontwikkeling van de intelligentie.

Zoals wij in het voorbeeld reeds lieten zien, wordt de diagnosticus voor een onoplosbaar probleem gesteld, indien hij het leervermogen met behulp van een intelligentietest wil bepalen. Whitely & Dawis (1975) geven echter aan hoe hij dit probleem, althans in theorie, kan oplossen:

"Direct assessment of aptitude (...) would imply equally favorable learning experiences for all individuals. Obviously, this is not the case, but it is theoretically possible to distinguish between individuals having the same ability but different aptitudes by directly measuring the modifiability of ability. That is, when two individuals within the same current level of ability are given equivalent intervention (e.g., practice), the individual with the greater aptitude should show a faster rate of change in ability than the individual with lesser aptitude" (p. 52).

Met andere woorden: door te bepalen in welke mate intelligentie te beïnvloeden (veranderen) valt als gevolg van oefening of training op intelligentietesttaken, lijkt het mogelijk om onderscheid te maken tussen intelligentie ('ability') en leervermogen ('aptitude'). Grenzen-toetsen is één van de twee klinische testprocedures - naast 'learning potential assessment' (cf. par. 12.2) - waarmee men dit onderscheid in de praktijk heeft trachten te realiseren.

9.3 Grenzen-toetsen

Voor zover wij konden nagaan, wordt de term 'limit testing' - met als equivalent 'testing the limits' of 'extension testing' (cf. Glasser & Zimmerman, 1967; Halberstadt-Freud, 1973) - voor het eerst in de vroeg-Amerikaanse literatuur over de Rorschach gebruikt (Klopfer & Kelley, 1942). Als testprocedure was limit testing echter veel eerder bekend, getuige enkele passages uit het in 1934 verschenen werk van Vigotsky (1962; p. 103). Het is dan ook niet verwonderlijk dat in de loop van de tijd vele varianten op limit testing - of grenzen-toetsen met een goed Nederlands woord - zijn ontwikkeld. Alleen al het bestaan van deze varianten maakt het geven van een definitie tot een tamelijk lastige opgave. Kornmann (1978) is nog het meest duidelijk wanneer hij grenzen-toetsen als volgt definieert:

"'Testing the limits' ist eine Test-Phase unter variierten Bedingungen, die bei manchen Versuchspersonen im Anschluss an den Standard-Durchgang vorgenommen werden kann. Dabei wird geprüft, ob sich unter den variierten Bedingungen andere Antworten ergeben als unter Standard-Bedingungen. Aus der Art der Unterschiede und der Art der Bedingungsvariation lassen sich oft wichtige Erkenntnisse für die Therapie ableiten" (p. 4 en 5).

De achtergronden van waaruit grenzen-toetsen werd (en wordt) toegepast zijn op en top klinisch te noemen. Zo voelden veel diagnostici zich wei-

nig gelukkig met de resultaten van een onderzoek waarin een zorgvuldig gestandaardiseerde psychologische test van het cognitieve type conform de standaard-testinstructies werd afgenomen. Vooral bij het onderzoek van zwakbegaafde, gedepriveerde of anderszins gehandicapte kinderen overheers- te het gevoel dat:

- resultaten in de vorm van zo objectief mogelijk verkregen testcores een te onnauwkeurige weergave vormen van het cognitief ontwikkelings-niveau, omdat " ... testing restrictions on time, testing conditions, etc. often require assigning failure scores to suboptimal though useful skills" (Mark, 1969; p. 73);
- vaak iets anders gemeten wordt dan de beoogde formele aspecten van de cognitieve ontwikkeling omdat soms testvragen of testopdrachten verkeerd begrepen worden (Volle, 1957), dan weer motivationele en/of emotionele factoren als bijvoorbeeld faalangst de testprestaties op een oneigen- lijke manier beïnvloeden (Moriarty, 1961; Zigler & Butterfield, 1968);
- de voorgeschreven standaard-testsituatie op gespannen voet staat met het a priori te verwachten gedrag van bijvoorbeeld fysiek gehandicapte of zwaar-emotioneel gestoorde kinderen (Mittler, 1970; Taylor, 1959);
- psychologische tests van het cognitieve type slechts meten wat vroeger geleerd is ('past learning') en/of nú gepresteerd wordt, waarbij geen of weinig inzicht verkregen wordt in de mogelijkheden van het kind om nieuwe dingen te leren of nieuwe problemen op te lossen (Mark, 1969, 1972; Vigotsky, 1962);
- testcores geen informatie verschaffen over de strategieën die het kind volgt bij de oplossing van de testproblemen (een opmerking die door vrijwel alle hier genoemde auteurs wel wordt gemaakt).

Om de al dan niet terechte bezwaren die aan de standaard-testprocedure kleven te omzeilen, heeft men in de praktijk op vaak zeer ingenieuze wijze allerlei wijzigingen aangebracht in het testmateriaal, de onder- zoeksituatie, afnameprocedure, instructie, scoring en uitwerking der antwoorden. Wij zullen deze wijzigingen hier niet verder bespreken: men raadplege daarvoor Aliotti (1977), Glasser & Zimmerman (1967), Lezak (1976), Mittler (1970), Sattler (1974), Schucman (1960), Taylor (1959) en Volle (1957). Wel willen wij opmerken dat de zo dadelijk te beschrij- ven 'optimizing conditions' van Mark grotendeels overeenstemmen met de hier bedoelde wijzigingen in de standaard-testsituatie.

Evenals andere onderzoekers die in de klinische praktijk werkzaam zijn, achten ook Mark & Mark (1971) het in een aantal gevallen noodzakelijk om

na afloop van een gestandaardiseerd testonderzoek tot grenzen-toetsen over te gaan, en wel om de volgende reden:

"One difficulty with identifying CNS (Central Nervous System) effects is that they frequently mimic and/or compound the effects of psychosocial factors such as cultural deprivation and emotional disturbance. Thus, CNS dysfunction just as cultural deprivation may result in evenly retarded development such as a 10 point lowering of I.Q. equivalent (I.Q.E.) scores all across the board - i.e., on all of the subscales of developmental or I.Q.-type batteries. Similarly, CNS dysfunction, just as emotional disturbance may give rise to "hyperactivity" or extreme fidgetiness which may result in evenly lowered performance scores on only those tasks requiring continuous attention.

For example, a mentally subnormal, hyperactive, young child may achieve successes on a block-design test under standard testing conditions which earn him an I.Q.E. of 50. This score will reflect the hyperactivity resulting in inattentiveness to the task. To find the relative importance of organic and psychosocial factors, we may then limit-test the child by refocusing his attention on the problem a number of times or by attempting different methods of motivation. Under such limit-testing conditions a small but important percentage of children will improve their scores by ten or more I.Q.E. points. A very small percentage of children (less than 1 in 100) might improve their performance by 30 or 40 I.Q.E. points. Thus, testing the limits (optimizing) and observing the changing score patterns give us some basis for making initial judgements about the child's attention. When the scores do *not* change under a variety of optimizing conditions, then there is some basis for a tentative clinical assumption that the child has a) biological (organic) limitations in paying continuous attention and solving problems and/or b) biological limitations in analyzing problems in sufficient detail required for solutions. In contrast, when the scores do change significantly, then it is more reasonable to assume that environmental factors tend to detract from the child's ability to pay continuous attention to the problem; such factors, then, may contribute toward other "secondary" failures and it may well be that the child could really do better and analyze problems in greater depth in many other areas if he could be trained in the mechanics of paying attention (...).

Testing the limits, thus, is a miniature teaching-learning paradigm or behavior-shaping effort in which quite a large number of common-sensible approaches to teaching and management can be pre-tested" (p. 43-44).

Met opzet hebben wij Mark tamelijk uitvoerig geciteerd; niet alleen om aan te geven wat grenzen-toetsen nu konkrete inhoudt, maar ook vanwege twee duidelijke parallellen met het voorafgaande. Daarnaast verschaft dit citaat een goed uitgangspunt voor de beschrijving van de verschillende optimaliserende kondities - 'Bedingungen' volgens Kornmann - die tijdens het grenzen-toetsen gehanteerd worden.

De eerste parallel slaat op de opmerking dat het moeilijk is om de effecten van biologische en psychosociale factoren uit elkaar te halen met behulp van een 'IQ type' testbatterij. In wezen zeggen de 'interaktionisten' van paragraaf 9.2 hetzelfde als zij konstateren dat het IQ of een soortgelijke cognitieve maat de resultante is van een complexe interactie van erfelijkheids- en omgevingsfactoren. De tweede parallel betreft Mark's

oplossing van dit probleem door de introductie van grenzen-toetsen als een miniatuur 'teaching-learning' paradigma. Ook dit zijn wij eerder tegengekomen bij Whitely & Dawis (1975), die in feite een pretest-interventie-posttest paradigma voorstelden ter oplossing van het probleem 'intelligentie ('ability') vs. leervermogen ('aptitude')', cf. par. 9.2. Vertaald in het paradigma van Mark betekent dit dat eerst een test (pretest) volgens de standaard-instructie wordt afgenomen. Vervolgens vindt een of andere vorm van interventie plaats, door Mark aangeduid met 'optimizing conditions'. Deze optimaliserende kondities kunnen variëren van de simpele herhaling van een bepaald testitem tot het opsplitsen van een item of testprobleem in deelproblemen, welke afzonderlijk ter oplossing aan de onderzochte worden aangeboden. De uiteindelijke testprestatie is dan de posttest.

In totaal onderscheidt Mark (1972)⁴⁸ een veertiental optimaliserende kondities welke - afhankelijk van de testtaak en het testgedrag van de onderzochte - systematisch gevarieerd worden tijdens het grenzentoetsend onderzoek:

1. *Increase stimulus size*. Het stimulus- of testmateriaal wordt in ver-grote vorm aangeboden, indien bijvoorbeeld ernstige visuele waarne-mingsstoornissen een rol spelen.
2. *Speak in louder than normal tone of voice*. Pl zal harder en duidelij-ker dan gewoonlijk spreken, indien een gehoorstoornis wordt vermoed.
3. *Increase figure-ground contrast*. Het stimulus- of testmateriaal - blokken bijvoorbeeld - wordt tegen een kontrasterende achtergrond aangeboden, indien een visuele waarnemingsstoornis daartoe aanlei-ding geeft.
4. *Reduce superfluous distractions*. Indien pp door visuele of auditieve prikkels snel is afgeleid, zal pl er zorg voor dragen dat de algehele testsituatie zo min mogelijk afleiding biedt.
5. *Alter rate of exposure*. Teststimuli, bijvoorbeeld de cijfers van een cijferspantest, worden versneld of vertraagd aangeboden.
6. *Increase frequency of exposure*. Testitems, bijvoorbeeld vragen, wor-den herhaalde malen gesteld.
7. *Increase frequency of paired S-presentation*. Gepaarde stimuli (woor-den, non-sense syllaben, etc.) worden herhaalde malen aangeboden; te vergelijken met een 'paired-associate' leertaak-experiment.
8. *Increase encouragement and feedback*. Pp wordt zoveel mogelijk aange-moedigd en gestimuleerd, bijvoorbeeld door bij voortduring positief te reageren op zijn testprestaties.

9. *Increase time on task.* Pp krijgt meer tijd voor de oplossing van een item dan officieel voorgeschreven staat in de testhandleiding.
10. *Use multiple choice tests.* Indien pp door een motorische handicap bepaalde testopdrachten niet kan uitvoeren, dan wordt gebruik gemaakt van meerkeuze-tests die qua inhoud zoveel mogelijk overeenkomen met de oorspronkelijke test.
11. *Focus attention on tasks.* Indien pp zich slecht kan concentreren of nogal snel afgeleid is, wordt telkenmale de aandacht gevestigd op de testtaak.
12. *Inhibit 'noisy' outputs.* Als pp last heeft van stotteren of andere spraakstoornissen, dan wordt hij zoveel mogelijk geholpen bij het zich verstaanbaar maken.
13. *Steady test materials.* Onhandige pp'n wordt hulp geboden bij performante testtaken, bijvoorbeeld door de blokken voor hen vast te houden.
14. *Represent parts of a problem.* Een item of testprobleem wordt opgesplitst in deelproblemen en afzonderlijk ter oplossing aan pp aangeboden.

Zoals uit bovenstaande lijst op te maken valt, zijn niet alle kondities die een optimale testprestatie moeten bevorderen, van toepassing op iedere pp. Zo hebben de kondities 1, 3, 7, 10 en 14 uitsluitend betrekking op moeilijk tot zeer moeilijk testbare pp'n met ernstige motorische, sensorische en/of cognitieve stoornissen, zoals spastici, afatici, imbecielen etc. De resterende kondities worden evenmin tegelijkertijd en in dezelfde mate bij iedere pp gevarieerd. Wat en wanneer pl dan wél doet tijdens het grenzen-toetsen is - zoals gezegd - niet alleen afhankelijk van de testtaak en het testgedrag van pp, maar ook van het klinisch inzicht en de opmerkingsgave van pl. Met dit laatste hebben wij het punt bereikt waarop enkele kritische kanttekeningen bij het grenzen-toetsen geplaatst moeten worden.

Het is niet moeilijk om vanuit het vertrouwde testpsychologische kader kritiek te leveren op grenzen-toetsen. Aan een belangrijke conditio sine qua non van een testonderzoek, namelijk 'de standaardisatie van het testmateriaal, de onderzoeksituatie, afnameprocedure, instructie, scoring en uitwerking der antwoorden' wordt immers allerminst voldaan. De marginale omschrijving van de optimaliserende kondities - overigens niet alleen kenmerkend voor Mark - wekt bovendien de indruk dat men slechts via de leermeester - leerling relatie de techniek van het grenzen-toetsen onder de knie kan krijgen. En dan nog: normen die een referentiekader zouden

moeten bieden bij de beoordeling van de geoptimaliseerde testcores (posttest), worden nergens vermeld, evenmin als betrouwbaarheids- en validiteitsgegevens. In dit licht bezien klinken de argumenten van diegenen die het traditionele psychologisch testonderzoek bekritisieren, minder overtuigend.

Op voorhand lijkt het echter niet volstrekt onaannemelijk dat grenzentoetsen - mits zo uniform mogelijk uitgevoerd na afloop van een standaard-testonderzoek - aanvullende informatie kan verschaffen over het cognitief functioneren van een kind. Daartoe dient dan wel de waarde van deze informatie nader onderzocht te worden. In het nu volgende zullen wij het doel, evenals de beperkingen van het door ons verrichte onderzoek kort toelichten.

9.4 Tweede doel van het onderzoek

Ofschoon in het voorgaande herhaalde malen werd verwezen naar het werk van andere onderzoekers, beschikten wij tijdens de voorbereidingen voor het grenzentoetsend onderzoek alleen over de informatie die Mark verstrekt had. Deze informatie is niet op alle punten even duidelijk, zoals wij zojuist konden konstateren. Desondanks verschaftte zij ons enkele aanwijzingen voor een nadere precisering van het tweede onderzoeksdoel, dat wij eerder globaal omschreven als 'de praktische beproeving van grenzentoetsen, in aansluiting op het standaard-testonderzoek met de LDT-E' (cf. par. 9.1).

De vraag is nu welke aspecten van grenzentoetsen zich voor een praktische - dat wil zeggen niet volgens experimentele onderzoeksmaatstaven opgezette - beproeving lenen. Deze vraag stellen wij daarom omdat de introductie van enkele optimaliserende kondities in ons onderzoek bij voorbaat weinig zinvol lijkt of zelfs geheel onmogelijk is, uitgaande van de volgende randvoorwaarden en/of gegevensheden:

- een onderzoeksgroep van 411 schoolverlatende kleuters zonder ernstige motorische, sensorische en/of cognitieve stoornissen;
- acht tests van de LDT-E, afgenomen volgens standaard-testinstructies en te beschouwen als pretests;
- een zo goed mogelijk gestandaardiseerde grenzentoetsende onderzoeksprocedure, welke (a) niet al te veel tijd in beslag neemt ($\pm$ drie kwartier) in verband met organisatorische beperkingen, en (b) relatief snel te leren valt door dezelfde groep van 20 proefleid(st)ers die de LDT-E eerder volgens standaard-testprocedures heeft afgenomen.

Deze punten in aanmerking genomen, vallen de optimaliserende kondities 1, 2, 3, 5, 7, 10, 12 en 14 direkt af (cf. par. 9.3). Blijven over de kondities 4 'Reduce superfluous distractions', 6 'Increase frequency of exposure', 8 'Increase encouragement and feedback', 9 'Increase time on task', 11 'Focus attention on tasks' en 13 'Steady test materials'. Van laatstgenoemde kondities is 6 'Increase frequency of exposure' - dat wil zeggen herhaalde aanbieding van tijdens het standaard-testonderzoek fout gescoorde items - het belangrijkste, omdat pp hierdoor in de gelegenheid wordt gesteld om zijn score te verbeteren.

Een verbeterde score, ook wel geoptimaliseerde of post-testscore genaamd, geeft volgens Mark (op.cit.) en vele andere voorstanders van deze vorm van grenzen-toetsen⁴⁹ een nauwkeuriger indruk van het cognitief ontwikkelingsniveau dan de pre-testscore; met andere woorden de post-testscore wordt betrouwbaarder geacht dan de pre-testscore. Veelal wordt hierbij impliciet aangenomen dat ook de prediktieve validiteit van de posttest hoger zou zijn. Beide veronderstellingen lijken een nader onderzoek waard, evenals de aanvullende veronderstelling dat winstcores (het verschil van post- en pre-testscores) - gegeven bepaalde pre-testscores - een indicatie zouden geven van het leervermogen van een kind.

Aangezien een scoreverbetering mede bereikt kan worden via de kondities 4, 8, 9, 11 en 13, zullen wij ook deze in ons onderzoek betrekken, te zamen met konditie 6. Ofschoon wij ons ervan bewust zijn dat de gezamenlijke introductie van deze kondities tot een - onderzoekstechnisch gezien - onwenselijke kontaminatie van variabelen leidt, hebben wij toch besloten het grenzentoetsend onderzoek volgens deze kondities uit te voeren; dit omdat wij primair geïnteresseerd zijn in de evaluatie van een grenzentoetsende onderzoeksprocedure, die zo nauw mogelijk aansluit bij wat in de praktijk door psychologen gedaan wordt. Uiteraard legt de gevolgde procedure de nodige beperkingen op ten aanzien van de resultaten die men van dit evaluatie-onderzoek kan verwachten. In de discussie (hfdst. 12) zullen wij hier echter nog op terugkomen.

Een laatste opmerking betreft de keuze van de tests uit de LDT-E die voor grenzen-toetsen in aanmerking komen. Ofschoon het in principe mogelijk is om alle acht tests volgens de hier aangeduide grenzentoetsende onderzoeksprocedure af te nemen, stoot dit ten aanzien van Plaatjes aanwijzen, Zinnen nazeggen, Verhaaltje vragen en Begrip en inzicht op enkele problemen. Deze houden onder andere verband met de gevolgde afnameprocedure en de scoring van de antwoorden. Vandaar dat wij genoemde vier

tests niet in ons onderzoek betrokken hebben⁵⁰. Van der Heiden (1976) heeft echter laten zien welke uitgebreide mogelijkheden (en problemen) er bestaan indien men de grenzen bij Zinnen nazeggen en Verhaaltje vragen wil toetsen.

Gegeven bovenstaande beperkingen kunnen wij nu het tweede doel van ons onderzoek omschrijven als *het verkrijgen van een antwoord op de vraag of (a) de post-testscores van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan - zoals verkregen door middel van grenzen-toetsen - een nauwere samenhang met schoolprestaties bezitten dan de pre-testscores, en (b) het verschil van post- en pre-testscores op dezelfde tests - gegeven bepaalde pre-testscores - een maat voor het leervermogen kan verschaffen.*

items	score ^{a,b}		
	(1)	(2)	(3)
1	1		
2	1		
3	1		
4	0		
5	1		
6	0		
7	0		
8	0		
9	-		
10	-		
n	-	-	-
totaal	4	6	7

^a 1 = goed item; 0 = fout item

^b (1) = eerste aanbieding, standaardonderzoek LDT-E
 (2) = tweede aanbieding, grenzentoetsend onderzoek
 (3) = derde aanbieding, grenzentoetsend onderzoek

^c p = proefitem

Figuur 8. Schema van aanbieding en scoring der items per test tijdens het standaard- en grenzentoetsend onderzoek

10. METHODE

Dezelfde groep van 411 schoolverlatende kleuters als beschreven in paragraaf 4.3, is na afname van de LDT-E nogmaals onderzocht op Blokpatronen, Vouwblaadjes, Natikken en Woordenspan. Hiertoe werden de items waarop het kind gefaald had tijdens het standaardonderzoek met de LDT-E - aangeduid met eerste aanbieding (1) - opnieuw aangeboden op een manier die enigszins te vergelijken valt met de 'up-and-down method' (Dixon & Massey, 1957; p. 319). In principe verliep het grenzen-toetsen als volgt (cf. Schroots, 1973).

Na aanbieding van een proefitem werd het eerste foute item van de eerste aanbieding (1) voor een tweede keer aangeboden (2). Indien pp dit goed deed dan werd het daaropvolgende foute item van (1) aangeboden, net zolang tot twee fouten tijdens de tweede aanbieding (2) verkregen waren. Op dit punt brak pl voorlopig af om het eerste foute item van (1) en (2) voor een derde keer aan te bieden (3). Indien deze derde aanbieding tot een goed resultaat leidde, dan werd weer een nieuw fout item van (1) voor de tweede maal aangeboden (2). Indien de derde aanbieding daarentegen geen succes had, dan werd het eerstvolgende foute item van (1) en (2) voor een derde keer aangeboden. Na twee opeenvolgende fouten bij (3) brak pl tenslotte definitief af. In figuur 8 wordt de gevolgde aanbiedingsprocedure en scoring per test schematisch weergegeven.

De pijlen in deze figuur geven - te beginnen met proefitem p - de volgorde aan waarin de foute items van (1) door pl werden aangeboden. Uit de figuur kan men verder aflezen dat pp voor ieder fout item tijdens het standaardonderzoek, in principe twee kansen kreeg ter korrektie tijdens het grenzentoetsend onderzoek. In principe, aangezien de mogelijkheid bestond dat een nog niet eerder aangeboden item van (1) - aangeduid met een streepje - pas tijdens het grenzentoetsend onderzoek voor de eerste keer werd aangeboden (cf. fig. 8, item 9); dit vanwege de gehanteerde afbreekregel van drie achtereenvolgende fouten tijdens het standaardonderzoek. De scoring is voorts zo eenvoudig mogelijk gehouden: voor ieder goed item tijdens het grenzentoetsend onderzoek kreeg pp een extra-punt, bovenop het reeds behaalde puntentotaal (score) van het standaardonderzoek. Het resultaat is drie scores per test, waarvoor geldt dat score (3) $\geq$ score (2) en score (2) $\geq$ score (1).

Tijdens het grenzentoetsend onderzoek manipuleerde pl een aantal aspecten van het testgedrag zodanig dat een optimale testprestatie verkregen werd. Tot de aspecten van dit gedrag behoorden de aandacht van pp voor zijn testtaak (Optimaliserende konditie 6. Focus attention on tasks), de inzet

en motivatie waarmee de items beantwoord werden (Optimaliserende konditie 8. Increase encouragement and feedback), allerlei gedragingen die interfereerden met het gewenste testgedrag (Optimaliserende konditie 4. Reduce superfluous distractions) en de motorische vaardigheid van pp (Optimaliserende konditie 13. Steady test materials).

De mate waarin pl naar zijn oordeel de eerste drie aspecten van het testgedrag heeft moeten manipuleren, moest worden aangegeven op een drietal 4-puntsschalen, aangeduid met respectievelijk *Aandacht*, *Aarmoediging* en *Interferentie*. De schaalwaarden varieerden van normaal (1), sterker (2) en vrij sterk (3) tot zeer sterk (4). Indien het duidelijk was dat de motorische vaardigheid van pp (vierde aspekt) hem parten speelde bij een zo goed mogelijke beantwoording van de items, dan kon pl bovendien assistentie bieden volgens bepaalde richtlijnen. Deze *Assistentie* moest worden aangegeven op een 2-puntsschaal: geen assistentie (1), wel assistentie (2). Tenslotte gaf pl nog eens zijn indruk van de fijnmotorische *Vaardigheid* weer op een 4-puntsschaal met als schaalwaarden: normaal (1), enigszins onhandig (2), vrij onhandig (3) en zeer onhandig (4).

Bovenstaande beschrijving van de testprocedure tijdens het grenzentoetsend onderzoek geldt voor Vouwblaadjes, Natikken en Woordenspan. Bij Blokpatronen werd een lichtelijk afwijkende procedure gevolgd, welke hieronder kort wordt toegelicht als wij per test de richtlijnen voor het interventie- of 'manipulatie'gedrag van pl bespreken. [Voor een uitgebreide instructie: zie Schroots, 1973].

Blokpatronen. Zoals bekend bestaat de standaardversie van deze test uit tien voorbeeldpatronen (items), die ieder binnen een bepaalde tijd met behulp van platte rood- en witgekleurde blokjes gereproduceerd moeten worden (cf. Wechsler, 1967; p. 73 e.v.). Afhankelijk van het feit of pp een bepaald voorbeeldpatroon binnen de gestelde tijd reeds bij de eerste poging korrekt reproduceert, pas bij de tweede poging of helemaal niet, worden respectievelijk 2, 1 of 0 punten toegekend.

In afwijking van het hiervoor beschrevene vond tijdens het grenzentoetsend onderzoek op Blokpatronen géén derde aanbieding plaats; dit vanwege het karakter van de test, waarbij twee pogingen per item zijn toegestaan. Daarentegen mochten tijdens de tweede aanbieding wel weer twee pogingen per item gedaan worden, onder *verdubbeling van de tijd* (Optimaliserende konditie 9. Increase time on task) die voor het standaardonderzoek (eerste aanbieding) gold. Zowel tijdens de eerste als tweede aanbieding werd 2 punten als een goed item beschouwd, 1 of 0 punten als

een fout item. Na tweemaal 0 punten moest worden afgebroken.

A a n d a c h t. In tegenstelling tot bijvoorbeeld Vouwblaadjes wordt bij Blokpatronen géén continue aandacht verlangd. Ook al kijkt pp tijdens het demonstreren of naleggen van een patroon even de andere kant op, dan hoeft dit voor het uiteindelijk resultaat weinig uit te maken. Ook het voorbeeldpatroon is tijdens het naleggen de hele tijd te raadplegen. De testsituatie met betrekking tot het aspect 'Aandacht' bij het grenzentoetsend onderzoek verschilt dan ook niet wezenlijk van de situatie bij het standaardonderzoek, waarbij pl de demonstratie inleidt met een opmerking als: "Kijk naar wat ik doe".

Er zijn echter pp'n die nauwelijks aandacht besteden aan de demonstratie van het patroon of tijdens het naleggen het voorbeeldpatroon onvoldoende raadplegen. Ook korrigerende opmerkingen, in verband met aansluitings- en rotatieproblemen, worden vaak niet gehoord. Het is daarnaast niet onmogelijk dat de verdubbelde testtijd per item - waardoor sommige kinderen 'trial-and-error' gedrag gaan vertonen of in foutieve, stereotype oplossingen vervallen - een rol speelt bij de aandachtsverslapping. In dergelijke gevallen leek het noodzakelijk dat pl méér dan normaal de aandacht van pp op zijn taak vestigde. Hiervoor stonden hem verschillende - voorgeschreven - opmerkingen ter beschikking als:

Kijk eens goed, het moet zó komen te liggen

Nu moet je heel goed opletten hoe ik het doe, dan kan jij het dadelijk ook, etc.

Onder geen voorwaarde mocht pl echter tijdens het naleggen van het patroon de aandacht vestigen op deelpatronen of -oplossingen.

A a n m o e d i g i n g. Bij het grenzentoetsend onderzoek op deze test kan het zeer frustrerend zijn dat pp patronen moet naleggen die tijdens het standaardonderzoek bij een tweede poging (1 punt) goed gelegd zijn. Daarnaast zijn er altijd pp'n die weinig gemotiveerd zijn en nauwelijks of geen interesse kunnen opbrengen voor de testtaak; zéker, naarmate de patronen moeilijker worden en pp gekonfronteerd wordt met eigen falen. In dergelijke gevallen leek het noodzakelijk dat pl méér dan normaal een aantal voorgeschreven, aanmoedigende opmerkingen maakte als:

Ik geloof vast dat jij de blokjes even mooi kunt naleggen als ik

Het was zojuist bijna goed; je mag het nog een keer proberen, dan lukt het vast wel, etc.

I n t e r f e r e n t i e. Sommige gedragingen interfereren met het gewenste testgedrag of hebben daar een storende invloed op. Tijdverdubbeling, plus de mogelijkheid van een tweede kans direct ná de eerste om

het patroon goed na te leggen, kunnen bepaalde ongewenste gedragingen des te sterker oproepen. Zo krijgt pp niet alleen alle gelegenheid om zijn score te verbeteren, maar óók om met de blokken te gaan spelen, te babbelen, onrustig op zijn stoeltje heen en weer te wiebelen, plotseling op te staan, etc. Het leek noodzakelijk om deze gedragingen door middel van een aantal standaardopmerkingen zoveel mogelijk af te remmen, bijvoorbeeld:

Je moet niet zo veel praten hoor als je bezig bent, dadelijk mag je weer vertellen

Laat eens zien hoe mooi je op je stoeltje kunt zitten, etc.

Daarnaast mocht pl een overbewegelijke pp gerust even vasthouden op zijn plaats door bijvoorbeeld zijn stoeltje dichterbij de tafel te schuiven of op een andere wijze proberen pp aan zijn testtaak te houden.

A s s i s t e n t i e. De handigheid bij het naleggen van een patroon kan van pp tot pp verschillen. Sommige pp'n stoten telkens het ene blokje weg met het andere of gooien door een onverwachte beweging het patroon in de war. Indien pp al te onhandig was, mocht pl hulp bieden in de vorm van een grijs rubber matje (20x20 cm) waarop het patroon nagelegd moest worden. De instructie luidde om een zo spaarzaam mogelijk gebruik te maken van dit matje.

Vouwblaadjes. Aanbieding en scoring der items verliep geheel overeenkomstig het schema in figuur 8 en de daarbij gegeven toelichting.

A a n d a c h t. Bij Vouwblaadjes wordt continue aandacht verlangd. Als pp tijdens de demonstratie van een vouwpatroon maar even niet oplet, dan zal hij niet in staat zijn om het blaadje goed na te vouwen. Pl was daarom geïnstrueerd om vóór de demonstratie van ieder vouwpatroon te zeggen: "Let nu goed op wat ik doe".

Er zijn echter pp'n die ondanks de vaste opmerking van pl toch moeite hebben om zich te concentreren. Pl werd daarom geïnstrueerd om zo goed mogelijk oog-kontakt te houden en daarnaast de aandachtsoptmerkingen te intensiveren met woorden als:

Nu moet je heel goed opletten hoe ik mijn blaadje vouw, anders kun je het niet nadoen, etc.

Onder geen voorwaarde mocht pl echter tijdens de demonstratie of het navouwen aandacht-trekkende opmerkingen maken.

A a n m o e d i g i n g. Hiervoor gelden dezelfde aanwijzingen als bij Blokpatronen. Typerende opmerkingen zijn:

Ik wil wel eens zien of je het nu nog beter kunt

Ja, dat is ook een heel moeilijke, maar als je heel goed je best doet kun je het vast wel, etc.

De instructie luidde dat pl als het enigszins kon stimulerende opmerkingen moest vermijden tussen het demonstreren en het navouwen.

I n t e r f e r e n t i e. Hiervoor geldt hetzelfde als bij Blokpatronen. *A s s i s t e n t i e.* Indien pl de indruk kreeg dat het kind het patroon wel kon navouwen, maar gehinderd werd door zijn onhandigheid, dan mocht hij assistentie bieden door bijvoorbeeld het blaadje of een stukje daarvan vast te houden, terwijl pp vouwde. Onder geen voorwaarde mocht pl echter richting of volgorde bij het vouwen aangeven. De instructie luidde om een zo spaarzaam mogelijk gebruik te maken van deze mogelijkheid.

Natikken. Aanbieding en scoring der items verliep geheel overeenkomstig het schema in figuur 8 en de daarbij gegeven toelichting.

A a n d a c h t. Hiervoor geldt hetzelfde als bij Vouwblaadjes.

A a n m o e d i g i n g e n I n t e r f e r e n t i e. Zie Blokpatronen en Vouwblaadjes.

A s s i s t e n t i e. Deze mocht niet gegeven worden.

Woordenspan. Aanbieding en scoring der items verliep geheel overeenkomstig het schema van figuur 8 en de daarbij gegeven toelichting.

A a n d a c h t. Zie Vouwblaadjes. In plaats van "Let nu goed op wat ik doe" moest pl zeggen: "Luister goed naar wat ik zeg".

A a n m o e d i g i n g e n I n t e r f e r e n t i e. Zie Blokpatronen en Vouwblaadjes.

A s s i s t e n t i e. Deze mocht niet gegeven worden.

Als proefleid(st)ers voor het grenzentoetsend onderzoek fungeerde dezelfde groep van 20 doctoraal-studenten psychologie of pedagogiek zoals beschreven in paragraaf 4.3. In aansluiting op hun training voor het standaardonderzoek met de LDT-E vond - eveneens in het kader van een VU-seminar over 'Neuropsychologie en leerstoornissen' - de training plaats voor het grenzentoetsend onderzoek. Het trainingsprogramma bestond onder andere uit een demonstratie met behulp van video, evenals twee verplichte proefonderzoeken op kinderen in de leeftijd van 5 tot 6 jaar, gevolgd door een nabespreking.

Tussen het standaard- en het grenzentoetsend onderzoek lag een tijdsinterval van één tot twee dagen. Organisatorische problemen verhinderden de doorvoering van een gelijk tijdsinterval voor alle kinderen uit de onderzoeksgroep. Enkele uitzonderingen daargelaten werd het grenzentoetsend onderzoek door dezelfde proefleid(st)er verricht die ook het

standaardonderzoek met de LDT-E had uitgevoerd.

Na afloop van het grenzentoetsend onderzoek bleek een aantal antwoordformulieren geheel of gedeeltelijk onbruikbaar voor verdere bewerking.

Het resultaat hiervan is dat de aantallen per test iets kleiner zijn dan aangegeven in paragraaf 4.3, namelijk N = 382 (Blokpatronen), N = 398 (Vouwblaadjes), N = 410 (Natiekken) en N = 408 (Woordenspan).

Tenslotte, alle gegevens van paragraaf 4.4 Schoolprestaties zijn onverminderd van toepassing op het grenzentoetsend onderzoek.

Tabel 27. Moeilijkheidsgraad (P) van de items uit Blokpatronen, Vouwblaadjes, Natikken en Woorden-span bij eerste (1)^a, tweede (2)^b en derde (3)^b aanbieding

item	Blok- patronen ^c			Vouw- blaadjes			Naticken			Woorden- span		
	P			P			P			P		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	100	100	100	100	100	100	95	99	100	100	100	100
2	100	100	100	100	100	100	93	99	100	93	98	99
3	95	100	100	99	100	100	85	97	99	95	98	99
4	87	98	100	96	99	100	64	85	90	95	98	99
5	96	100	100	88	98	99	56	80	85	93	97	98
6	85	96	100	72	93	96	53	78	85	63	76	82
7	95	99	100	81	96	98	33	62	73	53	68	73
8	84	97	100	56	81	88	45	69	74	52	66	71
9	32	59	100	41	76	87	30	52	62	44	60	61
10	45	71	100	44	75	86	13	29	40	3	7	10
11			100	34	61	74	10	24	29	6	11	11
12			100	17	41	57	14	24	30	2	4	5
13			100				1	3	4			
14			100				1	5	7			
15			100				0	1	2			
16			100				0	2	2			
17			100				0	0	0			
N	411	382	398	411	398	398	411	410	410	411	408	408

^a standaardonderzoek

^b grenzetoetsend onderzoek

^c bij Blokpatronen vond geen derde aanbieding plaats

11. RESULTATEN

11.1 Betrouwbaarheid

Itemkenmerken

In tabel 27 en 28 worden respectievelijk de moeilijkheidsgraad (P) en de discriminatie-index (r_{ip}) van de items uit Blokpatronen, Vouwblaadjes, Natikken en Woordenspan weergegeven bij eerste (1), tweede (2) en derde (3) aanbieding. De eerste aanbieding heeft betrekking op het standaard-onderzoek; voor een uitvoerige toelichting op de hierbij verkregen P-waarden en discriminatie-indices raadplege men paragraaf 5.1. De tweede en derde aanbieding hebben betrekking op het grenzentoetsend onderzoek. Opgemerkt moet worden dat bij Blokpatronen geen derde aanbieding heeft plaatsgevonden⁵¹.

Uit tabel 27 kan men aflezen wat de herhaalde aanbieding van items - onder voor pp gunstige kondities⁵² - voor effect heeft op de hoogte van de P-waarden: alle P-waarden nemen toe - voor zover dit mogelijk is - met de sterkste stijging tussen aanbieding (1) en (2). Tabel 28 laat zien wat de gevolgen zijn voor de discriminatie-indices (item-restkorrelaties): vrijwel alle indices zijn gestegen in waarde. Oppervlakkig beschouwd wijst dit laatste op een verhoging van de homogeniteit of interne consistentie van de tests bij tweede en derde aanbieding. Deze verhoging is des te opmerkelijker indien men in de beschouwing ook het toegenomen aantal extreme P-waarden (>50) betreft, die - zoals bekend - een ongunstige invloed uitoefenen op de betrouwbaarheid van een test. Alvorens nu de feitelijke betrouwbaarheidsgegevens te bekijken in het licht van de verkregen P-waarden en discriminatie-indices, lijkt het nuttig om deze waarden eerst afzonderlijk per test te bespreken.

Blokpatronen. Deze test bleek al bij de eerste aanbieding aan de gemakkelijke kant te zijn. Bij de tweede aanbieding leveren de testitems zo mogelijk nog minder problemen op: de eerste acht items worden door vrijwel alle pp'n foutloos gemaakt (P-waarden $\geq .96$). Alleen de laatste twee items differentiëren nog in redelijke mate. De bijbehorende discriminatie-indices van .49 en .54 kunnen verder niet ongunstig genoemd worden. Desondanks mogen wij nu al stellen dat deze versie van Blokpatronen, afgenomen volgens een grenzentoetsende procedure, nauwelijks of niet differentieert binnen de onderzochte leeftijdsgroep.

Tabel 28. Discriminatie-index (r_{ir})^d van de items uit Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste (1)^a, tweede (2)^b en derde (3)^b aanbieding

item	Blok- patronen ^c		Vouw- blaadjes			Naticken			Woorden- span		
	r_{ir}		r_{ir}			r_{ir}			r_{ir}		
	(1)	(2)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	. ^d	-	-	-	-	.08	-	-	-	-	-
2	-	-	-	-	-	.01	-	-	.27	-	-
3	.23	-	-	-	-	.17	-	-	.24	-	-
4	.20	-	.14	-	-	.40	.45	.44	.33	-	-
5	.38	-	.28	-	-	.47	.47	.50	.40	-	-
6	.33	.23	.27	.30	.33	.44	.50	.55	.59	.61	.57
7	.43	-	.31	.23	-	.46	.56	.58	.61	.63	.63
8	.36	-	.37	.34	.44	.58	.65	.68	.57	.66	.68
9	.45	.49	.36	.42	.55	.54	.61	.61	.60	.64	.66
10	.44	.54	.40	.39	.58	.40	.53	.58	-	.29	.35
11			.42	.47	.59	.35	.50	.56	.30	.36	.38
12			.36	.42	.51	.44	.52	.58	-	.25	.33
13						-	-	.28			
14						-	.28	.34			
15						-	-	-			
16						-	-	-			
17						-	-	-			
N	411	382	411	398	398	411	410	410	411	408	408

^a standaardonderzoek

^b grenzentoetsend onderzoek

^c bij Blokpatronen vond geen derde aanbieding plaats

^d discriminatie-indices van items met een $P > .96$ of $< .4$ zijn weggelaten vanwege hun onbetrouwbaar karakter (cf. tabel 27)

Vouwblaadjes. De discriminatie-indices van de items 6 t/m 12 bij tweede aanbidding verschillen niet wezenlijk van die bij eerste aanbidding, ondanks het feit dat de bijbehorende P-waarden gemiddeld zo'n 25 punten hoger liggen. De discriminatie-indices van dezelfde items (met uitzondering van item 7) laten daarentegen bij derde aanbidding beduidend hogere waarden zien, ook al nemen de P-waarden met nog gemiddeld zo'n 9 punten toe bovenop de eerdere toename van 25 punten. Wat het uiteindelijk effect hiervan is op de hoogte van de betrouwbaarheid valt moeilijk te voorspellen. Wel kan men uit de hoogte en het verloop van de P-waarden afleiden dat Vouwblaadjes na herhaalde aanbidding niet al te goed zal differentiëren.

Natikken. Gezien het feit dat Natikken over voldoende moeilijke items beschikt, biedt deze test nog het beste perspectief op differentiatie bij grenzen-toetsen. De hoogte en de verdeling van de P-waarden bij tweede en derde aanbidding vormen een ondersteuning voor dit perspectief. De stijging van de P-waarden laat het van Vouwblaadjes bekende patroon zien: een relatief sterke stijging van (1) naar (2) en een duidelijk minder sterke stijging van (2) naar (3). De discriminatie-indices, welke op een enkele uitzondering na (item 13 en 14) zeer gunstig te noemen zijn, volgen eenzelfde patroon; dit in tegenstelling tot het stijgingspatroon van de discriminatie-indices bij Vouwblaadjes. Afgaande op de verkregen P-waarden en discriminatie-indices onder (2) en (3), mogen we de verwachting uitspreken dat de betrouwbaarheid van Natikken zal toenemen bij een grenzentoetsende procedure.

Woordenspan. Vergeleken met de voorafgaande drie tests levert het patroon van P-waarden en discriminatie-indices onder (2) en (3) weinig nieuws op. Opmerkelijk is echter wel dat de stijging van de P-waarden van (1) naar (2), en van (2) naar (3), beduidend minder sterk is dan bij Vouwblaadjes. Gelet op de hoogte en verdeling van de P-waarden onder (2) en (3) mogen we verwachten dat Woordenspan bij een grenzentoetsende procedure nog redelijk zal differentiëren. Of de betrouwbaarheid van deze test erg veel zal toenemen is, gezien de combinatie van P-waarden en discriminatie-indices, nog maar de vraag.

Interne consistentie en homogeniteit

In paragraaf 5.2 merkten wij reeds op dat de moderne generaliseerbaarheidstheorie van Cronbach et al. (1972) langzamerhand de plaats inneemt van de klassieke betrouwbaarheidstheorie. Als nu één probleem zich leent voor oplossing door middel van de generaliseerbaarheidstheorie, dan is

Tabel 29. Maximumscore, gemiddelde (m), standaarddeviatie (s), koëfficiënt van interne consistentie (r_{11})^a en homogeniteitskoëfficiënt (r_{kk}) van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste (1)^b, tweede (2)^c en derde (3)^c aanbieding

tests ^e	max.- score	m			s			r_{11}			r_{kk}		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Blokpatronen ^d	20	16.38	18.40	-	2.42	1.61	-	.70	- ^f	-	.63	- ^f	-
Vouwblaadjes	12	8.27	10.20	10.85	1.95	1.63	1.54	.66	.66	.79	.63	.62	.70
Natikken	17	5.94	8.08	8.82	2.52	2.76	2.81	.83	.86	.87	.73	.80	.82
Woordenspan	12	6.99	7.81	8.08	2.04	1.93	1.88	.80	.83	.85	.75	.75	.76

^a gekorrigeerd voor testhalvering met behulp van de Spearman-Brown formule

^b standaardonderzoek

^c grenzentoetsend onderzoek

^d bij Blokpatronen vond geen derde aanbieding plaats

^e voor de aantallen (N) per test en per aanbieding: zie tabel 27

^f voor toelichting: zie tekst van paragraaf 11.1 onder de titel 'Interne consistentie en homogeniteit'

het wel de bepaling van de betrouwbaarheid, c.q. generaliseerbaarheid, van de pre- en post-testscores van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan, welke - evenals bijvoorbeeld de optimaliserende kondities en de pl's - opgevat kunnen worden als de verschillende facetten van een generaliseerbaarheidsonderzoek. Helaas is de door ons gevolgde opzet niet van dien aard dat hij de bepaling van generaliseerbaarheidskoëfficiënten in relatie tot duidelijk gedefinieerde universa toelaat. Aangezien wij toch een indruk van de betrouwbaarheid willen krijgen, zijn wij daarom genooddaakt om enkele klassieke betrouwbaarheidskoëfficiënten te berekenen, i.c. de koëfficiënt van interne consistentie, de homogeniteitskoëfficiënt en de stabiliteitskoëfficiënt. Op laatstgenoemde koëfficiënt zullen wij in de volgende paragraaf nog terugkomen; aan de twee andere koëfficiënten hebben wij in paragraaf 5.2 reeds enkele woorden gewijd.

In tabel 29 worden respectievelijk de maximumscore, het gemiddelde (m), de standaarddeviatie (s), de koëfficiënt van interne consistentie (r_{11}) en de homogeniteitskoëfficiënt (r_{kk}) van Blokpatronen⁵³, Vouwblaadjes, Natikken en Woordenspan weergegeven bij eerste (1), tweede (2) en derde (3) aanbieding. Alvorens de resultaten per test te bespreken, eerst een paar algemene opmerkingen. Uit de tabel kan men aflezen dat het gemiddelde van iedere test toeneemt en wel het sterkst van (1) naar (2); de spreiding neemt navenant af, met uitzondering van Natikken. Met het gemiddelde neemt ook de betrouwbaarheid van de tests toe, zowel gemeten naar de koëfficiënt van interne consistentie als naar de homogeniteitskoëfficiënt.

Blokpatronen. Zoals de P-waarden reeds deden vermoeden, zijn de resultaten onder (2) bepaald niet gunstig te noemen. Was er bij deze test - afgenomen volgens standaardprocedures (1) - al duidelijk sprake van een 'plafond'-effekt, bij grenzen-toetsen manifesteert dit effekt zich nog sterker: het gemiddelde is met iets meer dan twee punten toegenomen tot 18.40 (max.score = 20), terwijl de spreiding met iets minder dan één punt gezakt is tot 1.61. Met andere woorden: Blokpatronen, afgenomen volgens een grenzentoetsende procedure, differentieert nauwelijks meer tussen de onderzochte pp'n. Mede gelet op het uiterst geringe aantal nog betrouwbare discriminatie-indices leek ons daarom de vermelding van de koëfficiënt van interne consistentie en de homogeniteitskoëfficiënt van Blokpatronen onder (2) in tabel 29 weinig zinvol.

Vouwblaadjes. Grenzen-toetsen heeft op de betrouwbaarheid van deze test een wonderlijk effekt, te oordelen naar het 'verloop' van beide koëffi-

ciënten van (1) naar (3). Ondanks het feit dat het gemiddelde bij tweede aanbieding met 2 punten stijgt tot 10.20 (max.score = 12) en de spreiding iets naar beneden gaat tot 1.63, veranderen de koëfficiënten van interne consistentie en de homogeniteitskoëfficiënt niet ($r_{11} = .66$) of nauwelijks ($r_{kk} = .62$). Met een relatief geringe toename van het gemiddelde tot 10.85 bij derde aanbieding en iets kleinere spreiding van 1.54 gaat echter een aanzienlijke verhoging van de interne consistentie ($r_{11} = .79$) gepaard en een wat minder sterke verhoging van de homogeniteit ($r_{kk} = .70$). Een eenduidige verklaring voor dit effect kunnen wij vooralsnog niet geven. Wel willen wij opmerken dat ook Vouwblaadjes bij herhaalde aanbieding een duidelijk 'plafond'-effect te zien geeft.

Natikken. Zoals verwacht kon worden op basis van de P-waarden en discriminatie-indices, is de betrouwbaarheid toegenomen: de interne consistentie met 4 punten tot .87 van (1) naar (3) en de homogeniteit met 9 punten tot .82, eveneens van (1) naar (3). De sterkste stijging in betrouwbaarheid vond echter plaats van (1) naar (2), resulterend in een koëfficiënt van interne consistentie van .86 en een homogeniteitskoëfficiënt van .80 onder (2). In tegenstelling tot de bevindingen bij de overige drie tests is de spreiding van Natikken bij tweede en derde aanbieding toegenomen tot respectievelijk 2.76 en 2.83. Met andere woorden: grenzen-toetsen verhoogt het differentiërend vermogen van deze test. Gegeven een relatief geringe toename in gemiddelde tot 8.08 onder (2) en 8.82 onder (3) hoeft men verder geen 'plafond'-effect te vrezen binnen de onderzochte leeftijdsgroep (max.score = 17).

Woordenspan. Vergeleken met de resultaten onder (1), laten het gemiddelde, de spreiding en de betrouwbaarheid van Woordenspan bij tweede en derde aanbieding maar weinig verschil zien. Zo neemt het gemiddelde uiteindelijk met net iets meer dan één punt toe - $m = 8.08$ onder (3) - terwijl de spreiding met nauwelijks een kwart punt afneemt, $s = 1.88$ onder (3). Ook de homogeniteit gaat er met één punt niet noemenswaard op vooruit: $r_{kk} = .76$ onder (3). De interne consistentie laat met een stijging van 5 punten van (1) naar (3) tot .85 nog het grootste verschil zien. Gelet op de waarden voor het gemiddelde en de spreiding onder (3) enerzijds en de maximumscore (= 12) anderzijds, lijkt het tenslotte onaanvaardig dat een plafond-effect de onderzoeksresultaten heeft beïnvloed.

11.2 Structuur

Ofschoon wij in de vorige paragraaf de berekening van de stabiliteitskoëfficiënt (test-hertestkorrelatie) aankondigden binnen het kader van

de klassieke betrouwbaarheidstheorie, stuit de strikte interpretatie van deze koëfficiënt als betrouwbaarheidsmaat op een tweetal bezwaren, althans in dit onderzoek.

Het eerste bezwaar betreft het verschil in methode. Bij grenzen-toetsen worden alleen de fout gescoorde items van de eerste testafname opnieuw aangeboden. De stabiliteitskoëfficiënt is daarentegen gebaseerd op de test-hertestmethode, waarbij niet alleen de fout gescoorde maar ook de goed gescoorde items voor een tweede keer worden aangeboden. Het tweede bezwaar heeft betrekking op het verschil in afnamekondities. Bij grenzen-toetsen worden de kondities waaronder de items voor een tweede of derde keer worden aangeboden, geoptimaliseerd. Bij de test-hertestmethode waakt men er juist voor dat de (standaard)testkondities van de eerste en tweede afname zoveel mogelijk gelijk zijn.

Als wij in het vervolg dan ook van de stabiliteit van een test spreken, dan bedoelen wij hiermee niet zozeer zijn betrouwbaarheid maar eerder zijn betekenis of validiteit. De vloeiende overgang, c.q. het relatieve verschil tussen beide begrippen, wordt door Drenth (1975) als volgt uitgedrukt:

"'Ideaaltypisch' gedefinieerd komt het verschil hierop neer dat betrouwbaarheid gebaseerd is op variantie bij herhalingen van *dezelfde* meting van een bepaald attribuut, en validiteit op variantie bij *verschillende* metingen van hetzelfde attribuut" (p. 302).

Zagen wij in paragraaf 11.1 dat de betrouwbaarheid van Vouwblaadjes, Nattikken en Woordenspan op een enkele uitzondering na omhoog gaat als gevolg van grenzen-toetsen, dan is nu de vraag in hoeverre dit ook met een betekenisverschuiving van de tests gepaard gaat. In eerste instantie valt een verschuiving qua testbetekenis af te leiden uit de mate van samenhang tussen de tests bij eerste, tweede en derde aanbieding. Een geringe mate van samenhang kan erop duiden dat de test zeer gevoelig is voor een inbreuk op de afnamekondities. In deze zin geformuleerd zou men de test als minder stabiel kunnen kenschetsen, et vice versa. In tweede instantie kan men een betekenisverandering op het spoor komen door de samenhang tussen de tests bij eerste, tweede en derde aanbieding en de overige vier tests uit de LDT-E te bestuderen.

In tabel 30 worden de korrelaties tussen de tests weergegeven bij eerste, tweede en derde aanbieding. De korrelatiekoëfficiënten in de driehoekjes van deze tabel representeren de samenhang tussen de verschillende aanbiedingen (metingen) van één test.

Blokpatronen. De tweede aanbieding van deze test heeft - vergeleken met de tweede aanbiedingen van de overige drie tests - de minste variantie

Tabel 30. Korrelaties^a tussen de tests bij eerste (1)^b, tweede (2)^c en derde (3)^c aanbieding

N : 368 - 411

tests		Blok- patronen ^d		Vouw- blaadjes			Natikken			Woorden- span		
		(1)	(2)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Blokpatronen ^d	(1)											
	(2)	.69										
Vouwblaadjes	(1)	.30	.32									
	(2)	.34	.39	.73								
	(3)	.38	.39	.64	.89							
Natikken	(1)	.38	.33	.31	.34	.34						
	(2)	.37	.33	.32	.37	.39	.85					
	(3)	.35	.30	.30	.35	.36	.79	.95				
Woordenspan	(1)	.20	.24	.15	.16	.16	.35	.38	.34			
	(2)	.25	.27	.18	.21	.19	.36	.41	.38	.89		
	(3)	.26	.29	.18	.22	.20	.36	.40	.38	.86	.96	
Plaatjes aanwijzen ^e	(1)	.29	.23	.30	.32	.30	.45	.48	.44	.63	.66	.65
Zinnen nazeggen ^e	(1)	.24	.27	.29	.31	.32	.39	.43	.42	.55	.62	.61
Verhaaltje vragen ^e	(1)	.22	.23	.25	.25	.23	.32	.33	.36	.27	.31	.30
Begrip en inzicht ^e	(1)	.16	.19	.16	.23	.24	.29	.28	.28	.28	.35	.33

^a alle korrelaties zijn significant op 1%-niveau ($p < .01$)

^b standaardonderzoek

^c grenzentoetsend onderzoek

^d bij Blokpatronen vond geen derde aanbieding plaats

^e bij Plaatjes aanwijzen, Zinnen nazeggen, Verhaaltje vragen en Begrip en inzicht vond geen tweede of derde aanbieding plaats

gemeen met de eerste aanbieding (48%). Verhoudingsgewijs mag Blokpatronen dan ook als een weinig stabiele test bestempeld worden. Desondanks laat het patroon van korrelaties tussen Blokpatronen onder (1) en (2) en de overige tests onder (1), (2) en (3) slechts geringe verschuivingen zien. De grootste verschuiving treft men nog aan bij Plaatjes aanwijzen: Blokpatronen (2) korreleert 6 punten lager met Plaatjes aanwijzen (.23) dan Blokpatronen (1): .29.

Vouwblaadjes. De grootste samenhang (.89) treft men aan tussen Vouwblaadjes (2) en (3); de kleinste samenhang (.64) tussen (1) en (3). De korrelatie tussen Vouwblaadjes (1) en (2) ligt daar tussenin (.73). Een dergelijk korrelatiepatroon is te verwachten op grond van het verschil in afnamekondities, namelijk: standaard-testkondities onder (1) en optimaliserende kondities onder (2) en (3). Vergeleken met Natikken en Woordenspan mag Vouwblaadjes als de minst stabiele test gekenschetst worden. Zo heeft bijvoorbeeld de derde aanbieding slechts 41% variantie gemeen met de eerste aanbieding. Ondanks deze naar verhouding geringe stabiliteit vertoont het patroon van korrelaties tussen Vouwblaadjes onder (1), (2) en (3) en de overige tests onder (1), (2) en (3) nauwelijks opmerkelijke verschuivingen. Het meest opvallend nog is de sprong in korrelaties van .16 naar .23 en .24 voor Vouwblaadjes (1), (2) en (3) met Begrip en inzicht. Hetzelfde geldt - zij het in iets mindere mate - voor de korrelaties met Natikken (2).

Natikken. Deze test laat binnen de driehoek hetzelfde korrelatiepatroon zien als Vouwblaadjes. De korrelaties tussen Natikken (1), (2) en (3) zijn echter aanzienlijk hoger dan die tussen Vouwblaadjes (1), (2) en (3), namelijk: .79 voor Natikken (1) en (3), .85 voor (1) en (2), en .95 voor (2) en (3). In percentages gemeenschappelijke variantie uitgedrukt zijn de cijfers respectievelijk 62, 72 en 90. Verhoudingsgewijs mag Natikken dan ook als een redelijk stabiele test beschouwd worden. Opmerkelijke verschuivingen in korrelaties tussen Natikken (1), (2) en (3) en de overige tests doen zich verder niet voor.

Woordenspan. Van alle vier de tests is Woordenspan wel de meest stabiele, met interkorrelaties variërend van .86 voor (1) en (3), en .89 voor (1) en (2) tot .96 voor (2) en (3). Uitgedrukt in percentages gemeenschappelijke variantie zijn de cijfers respectievelijk 74, 79 en 92. Evenals bij Blokpatronen, Vouwblaadjes en Natikken doen zich ook bij Woordenspan (1), (2) en (3) geen opvallende verschuivingen voor in het patroon van korrelaties met de overige tests uit de LDT-E. Nog het meest opmerkelijk is de sprong in korrelaties van .55 naar .62 en .61 voor Woordenspan (1), (2)

label 31. Relatieve frekventieverdeling (%), gemiddelde (m) en standaarddeviatie (s) van de schaalwaarden van vijf schalen voor Blokpatronen, Vouwblaadjes, Natikken en Woordenspan na tweede en derde aanbieding^a

schalen			Blokpatronen ^b				Vouwblaadjes			Natikken			Woordenspan		
			(N = 382)				(N = 398)			(N = 410)			(N = 408)		
			waarde	%	m	s	%	m	s	%	m	s	%	m	s
Aandacht -	normaal	1	75.6	1.34	0.67		54.8	1.59	0.75	46.1	1.70	0.76	65.7	1.40	0.60
	sterker	2	16.0				34.2			39.5			28.7		
	vrij sterk	3	7.1				8.8			12.2			5.4		
	zeer sterk	4	1.3				2.2			2.2			0.2		
Aanmoediging -	normaal	1	49.7	1.66	0.77		56.5	1.56	0.73	55.6	1.53	0.68	68.9	1.35	0.55
	sterker	2	37.4				32.4			37.3			27.7		
	vrij sterk	3	9.9				9.3			5.4			3.2		
	zeer sterk	4	3.0				1.8			1.7			0.2		
Interferentie -	normaal	1	86.1	1.18	0.49		74.9	1.33	0.63	75.9	1.33	0.66	83.3	1.21	0.50
	sterker	2	10.2				18.3			16.3			13.2		
	vrij sterk	3	3.1				5.8			6.3			2.9		
	zeer sterk	4	0.6				1.0			1.5			0.6		
Assistentie ^c -	geen	1	97.1	1.03	0.17		71.9	1.28	0.45	-	-	-	-	-	-
	wel	2	2.9				28.1			-			-		
Vaardigheid ^c -	normaal	1	89.0	1.12	0.37		57.8	1.52	0.60	85.6	1.18	0.49	-	-	-
	enigsz. onhandig	2	9.7				33.7			11.5			-		
	vrij onhandig	3	1.3				7.3			2.2			-		
	zeer onhandig	4	0.0				1.2			0.7			-		

^a grenzentoetsend onderzoek

^b bij Blokpatronen vond geen derde aanbieding plaats

^c de schalen Assistentie en Vaardigheid zijn niet van toepassing op Woordenspan; de schaal Assistentie is evenmin van toepassing op Natikken

en (3) met Zinnen nazeggen.

11.3 Bewerking van de schaalgegevens

De resultaten die in de vorige twee paragrafen zijn weergegeven, vloeien niet alleen voort uit de herhaalde aanbieding van items, maar zijn ook een gevolg van het feit dat pl enkele aspecten van het testgedrag tijdens het grenzen-toetsen manipuleerde. De mate waarin pl naar zijn oordeel het testgedrag heeft moeten manipuleren, moest na afloop van het grenzen-toetsen per test worden aangegeven op de schalen Aandacht, Aanmoediging, Interferentie en - indien van toepassing - op de schaal Assistentie. Met uitzondering van Woordenspan moest pl tevens zijn indruk van pp's motorische vaardigheid per test weergeven op de schaal Vaardigheid (cf. hfdst. 10).

In tabel 31 worden de relatieve frekwentieverdeling (%), het gemiddelde (m) en de standaarddeviatie (s) gegeven van de verkregen schaalwaarden voor Blokpatronen, Vouwblaadjes, Natikken en Woordenspan. Aangezien deze tabel nauwelijks enige toelichting behoeft zullen wij alleen de belangrijkste resultaten per schaal kort bespreken.

Aandacht. Uit de percentages en gemiddelden van de schaalwaarden per test valt af te leiden dat Natikken de meest aandachtsgevoelige test is en Blokpatronen de minst aandachtsgevoelige. Zo moest pl bij meer dan 50% van de groep onderzochte kinderen de aandacht trekken om een optimale testprestatie op Natikken te verkrijgen, terwijl dit op Blokpatronen slechts in 25% van de gevallen naar het oordeel van pl noodzakelijk bleek.

Aanmoediging. De percentages en gemiddelden van de schaalwaarden per test ontlopen elkaar niet veel. Bij gemiddeld iets minder dan de helft van het aantal onderzochte kinderen besloot pl tot aanmoediging over te gaan. Woordenspan vormt hierop een uitzondering in die zin dat een derde van de kinderen meer dan normale aanmoediging behoefde.

Interferentie. Deze optimaliserende konditie heeft betrekking op al die (ongeremde) gedragingen van pp die interfereren met het gewenste (standaard-)testgedrag. Gemiddeld één vijfde van de kinderen in de onderzochte leeftijdsgroep vertoonde zodanig ongeremd gedrag dat pl zich ter verzekering van een optimale testprestatie genooddaakt voelde om in te grijpen.

Assistentie. Slechts 2.9% van de kinderen behoefde assistentie bij het naleggen van Blokpatronen in de vorm van een rubbermatje; dit in tegenstelling tot het niet onaanzienlijke percentage kinderen (28.1) dat

passief geholpen werd bij het navouwen op Vouwblaadjes. Dit laatste gegeven ondersteunt indirect de juistheid van een eerder geopperde interpretatie dat van alle functies die door Vouwblaadjes gemeten worden 'oog - hand coördinatie' - met het aksent op de fijne motoriek - niet het kleinste deel van de testvariantie in beslag lijkt te nemen (cf. par. 8.3; p. 131). Verdere ondersteuning van deze interpretatie verschaffen de cijfers voor Vaardigheid.

Vaardigheid. Gemiddeld 12.7% van de kinderen beschikt volgens het oordeel van pl over een minder dan normale fijnmotorische vaardigheid in het leggen van Blokpatronen en het natikken van de blokjes op Natikken. Bij het navouwen op Vouwblaadjes wordt daarentegen 42% van de kinderen als minder handig beschouwd. Ook dit wijst er op dat Vouwblaadjes een relatief sterk beroep doet op de 'oog - hand coördinatie', c.q. de fijnmotorische vaardigheid.

Direkte ondersteuning van deze zienswijze verkregen wij via toetsing van het geslachtsverschil op de schaal Vaardigheid, uitgaande van de volgende redenering (voor het gemak citeren wij uit deel A):

"In tegenstelling tot de bevindingen van Maccoby & Jacklin (1975) dat jongens van deze leeftijd in het algemeen iets betere prestaties dan meisjes leveren op tests voor visueel-ruimtelijke vaardigheden, staan de resultaten van ons onderzoek bij Vouwblaadjes: meisjes doen het duidelijk beter dan jongens op deze test. Gelet op de functies die door Vouwblaadjes gemeten zouden worden (cf. par. 3.2), evenals op de bevindingen van McGuiness en Pribram (geciteerd door Goleman, 1978; p. 59) dat meisjes van jongs af aan over een betere handvaardigheid en fijne motoriek beschikken, lijkt het aannemelijk dat de functie 'oog - hand coördinatie' het verschil in testprestaties tussen jongens en meisjes verklaart" (p. 130-131).

Met andere woorden: als Vouwblaadjes deels 'oog - hand coördinatie' meet, dan moeten jongens in het algemeen ook een minder handige indruk maken dan meisjes op deze test. Getoetst werd de nulhypothese 'er zijn geen verschillen in vaardigheid tussen jongens en meisjes op Vouwblaadjes' tegen de alternatieve hypothese dat er wel verschillen zijn. De verkregen χ^2 -waarde bleek zeer significant te zijn ($\chi^2 = 16.83$; $df = 3$; $p < 0.001$) zodat de nulhypothese verworpen moest worden. Inspectie van de waargenomen frequenties voor jongens en meisjes liet tenslotte zien dat de significantie veroorzaakt wordt door de grotere onhandigheid van jongens.

Reeds eerder - in paragraaf 9.4 - vestigden wij de aandacht op het feit dat de gezamenlijke introductie van een aantal optimaliserende kondities krachtens de door ons gevolgde methode (hfdst. 10) tot een onwenselijke kontaminatie van variabelen leidt. Zo is het bijvoorbeeld onmogelijk om

het afzonderlijke effect van de herhaalde aanbieding van items te bestuderen, net zo min als het mogelijk is om het afzonderlijke effect van een konditie als Aandacht vast te stellen. Deze - onderzoekstechnisch gezien - onwenselijke situatie wordt nog gekompliceerder als men bedenkt dat pl naar eigen inzicht zowel het moment kan bepalen waarop hij één of meer van de kondities Aandacht, Aanmoediging, Interferentie en Assistentie varieert als de mate waarin hij het testgedrag beïnvloedt overeenkomstig deze kondities.

Ondanks deze komplicerende factoren is het toch mogelijk een voorlopige indruk te krijgen van het effect van herhaalde aanbieding enerzijds en van het effect van herhaalde aanbieding plus overige optimaliserende kondities anderzijds, door de onderzoeksgroep per test in tweeën te splitsen, groep A en groep B. Groep A heeft betrekking op al die kinderen wier testgedrag bij herhaalde aanbieding niet of nauwelijks tot interventie noodde en die daarom een één (1) scoorden op de schalen Aandacht, Aanmoediging, Interferentie en Assistentie. Groep B betreft de resterende kinderen die wel een of andere vorm van hulp behoeften en daarom tenminste een twee (2) scoorden op een of meer van de schalen (uitgezonderd Vaardigheid).

In tabel 32 worden respectievelijk het gemiddelde (m) en de standaarddeviatie (s) weergegeven van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste (1), tweede (2) en derde (3) aanbieding voor de groepen A 'zonder hulp' en B 'met hulp' (cf. p. 182).

Het eerste dat opvalt in deze tabel is dat de gemiddelde testprestatie van groep B systematisch lager is dan die van groep A, zowel bij eerste, tweede als derde aanbieding. Laten wij alleen op de gemiddelde testprestaties bij eerste aanbieding - het standaard-testonderzoek - dan zijn de resultaten voor tweërlei uitleg vatbaar. Ten eerste: pl is eerder geneigd om kinderen met een relatief lage score op het standaard-testonderzoek te helpen dan kinderen met een relatief hoge score; en/of ten tweede: kinderen met een relatief lage score vertonen meer ongewenst testgedrag dan kinderen met een relatief hoge score.

Het tweede punt dat opvalt is dat de toename in gemiddelde testprestaties van (1) naar (2) en van (2) naar (3) per test vrijwel dezelfde is voor groep A en B. Groep B gaat weliswaar absoluut genomen iets meer vooruit dan groep A, maar het gemiddelde verschil van 0.38 punten over alle tests tussen eerste en laatste aanbieding is in eerste instantie te verwaarlozen. In eerste instantie, omdat ook hier de resultaten weer

Tabel 32. Maximumscore, gemiddelde (m) en standaarddeviatie (s) van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste (1)^a, tweede (2)^b en derde (3)^b aanbieding voor twee groepen: groep A 'zonder hulp' en groep B 'met hulp'^c

tests		max.- score	groep A			groep B		
			m	s	N	m	s	N
Blokpatronen ^d	(1)	20	17.05	1.93	165	15.41	2.30	217
	(2)	20	19.06	1.37	165	17.91	1.60	217
Vouwblaadjes	(1)	12	9.01	1.71	136	7.73	1.80	262
	(2)	12	10.90	1.41	136	9.84	1.62	262
	(3)	12	11.26	1.29	136	10.64	1.61	262
Naticken	(1)	17	6.64	2.57	143	5.58	2.41	267
	(2)	17	8.80	2.83	143	7.70	2.65	267
	(3)	17	9.41	2.74	143	8.50	2.80	267
Woordenspan	(1)	12	7.38	1.93	209	6.53	2.01	199
	(2)	12	8.21	1.86	209	7.38	1.92	199
	(3)	12	8.38	1.77	209	7.77	1.93	199

^a standaardonderzoek

^b grenzentoetsend onderzoek

^c voor toelichting: zie tekst van paragraaf 11.3

^d bij Blokpatronen vond geen derde aanbieding plaats

voor verschillende uitleg vatbaar zijn. Zo wil het verkregen minieme verschil niet zeggen dat dus de hulp aan groep B nauwelijks enig effect heeft gesorteerd. Het is immers best mogelijk dat groep B - gelet op haar testgedrag - aanzienlijk minder vooruitgegaan zou zijn, indien pl géén hulp had geboden. Daar staat weer tegenover dat groep B - gelet op haar lagere gemiddelde testprestaties bij eerste aanbieding - meer kan- sen heeft om vooruit te gaan bij herhaalde aanbieding dan groep A (het zgn. 'regression to the mean' effect). Onze proefopzet en dergelijke in aanmerking genomen, is het onmogelijk om een eenduidige verklaring te geven voor de verkregen resultaten. De cijfers van tabel 32 geven dan ook slechts een ruwe, voorlopige indruk van wat men kan verwachten bij grenzen-toetsen, en beslist niets meer.

11.4 Prediktieve validiteit

Tabel 33 geeft een overzicht van de korrelaties tussen Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste (1), tweede (2) en derde (3) aanbieding enerzijds en acht criteriumvariabelen anderzijds. Voor

een toelichting op de criteriumvariabelen raadplege men de paragrafen 4.4 en 7.1.

Slechts een vluchtige blik op tabel 33 is voldoende om te konstaten dat grenzen-toetsen - ondanks een overwegend hogere betrouwbaarheid van de tests bij tweede en derde aanbieding - geen unanieme verhoging van de prediktieve validiteit ten gevolge heeft. De toch al niet overdreven hoge prediktieve validiteitskoëfficiënten van Blokpatronen (1) zakken zelfs bij tweede aanbieding fors naar beneden over de gehele linie van criteriumvariabelen (mediane verlaging van de korrelaties: 6 punten). Een verklaring hiervoor is niet moeilijk te geven: het hoge gemiddelde en de geringe spreiding ('plafond'-effekt) - cf. par. 11.1 - oefenen een negatieve invloed uit op de korrelaties met de diverse criteriumvariabelen.

Ook bij Vouwblaadjes doet zich het verschijnsel voor dat grenzen-toetsen tot een verlaging van de prediktieve validiteitskoëfficiënten leidt en wel het sterkst ten aanzien van de Brus, Oordeel Lezen en Oordeel Taal (respektievelijk 6, 10 en 5 punten verschil tussen eerste en derde aanbieding). Het eerder gekonstateerde 'plafond'-effekt biedt onzes inziens een aannemelijke verklaring voor dit verschijnsel, ondanks het feit dat zowel de interne consistentie als de homogeniteit van Vouwblaadjes bij derde aanbieding fors is toegenomen (cf. par. 11.1). Ogenschijnlijk een paradoxaal resultaat dat men als volgt zou kunnen interpreteren: door middel van grenzen-toetsen meet men op redelijk betrouwbare wijze iets wat nog maar weinig verband houdt met de latere schoolprestaties van het kind. In dit licht bezien is het adjektief 'spurieuus' geen overdreven aanduiding voor de verkregen betrouwbaarheidskoëfficiënten van Vouwblaadjes onder (2) en (3).

Het effect van grenzen-toetsen op Natikken is vrijwel nihil: sommige korrelaties gaan 1 à 2 punten achteruit (Beelddiktee, Brus, Schiedamse rekentest, Oordeel Schrijven), andere gaan 1 à 2 punten vooruit (Oordeel Taal, Oordeel Rekenen) en weer andere blijven gelijk (Begrijpend lezen, Oordeel Lezen); dit ondanks de grotere spreiding en de hogere betrouwbaarheid van Natikken bij tweede en derde aanbieding (cf. par. 11.1). Een verklaring hiervoor kunnen wij vooralsnog niet geven, zeker niet in het licht van de resultaten bij Woordenspan met zijn iets kleinere spreiding en iets hogere betrouwbaarheid.

Laatstgenoemde test is namelijk de enige waarbij grenzen-toetsen wèl een positief effect oplevert, zij het niet overweldigend: alle prediktieve

Tabel 33. Korrelaties^a van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan bij eerste (1)^b, tweede (2)^c en derde (3)^c aanbieding met acht criteriumvariabelen

N : 290 - 337

criteriumvariabelen		Lezen / taal				Rekenen		Schrijven	
		Begrijpend lezen I	Beeld- diktee I	Brus-Eén- Minuut-Test vorm B	Oordeel Lezen	Oordeel Taal	Schiedamse rekentest	Oordeel Rekenen	Oordeel Schrijven
Blokpatronen ^d	(1)	.24	.22	.16	.13	.14	.23	.28	.29
	(2)	.18	.16	.11	.05	.11	.14	.25	.20
Vouwblaadjes	(1)	.27	.22	.23	.27	.29	.22	.30	.27
	(2)	.26	.21	.18	.19	.27	.21	.30	.27
	(3)	.24	.23	.17	.17	.24	.20	.28	.25
Natikken	(1)	.45	.35	.35	.32	.37	.30	.41	.27
	(2)	.45	.34	.35	.32	.38	.32	.45	.26
	(3)	.45	.33	.33	.32	.38	.29	.43	.25
Woordenspan	(1)	.38	.36	.32	.31	.29	.18	.31	.28
	(2)	.44	.41	.37	.32	.31	.19	.34	.30
	(3)	.42	.39	.38	.33	.32	.20	.34	.28

^a korrelaties $\geq .13$ zijn significant op 1%-niveau ($p < .01$)

^b standaardonderzoek

^c grenzentoetsend onderzoek

^d bij Blokpatronen vond geen derde aanbieding plaats

validiteitscoëfficiënten stijgen van (1) naar (2) met 1 tot 6 punten, afhankelijk van de betrokken criteriumvariabelen. De grootste stijging treft men aan bij Begrijpend lezen, Beelddiktee en de Brus (5 à 6 punten). De overige criteriumvariabelen laten slechts een marginale verhoging van 1 tot 3 punten zien. Opvallend is tenslotte dat de korrelaties van (2) naar (3) nauwelijks verder stijgen (Brus, Oordeel Lezen, Oordeel Taal, Schiedamse rekentest) en in sommige gevallen zelfs dalen (Begrijpend lezen, Beelddiktee, Oordeel Schrijven) of gewoon gelijk blijven (Oordeel Rekenen).

In het voorgaande konden wij konstateren dat de post-testscores van Blokpatronen en Vouwblaadjes - dat wil zeggen de scores behaald bij tweede of derde aanbieding onder optimaliserende kondities - een geringere samenhang met latere schoolprestaties vertonen dan de pre-testscores, en dat de post-testscores van Natikken en Woordenspan respectievelijk eenzelfde of iets hogere samenhang laten zien. Alvorens nu de konklusie te trekken dat men het grenzen-toetsen op Blokpatronen en Vouwblaadjes maar beter kan laten en dat grenzen-toetsen op Natikken en Woordenspan meer tijdverdrijf dan nuttige tijdsbesteding betekent, lijkt het zinvol de onderzoeksgegevens nog op een andere manier te bekijken.

Men zal zich herinneren dat één van de twee vragen die wij in dit onderzoek wilden beantwoorden, als volgt luidde: "Kan het verschil van post- en pre-testscores (zgn. winst-, toename- of veranderingsscores) - gegeven bepaalde pre-testscores - een maat voor het leervermogen verschaffen?" (cf. par. 9.4). Ter beantwoording van deze vraag zullen wij nagaan in hoeverre de toenamescores per test een modererende invloed uitoefenen op de korrelatie tussen prediktor (pre-testscores) en criterium (schoolprestaties). Indien namelijk toenamescores als moderator fungeren - dat wil zeggen indien er een significante interactie bestaat tussen pre- en post-testscores bij de voorspelling van schoolprestaties - dan kunnen deze niet alleen gebruikt worden voor een verbetering van de voorspelling, maar ook voor het opstellen van een maat voor het leervermogen. Daarnaast zullen wij nagaan in hoeverre een variabele als 'hulp' (cf. par. 11.3) eveneens een modererende invloed uitoefent op de relaties tussen de prediktoren (pre- en post-testscores), en criterium (schoolprestaties); de kans is immers niet denkbeeldig dat wat voor groep A 'zonder hulp' geldt, niet opgaat voor groep B 'met hulp', et vice versa. Mellenbergh en Pannekoek (1979) geven een overzicht van de verschillende methoden waarmee men in de loop der tijd het optreden van een of meer

moderator-effekten heeft trachten te signaleren. De meest bekende zijn wel regressie-, (co)variantie- en discriminantanalyse. In hetzelfde overzicht laten zij zien aan welke voorwaarden de gegevens van psychologisch onderzoek moeten voldoen, indien men van genoemde methoden een verantwoord gebruik wil maken, met name waar het de analyse van kwalitatieve (ordinale/nominale), diskrete gegevens betreft. Na een bespreking van de argumenten vóór en tegen de verschillende methoden, komen Mellenbergh en Pannekoek (op.cit.) tot de volgende konklusie:

"These arguments lead to the consideration of methods with weak distributional assumptions for the analysis of qualitative discrete data. An example is the log-linear model which only assumes a multinomial distribution, and which is suitable for the analysis of qualitative discrete data. It is also possible to include quantitative, discrete variables in the model. It was said earlier that in moderator research a subject is generally considered to be drawn randomly from a total population. The appropriate assumption for the log-linear model is then the multinomial distribution. The model applies also if the sampling is done from fixed predictor or fixed moderator categories (Fienberg, 1977, p. 15). For these reasons the log-linear model can be applied to a wide variety of data. Moreover it has the desirable properties that the fit of a model can be tested, tested models can be compared statistically, and parameters can be estimated" (p. 21 en 22).

Om twee redenen nu lijkt toepassing van het log-lineaire model in ons onderzoek de meest geschikte methode ter opsporing van enigerlei moderatoreffekt. Ten eerste: vergeleken met de meer konventionele regressiemethoden stelt het log-lineaire model de minst strenge eisen aan de verdelings- en schaalkenmerken van onze onderzoeksvariabelen. Ten tweede: ook al hebben wij tot nog toe alle test- en toetsscores als kwantitatieve (intervalschaal-)gegevens beschouwd, in feite zijn het kwalitatieve (ordinale) gegevens die slechts met behulp van daartoe geëigende schaaltechnieken in kwantitatieve gegevens omgezet kunnen worden; en dit laatste is noodzakelijk wil men bijvoorbeeld van covariantie-analyse - als de 'next best' methode - een korrekt gebruik maken.

Omdat een begrijpelijke, niet-mathematische uiteenzetting van een relatief onbekende methode als het log-lineaire model nogal wat ruimte vraagt, zullen wij deze hier niet geven. Voor een eenvoudige introductie tot dit model zijn echter de publikaties van Everitt (1977) en Fienberg (1977) zeer geschikt. Beide publikaties vormen weer een inleiding tot het standaardwerk van Bishop et al. (1975) waarin uitvoerig de analyse van meergeweg kruistabellen wordt behandeld. Hoogstraten en Mellenbergh (1978) geven verder een verhelderend, niet-mathematisch voorbeeld van de praktische toepassing van het log-lineaire model.

De vierweg kruistabellen 34, 35, 36 en 37 (Bijlage; p. 262 t/m 265) vormen het uitgangspunt voor de analyse van respectievelijk Blokpatronen, Vouwblaadjes, Natikken en Woordenspan in relatie tot het samengestelde criterium Lezen/taal. Alvorens de resultaten van deze analyse te presenteren, zullen wij eerst een korte toelichting op de samenstelling van deze tabellen geven.

Tabel 34 geeft de frekwenties weer van de samengestelde criteriumvariabele Lezen/taal, uitgesplitst naar de prediktorvariabele Blokpatronen - pretest en de moderatorvariabelen Blokpatronen - posttest (tweede aanbieding) en Blokpatronen - hulp. De criteriumvariabele Lezen/taal is samengesteld uit de som van de z-scores op Begrijpend lezen, Beelddiktee, Brus, Oordeel Lezen en Oordeel Taal, en vervolgens opgesplitst in drie categorieën (laag - midden - hoog) met een vrijwel gelijk aantal pp'n per categorie. Gekozen is voor de criteriumvariabele Lezen/taal, aangezien de betrouwbaarheid van deze variabele zeer groot mag worden geacht (cf. par. 7.1) en wij primair enig resultaat ten aanzien van dit criterium mogen verwachten.

De pre-testscores van Blokpatronen zijn eveneens opgesplitst in drie categorieën (laag - midden - hoog) met een ongeveer gelijk aantal pp'n per categorie. De pre-testscores van de lage categorie variëren van 0 - 15, die van de midden-categorie van 16 - 17 en die van de hoge categorie van 18 - 20. Voor de post-testscores van Blokpatronen is eenzelfde indeling in categorieën aangehouden, respectievelijk: 0 - 15 (laag), 16 - 17 (midden) en 18 - 20 (hoog). Voor de moderatorvariabele 'hulp' is de reeds eerder vermelde indeling in twee categorieën of groepen gehanteerd: groep A 'zonder hulp' en groep B 'met hulp' (cf. par. 11.3). Het resultaat is een 3 (pretest) $\times$ 3 (posttest) $\times$ 2 (hulp) $\times$ 3 (Lezen/taal) kruistabel voor Blokpatronen - Lezen/taal, zoals weergegeven in tabel 34.

De tabellen 35, 36 en 37 zijn op overeenkomstige wijze samengesteld, zij het dat de pre- en post-testscore-indeling van respectievelijk Vouwblaadjes, Natikken en Woordenspan afwijkt van die van Blokpatronen. Voor Vouwblaadjes (tabel 35) is de volgende indeling aangehouden: 0 - 7 (laag), 8 - 9 (midden) en 10 - 12 (hoog). Voor Natikken (tabel 36) geldt: 0 - 4 (laag), 5 - 7 (midden) en 8 - 17 (hoog), terwijl voor Woordenspan (tabel 37) de driedeling 0 - 6 (laag), 7 - 8 (midden) en 9 - 12 (hoog) gehanteerd werd. Daarnaast hebben de post-testscores van Vouwblaadjes, Natikken en Woordenspan alleen betrekking op de derde aanbieding.

Vergeleken bij de theoretische spreidingsbreedte van de test-, schaal- en toetsscores, mag een indeling in twee of drie categorieën per varia-

bele tamelijk grof genoemd worden. Het relatief geringe aantal pp'n per kruistabel - met alle gevolgen voor een adequate celvulling vanden - weerhield ons echter van een meer verfijnde indeling, bijvoorbeeld in vier of vijf categorieën of zelfs evenveel categorieën als er items zijn. Bij voorbaat kunnen wij daarom enig informatieverlies verwachten. Tenslotte: de nu volgende berekeningen zijn uitgevoerd met behulp van het programma ECTA (Goodman & Fay, 1974), uitgaande van de tabellen 34, 35, 36 en 37. Aangezien in al deze tabellen randfrequenties van 0 'kunnen' voorkomen - met als gevolg dat het aantal vrijheidsgraden gewijzigd wordt - zijn eerst alle celfrequenties met 0.01 verhoogd. Met opzet schreven wij 'kunnen' voorkomen, aangezien randfrequenties van 0 - afhankelijk van de wijze waarop de tabellen 34 t/m 37 bij verdere analyse samengeklapt worden over één of meer variabelen - niet hoeven voor te komen.

Als eerste stap in de analyse hebben wij onderzocht welk log-lineaire model een goede beschrijving oplevert van de gegevens per tabel, c.q. test. Volgens een algemeen gebruikelijke strategie wordt hiertoe dat model gekozen dat het minste, nog zinnige aantal parameters telt (Everitt, 1977; p. 99); i.c. het model met uitsluitend 'hoofdeffekten'. In sterk versimpelde notatie ziet de 'formule' voor dit model er als volgt uit:

model 1: 12, 13, 14, 23, 24, 34

waarin de cijfers 1, 2, 3, en 4 betrekking hebben op variabele 1 (pre-test), variabele 2 (posttest), variabele 3 (hulp) en variabele 4 (Lezen/taal); de combinaties 12, 13, 14 etc. representeren dan de 'hoofdeffekten' (naar analogie van de variantie-analyse).

Toetsing van model 1 vond plaats via berekening van χ^2_L (chi-kwadraat van de aannemelijkheidsverhouding) per test. Geen enkele χ^2_L bleek echter significant te zijn ($p > .5$). In tabel 38 worden onder andere de resultaten voor model 1 weergegeven.

De interpretatie van de verkregen resultaten luidt als volgt. Een model met uitsluitend hoofdeffekten blijkt een zeer goede beschrijving van de gegevens op te leveren. Dit betekent dat in eerste instantie geen enkel interactie- of moderator-effekt aangetoond kan worden, noch bij Blokpatronen, noch bij Vouwblaadjes, Natikken en Woordenspan. Het is echter mogelijk dat door de toevoeging van een combinatie van telkens twee variabelen aan model 1 - met andere woorden: door de toevoeging van interactieparameters aan het model - alsnog een moderator-effekt gedemonstreerd kan worden.

Tabel 38. Chi-kwadraten aannemelijkheidsverhouding (χ^2_L)^a met aantal vrijheidsgraden (df) voor vier modellen per test^b

model ^b	df	χ^2_L			
		Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span
I: 12, 13, 14, 23, 24, 34	28	11.86	10.86	14.88	5.14
II: 124, 13, 23, 34	20	10.43	9.10	9.40	4.01
III: 134, 23, 24, 12	24	9.03	10.61	14.02	3.27
IV: 234, 12, 13, 14	24	9.45	7.21	10.97	4.13
N		284	315	304	301

^a geen enkele χ^2_L is significant ($p > .5$)

^b 1 = pretest; 2 = posttest; 3 = hulp; 4 = Lezen/taal

Als tweede stap in de analyse hebben wij daarom al die modellen uitgeschreven die extra-parameters omvatten, namelijk:

model II: 124, 13, 23, 34

model III: 134, 23, 24, 12

model IV: 234, 12, 13, 14

model V: 123, 14, 24, 34

waarin de combinaties 123, 134 en 234 te vergelijken zijn met de interacties van telkens twee variabelen (1 en 2, 1 en 3, 2 en 3) bij de voor spelling van variabele 4 (Lezen/taal). Aangezien variabele 4 niet in de interactie-term van model V zit opgesloten, hebben wij dit model bij verdere analyse weggelaten.

Toetsing van de modellen II t/m IV vond plaats via berekening van χ^2_L per test. In tabel 38 worden de verkregen waarden weergegeven. Geen enkele χ^2_L bleek significant te zijn ($p > .5$); met andere woorden alle modellen leveren een goede beschrijving van de gegevens op. De vraag is nu of de passing van een van de modellen II t/m IV beter is dan die van model I. Everitt (1977) merkt hierover het volgende op:

"In many cases (...) it will be found that several models provide an adequate fit to the data as indicated by the non-significance of the likelihood ratio criterion. In general the preferred model will be that with fewer parameters. In some cases, however, a test between rival models may be required to see which gives the most parsimonious summary of the data. Goodman (1971) and Fienberg (1970) show that (...) such a test may be obtained by subtracting the χ^2_L values for the two models to assess the 'change' in goodness of fit which results from adding further parameters. (...) The model including the extra parameter would

only be preferred if it provided a significantly *improved* fit over the simpler model" (p. 99).

Het antwoord op de vraag luidt dus dat het verschil in χ^2_L tussen model I en één der modellen II t/m IV een indicatie vormt voor het best passende model. Dit verschil volgt in grote steekproeven bij benadering een χ^2 -verdeling met een aantal vrijheidsgraden gelijk aan het verschil in vrijheidsgraden tussen beide modellen. Zoals in een oogopslag uit tabel 38 af te lezen valt, blijkt geen enkel verschil in χ^2_L significant te zijn ($p > .05$); met andere woorden: geen van de modellen II t/m IV geeft een verbetering ten opzichte van model I, noch voor Blokpatronen, noch voor Vouwblaadjes, Natikken en Woordenspan.

Onze konklusie moet dan ook luiden dat model I - het model met uitsluitend hoofdeffekten - de best passende beschrijving oplevert van de gegevens per test. Eenzelfde konklusie moesten wij eveneens trekken na analyse van dezelfde gegevens per test, maar dan met een meer verfijnde indeling van de variabelen in vier of vijf categorieën. Gegeven deze resultaten leek het tenslotte weinig zinvol op zoek te gaan naar eventuele moderatoreffekten bij de voorspelling van andere, waarschijnlijk minder betrouwbare criteria dan Lezen/taal.

12. DISKUSSIE

"Op voorhand lijkt het (...) niet volstrekt onaannemelijk dat grenzen-toetsen - mits zo uniform mogelijk uitgevoerd na afloop van een standaard-testonderzoek - aanvullende informatie kan verschaffen over het cognitief functioneren van een kind. Daartoe dient dan wel de waarde van deze informatie nader onderzocht te worden", zo merkten wij aan het eind van paragraaf 9.3 op. In de hoofdstukken 10 en 11 deden wij achtereenvolgens verslag van de methode en de resultaten van ons onderzoek dat in laatste instantie gericht was op de ontwikkeling van een praktische maat voor het leervermogen. Dat wij hier niet in geslaagd zijn, staat - gelet op de onderzoeksresultaten - buiten kijf. Of dit nu meteen betekent dat grenzen-toetsen onder het hoofdstuk 'klinische folklore' gerangschikt moet worden, is nog maar de vraag. Het is bijvoorbeeld best mogelijk dat de door ons gevolgde methode van onderzoek minder adequaat is dan zij aanvankelijk leek.

In de nu volgende paragraaf 12.1 zullen wij enkele kanttekeningen plaatsen bij de methode van onderzoek, uitgaande van onze onderzoeksbevindingen. In de daaropvolgende paragraaf 12.2 "Learning potential assessment" zullen wij ingaan op een alternatieve testprocedure, die nauw verwant is aan grenzen-toetsen. De derde en laatste paragraaf van dit hoofdstuk behelst enkele suggesties voor verder onderzoek.

12.1 Kanttekeningen bij de methode

In paragraaf 11.4 konden wij konstateren dat de post-testscores van Blokpatronen en Vouwblaadjes een geringere samenhang met schoolprestaties in de eerste klas GLO vertonen dan de pre-testscores, en dat de post-testscores van Natikken en Woordenspan respectievelijk eenzelfde of iets hogere samenhang laten zien. Op basis van een analyse van de gegevens volgens het log-lineaire model mochten wij voorts konkluderen dat nòch de post-testscores nòch hulp een modererende invloed uitoefenen op de samenhangen tussen de pre-testscores van Blokpatronen, Vouwblaadjes, Natikken en Woordenspan enerzijds en het criterium Lezen/taal in de eerste klas GLO anderzijds.

Is een dergelijk resultaat niet bij voorbaat te verwachten? Zuiver en alleen gelet op de verhoogde koëfficiënten van interne consistentie van de tests bij laatste aanbieding (Vouwblaadjes $r_{11} = .79$, Natikken $r_{11} = .87$, Woordenspan $r_{11} = .85$), zijn wij geneigd om deze vraag ontkennend te beantwoorden; althans voor zover het Natikken en Woordenspan betreft.

Ten aanzien van Vouwblaadjes luidt ons antwoord duidelijk bevestigend. Niet alleen vanwege het duidelijke 'plafond'-effekt dat wij bij deze test konden konstateren (cf. par. 11.1), maar ook vanwege de lage korrelatie tussen de eerste en derde aanbieding ($r_{(1)(3)} = .64$), cf. par. 11.2. In het licht van deze gegevens kunnen wij nog maar weinig waarde hechten aan de interne consistentie van Vouwblaadjes bij derde aanbieding, evenals aan zijn homogeniteit ($r_{kk} = .70$).

Dezelfde argumenten die voor Vouwblaadjes gelden, kunnen ook aangevoerd worden voor Blokpatronen: enerzijds een overduidelijk 'plafond'-effekt, anderzijds een lage korrelatie tussen eerste en tweede aanbieding ($r_{(1)(2)} = .69$). Het 'plafond'-effekt was bovendien zo sterk dat wij het weinig zinvol achtten om betrouwbaarheidscoëfficiënten voor Blokpatronen te vermelden.

Het moge duidelijk zijn dat bovenstaande gegevens weinig goeds voorspellen voor een prediktief valideringsonderzoek, evenals voor het aantonen van enigerlei moderatoreffekt. Hadden wij bovendien de beschikking gehad over de psychometrische kenmerken van Blokpatronen en Vouwblaadjes, zoals besproken in paragraaf 8.3, dan zouden deze tests - althans in de hier gebruikte versie - niet in aanmerking gekomen zijn voor (exploratief) grenzentoetsend onderzoek.

Ten aanzien van Natikken en Woordenspan liggen de zaken duidelijk anders. Niet alleen beschikken beide tests over voldoende moeilijke items met het oog op grenzen-toetsen (geen 'plafond'-effekt), maar ook zijn de korrelaties tussen eerste en derde aanbieding redelijk hoog, namelijk $r_{(1)(3)} = .79$ voor Natikken en $r_{(1)(3)} = .86$ voor Woordenspan. Er is dan ook weinig reden om aan de geldigheid van de verhoogde coëfficiënten van interne consistentie en homogeniteitscoëfficiënten bij derde aanbieding te twijfelen (Natikken $r_{kk} = .82$, Woordenspan $r_{kk} = .76$). Desondanks zijn niet of nauwelijks hogere prediktieve validiteitscoëfficiënten verkregen, evenmin als zelfs maar een spoor van een moderatoreffekt. Het is duidelijk dat naar een andere verklaring gezocht moet worden; zowel Bersoff (1973) als Feuerstein (1972b) geven aan in welke richting wij deze moeten zoeken.

Het kernprobleem bij de validering van tests die volgens een grenzen-toetsende procedure zijn afgenomen, ligt in de keuze van de criteria. Volgens Feuerstein (op.cit.) heeft het weinig zin konventionele criteria als schoolvorderingentoetsen etc. te hanteren, indien de schoolvorderingen in klasverband gemaakt zijn zonder dat extra aandacht, individuele hulp of aanmoediging gegeven werd; en dit laatste behoort nu juist tot die aspecten die bij grenzen-toetsen aan de orde komen.

Het enige dat men dan kan verwachten zijn spurieus lage prediktieve validiteitscoëfficiënten. In de woorden van Feuerstein (1972b):

"To make the validity of estimated potential contingent upon the extent of its realization would be similar to denying the wealth of a proven miser, merely because he never spends his money! The concept of capacity, indeed, depends on the crucial difference between potential and actualized achievement level, for if capacity for achievement did not require special conditions for its actualization, what difference would remain between "capacity" and manifest level of functioning? Capacity of change requires intervention in order to become actualized (...)." p. 151.

Bersoff (1973) doet er in dit verband nog een schepje bovenop door te stellen dat tests als de WISC en de Stanford-Binet - afgenomen volgens standaard-testprocedures - eerder het cognitief functioneren onder optimale kondities meten, dan onder de reële, feitelijke kondities die men in de klas aantreft. Als bewijs hiervoor dient niet alleen de instructie voor genoemde tests maar ook een onderzoek van Rigsby (1972) naar het positieve bekrachtigingsgedrag van leerkrachten en schoolpsychologen.

"The authors of the Stanford-Binet and Wechsler Intelligence Scale for Children (WISC) manuals specifically instruct testers to establish rapport, to be sympathetic and understanding, to watch for fatigue, to minimize physical defects, and more, all in order to "maintain both high motivation and optimal performance during administration [Terman & Merrill, 1960, p. 50]." (...) In a recent study (Rigsby, 1972), for example, it was found that school psychologists, during administration of the verbal portion of the WISC, emitted an average rate of 1.38 positive reinforcements per minute to 26 children tested individually while the teacher emitted less than 1 every 10 minutes to these same children in a classroom instructional situation. Thus, each child received more than 15 times as many reinforcements in a testing situation as he did in a classroom. In fact, the average reinforcement emission by a teacher to a whole classful of children was less than that for one child by a psychologist. This hardly makes the two situations comparable, and extrapolation from one to the other is very risky, if not absurd" (p. 894).

Men hoeft het met Bersoff's konklusie niet volstrekt eens te zijn om in te zien dat ook hij een wezenlijk punt aanroert, namelijk: de standaard-testsituatie is - vergeleken met de leersituatie in de klas - al zo optimaal dat men van de verschillende optimaliserende kondities bij het grenzen-toetsen geen wezenlijke bijdrage aan de gemeenschappelijke variantie van prediktoren (pre-testscores) en criteria kan verwachten, eerder het tegendeel. In dit licht bezien is het bijna verbazingwekkend dat de post-testscores van Naticken en Woordenspan nog respectievelijk eenzelfde of iets hogere samenhang met de door ons gehanteerde criteria vertonen als de pre-testscores. Uitdrukkelijk schreven wij 'bijna', aangezien het aannemelijk lijkt dat het mogelijk negatieve effect van de

toegepaste interventie op de prediktor - criterium relatie gekompenseerd wordt door het positieve effect van de hogere betrouwbaarheid van de post-testscores van Natikken en Woordenspan op dezelfde relatie. Betekent dit alles nu dat het principieel onmogelijk is om door middel van grenzen-toetsen het leervermogen van een kind vast te stellen? Op basis van ons onderzoek kunnen wij geen afdoende antwoord op deze vraag geven. In de eerste plaats omdat het criterium 'schoolprestaties' bij nader inzien minder gelukkig gekozen lijkt. Criteria, waarin op een of andere manier een doelgerichte interventie verwerkt is, bieden misschien méér kans op sukses. Ten tweede: de door ons toegepaste interventie - maximaal twee aanbiedingen per item onder optimaliserende kondities - is wel van heel korte duur, zeker in vergelijking met de interventie die bij 'learning potential assessment' wordt toegepast (cf. par. 12.2). Een vergroting van het aantal aanbiedingen per item kan mogelijk wél het geringste moderator-effect sorteren, ook al zijn er duidelijke aanwijzingen dat dit effect bij vier à vijf aanbiedingen van een en dezelfde test onder standaardkondities zijn maximum bereikt (cf. Sandbergen, 1968). Ten derde: het is in het algemeen niet uitgesloten dat de gebruikte statistische procedures - onder andere door de assumpties waarop zij gebaseerd zijn of door gebrek aan pp'n - te ongevoelig zijn voor de registratie van (minieme) veranderingen, die mogelijk wél door een ervaren praktizerend psycholoog op klinisch-intuïtieve wijze geobserveerd en geïnterpreteerd kunnen worden. [Op de gevaren hieraan verbonden komen wij aan het slot van deze paragraaf nog terug.]

Dat de situatie niet denkbeeldig is dat de voorspelling van een ervaren praktizerend psycholoog superieur is aan die welke met behulp van multivariate technieken gedaan wordt, hebben wij reeds in deel II (p. 15) kunnen konstateren. Op deze plaats zouden wij er nog het volgende aan willen toevoegen. De psychometrische problemen met betrekking tot toenamescores, zoals onder andere uiteengezet in het artikel van Cronbach en Furby (1970) en de boeken van Harris (1963) en De Gruyter en Van der Kamp (1976), zijn nog steeds zeer groot en praktisch gesproken misschien wel onoplosbaar. Dit betekent dat psychometrische criteria niet alleen de doorslag kunnen geven bij de beantwoording van de vraag of een bepaalde procedure het beoogde effect gehad heeft of niet. Om genoemde redenen moeten wij dan ook vooralsnog het voordeel van de twijfel gunnen aan grenzen-toetsen als praktisch-klinische testprocedure ter bepaling van het leervermogen.

Leervermogen vormt echter slechts één kant van de medaille die grenzen-toetsen heet. De andere kant wordt namelijk in beslag genomen door zaken als observatie, standaardisatie en betrouwbaarheid. Zoals bekend is maximale standaardisatie van onder andere de onderzoekssituatie, afnameprocedure en instructie het ideaal waarnaar een praktizerend psycholoog moet streven; dit met het oog op een zo hoog mogelijke betrouwbaarheid van van zijn metingen. Bij een groot aantal psychologen overheerst echter het gevoel dat aan de betrouwbaarheid van de metingen bij zwakbegaafde, gedepriveerde of anderszins gehandicapte kinderen nogal wat af te dingen valt (cf. par. 9.3), juist omdat men zich strikt aan de eisen van de standaard-testsituatie houdt. Eisen die bij tests als de WPPSI en de ITPA - maar ook de LDT-E - eerder afgestemd zijn op een normale populatie kinderen dan op bovengenoemde groepen kinderen. Tegen deze achtergrond bezien getuigt het van niet meer dan van gezond verstand dat de psycholoog die aan zijn eigen waarnemingen twijfelt tot grenzen-toetsen overgaat.

In ons onderzoek hebben wij een voorbeeld gegeven van de manier waarop men grenzen-toetsen zou kunnen toepassen. Tevens hebben wij laten zien - aan de hand van p-waarden, item-restkorrelaties, gemiddelden, spreiding etc. - welke uitkomsten men kan verwachten indien men tot grenzen-toetsen op Blokpatronen, Vouwblaadjes, Natikken en Woordenspan overgaat bij een in doorsnee normale populatie kinderen van vijf tot zes jaar. Het laatste neemt echter niet weg dat de onderzoeksresultaten wellicht interessanter zouden zijn geweest, indien wij ons beperkt hadden tot de groep minder begaafde kinderen - met een IQ <80 bijvoorbeeld - onder gelijktijdige uitbreiding van het aantal en type observaties, c.q. interventies per test. Uiteraard zou dit een geheel andere onderzoeksopzet gevergd hebben, onder andere gebaseerd op de principes van de generaliseerbaarheidstheorie van Cronbach et al. (1972) en op het werk van Cook en Campbell (1976) met betrekking tot quasi-experimentele designs. Maar dan nòg: Soudijn (1979) wijst erop dat rechtvaardiging van psychotherapie via het gebruikelijke empirisch onderzoek - en grenzen-toetsen heeft bijzonder veel weg van een therapeutische interventie - praktisch onmogelijk is. Hij voert daartoe een aantal argumenten aan, waarvan het argument over de eis van standaardisatie ons wel bijzonder aanspreekt. Volgens Soudijn (1979) impliceert

"Veel onderzoek naar empirische regelmaat (...) dat onderzochte technieken op een *standaard-wijze* toegepast moeten worden. De practicus lijkt er echter verstandig aan te doen om veel interventies juist *niet* op een standaard-wijze toe te passen: door de interventie te variëren, lijkt

deze beter aan te passen te zijn aan de persoonlijke eigenaardigheden van de cliënt" (p. 124).

De paradoxale belangen van onderzoeker en practicus worden in dit citaat duidelijk weergegeven. Rest ons de hoop uit te spreken dat de onderzoeker - op grond van onze onderzoeksresultaten - er vooralsnog niet toe overgaat om grenzen-toetsen tot één van de idolen van de psycholoog te verklaren en omgekeerd dat de practicus zelfs maar de schijn vermijdt om via grenzen-toetsen het vlees geworden idool van de psycholoog te vertegenwoordigen. Om de practicus hierbij te helpen willen wij daarom tot slot van deze paragraaf nog een enkel 'caveat' laten horen.

Ten eerste: ook al acht men de standaard-testsituatie in sommige gevallen minder natuurlijk, toch is het noodzakelijk eerst de test volgens de standaardinstructies af te nemen en dan pas tot grenzen-toetsen over te gaan. Normen en overige gegevens die in een testhandleiding worden vermeld, verliezen aan betekenis indien standaard- en grenzentoetsende procedures door elkaar worden gehaald.

Ten tweede: de door ons gevolgde 'up-and-down method' bij de herhaalde aanbieding van items (cf. hfdst. 10) bleek in de praktijk van het grenzentoetsend onderzoek nogal ingewikkeld te zijn; er werden te veel aanbiedings- en/of scoringsfouten gemaakt. Een aantal antwoordformulieren kwam om deze reden dan ook niet in aanmerking voor verdere bewerking. Aanzienlijk eenvoudiger lijkt ons de methode waarbij men niet meer 'terugspringt' van derde naar tweede aanbieding (cf. fig. 8).

Ten derde: naarmate items vaker worden herhaald onder al dan niet optimaliserende kondities, wordt steeds onduidelijker wat men aan het meten is. Door oefening op de testtaak kunnen namelijk andere cognitieve functies worden aangesproken die men niet bedoelt te meten.

Ten vierde: bij de introductie van optimaliserende kondities dient men bijzonder voorzichtig te zijn; niet elke test is wat zijn betekenis betreft even gevoelig voor iedere optimaliserende konditie. Bij sommige tests - Blokpatronen bijvoorbeeld - is een konditie als 'Aandacht' minder interessant, terwijl het bij Vouwblaadjes of Natikken daarentegen weer wel veel kan uitmaken of het kind aandacht besteedt aan de demonstratie van pl. Als men bijvoorbeeld de onmiddellijk geheugenspan voor ruimtelijke transformaties (Vouwblaadjes) of ruimtelijke sekwenties (Natiekken) nauwkeurig wil meten, dan dient men een konditie als 'Aandacht' goed onder controle te houden. Bij Blokpatronen lijkt verder een optimaliserende konditie als 'Tijdverdubbeling' een sterk inflatoire invloed uit te oefenen op de uiteindelijke testprestaties: vrijwel ieder kind

schijnt elk patroon foutloos te kunnen naleggen, mits hem maar voldoende tijd wordt gegund. Tijdverdubbeling is duidelijk te veel van het goede, zoals wij ook naderhand in de literatuur terugvonden (Greene, 1937). Tenslotte: bij de beoordeling van de vorderingen die een kind maakt als gevolg van grenzen-toetsen, dient men op zijn hoede te zijn voor het regressiefenomeen (cf. par. 11.3), dat wil zeggen:

"Wanneer een meting een resultaat oplevert dat k eenheden van het gemiddelde ligt dan is de kans dat een tweede meting een resultaat oplevert dat minder dan k eenheden van het gemiddelde ligt groter dan de kans dat het resultaat meer dan k eenheden van het gemiddelde ligt" (deel II; p. 25).

Ceteris paribus geldt deze waarschuwing ook voor tweede en derde aanbieding, derde en vierde, etc. Vrij vertaald komt het regressiefenomeen hierop neer dat als de gemiddelde vooruitgang van een groep kinderen op een bepaalde test 3 punten bedraagt, het te verwachten valt dat kinderen die bij eerste aanbieding laag scoorden 4 punten vooruit gaan na tweede of derde aanbieding en kinderen die hoog scoorden bij eerste aanbieding slechts 2 punten. Met andere woorden: bij toenamescores dient men altijd het beginniveau (eerste aanbieding) te verdisconteren in de beoordeling of er sprake is van werkelijke vooruitgang dan wel van een statistisch artefact. Yule et al. (1969; p. 3) illustreren op fraaie wijze hoe Wechsler (1967; p. 11) in dit verband misleidende uitspraken doet over de Animal House (her)test als een index voor het leervermogen.

12.2 Learning potential assessment

In het voorgaande is al verschillende malen de term 'Learning potential assessment' gevallen als aanduiding voor één van de twee klinische testprocedures - naast grenzen-toetsen - waarmee men in de praktijk van het onderzoek onder andere het onderscheid intelligentie ('ability') en leervermogen ('aptitude') heeft trachten te realiseren. Aangezien 'Learning potential assessment' - afgekort tot LPA - opgevat kan worden als een verdere verfijning en uitbreiding van grenzen-toetsen, lijkt het nuttig hieraan een korte bespreking te wijden.

Aan LPA zijn primair de namen verbonden van Budoff en Feuerstein. In de nodige artikelen hebben deze twee auteurs hun veelomvattende bedoelingen en - vaak ingewikkelde - testprocedures uiteengezet (onder andere Babad & Budoff, 1974; Budoff, 1973; Budoff et al., 1971; Feuerstein, 1970; Feuerstein, 1972a, b; Feuerstein & Krasilowsky, 1972). In essentie gaan echter zowel Budoff als Feuerstein uit van het pretest - interventie - posttest paradigma. Dat wil zeggen: eerst wordt met behulp van een

non-verbale test het cognitieve ontwikkelingsniveau op objectieve wijze vastgesteld, overeenkomstig de standaardinstructie die voor deze test geldt. Vervolgens vindt een of andere vorm van interventie plaats; deze kan variëren van oefening op de pretest zelf tot training of 'coaching' op een met de pretest vergelijkbare testtaak of onderdelen daarvan. Tenslotte wordt een parallelvorm van de pretest als posttest afgenomen, eveneens volgens standaard-testinstructies. De gebruikte tests zijn onder andere de Kohs Block Designs, Raven's Progressive Matrices, Stencil Design en de Series Learning Potential Test, welke alle een beroep doen op non-verbaal redeneren.

Het belangrijkste verschil tussen Budoff en Feuerstein betreft de aard en het doel van de training. Zo bestaat de training volgens Budoff uit het aanleren van strategieën die een zo goed mogelijke oplossing van de posttest-problemen bevorderen. Bijvoorbeeld:

1. The students learn to identify each concept that changes.
2. They learn to "sing the tune" for each concept, one at a time as an organizational or chunking aid to identify the pattern of each concept.
3. To reduce the memory load in a multiconcept item, the student crosses out the wrong choices for each concept.
4. The child learns to reverse the direction in which the tune is sung when the location of the blank space calls for such reversal, for example, blank at the beginning of a series.
5. The child learns to identify the starting point of a concept when the series starts in the middle of a pattern" (Babad & Budoff, 1974; p. 442).

Het doel van de training wordt als volgt beschreven:

"The training helps the child to narrow the cognitive gap between his previously learned problem-solving strategies and these implicit to the problems he must ordinarily solve on the middle-class-bias tests he encounters. The test-train-test paradigm also minimizes the artificiality of the test situation. (...). The pretest reflects the subject's present level of functioning and his existing ability to work with the problems. Posttraining scores reflect the child's ability under optimized conditions in which all subjects are familiar with the task and its demands; have had success in solving problems similar to these on the test, and have had the opportunity to learn and apply relevant strategies" (Babad & Budoff, 1974; p. 440).

Met andere woorden: via LPA wordt een indruk verkregen van de 'trainability', de leerzaamheid of het leervermogen van een kind. In een beknopt overzicht van zowel Budoff's als Feuerstein's werk wijzen Haywood et al. (1975) er verder op dat training van probleemoplossingsstrategieën op bepaalde testtaken aanwijzingen kan verschaffen voor een gerichte hulpverlening aan het kind met leermoeilijkheden.

Het doel en de aard van LPA volgens Feuerstein is aanzienlijk lastiger te beschrijven dan die volgens Budoff; niet alleen omdat de toegepaste

interventie de vorm en het verloop kan aannemen van een (kortdurend) re-mediaal programma, maar ook omdat de pretenties omtrent het uiteindelijk resultaat veelomvattend zijn. Volgens Narrol en Bachor (1975) - twee leerlingen van Feuerstein - levert LPA echter de volgende gegevens op:

- "(1) the extent of modifiability and the generalization from task to task;
- (2) the amount of investment needed to produce the noted modification and generalization;
- (3) the nature of the investment required and whether it was predominantly focused on any part or parts of the cognitive process (i.e. input, elaborational, output, or affective-emotional);
- (4) the presence of preferred communicational and learning modalities, the use of which appears to make teaching and learning more effective" (p. 11).

Voor het verwerven van deze gegevens streeft Feuerstein primair naar een verbetering van het cognitief functioneren door middel van training op een groot aantal testtaak-variabelen (strategieën) enerzijds en beïnvloeding van het affectief-emotionele taakgedrag anderzijds. Aan Narrol en Bachor (op.cit.) ontleen wij een voorbeeld van enkele strategieën die op de Organization of Dots Test geobserveerd en getraind worden:

- "(1) the ability to project variously oriented figures over the available dots in order to discover the embedded figures;
- (2) a process of analysis of length of lines;
- (3) the adaption of an optimal plan of attack;
- (4) possession of sufficient restraint to keep from impulsively scrabbling lines between dots based on perceptual hunches unaided by analytic pretesting" (p. 8).

Beïnvloeding van het affectief-emotionele gedrag vindt plaats doordat Feuerstein de onderzochte aanmoedigt om de strategie waarlangs hij tot een juiste oplossing van het probleem kwam, te bepraten.

Op het werk van Budoff en vooral van Feuerstein is nogal wat commentaar mogelijk; wij zullen ons hier echter tot enkele hoofdlijnen beperken.

Ten eerste: een positief punt van LPA is dat de 'cognitieve proces'kant - dat wil zeggen de manier waarop een kind tot de oplossing van een probleem komt - in meer (Feuerstein) of mindere mate (Budoff) de nadruk krijgt. Hierdoor wordt tegemoet gekomen aan de kritiek van diegenen die vinden dat de traditionele psychologische test geen informatie verschaft over de strategieën die het kind volgt bij de oplossing van testproblemen (cf. par. 9.3). Wij zouden hieraan willen toevoegen dat een en ander afhangt van de manier waarop een psychologische test gehanteerd wordt. Biedt grenzen-toetsen in dit opzicht minder rijke mogelijkheden, de mogelijkheden die LPA verschaft zijn - zoals Feuerstein laat zien - vrijwel onuitputtelijk.

Ten tweede: juist omdat de mogelijkheden vrijwel onuitputtelijk zijn, rijzen er ernstige problemen met betrekking tot de standaardisatie van de interventie. Ofschoon deze problemen in principe oplosbaar zijn - zowel Budoff als Feuerstein maken gebruik van relatief simpele en beperkte standaardinterventies bij groepsgewijs onderzoek volgens de LPA-methode - wordt met de oplossing wel weer een nieuw probleem geschapen. Zo verhindert al te rigide standaardisatie in het individuele geval de opstelling van een maat voor de veranderbaarheid, c.q. het leervermogen van een kind, althans volgens Feuerstein (1972b):

"(...) the establishment of an index of modifiability as implied by the LPAD model (= LPA, Schr.), requires assessment of the nature, and measuring of the amount of investment needed in order to reach a given level of functioning, starting from a given baseline level. That is, we need to find and employ efficient measures of change. Thus not only has one to establish the baseline level of functioning with all the difficulties we have alluded to in doing so, but, in addition, one has to categorize the nature of input and ascribe to it a certain value and, then, quantify the amount of this investment and use it as a control variable in assigning a quantitative weight to the achieved modification. One can immediately sense the contradiction in sensu, between such a goal and the clinical nature of our method. In order to establish an exact measure of investment one might have to standardize the intervention strategies to such an extent that this could greatly reduce the effectiveness of the clinical method, since one will no longer be able to flexibly consider the differences inherent in individuals assessed under this method, but will have to follow a rather rigid, confining prescription" (p. 149).

Feuerstein's argumenten tégen standaardisatie van de interventie in het individuele geval doen herinneren aan het citaat van Soudijn (par. 12.1), waarin de paradoxale belangen van onderzoeker en practicus worden weergegeven. Het laatste neemt echter niet weg dat zich nu wel een aantal vervelende vragen aan ons begint op te dringen, namelijk: als standaardisatie van een uitgebreide en individuele interventie welhaast een contradictio in terminis is hoe kan men dan a) de beoefening van de interventietechniek overdragen aan praktizerende psychologen anders dan via een misschien wel zeer langdurige leermeester - leerling relatie, b) het effect van een interventie (welke interventie?) evalueren, en c) een maat voor het leervermogen of de veranderbaarheid opstellen die meer dan een hoogst individuele en daardoor unieke betekenis heeft. Helaas moeten wij het antwoord op deze vragen schuldig blijven. Ook Feuerstein zelf heeft geen pasklaar antwoord, ondanks de preliminaire evidentie die hij - in de vorm van een aantal gevalsbeschrijvingen - aandraagt voor de superioriteit van LPA boven het traditionele standaard-testonderzoek. Onze indruk is dan ook dat LPA weliswaar veel 'zachte' informatie oplevert over

het cognitief functioneren, maar weinig 'harde' gegevens over de betrouwbaarheid en validiteit van de LPA-onderzoeksresultaten.

Tenslotte: als de interaktionisten gelijk hebben met hun stelling dat de traditionele intelligentietest gedepriveerde kinderen benadeelt (cf. par. 9.2) dan is het zaak om aan te tonen dat de gepropageerde andere onderzoeksinstrumenten en/of methoden dat niet doen, of in mindere mate. De gegevens die Babad en Budoff (1974) en Sewell en Severson (1974) publiceerden over LPA zijn in dit opzicht zowel hoopgevend als teleurstellend. Teleurstellend, omdat Sewell en Severson (op.cit.) na onderzoek op 62 zesjarigen van laag sociaal-ekonomisch niveau moesten konkluderen dat Raven's Coloured Progressive Matrices (posttest) - afgenomen volgens Budoff's LPA-methode - niet hoger korreleert met schoolprestaties dan de WISC, een traditionele intelligentietest. Hoopgevend, omdat Babad en Budoff (op.cit.) in een soortgelijk onderzoek met de groepsgewijs afgenomen Series Learning Potential Test (posttest) vonden dat deze testmethode een significant betere voorspeller is van schoolprestaties dan een (groeps)intelligentietest, i.c. Flanagan's Test of General Ability; althans voor kinderen met een IQ <100. Dit laatste is een belangrijk gegeven en verklaart misschien ook waarom Sewell en Severson teleurstellende resultaten behaalden, namelijk: intelligentietests zijn niet zozeer 'bias-middle-class' tests als wel minder gevoelig en betrouwbaar aan de onderkant van de meetschaal, met alle gevolgen voor de prediktieve validiteit van deze tests voor minder begaafden en gedepriveerden vanden. Ongetwijfeld zou Feuerstein hier tegenin gebracht hebben dat zowel Babad en Budoff als Sewell en Severson in hun onderzoek uitgegaan zijn van een verkeerd criterium, namelijk 'schoolprestaties' (cf. par. 12.1).

12.3 Suggesties voor verder onderzoek

Ofschoon mogelijk de indruk gewekt is dat grenzen-toetsen in een experimenteel-psychologisch en psychometrisch vacuüm beoefend werd, is dit niet het geval. Sinds 1938 valt namelijk een duidelijke lijn te bespeuren in wat wij voor het gemak maar het 'ability-aptitude' onderzoek zullen noemen; deze lijn begint met Woodrow (1938) en eindigt voorlopig met Cronbach & Snow (1977). In de tussentijd zijn enkele belangrijke publikaties verschenen van onder anderen Ferguson (1954), Anastasi (1958) en Vernon (1969). In deze publikaties wordt - deels onder de noemer 'intelligentie en omgeving' - het effect van oefening, 'coaching' en leren op intelligentietesttaken besproken. Aangezien Sandbergen (1968) een gedeel-

telijk overzicht en Cronbach & Snow (1977) een vrijwel volledig overzicht geven van het desbetreffende onderzoek, zullen wij er hier niet verder op ingaan. Wel willen wij opmerken dat de resultaten van het 'ability-apptitude' onderzoek vrijwel volledig genegeerd worden in de overwegend klinische literatuur over grenzen-toetsen. Hiervoor zijn verschillende verklaringen te geven.

Ten eerste: het is denkbaar dat praktizerend psychologen domweg geen kennis nemen van empirische onderzoeksresultaten, bijvoorbeeld omdat ze in een taal gesteld zijn die de practicus niet begrijpt (cf. deel II; p. 13 e.v.). Ten tweede: veel onderzoeksresultaten bieden weinig houvast voor toepassing in de praktijk of zijn daar zelfs in het geheel niet op afgestemd. Ten derde: de uitkomsten van onderzoek - bijvoorbeeld het Aptitude x Treatment Interactions (ATI) onderzoek - zijn nogal magertjes (cf. Cronbach, 1975b), terwijl de practicus het overtuigende gevoel heeft dat wat hij doet wel degelijk effect heeft.

Voor al het laatste punt achten wij belangrijk, al is het maar omdat de kennismaking met psychometrische onderzoeksbevindingen en de resultaten van individueel-klinisch onderzoek bij ons een sterk gevoel van cognitieve dissonantie veroorzaakt. Nu is het negeren van één van de twee informatiebronnen die aan de cognitieve dissonantie ten grondslag liggen nòch voor psychometrici, nòch voor praktici een fraaie manier om de dissonantie op te lossen. Men doet er beter aan om van beide bronnen zo goed mogelijk kennis te nemen: de mogelijkheid is tenslotte niet uitgesloten dat het bij nader inzien om een schijnbare dissonantie gaat.

In het nu volgende zullen wij enkele eerder gemaakte opmerkingen verder uitwerken, in de hoop hieraan enige suggesties te kunnen ontleen voor verder onderzoek.

Tussen de regels door suggereerden wij in paragraaf 12.1 een opmerkelijke overeenkomst tussen psychotherapie en grenzen-toetsen. In plaats van psychotherapie hadden wij ook een andere interventietechniek kunnen noemen, bijvoorbeeld de 'diagnostische dialoog' (Groenendaal, 1978). Iedere techniek zou in feite goed geweest zijn als zij maar een interventie impliceerde gericht op een verandering van het gedrag. Wie van interventie spreekt, heeft het echter ook over interactie tussen therapeut en cliënt of tussen proefleider en proefpersoon. Als nu het effect van interventies, zoals die bij grenzen-toetsen en LPA worden toegepast, niet of nauwelijks is vast te stellen ondanks - of dankzij? - de gebruikte tamelijk sophisticated analysetechnieken, dan lijkt het misschien wel nuttig

een stap terug te doen in het huidige evaluatie-onderzoek en de interactie zelf aan een nadere analyse te onderwerpen. Groenendaal (op.cit; p. 145) maakt een soortgelijke opmerking naar aanleiding van de summatieve evaluatie van zijn programma ter bevordering van het abstracte denken bij zwakfunktionerende kleuters. Ter illustratie van wat bedoeld wordt, citeert hij vervolgens Sigel et al. (1973):

"What is needed is a careful monitoring of transactions that occur from Time 1 to Time 2. To carry out such a process involves a heavy input of observational data-gathering because we are observing a complex systematic network in which the teacher, child, other children and materials interact in varying degrees and on various levels" (p. 59).

Ook Feuerstein (1972b; p. 150) verwacht voor de nabije toekomst meer van een analyse van het interactieproces tussen proefleider en proefpersoon dan van het traditionele evaluatie-onderzoek. De konsekwentie hiervan moge duidelijk zijn: een uitgebreid interactie-analysesysteem plus bijbehorende observatieschalen dient ontwikkeld te worden.

Wij kunnen echter nog een stap terug doen en de vraag stellen of het interactie-sop de kool wel waard is; met andere woorden of de kosten wel tegen de baten opwegen. In dit verband willen wij een passage uit paragraaf 9.1 in herinnering brengen, waarin we stelden dat grenzen-toetsen een betrekkelijk weinig tijd vergende klinische methode is. Van een aantal klinische methoden is echter bekend dat zij de toets der psychometrische kritiek niet kunnen doorstaan. Zelfs als dit zo is dan is dit nog geen reden om voortaan van deze methoden af te zien; men dient alleen andere verwachtingen te koesteren over het resultaat dat men met deze methoden tracht te bereiken. Vanuit dit perspectief bezien zou men grenzen-toetsen kunnen opvatten als een klinische methode, c.q. testprocedure waarmee men hypothesen omtrent een groot aantal leer- en testgedragkenmerken van de onderzochte kan oproepen, welke nadere toetsing behoeven door vervolgonderzoek met andere instrumenten en technieken (cf. Glasser & Zimmerman, 1967; p. 124). Onzes inziens zou het daarbij aanbeveling verdienen om de ontwikkeling van een uitgebreider aantal observatieschalen ter hand te nemen dan wij in ons onderzoek konden realiseren. De studie van Moriarty (1961) lijkt wat dit betreft goede aanknopingspunten te bieden, althans volgens Cronbach (1970):

"The Moriarty study provides one fact that strongly supports the use of the test as a basis for observation even when the validity of the score itself is doubtful. Ratings were made of adjustment and resourcefulness during the test; these did not correlate highly with the test score, but they correlated 0.50 to 0.70 with an overall rating of "coping capacity" based on a thorough study of each child" (p. 264).

Nu wij twee stappen terug gedaan hebben, kunnen we proberen drie stappen vooruit te doen. Dezelfde Cronbach die - blijkens bovenstaand citaat - bepaald niet afwijzend staat tegenover systematische observatie, heeft een belangwekkend artikel geschreven "On the design of educational measures" (1976). In dit artikel probeert hij zijn Generaliseerbaarheids-theorie - als opvolger van de klassieke betrouwbaarheidstheorie - en zijn Decisie)theorie van aanmerkelijk oudere datum te integreren tot één theorie waarin de sleutelzin ongetwijfeld het ontwerp is van een "multifacet sequential design" voor de evaluatie van onderwijsprogramma's. Zowel grenzen-toetsen als LPA kunnen als miniatuur-onderwijsprogramma's beschouwd worden en lenen zich in die zin dan ook voor toepassing van deze G/D-theorie. Zonder op de theorie in te gaan kunnen wij toch enkele punten noemen die ons bijzonder aanspreken.

In de eerste plaats het kosten-batenaspect: "To what degree if any does collecting additional information improve a decision?" (p. 200). Dit citaat verwijst niet alleen naar het bovenstaande, maar ook naar paragraaf 9.3 waarin wij schreven dat het op voorhand niet volstrekt onaannemelijk lijkt dat grenzen-toetsen aanvullende informatie kan verschaffen over het cognitief functioneren van een kind. Kijken we nu naar de praktijk van het grenzen-toetsen zoals die zich in de klinische literatuur aan ons voordoet, dan wordt zelden of nooit tot grenzen-toetsen overgegaan bij hoog-scorende proefpersonen. Kennelijk beschikt de psychodiagnosticus na de standaardafname van een test reeds over voldoende diagnostische informatie om op basis daarvan een beslissing te kunnen nemen. Toepassing van de D-theorie - zoals onder meer uitgewerkt in deel II, p. 29 e.v. - zou kunnen helpen bij de oplossing van het probleem wanneer het nuttig is om tot verdere informatieverzameling over te gaan.

Het tweede punt dat wij willen noemen is de integratie van G-theorie - waarin het aksent valt op de identifikatie en toetsing van verschillende variantiebronnen - met D-theorie, zoals aangestipt in het nu volgende citaat:

"Error variance, for example, is appraised as if an error of (say) 5 score points is equally distressing everywhere along the score scale. Yet most educational uses of tests are concerned with distinguishing the adequate from the inadequate. Errors in judging persons or groups close to the borderline are far more important than distant errors. Positive errors may be more important than negative errors, especially if a score called adequate signals us to let the student graduate from school or to pass from one unit of instruction to a more advanced one" (p. 200 en 201).

In dit citaat wordt duidelijk gemaakt dat bepaalde variantie, in dit geval foutenvariantie, niet altijd even belangrijk hoeft te zijn. In een

G-studie wordt echter over het belang - dat wil zeggen de kosten - van bepaalde fouten niet gerept. Door nu G- en D-theorie met elkaar te combineren, kan men ook aan dit aspect aandacht besteden. Zoals Cronbach (op.cit) laat zien, biedt een geïntegreerde G/D-theorie vele toepassingsmogelijkheden, bijvoorbeeld bij de opstelling van een 'design'. Wij zullen deze hier niet verder bespreken. Als echter onze gedachten uitgaan naar verder onderzoek op het gebied van grenzen-toetsen, dan is dat in deze richting.

Tot slot van deze paragraaf willen wij terugkomen op grenzen-toetsen als klinische testprocedure voor de vaststelling van het leervermogen. Zoals uit het voorgaande gebleken is, heeft het er alle schijn van dat wij in dit opzicht op zoek zijn geweest naar een witte raaf. De vraag rijst nu of het principieel onmogelijk is een onderscheid te maken tussen 'ability' en 'aptitude'. Whitely & Davis (1975) beweren van niet. In een ingenieus en goed doordacht experiment, waarbij zij proefpersonen lieten oefenen op een 'spatial reversal motor'taak en de data vervolgens analyseerden volgens Jones' 'molar correlational analysis' (1970) kwamen zij uiteindelijk tot de volgende konklusie:

"The spatial reversal data indicate that the approach suggested here to distinguish aptitude from ability psychometrically has some feasibility. It was shown that predictability could be increased by adding an index of modifiability to initial status from a test-intervention-retest sequence. However, the data also indicate the crucial importance of the stage of practice or degree of intervention over which modifiability is computed. The degree of success with which aptitude can be distinguished from ability depends directly on the intervention or amount of practice which is selected" (p. 64).

Ofschoon deze konklusie voor zichzelf spreekt, willen wij er toch nog een paar opmerkingen aan toevoegen. Ten eerste: de toepassing van 'molar correlational analysis' vereist een aanzienlijk groter aantal 'practice trials' (± 9 aanbiedingen in totaal) dan bij grenzen-toetsen in een klinische testsituatie gerealiseerd kan worden. Ten tweede: de relevantie van een relatief simpele 'spatial reversal motor' taak voor de voorspelling van schoolprestaties lijkt op het eerste gezicht zeer gering.

Is het dus principieel mogelijk om met behulp van een typische laboratoriumtaak onderscheid te maken tussen 'ability' en 'aptitude', praktisch gesproken lijkt het vooralsnog onuitvoerbaar om met behulp van grenzen-toetsen op qua functies relatief complexe psychologische tests het onderscheid tussen intelligentie en leervermogen te realiseren (cf. par. 9.2). Gegeven deze situatie is het misschien verstandiger om het toekomstig onderzoek te richten op betrouwbaarder en valider meetinstrumenten dan

de traditionele psychologische tests van het cognitieve type. Wij denken hierbij primair aan de verdere ontwikkeling en verfijning van zogenaamde neurometrische tests (John et al., 1977), welke gebaseerd zijn op de elektrofysiologische techniek van de 'evoked potentials'. De reden hiervoor is de volgende.

Het gehele 'ability-aptitude' onderzoek vindt zijn oorsprong in de idee dat psychologische tests van het cognitieve type gedeeltelijk de individuele leerervaringen uit het verleden reflektieren. Ter 'korrektie' van deze leerervaringen heeft men de nodige ingewikkelde methoden, technieken en onderzoeksprocedures bedacht welke ieder voor zich op de lange duur best resultaten kunnen afwerpen, maar die niets veranderen aan het feit dat de psychologische test in bepaalde opzichten een gebrekkig meetinstrument is en waarschijnlijk zal blijven. Wat is dan meer voor de hand liggend dan dat men aan de psychologische test - net zoals aan de sinds 150 jaar nauwelijks veranderde stethoscoop van de dokter - de uitermate nuttige, klinische status geeft die hem toekomt en vervolgens op zoek gaat naar andere meetinstrumenten die de individuele leerervaringen nauwelijks of niet reflektieren. Te oordelen aan de reeds behaalde resultaten - onder andere op het gebied van het intelligentie-onderzoek - lijken ons de bovengenoemde neurometrische tests hiervoor zeker in aanmerking te komen.

SLOTBESCHOUWING ⁵⁴

Wanneer de psychodiagnosticus gereed is met het analyseren, interpreteren en rapporteren van zijn bevindingen met de LDT, dan luidt maar al te vaak de verzuchting: "En wat moet ik nu doen met al die gegevens?". In feite wordt hiermee de relatie cognitieve ontwikkelingsdiagnostiek / behandeling aan de orde gesteld. Deze relatie is allerminst een eenvoudige één-op-één relatie tussen score en interventie, zoals nog dikwijls wordt gedacht. Immers, al naar gelang de aard van de diagnostische bevindingen, de prognose en de therapeutische mogelijkheden, kan de behandeling van de cliënt, i.c. het leergestoorde kind, verschillen. Belangrijk is dat men zich realiseert op welke vragen het psychologisch testonderzoek een antwoord kan geven. Alvorens deze vragen te formuleren, zullen wij eerst de mogelijkheden van de psychodiagnostiek zelf kort bespreken.

Gough (1971) verdeelt de diagnostiek in drie niveau's van 'oorzakelijkheid'. Op het eerste diagnostische niveau is sprake van de klontering van fenotypische of symptomatische gegevens, i.c. de observaties van het problematische leergedrag. Diagnostiek op niveau 1 wordt voorname-lijk bedreven met behulp van observatieschalen, toetsen en gedragsbeoor- delingslijsten. De vraag naar de onderliggende 'oorzakelijke' factoren die een (gedeeltelijke) verklaring vormen voor het probleemgedrag kan volgens Gough op dit niveau niet beantwoord worden. Een behandeling zal daarom op zijn hoogst tot 'verlichting' van de symptomen leiden. Bij- voorbeeld: als door middel van een criteriumtoets gekonstateerd wordt dat een kind een specifieke vaardigheid als essentieel onderdeel van een bepaalde leertaak niet beheerst, dan zou men als aansluitende behande- lingsmaatregel extra training kunnen overwegen in de hoop dat het daar- door beter gaat. Als het niet beter gaat, is het zaak een verklaring te vinden; deze komt meestal hierop neer dat in de hiërarchie voorafgaande vaardigheden nog niet beheerst worden door onvoldoende training (cf. Smead, 1977; p. 115). Indien nu ná training van deze voorafgaande vaar- digheden blijkt dat de betrokken vaardigheid alsnog niet beheerst wordt, dan dient een meer fundamentele verklaring gegeven te worden, aangenomen dat de opgestelde vaardigheidshiërarchie juist is.

Deze verklaring wordt geleverd door de pathologie, het tweede niveau waarop de diagnostiek zich beweegt. Hierbij gaat men op zoek naar de onderliggende oorzakelijke factoren van de problematische vaardigheid. Binnen het werkterrein van de psycholoog zijn psychologische tests van het cognitieve type de belangrijkste instrumenten waarmee deze onder- liiggende factoren opgespoord kunnen worden. Op dit niveau wordt de kans

op sukses bij een behandeling al aanmerkelijk vergroot. Niet het symptoomgedrag of de vaardigheid zelf, maar de onderliggende functies ('abilities') zijn primair bepalend voor de keuze van een behandeling.

Het derde en meest basale niveau in de diagnostiek is het gebied van de etiologie. Op dit niveau probeert men de feitelijke oorzaak van het probleemgedrag te achterhalen, gelegen in de levensgeschiedenis (medisch, psychologisch en/of sociaal) van de onderzochte. Vanuit de verzamelde anamnestiche informatie over vele kinderen met leerstoornissen is het mogelijk tot primaire preventie over te gaan, namelijk het voorkómen van leerstoornissen. Op dit niveau ook is de kans dat een volledig succesvolle behandeling ontwikkeld wordt, het grootst.

Keren wij thans terug naar het probleem op welke vragen het psychologisch testonderzoek een antwoord kan geven. Met behulp van bovenstaande indeling in diagnostische niveau's valt de eerste vraag - die naar het diagnostisch niveau van 'oorzakelijkheid' - eenvoudig te beantwoorden. Diagnostiek door middel van psychologische tests kan een verklaring geven op het niveau van de onderliggende 'oorzakelijke' factoren van de leerstoornis, zodat de kans op sukses bij behandeling - in vergelijking tot een benadering door middel van taakanalyse en daarbijbehorende criteriumtoetsen - al aanzienlijk wordt vergroot. In concreto: men mag van een psychologisch testonderzoek verlangen dat een zo nauwkeurig mogelijke beschrijving gegeven wordt van de functies en dysfuncties die aan de leerstoornis ten grondslag liggen. Mede op basis van deze beschrijving dient een bepaalde behandeling gekozen te worden.

De tweede vraag heeft betrekking op de prognose; dat wil zeggen wat gebeurt er met de (dys)functies en leerstoornis indien géén vorm van interventie plaatsvindt. Treedt er een verslechtering op, vindt er na verloop van tijd spontaan herstel plaats of mag men een verbetering noch een verslechtering verwachten? Deze vraag is vooral interessant voor de beslissing of tot therapeutische interventie zal worden overgegaan. Prognostische informatie treft men echter zelden aan in psychologische rapporten, vooral omdat het hiervoor benodigde longitudinale onderzoek nog te weinig verricht wordt (om respectabele redenen overigens). Dit betekent dat men zelf, op basis van ervaring, literatuuronderzoek en gezond verstand, een schatting moet maken van de prognose.

De keuze van de behandeling wordt nu bepaald door de antwoorden op vraag 1 en 2, de behandelingsmogelijkheden en de kennis over het effect van een behandeling, gegeven een bepaalde diagnose en prognose. In principe kunnen de volgende behandelingsvormen onderscheiden worden:

a. Géén behandeling, waarbij men het kind op niet-traumatische wijze laat rijpen. Hiertoe gaat men over indien spontaan herstel van de dysfunctie binnen redelijke termijn te verwachten valt of indien men meent dat het kind in een ontwikkelingsfase verkeert, waarin het zeer snel een vertraagde ontwikkeling zal inlopen. Geen behandeling vindt uiteraard plaats in alle gevallen waarin reeds eerder genomen therapeutische maatregelen geen effect bleken te hebben of omdat therapeutische maatregelen niet gerealiseerd konden worden. Het laatste klinkt als een 'open deur' maar de bedoeling van deze opmerking is erop te wijzen dat leerstoornissen tot secundaire problemen als faalangst, verminderde motivatie etc. kunnen leiden welke, zeker in geval van 'géén behandeling', zoveel mogelijk voorkomen dienen te worden door het geven van extra aandacht, door te wijzen op wat het kind wèl kan, etc.

b. Training van dysfuncties door middel van speciale funktietrainingsprogramma's. Hiertoe gaat men over indien ten eerste: het ernstige vermoeden bestaat dat een bepaalde dysfunctie de oorzaak van de leerstoornis is of tot leerstoornissen aanleiding kan geven (preventie), en ten tweede: een nuttig effect van een dergelijk type training verwacht wordt. Kijkt men naar de praktijk dan blijkt dat 'training van dysfuncties' de meest favoriete behandelingsvorm is. Talloze funktietrainingsprogramma's zijn de laatste twee decennia ontworpen, maar helaas nauwelijks of niet beproefd. Voor zover wèl een systematische evaluatie van programma-effecten plaatsvond, zijn de resultaten op zijn minst teleurstellend (voor een overzicht: zie Rispens, 1977). Naast de argumenten die Rispens aanvoert ter verklaring van de magere behandelingsresultaten, zouden wij nog het volgende willen opmerken. Zelden zijn de kinderen die voor funktietraining in aanmerking komen adequaat gediagnostiseerd; adequaat wil zeggen dat men zorgvuldig het algemeen cognitief ontwikkelingsniveau vaststelt en nauwkeurig aangeeft welke functies binnen welke leerkanalen gestoord zijn of in hun ontwikkeling vertraagd. Te heterogene groepen kinderen zijn aldus onderworpen aan funktietraining, waarvan men evenzeer als voor de diagnostiek mag stellen dat deze weinig adequaat lijkt: de meeste funktietrainingsprogramma's zijn namelijk weinig specifiek ten aanzien van de kanaal- en functieterkenmerken waarop zij zich richten, zowel qua trainingsmateriaal als qua trainingsmethode. Voor het overige mag men bij funktietraining best bedenken dat sommige (dys)functies zich misschien wel helemaal niet laten trainen. Met dit laatste komen we op de volgende behandelingsvorm.

c. Circumventie; dat wil zeggen men probeert hetzelfde leerdoel langs alternatieve wegen te bereiken door middel van instructiemethoden en

leermiddelen die een beroep doen op nog intacte functies (de zgn. 'sterke punten' in de cognitieve ontwikkeling van een kind). Hiertoe gaat men over indien specifieke dysfuncties nauwelijks of niet te beïnvloeden, remediëren c.q. trainen zijn. Nog te weinig wordt onzes inziens van deze behandelingsvorm gebruik gemaakt.

d. Aanpassing van het curriculum of de leertaak aan de cognitieve ontwikkeling van het kind. Indien de aanpassing een vorm van buitengewoon onderwijs betreft dan zal deze behandelingsvorm vaak mede geïndiceerd worden door testgegevens welke betrekking hebben op de algemeen cognitieve ontwikkeling (intelligentie) van het kind. Indien de aanpassing betrekking heeft op het aantal (extra) leerstappen, dan zal deze behandelingsvorm primair geïndiceerd worden door de uitslagen van een criteriumtoets.

e. Extra training op de leertaak zelf of onderdelen daarvan; hiertoe gaat men over indien men meent dat het kind met wat extra-hulp zijn moeilijkheden te boven kan komen.

Lang niet alle behandelingsmogelijkheden zijn met dit rijtje opgesomd. Evenmin zijn de talloze combinaties van behandelingen in dit rijtje opgenomen. Eerder wilden wij aangeven dat er meer behandelingsvormen zijn dan men vaak praktizeert en dat alleen al om die reden de keuze van een behandeling niet zo eenvoudig en voor de hand liggend is als soms voetstoots wordt aangenomen. Dit geldt eens te meer indien men bedenkt dat het wetenschappelijk onderzoek naar leerstoornissen nog niet in die fase is gekomen dat men met zekerheid op bepaalde 'oorzaken' kan wijzen. Waar in de diagnostiek nog zoveel onzekerheid heerst, hoeveel meer onzekerheid mag men dan niet ten aanzien van de behandeling verwachten?

De konsekwentie hiervan is echter wel dat de eenmalige vaststelling van de aard van de leerstoornis en - in aansluiting daarop - de eenmalige keuze van een behandeling die de meeste kans op sukses biedt, een vooralsnog onbekend aantal mislukkingen zullen opleveren. Om dit risico zoveel mogelijk te verkleinen, willen wij daarom een pleidooi houden voor een continue terugkoppeling van informatie tussen de diagnosticus en de gene die de behandeling uitvoert (leerkracht, 'remedial teacher' of orthopedagoog). Op grotere schaal betekent dit dat de vroegtijdige opsporing, diagnostiek en behandeling van kinderen met potentiële of feitelijke leermoeilijkheden een continu, longitudinaal en interaktief proces dient te zijn (Groenendaal, 1979; Mark, 1972) dat niet vroeg genoeg - maar in ieder geval al op de kleuterschool - in beweging gebracht kan worden.

SUMMARY

1. General introduction

Since 1972 research has been carried out at the Netherlands Institute for Preventive Health Care, department of Child Health, under the title 'Early detection of learning disabilities'. The aim of this investigation is the development of instruments and techniques for the screening and diagnosis of children (four to eight year olds) with (potential) learning disabilities.

An extensive survey of the literature showed that there is no consensus about a definition of learning disabilities. The same survey, however, provided a number of starting points for the development of an instrument which was to be called Leyden Diagnostic Test (Schroots & Van Alphen de Veer, 1976).

Four experimental versions preceded the final version of the LDT. In this study the construction and the investigation of one of these versions, the LDT-E, is discussed. For the sake of clarity we have divided the study into two parts: A 'Cognitive ability and school achievement' and B 'Learning aptitude and school achievement'.

The object of part A is the development of a measuring instrument for the assessment of intelligence (general cognitive ability) as well as a number of specific cognitive abilities which may be of importance for learning in primary school. This measuring instrument, the LDT-E, aims at an individual psychological examination of children leaving nursery school (five to six year olds) in approximately one hour.

A. COGNITIVE ABILITY AND SCHOOL ACHIEVEMENT

2. Mark model

The LDT-E is based on a heuristic model, derived from Mark's theory of a hierarchical model of information-processing abilities (Mark, 1962; 1969). The model consists of four elements. Firstly, through a *channel* information is processed so that learning can take place. A large number of channels is available for the processing of information. These channels are defined by input and output. The input can be visual, auditory, tactile or kinesthetic. The output may be fine motor, gross motor or speech motor. As many channels can be distinguished as there are input-output combinations. Secondly, the central nervous system (CNS) performs

a number of *functions* within every channel. With the word 'function' (operation or system) Mark refers to an information processing activity of nerve cells. A total of 34 functions can be distinguished. Thirdly, there is the *hierarchical organization* of the set of functions. A function can only be activated if all other, less complex, functions have successively been activated. Fourthly and lastly, there is the *ontogenesis* of the set of functions within the CNS.

From the cognitive psychologist's point of view the weak spot in this model is the set of functions and their hierarchical organization. However, the importance of the model lies in its heuristic basis; that is to say, the Mark model may provide the psychodiagnostician with a guideline in composing and analyzing a battery of tests, with which the cognitive functioning of a child can be described.

3. Composition of the LDT-E

The final composition of the LDT-E is based on the following considerations.

Firstly, the most important channels should be represented. Because of the short time available for examination, the auditory-speech motor (A-S), the visual-fine motor (V-F) and the auditory/visual-fine motor (AV-F) channels seem to be most important for the diagnosis of (potential) learning disabilities.

Secondly, the most important functions should be represented in the LDT-E. It is difficult to give objective criteria for the selection of functions in advance. Some guidelines can be given, such as the hierarchical organization of functions. For each channel tests should be selected which measure more or less complex functions. It is also desirable to have the hierarchical organization of functions in the A-S channel parallel that of the V-F channel as much as possible. Attention, short term memory and simple reasoning seem to be the most important functions within the A-S channel as well as the V-F channel.

Thirdly, the LDT-E should enable us to estimate intelligence (general cognitive ability level). To this end the most important factors, as identified by Thurstone (1938), should be present in the LDT-E whether combined or not, except the factors 'numerical' and 'word fluency'. Fourthly, the LDT-E should be administered easily. This partly overlaps with the consideration that the total time of administering the LDT-E should not be much longer than one hour.

Given these considerations the LDT-E was composed of the following tests. *Block Design*, identical to the subtest Block Design from the WPPSI (Wechsler, 1967). In this test, selected for the V-F channel, the following functions can be distinguished: perceptual organization, matching, analysis, synthesis, abstract conceptualization ability, visual-motor co-ordination.

Paper Folding, an extended version of the subtest Paper Folding from the Hiskey-Nebraska Test of Learning Aptitude (Hiskey, 1966). A total of five new items were added. In this test, selected for the V-F channel, the following functions can be distinguished: continuous attention, memory for spatial transformations, reasoning and visual-motor co-ordination.

Knox Cubes, an abbreviated and slightly modified version of the Revised Knox Cube Test (Arthur, 1947). In this test, selected for the V-F channel, the following functions can be distinguished: continuous attention, short term memory for spatial sequences, visual-motor co-ordination.

Wordspan, based on Bakker's (1972) Temporal Order Auditory-Visual Test (TOAV). The TOAV was modified and ameliorated in several ways. Most important is the division into two separate tests: Wordspan and Picture Matching. As to Wordspan the subject had to repeat a series of nouns spoken by the examiner. After two practice trials (two words), twelve trials of respectively two (series 1), three (series 2-5), four (series 6-9) and five (series 10-12) words are presented. In this test, selected for the A-S channel, the following functions can be distinguished: continuous attention and short term memory for words.

Picture Matching. Immediately following the recall of a Wordspan series of nouns the subject is required to match the words - just recalled - with pictures, randomly arranged on a card. Besides the word matches each card shows one additional picture without a match. Matching has to be done in correct order. In this test, selected for the AV-F channel, the following functions can be distinguished: attention, memory for word sequences, matching.

Sentence Imitation, an unpublished 10-sentence story about a mother who is shopping with her child. Subsequent to reading the story at length each sentence, again being read, has to be recalled by the subject. In this test, selected for the A-S channel, the following functions can be distinguished: continuous attention and short term memory for sentences.

Story Questions. Following the recall of the sentences from Sentence

Imitation, the subject is required to answer ten questions about the story. In this test, selected for the A-S channel, the following functions can be distinguished: comprehension of questions, semantic memory and (re)productive verbal ability.

Comprehension, identical to the subtest Comprehension from the WPPSI (Wechsler, 1967). In this test, selected for the A-S channel, the following functions can be distinguished: social judgement, practical knowledge, logical reasoning, linguistic skill.

4. Method

Subjects. Four hundred and eleven children, sampled by age (5;05-6;04), sex and socio-economic status (SES-I = high; SES-II = middle; SES-III = low) from 19 nursery schools in 's-Hertogenbosch and Rosmalen, participated in the study.

Predictive measures. Block Design, Paper Folding, Knox Cubes, Wordspan, Picture Matching, Sentence Imitation, Story Questions, Comprehension.

Concurrent validity measures. Frankforter Schoolrijpheidstest (Gerritsen & Hulsman, 1974) and Nijmeegse Schoolbekwaamheidstest (Möns et al., 1969) are grouptests for the assessment of the school readiness of infants. The FST measures mainly 'Visual-motor differentiation' and the NST 'Visual-motor differentiation' as well as 'Perception and Verbal Comprehension'.

Criteria measures. School achievement as assessed by four tests and four teacher ratings at the end of the first grade of primary school.

Four achievement tests: Reading Comprehension (Van Calcar et al., 1974a), Spelling (Van Calcar et al., 1974b), Technical Reading (Brus & Voeten, 1974), and Technical Arithmetic (Heesen et al., 1969). Four teacher ratings: R(ating) Reading, R. Arithmetic, R. Language, and R. Writing. Ratings on 5-point scales, each scale varying from very weak (1) to excellent (5) performance as compared with class-mates.

Procedure. Nursery school testing was done by experienced graduate students in psychology (LDT-E) and trained nursery school teachers (FST and NST).

Primary school testing (achievement tests) was done some 15 months later by psychological assistants of the local School Advisory Center. Due to house-moving, non-admittance to primary school and illness, only 327 children out of the original 411 were tested at the end of the first year. A small 'restriction-of-range' effect seems likely as 25 drop-outs

were not admitted to the first year because they were not intelligent enough.

5. Reliability

Split-half reliabilities vary from .93 for LDT-E total (sum z-scores) and .87 for Sentence Imitation to .70 for Block Design and .66 for Paper Folding. Coefficients α are slightly lower, varying from .92 for LDT-E total and .85 for Sentence Imitation to .63 for Block Design and Paper Folding. The other coefficients are between those values.

The two reliability measures were interpreted in the light of p-values and item-rest correlations per test with reference to the theoretical range, mean and standard deviation. Some remarkable findings are: the low reliability of Block Design and Paper Folding, the ceiling-effect of Block Design, the discrepancy between the (good) internal consistency (split-half reliability) and (moderate) homogeneity (coeff. α) of Knox Cubes and the very high reliability of Sentence Imitation.

6. Structure

Age, sex and socio-economic status. Factorial analysis of variance was applied to the predictive measures as dependent, and age (5;05-5;10 vs. 5;11-6;04 yr. olds), sex (boys vs. girls) and socio-economic status (SES-I vs. SES-II vs. SES-III) as independent variables. No significant age effect has been found for Paper Folding, Wordspan, Sentence Imitation and Story Questions, this in contrast to the significant age effect for all the other tests and the LDT-E total.

With the exception of Paper Folding, sex affects none of the test performances; on Paper Folding, however, girls perform better than boys. The SES-effect shows most clearly in Sentence Imitation, to a lesser degree in Knox Cubes, Comprehension and LDT-E total, to an even smaller degree in Wordspan and Picture Matching, and shows hardly or not at all in Block Design, Paper Folding and Story Questions. Interactions appear to be non-significant.

Correlations between the tests vary from .15 (Wordspan - Paper Folding) to .63 (Wordspan - Picture Matching); correlations between the tests and the LDT-E total (corrected for contamination) vary from .37 (Block Design) to .70 (Sentence Imitation).

Comparing the attenuated correlations between the tests for the A-S and V-F channels, led to the conclusion that the correlations within a

channel are higher than those between channels; in other words, the functions - as measured by the tests from the LDT-E - are channel-linked to a certain degree.

Correlations between Knox Cubes and the other tests suggested that Knox Cubes not only measures short term memory for spatial sequences but also a relatively central, mediating function, not linked to a channel.

Finally, a systematic change in the order of the tests or intercorrelations gave a perfect simplex for the V-F channel and a quasi-simplex for the A-S channel. The hierarchical order of the tests for the V-F channel is as follows: Knox Cubes (least complex test), Block Design and Paper Folding (most complex test); for the A-S channel the order is as follows: Wordspan (least complex), Sentence Imitation and (ex aequo) Story Questions and Comprehension (most complex).

Factor structure. Several factor analyses were applied to the predictive measures in accordance with Jøreskog's maximum likelihood method. Tests were carried out for 0-, 1-, 2-, 3- and 4-factor solutions. Firstly, a 4-factor solution was selected. On the basis of the percentage extracted variance of the first unrotated factor (38%) in relation with the total percentage extracted variance (53%), it was concluded that the observed variance of the tests in the LDT-E could be described in one general ('g') dimension. After varimax rotation three of the four factors were interpreted easily as 'verbal comprehension', 'verbal memory' and 'spatial-analytic ability' respectively. It was difficult to interpret the fourth factor, but it could be interpreted tentatively within the Mark model. Secondly, also because of the risk that the fourth factor might only represent error variance (extracted variance 2%), a 3-factor solution was selected. This solution led also to the conclusion that the tests in the LDT-E could be described in one dimension. A subsequent varimax rotation yielded no dramatic change in the factor structure of the LDT-E as compared with the 4-factor solution: the fourth factor, which had been difficult to interpret, dissolved in the third factor 'spatial-analytic ability' (Block Design, Paper Folding and Knox Cubes). The first and second factor, 'verbal comprehension' (Sentence Imitation, Story Questions and Comprehension) and 'verbal memory' (Wordspan, Picture Matching and Sentence Imitation) respectively, turned out to be identical with the first two factors of the 4-factor solution.

Factor structure by age, sex and socio-economic status. Varimax rotation of 3-factor solutions by age (5;05-5;10 vs. 5;11-6;04), sex (boys vs. girls) and SES (SES-I vs. SES-II vs. SES-III) did not - apart from a few

exceptions - lead to remarkable changes in the factor structure of the LDT-E as compared with the 3-factor structure for the entire sample investigated. The factor structure of the LDT-E can be considered very stable, apart from the fact that some changes in the factor loadings of Knox Cubes and Picture Matching - the exceptions mentioned above - led to a reinterpretation of the second factor as 'memory'.

Factor structure in the light of FST and NST. A factor solution of the LDT-E, FST and NST together (3-factor solution, varimax rotation) largely confirmed previous results with respect to the psychological meaning and factor structure of the LDT-E. All the NST subtests, measuring 'perception and verbal comprehension', loaded on one and the same factor (I). A comparable result was found for the other subtests of the NST and the three subtests of the FST, measuring 'visual-motor differentiation'; these subtests also loaded on the same factor (III) as the three tests in the LDT-E, representing the factor 'spatial-analytic ability', mentioned above. Only Wordspan, Picture Matching and Sentence Imitation in the LDT-E, which are part of the factor 'memory' (identified above) loaded on the remaining factor (II) too. Taking into account the more limited factor structure and meaning of the FST and NST, it can be stated that the results of this analysis support the 3-factor structure and meaning of the LDT-E reasonably well.

7. Predictive validity

Correlations between LDT-E and criteria. The LDT-E total shows the highest correlation with all criteria; correlations vary from .40 (Technical Arithmetic) to .59 (Reading Comprehension). The least complex tests from the LDT-E - i.e. the 'memory' tests Sentence Imitation, Picture Matching, Knox Cubes and Wordspan - are the best separate predictors of reading and language in the order mentioned. [Reading and language is the abbreviation for Reading Comprehension, Spelling, Technical Reading, R. Reading and R. Language]. Sentence Imitation and Knox Cubes (ex aequo), followed by Picture Matching and Comprehension (ex aequo), are again the best separate predictors of arithmetic [= Technical Arithmetic and R. Arithmetic]. The best predictors of writing are Sentence Imitation, Block Design, Wordspan and (ex aequo) Paper Folding and Knox Cubes respectively.

Multiple correlations. In almost all cases it appeared that three tests were enough to obtain a maximum multiple correlation with the criterion measures in question. For reading and language these were Sentence

Imitation, Knox Cubes and Picture Matching (R varied from .49 to .61). No unequivocal result was found for arithmetic, probably because of the low validity of the criterion measure R . Arithmetic. There was no reasonable doubt, however, that Knox Cubes could be identified as a fairly good predictor of arithmetic. Combined with Comprehension, Knox Cubes gave a multiple correlation with Technical Arithmetic of .37. It appeared that Sentence Imitation, Block Design and Paper Folding were the best predictors of writing ($R = .44$). Stepwise multiple regression analysis by age, sex and socio-economic status gave only one surprising result, namely that Block Design, as well as Knox Cubes, is a good predictor of arithmetic in SES-III.

Canonical correlations. In order to find out the underlying relations between the LDT-E and school achievement, canonical correlations were calculated between the eight predictive measures on the one hand and three different combinations of criterion measures on the other hand, namely Achievement tests plus Teacher Ratings (I), Achievement Tests (II) and Teacher Ratings (III). The first canonical correlations are all significant ($p < .01$) and vary from .61 for combination II to .64 and .69 for combination III and I respectively. Of all variables Knox Cubes, Sentence Imitation, Reading Comprehension and R . Arithmetic consistently have the largest weight on the corresponding first canonical variates. On the basis of these results it was hypothesized that Sentence Imitation and Knox Cubes not only measure the rather simple 'short term memory' function, but also a more complex function, defined as 'coding'. Only the second canonical correlation of combination III turned out to be significant at the 1% level (.27). Block Design, Knox Cubes, Picture Matching, Story Questions, R . Language and R . Writing were the variables with the largest weight on the corresponding second canonical variate. An attempt to interpret the second canonical correlations and their variates did not give a clear result. Finally, the canonical correlations were calculated for younger and older infants separately, as well as for boys and girls and the various SES-groups. It appeared that of all the predictor and criterion measures Knox Cubes and Reading Comprehension had the largest weight on the first canonical variate. For the other measures no consistent pattern of weights was found for the various subgroups, neither on the first nor on the second canonical variate.

8. Discussion

A useful starting point of the discussion is the question in how far the results of this study justify the claims we made for the LDT-E, namely that it can measure the intelligence of school leaving infants, as well as a number of specific cognitive abilities which may be of importance for learning in primary school. Before answering this question we make some observations about the Mark model.

Mark model. We stated above that the set of 34 functions and their hierarchical organization are the weak spot in this model, not only because it is difficult to describe psychological tests in terms of these functions, but also because the various functions are too few for an adequate description of tests, administered to five and six year olds. In order to solve this problem, we resorted to the available literature on tests in which the various parts of the psychological content and meaning of the tests from the LDT-E are described somehow or other. Unfortunately, almost all these descriptions are traditional in character - that is originated in psychometrics - and hardly founded on some cognitive theory in which all kinds of internal processes and functions relating to human information processing, are expounded. As far as this last point is concerned, Carroll's views in his paper "Psychometric tests as cognitive tasks: a new 'Structure of Intellect'" (1976) offer a better perspective for the future. Therefore it is preferable to reformulate or extend the Mark model in the direction indicated by Carroll, particularly if this is combined with the construction of a taxonomy of functions in accordance with the principles of causal model analysis (Madaus et al., 1973).

LDT-E as intelligence test. All the results of this exploratory investigation into the significance and predictive validity of the LDT-E suggest directly or indirectly that the LDT-E can be regarded as an intelligence test. Age, sex and SES effects on the LDT-E total agreed largely with the effects of these variables on other intelligence tests. Another indication that the LDT-E measures intelligence, was obtained through factor analysis. In principle the observed variance of the tests in the LDT-E can be described by means of one general factor; besides, varimax rotation resulted in three factors which were interpreted easily: 'verbal comprehension', 'memory' and 'spatial-analytic ability'. Support for two of these three factors was obtained through the FST and NST, which can be regarded as (group)intelligence tests. The close relation between the LDT-E total, the FST and NST, was another indication that

the LDT-E is an intelligence test. A clear indication for the predictive validity of the LDT-E total was the positive correlation between the scores of school leaving infants on the LDT-E on the one hand and their achievements in reading, arithmetic, language and writing in the first grade of primary school on the other hand (median $r = .50$).

LDT-E as diagnostic test. The results of this investigation into the significance and predictive validity of the LDT-E as a diagnostic test were not satisfactory in every way. This is no doubt due to the lower reliability of four tests, namely Block Design, Paper Folding, Story Questions and Comprehension, as well as a lack of adequate reference variables. For the psychological meaning of these tests we have to refer to the results of the content analysis (literature survey) and factor analysis, although some findings on Paper Folding suggested that the function 'eye-hand coordination' is at least as important as the function 'memory for spatial transformations'. As the reliability of these four tests is rather low, their predictive validity was also rather low, as could be expected. By combining the data with respect to the reliability and validity obtained in this study with those of the tests in the LDT, we were able to formulate some suppositions about the differential validity of the tests as predictors for reading, arithmetic, language and writing in the first grade of primary school.

The reliability of Knox Cubes, Wordspan, Picture Matching and Sentence Imitation turned out to be much higher. Although there was an appreciable lack of reference variables for these tests as well, the results with respect to their validity were satisfactory. Wordspan, Knox Cubes, Picture Matching and Sentence Imitation turned out to be fair to good predictors of reading and language in the first grade of primary school. Sentence Imitation and Knox Cubes also appeared to predict arithmetic reasonably well. The findings with respect to Knox Cubes suggested moreover that this test measures not only the function 'short term memory for spatial sequences', but also the function 'verbal coding'. Therefore the description of the meaning of this test has to be supplemented. For the other three tests - namely Wordspan, Picture Matching and Sentence Imitation - a change or supplement of their meaning, based on content analysis and factor analysis, appeared unnecessary.

In view of the 'restriction of range' effect on the one hand and the relative wide time interval between the administration of the LDT-E and the various criterion measures on the other hand, it seems reasonable that all the eight tests from the LDT-E will show - separately or in

changing combinations - a greater relation with concurrent criteria like reading, arithmetic, language and writing than was stated in this study. Also considering the high degree of correspondence between the LDT-E and LDT, we may suppose that the greater relation expected, is true for the tests in the LDT and certainly for the less reliable tests in the LDT-E with the same name. Assuming this supposition to be true, it seems justified to use the LDT as a diagnostic test for the assessment of specific cognitive abilities which are of importance for learning in primary school. The question how important these abilities are for the acquisition of elementary skills such as reading, arithmetic, language and writing, can only be answered with more certainty after continued validation of the LDT.

B. LEARNING APTITUDE AND SCHOOL ACHIEVEMENT

9. Learning aptitude and limit testing

A major problem which has concerned many psychologists since the beginning of the twentieth century is the question whether the child's intelligence can be modified. At the moment three points of view can be distinguished: (1) genetic, (2) environmental and (3) interactionist. The last point of view, which most psychologists accept, is that IQ-type or cognitive ability measures reflect a complex interaction of environmental and genetic factors. This means that a cognitive ability test like the LDT-E is not a good measuring instrument for the assessment of a person's learning potential or learning aptitude, because part of the outcome of this test reflects the individual past learning experiences as well. However, according to Whitely & Dawis (1975),

"(...) it is theoretically possible to distinguish between individuals having the same ability but different aptitudes by directly measuring the modifiability of ability. That is, when two individuals within the same current level of ability are given equivalent intervention (e.g. practice), the individual with the greater aptitude should show a faster rate of change in ability than the individual with lesser aptitude" (p. 52).

It seems possible to realize the distinction 'cognitive ability vs. learning aptitude' in day-to-day psychological practice as well by means of limit testing.

Testing the limits or extension testing is a clinical test procedure, the essence of which is as follows. First an ability test is administered in accordance with standard procedures (pre-test). Then there is some

form of intervention in order to obtain an optimized test achievement. Intervention is also referred to as 'optimizing conditions': these conditions can vary from simple repetition of a certain test item to dividing a certain test item or test problem into subproblems, which are administered separately. Here the optimized score is called the post-test.

Mark (1972) distinguishes a total of fourteen optimizing conditions, the most important of which are: Increase frequency of exposure, Focus attention on tasks, Increase encouragement and feedback, Reduce superfluous distractions, Steady test materials, and Increase time on tasks. It depends not only on the child's test task and test behavior, but also on the examiner's clinical insight and attentiveness, which conditions are introduced during limit testing.

Testing the limits' is a procedure, which is far from standardized and against which many objections can be made. Norms for the evaluation of the post-test scores are given nowhere in the literature on tests; reliability and validity data are not mentioned either. These objections are cogent, the more so because those in favor of limit testing suppose the post-test scores not only to be more reliable than the pre-test scores, but also to have a greater predictive validity. Both suppositions seem worth looking into, as well as the related supposition that gain scores - given certain pre-test scores - give an indication of learning aptitude.

Taking into account a number of organisational and technical limitations, we can describe the object of part B of our study as obtaining an answer to the question whether (a) the post-test scores of Block Design, Paper Folding, Knox Cubes and Wordspan - as obtained by means of limit testing - have a closer relation with school achievement than the pre-test scores, and whether (b) the difference between post- and pre-test scores on the same tests - starting from given pre-scores - can supply us with an index of learning aptitude.

10. Method

After the administration of the LDT-E the same group of 411 children leaving nursery school as described in chapter 4 was examined once more as to Block Design, Paper Folding, Knox Cubes and Wordspan. Now the items the child had failed during the standard examination with the LDT-E - referred to as the first presentation (1) - were presented

once more in a manner more or less comparable with the up-and-down method (Dixon & Massey, 1957). In principle testing the limits proceeded as follows:

After presentation of a trial item the first item failed during the first presentation (1) was given a second time (2). When the child was successful, the next item failed on (1) was given until there were two failures during the second presentation (2). At this point the examiner stopped in order to present the first item failed on (1) and (2) for a third time (3). When this third presentation was successful, another item failed on (1) was offered for a second time (2). When the third presentation was unsuccessful, however, the next item failed on (1) and (2) was offered for a third time. After two consecutive failures on (3), the examiner stopped completely. For every successful item during limit testing the child received an extra point in addition to the score obtained during the standard examination. The result is three scores per test, score (3) being $\geq$ (2), and score (2) being $\geq$ score (1). Block Design was exceptional in that it was not presented for a third time, the time on task being doubled during the second presentation. While repeatedly presenting the items, the examiner varied a number of aspects of the test behavior in accordance with instructions in such a way that an optimized test achievement was obtained. The optimizing conditions were 'focus attention on tasks' (Attention), 'increase encouragement and feedback' (Encouragement), 'reduce superfluous distractions' (Distractions), and 'study test materials' (Assistance). The examiner had been instructed to rate the degree to which he varied the conditions Attention, Encouragement and Distractions on three 4-point scales - each scale varying from normal (1) to very strong (4) - for Block Design and Paper-Folding as well as for Knox Cubes and Wordspan. Only for Block Design and Paper Folding the condition 'steady test materials' had to be rated on a 2-point scale: no Assistance (1) and Assistance (2). Finally the examiner had to rate his impression of the child's fine motor skills for Block Design, Paper Folding and Knox Cubes, again on a 4-point scale (Skill), varying from normal (1) to very clumsy (4).

The examiners were the same group of experienced graduate students in psychology as described in chapter 4. There was an interval of one to two days between the standard and limit testing examination. Limit testing was carried out by the same examiners who had also carried

out the standard examination with the LDT-E. Finally, all the data regarding school achievements of the group of children examined (chapter 4), are relevant for the limit testing examination.

11. Results

Reliability. In order to obtain an impression of the reliability of Paper Folding, Knox Cubes and Wordspan on the second (2) and third (3) presentation, the split-half reliability coefficient and coefficient α were calculated. All coefficients increase in value from the first (1) to the third (3) presentation. Split-half reliabilities for Paper Folding, Knox Cubes and Wordspan on the third (3) presentation are .79, .87 and .85, respectively; coefficients α vary from .70 (Paper Folding) and .76 (Wordspan) to .82 (Knox Cubes), also on (3).

The two reliability measures were interpreted in the light of p-values and item-rest correlations per test and per presentation with reference to the theoretical range, mean and standard deviation. Some remarkable findings are: the ceiling-effect of Paper Folding and the greater standard deviation of Knox Cubes on (3). Considering the very high p-values and the very small number of reliable item-rest correlations of Block Design on (2), it appeared to make little sense to calculate the reliability for this test.

Structure. The correlations between Block Design, Paper Folding, Knox Cubes and Wordspan on first, second and third presentation on the one hand and the other four tests from the LDT-E on the other, were calculated.

The second presentation of Block Design has - compared with the second presentations of the other three tests - least variance in common with the first presentation (48%). Block Design can therefore be regarded as a relatively unstable test.

The correlations between Paper Folding on the first (1), second (2) and third (3) presentation are as follows. The greatest relationship (.89) is between Paper Folding (2) and (3); the smallest relationship (.64) between (1) and (3). The correlation between Paper Folding (1) and (2) lies in between (.73). Such a pattern can be expected on the basis of differences in administration, viz.: standard test conditions on (1) and optimizing conditions on (2) and (3). Compared with Knox Cubes and Wordspan, Paper Folding can be regarded as the least stable test. The correlations between Knox Cubes (1), (2) and (3) are much higher

than those between Paper Folding (1), (2) and (3), viz.: .79 for Knox Cubes (1) and (3), .85 for (1) and (2), and .95 for (2) and (3). Relatively speaking Knox Cubes can therefore be regarded as a reasonably stable test.

Of all four tests Wordspan is the most stable with correlations varying from .86 for (1) and (3), and .89 for (1) and (2) to .96 for (2) and (3).

The pattern of correlations between Block Design, Paper Folding, Knox Cubes and Wordspan on (1), (2) and (3) and the other tests from the LDT-E on (1), (2) and (3) shows no remarkable changes.

Ratings. Attention. From the percentages and means of rating scale values per test can be concluded that Knox Cubes is the most attention sensitive test and Block Design the least attention sensitive. The examiner had to attract the attention of more than 50% of the group of children examined in order to obtain an optimized test achievement, while this was only necessary in 25% of the cases in Block Design.

Encouragement. The examiner decided to use encouragement on Block Design, Paper Folding and Knox Cubes with slightly less than half of the children examined. Wordpan was exceptional in that one third of the children needed more than normal encouragement.

Distractions. One fifth of the children on an average showed such unbridled behavior that the examiner felt forced to intervene in order to obtain optimized test results.

Assistance. Only 2.9% of the children needed assistance in making Block Design; this in contrast to the fairly high percentage of children (28.1) who were helped passively imitating Paper Folding.

Skill. An average of 12.7% of the children has less than a normal fine motor skill in making Block Design and in imitating Knox Cubes. In imitating Paper Folding 42% of the children were regarded as clumsy, however. This indicates that Paper Folding calls on the child's fine motor skills to a relatively high degree. Direct support for this view was obtained by testing the sex difference on the 'Skill' scale on the basis of McGuiness' and Pribram's thesis (as quoted by Goleman, 1978) that girls are less clumsy than boys.

In order to obtain a tentative impression of the effect of repeated presentation on the one hand and of the effect of repeated presentation plus other optimizing conditions on the other, the group investigated was divided into two parts per test, group A 'no help' and group B

'help'. Group A consisted of those children whose test behavior on repeated presentation did not or hardly necessitate intervention and who therefore scored one (1) on the scales Attention, Encouragement, Distractions and Assistance. Group B consisted of the remaining children, who did need one form of help or other and who therefore scored at least two (2) on one or more scales (except Skill). The average test achievement of group B appears to be systematically lower per test than that of group A on the first, the second as well as the third presentation. In addition the increase in average test achievements from the first to the second presentation and from the second to the third presentation per test is almost identical for group A and B. Absolutely speaking, group B makes a little more progress than group A, it is true, but the average difference of 0.38 on all tests between first and last presentation is negligible.

Predictive validity. From the correlations between Block Design, Paper Folding, Knox Cubes and Wordspan on (1), (2) and (3) on the one hand and the eight criterion variables (cf. chapter 4 and 7) on the other, it appears that limit testing - in spite of a generally higher reliability of the tests on (2) and (3) - does not raise the predictive validity in all cases.

The predictive validity coefficients of Block Design, which are not exceptionally high anyway, clearly go down for all criterion variables on (2), (median lowering of correlations: .06).

Also on Paper Folding it appears that limit testing lowers the predictive validity coefficients, most strongly so regarding Technical Reading, Rating Reading and Rating Language (a difference of .06, .10 and .05 between (1) and (3) respectively).

The effect of limit testing on the predictive validity of Knox Cubes is virtually nil: some correlations go up or go down by .01 or .02 and other correlations stay the same.

Wordspan is the only test on which limit testing has a positive effect although not overwhelmingly so: all predictive validity coefficients go up by .01 to .06 from (1) to (2). The correlations with Reading Comprehension, Spelling and Technical Reading show the highest rise (.05 to .06). Moreover it is striking that the correlations from (2) to (3) hardly rise any further and that they even go down a little or remain the same in some cases.

In order to answer the question in how far gain scores - starting from

given pre-test scores - can provide an indication of learning aptitude, we applied the log linear model so as to find out in how far the gain scores per test can be regarded as moderators of the relations between predictor (pre-test scores) and criterion, in this case the composite variable 'Reading and language' (= sum z-scores of Reading Comprehension, Spelling, Technical Reading, Rating Reading and Rating Language). We also studied the moderator effect of a variable like 'help' (cf. Ratings) on the relations between the predictors (pre- and post-test scores, first and last presentations respectively) and the composite criterion Reading and language; for there is a chance that what goes for group A 'no help', does not go for group B 'help' and vice versa. Starting from $3 \text{ (pre-test)} \times 3 \text{ (post-test)} \times 2 \text{ (help)} \times 3 \text{ (Reading and language)}$ contingency tables for Block Design, Paper Folding, Knox Cubes and Wordspan, it appeared that the log linear model with only 'main effects' shows the best fit per test. This means that it was impossible to demonstrate any moderator effect and that it therefore does not make sense to construct an index of learning aptitude by means of limit testing.

12. Discussion

Some critical remarks. Given the clear ceiling effect on Block Design and Paper Folding (last presentation) as well as the spurious reliability of Paper Folding (it seemed to make little sense to even calculate the reliability of Block Design), the small correlation between the post-test scores of these tests and the criterion school achievement at the end of the first grade in primary school can hardly be called surprising. The same goes for the complete lack of any moderator effect. As the higher reliability of the post-test scores of Knox Cubes and Wordspan can be taken seriously, one wonders why the predictive validity of these tests has hardly gone up or not at all and why it was not possible to show any moderator effect either.

Feuerstein (1972b) would explain these results in the following manner. It makes little sense to use conventional criteria like achievement tests and ratings, if the school achievements were made in class without extra attention, individual help or encouragement, which are the aspects found in limit testing. The only thing one can expect in such a situation is spuriously low predictive validity coefficients. Bersoff (1973) gives an additional explanation when he states that the standard

test situation as compared with the learning situation in class has been optimized to such an extent that no real contribution to the common variance of predictors (pre-test scores) and criteria can be expected on limit testing.

In the light of these statements it is almost surprising that the post-test scores of Knox Cubes and Wordspan show respectively the same or slightly higher correlations with the criteria used as the pre-test scores. Almost, for it is likely that the possible negative effect of the intervention used on the predictor-criterion relation is compensated by the positive effect of the higher reliability of the post-test scores of Knox Cubes and Wordspan on the same relation.

There are two other possible explanations for our negative findings. Firstly: the intervention used - at most two presentations per item under optimizing conditions - is very short, the more so if compared with the intervention used in learning potential assessment. Secondly: in general one cannot exclude the possibility that the statistical procedures used - among other things because of the underlying assumptions or because of a lack of subjects - are not sensitive enough to register such (minimal) changes as may possibly be observed and interpreted in a clinical-intuitive manner by an experienced, practising psychologist.

For these reasons we wish to give limit testing as a practical-clinical test procedure for the assessment of learning aptitude the benefit of the doubt. Limit testing does seem to be a useful, little time consuming testing procedure, only for the raising of the reliability of measurements of less able or deprived children, the more so if it is accompanied by a greater number of observations or interventions per test.

Learning potential assessment. This testing procedure can be regarded as a further refinement and extension of limit testing. Budoff (1973) and Feuerstein (1972b) are the first names associated with learning potential assessment (LPA). Essentially both authors proceed from the pre-test - intervention - post-test paradigm. That is to say: first the cognitive ability level is assessed objectively by means of a non verbal test, in accordance with the standard test instruction. Then one form of intervention or other is used; this form can vary from practice on the pre-test itself to training or coaching on a test task comparable to the pre-test or parts of the pre-test. Finally a parallel

form of the pre-test is administered as a post-test, also in accordance with the standard test instructions.

A point in favor of LPA is that the cognitive process aspect - i.e. the way in which the child solves a problem - is emphasized more (Feuerstein) or less (Budoff). In this way those critics are accommodated who believe that the traditional cognitive ability test gives no information about the strategies used by the child in solving a test problem.

LPA, as used by Feuerstein, shows a rich variety of interventions. This causes problems in the standardisation of the testing procedure. Feuerstein states that he is opposed to too rigid a standardisation in individual cases, as this prevents the establishment of an index of modifiability (learning aptitude).

The criticism of LPA concentrates on the following three points. If standardisation of an extensive and individual intervention is undesirable, how can one a) convey the practice of intervention to practicing psychologists but through a very long teacher-pupil relationship, b) evaluate the effect of an intervention (which intervention?), and c) construct an index of learning aptitude or modifiability which is of more than individual and therefore unique significance?

Our final impression is that LPA provides much 'soft' information about the child's cognitive functioning, but has given very few 'hard' reliability and validity data.

Suggestions for further research. If it is hardly or not possible to determine the effect of interventions as used in limit testing and LPA, then it may seem useful to make a step backward in the present evaluation research and to submit the interaction between testee and examiner to further analysis. To this end an extensive interaction-analysis system plus corresponding rating scales should be developed. However, we can make another step backward and wonder whether the costs of developing such a system are outweighed by the benefits; certainly with respect to such a little time consuming procedure as limit testing. From this point of view it may be more interesting to further refine limit testing - also by developing a greater number of rating scales than were realised in this study - , simultaneously modifying the expectations about the outcomes of limit testing. In this way limit testing seems an extremely useful clinical testing procedure by which one can generate hypotheses about a large number of

the child's learning and behavioral characteristics, in need of further testing in a follow up study with other instruments and techniques. Now that we have made two steps backward, we can try to make three steps forward. Cronbach (1976) wrote a most interesting article 'On the design of educational measures'. In this article he tries to hybridize his G - and D theory so that the key note is the multifacet sequential design for the evaluation of educational programs. Limit testing as well as LPA can be regarded as mini educational programs and are suitable for the application of this G/D theory. If we have in mind any further study in the field of limit testing or if necessary LPA, then it would be in this direction.

Finally: if it should in principle be possible - according to Whitely & Dawis (1975) - to make a distinction between learning ability and learning aptitude by means of a simple laboratory task, then it still seems impracticable to realize this distinction by means of limit testing on relatively complex psychological tests. In the light of this situation it would be wiser to aim future research at the further development of measuring instruments which hardly reflect the individual past learning experiences, this in contrast to the traditional cognitive ability test. Here we mainly have in mind the so-called neurometric tests (John et al., 1977) which are based on the electrophysiological technique of the evoked potentials. Judging by the results obtained so far - also in the field of the study of intelligence - these tests should certainly be taken into serious consideration.

NOTEN

- 1 NIPG-TNO projekt 529 'Vroegtijdige herkenning van leerstoornissen' maakt deel uit van het door SVO (Stichting voor Onderzoek van het Onderwijs) gesubsidieerde projekt 0274 'Vroegtijdige herkenning van en hulp aan kinderen (4-8 jaar) die op enigerlei wijze vastlopen in het onderwijs'. Laatstgenoemd projekt is een samenwerkingsverband tussen de vakgroep Ontwikkelingspsychologie, Pedagogie en Speciale Pedagogiek van de Vrije Universiteit, het Gewestelijk Onderwijs Advies Centrum 's-Hertogenbosch e.o. en de afdeling Jeugdgezondheidszorg van het NIPG-TNO. SVO-projekt 0274 is inmiddels afgesloten in de vorm van een beknopt eindrapport (Groenendaal et al., 1978).
- 2 Elders zijn wij uitvoeriger op deze materie ingegaan (Schroots, 1976).
- 3 Keogh et al. (1974) halen in dit verband Faust (1970) aan die opmerkt dat:
"whereas correlation between intelligence test scores and later achievement test scores are high when the two measures are administered close together in time, the correlations become poorer as the time between testing grows larger. This is due, in part to the fact that intelligence is a developing characteristic and learning requirements and learning setting change" (p. 4).
- 4 Deze passage is door ons - gelet op Mark's specifieke terminologie - tamelijk vrij vertaald. Letterlijk staat er:
"At all levels involving more complex operations than simple relating (...), the reasoning power used by the system in any one operation may be estimated by considering the reasoning power an inverse function of the probability that the system guessed or was prompted (or rigidly programmed) in relating an answer to a question" (p. 79).
- 5 Met betrekking tot de logische vergelijking

$$x_n \longrightarrow y_n \wedge y_{n-1} \cdots y_1$$

merkt Mark (1962) het volgende op:

"Logical equation only represent what Ashby (3) calls a 'logic of mechanisms'. It may be seen that the mere presence of these factors, as shown by successful answer finding at time t_1 , does not assure subsequent successful answer finding at time t_2 . Thus, it is not true that

$$y_1 \wedge y_2 \cdots y_n \longrightarrow x_n$$

If it were true, we would in effect be able to predict, on the basis of the previous established presence of these brain operations, that these systems will again function correctly - that is, that they will go through a predetermined relating process and select the correct answer. The hierarchy, then, would have predictive characteristics such as Guttman scales (19). Prototypes of these hierarchies have indeed been used clinically because the hierarchies do have useful predictive characteristics (20). In biological forms, however, it is quite possible to have normally or pathologically time-dependent variable or changing systems. Such systems may well be able to find an answer at t_1 and be unable to find the same answer to the same questions at t_2 . Expletive speech in aphasic patients may be cited as an example of a condition where only extensive and often

irrelevant stimulation occasionally permits the patient to activate his speech mechanism. In general, individuals with recent traumatic head injuries, or individuals who have experienced recent cerebral vascular accidents may be cited as examples of physiologically unstable systems" (p. 80).

- 6 In Mark's 1962-artikel wordt een vrij gedetailleerde beschrijving gegeven van 34 verschillende informatieverwerkende systemen welke te zamen het controlesysteem vormen. Opvallend in deze beschrijving is de konsekwent volgehouden terminologie, ontleend aan de systeemtheorie van informatieverwerking.
Psychologische tests die deze systemen zouden moeten meten, worden echter beschreven in termen en begrippen die weinig van doen hebben met de systeemtheorie van informatieverwerking. Bijvoorbeeld: een test die bedoeld is om het vermogen tot 'abstraheren' te meten zou - in systeem- en informatietheoretische termen uitgedrukt - geëti-ketteerd worden als 'continuous problem solving by question-pattern manipulation where the original problem has a within-system origin'. De vertaling van konventionele testtermen en begrippen in Mark's systeem- en informatietheoretische terminologie vormde voor ons dan ook een onoverkomelijke moeilijkheid.
- 7 Eerder (p. 26) merkten wij reeds op dat de funktie-indeling van Mark slechts heuristische waarde heeft.
- 8 Men bedenke dat de einduitkomst van Cronbach's redenering alleen be-trekking heeft op de hiërarchische rangschikking van factoren en niet op hun ontogenese.
- 9 Soudijn (1977) merkt hierover het volgende op:
"Heterarchie en hiërarchie zijn geen los van elkaar staande begrip-pen. Heterarchie impliceert een verschuivende hiërarchie. Het 'mees-terplan' dat aangeeft dat je naar iets anders moet overgaan, moet zèlf weer kenmerken van een hiërarchie hebben" (p. 426).
- 10 Voor de LDT-E zijn geen normeringsgegevens verzameld; dientengevolge heeft de konstruktie van een deviatie IQ als maat voor de algemeen cognitieve ontwikkeling weinig zin. Om een redelijke vergelijking mogelijk te maken met het deviatie IQ is daarom de som van de stan-daardscores (z-scores) op de LDT-E berekend, aangeduid met 'totaal' (cf. hfdst. 5 e.v.).
- 11 Sommige tests uit de LDT-E heetten oorspronkelijk anders. Bijvoor-beeld: Zinnen nazeggen werd aangeduid met Verhaaltje nazeggen en Woordenspan met LPT-a (Lees Prediktie Test-a), respektievelijk LPT-nazeggen. Ter voorkoming van enig misverstand worden echter in dit deel de namen van de tests uit de handleiding voor de LDT (deel I) aangehouden.
- 12 Gegeven de manier waarop Woordenspan en Plaatjes aanwijzen moeten worden afgenomen, is de kans bijzonder gering dat pp de bij een se-rie woorden horende plaatjes goed zal aanwijzen indien deze serie eerder foutief is herhaald. In die zin zijn de behaalde resultaten op Plaatjes aanwijzen direkt afhankelijk van de eerder behaalde re-sultaten op Woordenspan. Deze afhankelijkheidsrelatie komt direkt tot uiting in het bij Woordenspan vergeleken lagere gemiddelde van Plaatjes aanwijzen (cf. par. 5.2). Dezelfde afhankelijkheidsrelatie verklaart ook gedeeltelijk de relatief hoge samenhang tussen Woor-denspan en Plaatjes aanwijzen (cf. par. 6.2).

- 13 Aangezien voor de LDT-E - in tegenstelling tot de LDT - geen genormaliseerde standaardscores (IQ Equivalenten) zijn berekend, is de hier voorgestelde analyse van het LDT-E profiel in feite onmogelijk. Vandaar dat wij in deze en de volgende passage konsekvent de naam LDT hanteren, ook al maakt het onderscheid LDT-E / LDT voor de gedachtengang achter de analyse van het testprofiel weinig uit.
- 14 Eerder verwoordden wij de opvatting dat een stoornis in een of meer cognitieve functies tot moeilijkheden bij het leren kan leiden en dat met behulp van psychologische tests die deze functies meten de stoornissen of dysfuncties opgespoord kunnen worden (cf. par. 2.1). Deze opvatting impliceert een voorzichtig geformuleerd causaal verband tussen cognitieve dysfuncties enerzijds en moeilijkheden bij het leren anderzijds. Voorzichtig geformuleerd, omdat de relatie tussen de vaardigheden die op school bijgebracht worden en de daarvoor benodigde cognitieve functies nog op vele punten onduidelijk is (Keogh & Becker, 1972). Hoe onduidelijk is in paragraaf 3.1 wel gebleken, waarin de samenstelling van de LDT-E besproken werd. In feite beschikt men alleen over globale aanwijzingen voor de keuze van relevante functies en kanalen. Een model waarin de relatie tussen schoolvaardigheden zoals lezen, rekenen, taal en schrijven en de verzameling functies en kanalen uitvoerig beschreven wordt, zijn wij dan ook in de literatuur niet tegengekomen.
- 15 Konfirmeren te verstaan als 'ondersteuning vinden voor'.
- 16 Voor een uitvoerige toelichting op de door de VU getrokken steekproef raadplege men Groenendaal (1978; p. 44 e.v. en p. 66 e.v.).
- 17 Ten tijde van het onderzoek werd de toenmalige experimentele versie (1971) gehanteerd.
- 18 Mondelinge mededeling.
- 19 Met dank aan mw drs F.M.E. Gerritsen.
- 20 De kleuters bij wie de FST en NST om een of andere reden niet zijn afgenomen, vormen geen aselekte steekproef uit de gehele groep van 411 schoolverlatende kleuters. Als zodanig zou dit tot een vertekening van de onderzoeksresultaten kunnen leiden. A priori achten wij een dergelijke vertekening echter weinig aannemelijk, aangezien er geen duidelijk aanwijsbare reden is voor de veronderstelling dat ziekte of verhuizing enerzijds en organisatorische problemen anderzijds op een of andere wijze samenhangen met de in dit onderzoek bestudeerde variabelen. Ten overvloed hebben wij echter onderzocht in hoeverre de frekwentieverdeling, zoals weergegeven in tabel 5, verschilt van de feitelijke verdelingen van FST en NST. Geen enkele uitgevoerde χ^2 -toets bleek significant te zijn.
- 21 In het kader van een routine-onderzoek op een aantal scholen in 's-Hertogenbosch en Rosmalen zijn door het GOAC op gedeeltelijk dezelfde groep kinderen als beschreven in paragraaf 4.3 nog drie andere toetsen afgenomen, namelijk:
 Leesvoorwaardentoets (Sixma, 1971): een individuele toets, samengesteld uit acht subtests, welke de belangrijkste leesvoorwaarden, zoals voorwaarde der zingeving, taalbeheersing, objectivatie, auditieve en visuele discriminatie, onderzoekt. Afnametijdstip: eind augustus 1973. N = 224.
 Kalom II (Van Calcar et al., 1973): een klassikale toets voor het onderzoek naar de mechanische leesvaardigheid. Afnametijdstip: november 1973. N = 332.

Hilversumse Begrippentoets Rekenen (School-pedagogische en -Psychologische Dienst Hilversum, 1973): een klassikale toets, welke de belangrijkste rekenvoorwaarden, zoals inzicht in kwalitatieve en ruimtelijke relaties, structurering en geleiding evenals een aantal kwantiteitsaspecten onderzoekt. Afnametijdstip: november 1973. N = 308.

Deze drie toetsen zijn in feite leervoorwaarden-toetsen en als zodanig minder interessant voor ons onderzoek. Desondanks hebben wij deze toetsvariabelen uit pure nieuwsgierigheid meegenomen in de analyse. Indien daartoe aanleiding bestaat zullen wij de resultaten hiervan kort weergeven in dit notenapparaat. Voor geïnteresseerden liggen de volledige analysesresultaten ter inzage op het NIPG-TNO.

- 22 Ten aanzien van het aantal onderzochte kinderen per criteriumvariabele gelden in principe dezelfde opmerkingen en argumenten als voor de FST en NST (zie noot 20). Ook voor deze variabelen hebben wij onderzocht in hoeverre de frekwentieverdeling, zoals weergegeven in tabel 5, verschilt van de feitelijke verdelingen per criteriumvariabele. Geen enkele uitgevoerde χ^2 -toets bleek echter significant te zijn.
- 23 In dezelfde vragenlijst wordt ook het oordeel van de leerkracht gevraagd over het schoolgedrag van de leerling. In dit gedrag worden 23 aspecten onderscheiden, variërend van 'aandacht', 'motoriek', 'taalgebruik' en 'interesse' tot diverse 'persoonlijkheidskenmerken'. Voor ieder aspect is een apart item opgesteld dat een korte beschrijving omvat van het bedoelde gedrag, gevolgd door een vijfpuntsschaal, waarop het oordeel van de leerkracht moest worden aangegeven. De 23 items zijn vervolgens a priori ingedeeld in een zevental categorieën, mede geleid op de praktische noodzaak om tot een reductie van het aantal criteriumvariabelen te komen. Om na te gaan in hoeverre onze verwachtingen omtrent de samenhang van de items - tot uitdrukking gebracht in de a priori-indeling - door de resultaten bevestigd worden, is tenslotte een faktoranalyse (varimaxrotatie) uitgevoerd. Een goed interpreteerbare faktorstructuur werd echter niet verkregen. Evenmin viel de a priori-indeling gemakkelijk terug te vinden in het wisselende beeld dat de faktorladingen ons verschaften. Naar aanleiding van deze bevindingen hebben wij dan ook besloten geen criteriumvariabelen met betrekking tot het gedrag van de leerling in ons onderzoek op te nemen. Op aanvraag zijn echter alle gegevens, inclusief de vragenlijst, beschikbaar.

24

$$P_i = \frac{\sum_{pp=1}^N i}{N \cdot i_{\max}} \times 100$$

waarin:

P_i = moeilijkheidsgraad van item i
 pp = proefpersoon
 N = aantal proefpersonen
 i = score op item i
 $i_{\max}$ = maximaal te behalen score op item i

- 25 Drenth's gedachtengang (1975; p. 203 e.v.) met betrekking tot de factoren die de variantie van een test bepalen, wordt het best geïllustreerd in zijn toelichting op de koëfficiënt van a-specificiteit als maat voor de betrouwbaarheid van een test. Deze koëfficiënt wordt verkregen door twee paralleltests met elkaar te korreleren onder de assumptie dat de gemeenschappelijke variantie van twee paralleltests

gelijk is aan de gezochte proportie ware variantie in één test, welke - zoals bekend - gelijk is aan zijn betrouwbaarheid. De koëfficiënt van interne consistentie, evenals de homogeniteitskoëfficiënt, mag nu als een verbijzondering opgevat worden van de koëfficiënt van a-specificiteit voor die gevallen waarin men slechts over één test beschikt en toch zijn betrouwbaarheid wil bepalen. Uiteraard leidt dit tot de nodige beperkingen bij de toepassing en interpretatie van deze koëfficiënten. Zie hiervoor echter Drenth (1975; p. 186 e.v.) en Nunnally (1967; p. 172 e.v.).

26 Zie ook paragraaf 3.1

27 De betrouwbaarheid (interne consistentie) van de definitieve versie van Blokpatronen en alle overige tests uit de LDT is vastgesteld op een aselekte steekproef uit de Nederlandse populatie kinderen van 4 tot 8 jaar, die gedurende de maand oktober en november 1975 een kleuterschool of een school voor gewoon lager onderwijs bezochten (deel I; Schroot's & Van Alphen de Veer, 1976; p. 29). Van alle berekende koëfficiënten van interne consistentie worden hier ter vergelijking alleen die voor de leeftijdsgroepen IV (5;06-5;11) en V (6;00-6;05) weergegeven (cf. deel I, p. 42).

28 De LDT-E is één van de vier experimentele versies van de LDT die aan de definitieve samenstelling zijn voorafgegaan. Afname van deze experimentele versies vond onder meer plaats bij jongere en oudere kinderen dan respectievelijk 5 en 6 jaar.

29 Deze nadelen kunnen echter grotendeels omzeild worden indien men gebruik maakt van per leeftijdsgroep ontwikkelde normen, waarbij de ruwe scores omgezet worden in genormaliseerde standaardscores, bijvoorbeeld in de vorm van IQ Equivalenten (cf. deel I; p. 35 e.v.). De verkregen IQ Equivalenten zijn dan vervolgens vrij eenvoudig om te zetten in Leeftijdsequivalenten volgens de formule:

$$LE = \frac{IQE \times CL}{100}$$

waarin:

LE = Leeftijdsequivalent

IQE = IQ Equivalent

CL = Chronologische Leeftijd

30 Het hier gepresenteerde probleem met betrekking tot de konstruktie van gedifferentieerde leeftijdsnormen en de daaraan verbonden konklusie is minder eenvoudig dan het lijkt. Zo heeft de gevolgde redenering uitsluitend betrekking op die situatie waarin de regressielijnen van onderscheiden groepen samenvallen. Het al dan niet samenvallen van regressielijnen doet echter weinig af aan de strekking van ons betoog.

31 Met dank aan de auteurs van het computerprogramma VARIAN / 02 (Kwaai-taal & Roskam, 1968).

32 De vergelijking van korrelaties met betrekking tot het AV - F kanaal is onmogelijk, aangezien dit kanaal slechts door één test, i.e. Plaatjes aanwijzen, gerepresenteerd wordt.

33 Praktische overwegingen bij de konstruktie van het LDT Stroomdiagram bepaalden indertijd hoofdzakelijk de volgorde waarin de tests in tabel 3 geplaatst zijn.

34 Voor deze analyse is gebruik gemaakt van het computerprogramma FAKTO / 80 (Psychologisch laboratorium, Universiteit Nijmegen). De toegepaste

optie 'listwise deletion' verklaart - in samenhang met de uitval van sommige pp'n op enkele tests uit de LDT-E - het geringere aantal pp'n van tabel 12 (N=392), vergeleken met het aantal pp'n in de steekproef (tabel 5, N=411).

- 35 Getoetst wordt de hypothese 'm factoren zijn voldoende om de samenhang tussen p variabelen te verklaren' tegen het alternatief dat er meer factoren nodig zijn.
Een aantal door Bethlehem et al. (1978) aanbevolen controlemaatregelen met betrekking tot de 'nauwkeurigheid van de schatters van de faktorladingen', de 'gedetermineerdheid van de factoren' etc., kon niet worden uitgevoerd, aangezien het gebruikte computerprogramma³⁴ de hiervoor benodigde gegevens niet verschafte.
- 36 Om de kans op interpretatie van foutenvariantie zo klein mogelijk te houden, hanteren we een tamelijk konservatieve vuistregel voor de inspectie van faktorladingen.
- 37 Dit is niet helemaal korrekt. Zo heeft de tweede faktor - via Plaatjes aanwijzen - ook betrekking op het AV-F kanaal, dat qua invoer van informatie (auditief/visueel) gelijkenis vertoont met het A-S kanaal en qua uitvoer (fijnmotorisch) met het V-F kanaal. Deze 'gemengde' kanaaleigenschappen worden duidelijk gedemonstreerd in de ladingen van Plaatjes aanwijzen op de tweede (.68) en de derde (.32) faktor. De toegepaste vuistregel³⁶ laat echter niet toe dat wij enige aandacht besteden aan de lading op de derde faktor. Om ons betoog verder zo helder mogelijk te houden doen we nu net alsof de tweede faktor - met de hoge lading van Plaatjes aanwijzen hierop - alleen betrekking heeft op het A-S kanaal.
- 38 De vraag of ladingverschil in subpopulaties te wijten kan zijn aan de door ons gekozen rotatietechniek (varimaxrotatie) hebben wij voorgelegd aan Dr J. ten Berge van de Rijksuniversiteit Groningen. Volgens zijn zeggen zijn er inderdaad gevallen bekend waarin varimaxrotatie verschillen te zien geeft die door verdere rotatie kunnen worden weggewerkt, bijvoorbeeld via een Procrustesrotatie met het programma Comma (Ten Berge, 1977; p. 272). Een dienovereenkomstig door Ten Berge uitgevoerde bewerking van ons onderzoeksmateriaal leverde echter als konklusie op dat de gekonstateerde ladingverschillen serieus genomen mogen worden. Bij deze willen wij Dr J. ten Berge hartelijk danken voor zijn belangeloze medewerking aan ons onderzoek.
- 39 Opmerkelijk is het beeld dat de korrelaties met de Sixma leesvoorwaardentoets en de Hilversumse Begrippentoets Rekenen ons verschaffen. Zo is het totaal van de Sixma leesvoorwaardentoets - in tegenstelling tot wat de naam van deze toets doet vermoeden - het hoogst gekorreleerd met de beide rekenvariabelen, respectievelijk .51 (Schiedamse Rekenetest) en .59 (Oordeel Rekenen) en het laagst met de Brus (.41) en Lezen volgens het oordeel van de leerkracht (.44). Met enige voorzichtigheid zouden wij hieruit kunnen konkluderen dat de leesvoorwaardentoets eerder reken- dan leesvoorwaarden meet. Voor de Hilversumse Begrippentoets Rekenen geldt dat deze weinig specifiek is: met uitzondering van het Oordeel Schrijven (.19) treffen we over de gehele linie van criteriumvariabelen ongeveer gelijkwaardige verbanden aan, variërend van .45 (Schiedamse Rekenetest) tot .50 (Begrijpend lezen).
- 40 Voor de berekening van de kumulatieve multipele korrelaties werd gebruik gemaakt van het computerprogramma NIPG8064(E) Ref.: VW.

LITERATUUR

- ALIOTTI, N.C. Covert assessment in psychoeducational testing. *Psychology in the Schools* 14 (1977) 438-43
- ALPHEN DE VEER, R.J. VAN, F.M.E. GERRITSEN, W.L.L. HULSMAN & J.J.F. SCHROOTS. Betekenis en voorspellende waarde van de Nijmeegse Schoolbekwaamheidstest, Frankforter Schoolrijpheidstest en Profieltest. Nijmegen, Berkhout Testmateriaal, 1978
- ALPHEN DE VEER, R.J. VAN, & J.W. PLOMP. De Verhaaltjestest; scoring van Zinnen nazeggen. Leiden, NIPG/TNO, 1976
- ANASTASI, A. *Differential psychology; individual and group differences in behavior*; 3rd ed. New York, MacMillan, 1958
- ANASTASI, A. *Psychological testing*; 2nd ed. New York, MacMillan, 1961
- ANGOFF, W.H. Scales, norms and equivalent scores. In: R.L. Thorndike (ed.). *Educational measurement*; 2nd ed. Washington, D.C., American Council on Education, 1971. pp. 508-600
- APOSTEL, L. The cognitive point of view: introduction to a discussion. In: M. de Mey, R. Pinxten, M. Poriau & F. Vandamme (eds). *International workshop on the cognitive viewpoint*. Gent, University of Ghent, 1977. pp. IX-XV
- ARLMAN-RUPP, A.J.L. Receptief taalgebruik van vijfjarige kinderen. ZWO-project 59001, 1971
- ARTHUR, G. *A point scale of performance tests; revised Form II; manual for administering and scoring the tests*. New York, Psychological Corporation, 1947
- AUSTIN, J.J., & R. CARPENTER. The use of the WPPSI in early identification of mental retardation and preschool special education. In: R.S. Morrow (Chairman). *Diagnostic and educational application of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI)*; symposium presented at the American Psychological Assoc., Miami, Sept. 1970
- BABAD, E.Y., & M. BUDOFF. Sensitivity and validity of learning-potential measurement in three levels of ability. *J.educ.Psychol.* 66 (1974) 439-47
- BAKKER, D.J. Temporal order, meaningfulness and reading ability. *Percept. Motor Skills* 24 (1967) 1027-30
- BAKKER, D.J. Temporal order in disturbed reading; developmental and neuropsychological aspects in normal and reading-retarded children. Rotterdam, University Press, 1972
- BANNATYNE, M., & A. BANNATYNE. Training van het lichaamsschema door beweging-taal-concentratie. Rotterdam, Lemniscaat, 1973

- 48 Niet alle veertien optimaliserende kondities worden in Mark (1972) vermeld. Dr U. Wälti, kinderpsycholoog te Bern, was echter zo vriendelijk om ons een volledige lijst met kondities (plus toelichting) toe te sturen.
- 49 In paragraaf 9.3 verwezen wij reeds direkt en indirekt naar de vele voorstanders van grenzen-toetsen.
- 50 Overigens met uitzondering van Begrip en inzicht. De scoring van de antwoorden leverde echter dermate grote problemen op dat wij afgezien hebben van analyse en uitwerking der resultaten.
- 51 De opmerking dat bij Blokpatronen geen derde aanbieding heeft plaatsgevonden, zullen wij in het vervolg weglaten, aangezien deze als bekend wordt verondersteld.
- 52 Bedoeld wordt de door pl geïntroduceerde optimaliserende kondities, c.q. het 'manipulatie'gedrag van pl tijdens het grenzen-toetsen. Ook deze opmerking zullen wij in het vervolg bekend veronderstellen en daarom weglaten.
- 53 De koëfficiënt van interne consistentie en de homogeniteitskoëfficiënt van Blokpatronen worden - zoals verderop in de tekst wordt uitgelegd - niet vermeld in tabel 29.
- 54 De slotbeschouwing is gedeeltelijk ontleend aan een lezing voor het Congres Ortho '78 van de Vereniging voor Ortho-Agogische Activiteiten O. en A. (Schroots, wordt gepubliceerd).

LITERATUUR

- ALIOTTI, N.C. Covert assessment in psychoeducational testing. *Psychology in the Schools* 14 (1977) 438-43
- ALPHEN DE VEER, R.J. VAN, F.M.E. GERRITSEN, W.L.L. HULSMAN & J.J.F. SCHROOTS. Betekenis en voorspellende waarde van de Nijmeegse Schoolbekwaamheidstest, Frankforter Schoolrijpheidstest en Profieltest. Nijmegen, Berkhout Testmateriaal, 1978
- ALPHEN DE VEER, R.J. VAN, & J.W. PLOMP. De Verhaaltjestest; scoring van Zinnen nazegen. Leiden, NIPG/TNO, 1976
- ANASTASI, A. Differential psychology; individual and group differences in behavior; 3rd ed. New York, MacMillan, 1958
- ANASTASI, A. Psychological testing; 2nd ed. New York, MacMillan, 1961
- ANGOFF, W.H. Scales, norms and equivalent scores. In: R.L. Thorndike (ed.). Educational measurement; 2nd ed. Washington, D.C., American Council on Education, 1971. pp. 508-600
- APOSTEL, L. The cognitive point of view: introduction to a discussion. In: M. de Mey, R. Pinxten, M. Poriau & F. Vandamme (eds). International workshop on the cognitive viewpoint. Gent, University of Ghent, 1977. pp. IX-XV
- ARLMAN - RUPP, A.J.L. Receptief taalgebruik van vijfjarige kinderen. ZWO-project 59001, 1971
- ARTHUR, G. A point scale of performance tests; revised Form II; manual for administering and scoring the tests. New York, Psychological Corporation, 1947
- AUSTIN, J.J., & R. CARPENTER. The use of the WPPSI in early identification of mental retardation and preschool special education. In: R.S. Morrow (Chairman). Diagnostic and educational application of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI); symposium presented at the American Psychological Assoc., Miami, Sept. 1970
- BABAD, E.Y., & M. BUDOFF. Sensitivity and validity of learning-potential measurement in three levels of ability. *J.educ.Psychol.* 66 (1974) 439-47
- BAKKER, D.J. Temporal order, meaningfulness and reading ability. *Percept. Motor Skills* 24 (1967) 1027-30
- BAKKER, D.J. Temporal order in disturbed reading; developmental and neuropsychological aspects in normal and reading-retarded children. Rotterdam, University Press, 1972
- BANNATYNE, M., & A. BANNATYNE. Training van het lichaamsschema door beweging-taal-concentratie. Rotterdam, Lemniscaat, 1973

- 48 Niet alle veertien optimaliserende kondities worden in Mark (1972) vermeld. Dr U. Wälti, kinderpsycholoog te Bern, was echter zo vriendelijk om ons een volledige lijst met kondities (plus toelichting) toe te sturen.
- 49 In paragraaf 9.3 verwezen wij reeds direkt en indirekt naar de vele voorstanders van grenzen-toetsen.
- 50 Overigens met uitzondering van Begrip en inzicht. De scoring van de antwoorden leverde echter dermate grote problemen op dat wij afgezien hebben van analyse en uitwerking der resultaten.
- 51 De opmerking dat bij Blokpatronen geen derde aanbieding heeft plaatsgevonden, zullen wij in het vervolg weglaten, aangezien deze als bekend wordt verondersteld.
- 52 Bedoeld wordt de door pl geïntroduceerde optimaliserende kondities, c.q. het 'manipulatie'gedrag van pl tijdens het grenzen-toetsen. Ook deze opmerking zullen wij in het vervolg bekend veronderstellen en daarom weglaten.
- 53 De koëfficiënt van interne consistentie en de homogeniteitskoëfficiënt van Blokpatronen worden - zoals verderop in de tekst wordt uitgelegd - niet vermeld in tabel 29.
- 54 De slotbeschouwing is gedeeltelijk ontleend aan een lezing voor het Congres Ortho '78 van de Vereniging voor Ortho-Agogische Activiteiten O. en A. (Schroots, wordt gepubliceerd).

- DUMONT, J.J. Leerstoornissen. Rotterdam, Lemniscaat, 1976. 2 dln.
1. Theorie en model
2. Diagnostiek en behandeling
- DYKMAN, R.A., P.T. ACKERMAN, S.D. CLEMENTS & J.E. PETERS. Specific learning disabilities: an attentional deficit syndrome. In: H.R. Myklebust (ed.). Progress in learning disabilities; vol. 2. New York etc., Grune & Stratton, 1971. pp. 56-93
- EBEL, R.L. Measuring educational achievement. Englewood Cliffs, N.J., Prentice Hall, 1965
- EVERITT, B.S. The analysis of contingency tables. London, Chapman & Hall, 1977
- FAUST, M. Cognitive and language factors. J.spec.Educ. 4 (1970) 335-46
- FERGUSON, G.A. On learning and human ability. Can.J.Psychol. 8 (1954) 95-112
- FERGUSON, G.A. On transfer and the abilities of man. Can.J.Psychol. 10 (1956) 121-31
- FEUERSTEIN, R. A dynamic approach to the causation, prevention, and alleviation of retarded performance. In: H.C. Haywood (ed.). Socio-cultural aspects of mental retardation. New York, Appleton-Century-Crofts, 1970. pp. 341-77
- FEUERSTEIN, R. Cognitive assessment of the socioculturally deprived child and adolescent. In: L.J. Cronbach & P.J.D. Drenth (eds). Mental tests and cultural adaptation. The Hague, Mouton, 1972a. pp. 265-75
- FEUERSTEIN, R. Studies in cognitive modifiability; the dynamic assessment of retarded performance; vol. 1. Clinical LPAD battery. Jerusalem, Hadassah WIZO Canada Research Institute, 1972b
- FEUERSTEIN, R., & D. KRASILOWSKY. Interventional strategies for the significant modification of cognitive functioning in the disadvantaged adolescent. J.Amer.Acad.Child Psychiatry 11 (1972) 572-82
- FIENBERG, S.E. The analysis of multi-dimensional contingency tables. Ecology 51 (1970) 419-33
- FIENBERG, S.E. The analysis of cross-classified categorical data. Cambridge, Mass., MIT Press, 1977
- FLOWER, R. The evaluation of auditory abilities in the appraisal of children with reading problems. In: A. Figurell (ed.). Perception and reading. Newark, I.R.A., 1968
- FRENCH, J.W., R.B. EKSTROM & L.A. PRICE. Manual for Kit of Reference Tests for Cognitive Factors; rev. ed. Princeton, N.J., Educational Testing Service, 1963
- FROSTIG, M. Frostig Visual Development Program. Chicago, Follett Publ., 1964

- COSTER, W. DE, I. MERVIELDE, R. PARMENTIER & A. VANDIERENDONCK. Cognition, cognitive science and psychology. In: M. de Mey, R. Pinxten, M. Poriau & F. Vandamme (eds). International workshop on the cognitive viewpoint. Gent, University of Ghent, 1977. pp. 229-38
- CRONBACH, L.J. Essentials of psychological testing; 3rd ed. New York etc., Harper & Row, 1970
- CRONBACH, L.J. Five decades of public controversy over mental testing. Amer.Psychol. 30 (1975a) 1-14
- CRONBACH, L.J. Beyond the two disciplines of scientific psychology. Amer. Psychol. 30 (1975b) 116-27
- CRONBACH, L.J. On the design of educational measures. In: D.N.M. de Gruyter & L.J.Th. van der Kamp (eds). Advances in psychological and educational measurement. London etc., Wiley, 1976. pp. 199-208
- CRONBACH, L.J., & L. FURBY. How we should measure "change" - or should we? Psychol.Bull. 74 (1970) 68-80
- CRONBACH, L.J., & G.C. GLESER. Psychological tests and personnel decisions; 2nd rev. ed. Urbana, Ill., University of Illinois Press, 1965
- CRONBACH, L.J., G.C. GLESER, N. HARINDER & N. RAJARATNAM. The dependability of behavioral measurements. New York, Wiley, 1972
- CRONBACH, L.J., & R.E. SNOW. Aptitudes and instructional methods; a handbook for research on interactions. New York, Irvington, 1977
- DAVIE, R. Children at increased educational risk; some results and some reservations. In: K. Wedell & E.C. Raybould (eds). The identification of educationally 'at risk' children. Birmingham, School of Education, University of Birmingham, 1976. pp. 30-5
- DELACATO, C.H. The diagnosis and treatment of speech and reading problems. Springfield, Ill., Thomas, 1963
- DIXON, W.J., & F.J. MASSEY. Introduction to statistical analysis; 2nd ed. New York etc., McGraw-Hill, 1957
- DRENTH, P.J.D. Protesten contra testen. Amsterdam, Swets & Zeitlinger, 1967.
Inaugurele rede Amsterdam VU
- DRENTH, P.J.D. Inleiding in de testtheorie. Deventer, Van Loghum Slaterus, 1975
- DRENTH, P.J.D., J.F. PETRIE & N. BLEICHRODT. Handleiding bij de Amsterdamse Kinder Intelligentie Test (AKIT). Amsterdam, Swets & Zeitlinger, 1968

- DUMONT, J.J. Leerstoornissen. Rotterdam, Lemniscaat, 1976. 2 dln.
1. Theorie en model
2. Diagnostiek en behandeling
- DYKMAN, R.A., P.T. ACKERMAN, S.D. CLEMENTS & J.E. PETERS. Specific learning disabilities: an attentional deficit syndrome. In: H.R. Myklebust (ed.). Progress in learning disabilities; vol. 2. New York etc., Grune & Stratton, 1971. pp. 56-93
- EBEL, R.L. Measuring educational achievement. Englewood Cliffs, N.J., Prentice Hall, 1965
- EVERITT, B.S. The analysis of contingency tables. London, Chapman & Hall, 1977
- FAUST, M. Cognitive and language factors. J.spec.Educ. 4 (1970) 335-46
- FERGUSON, G.A. On learning and human ability. Can.J.Psychol. 8 (1954) 95-112
- FERGUSON, G.A. On transfer and the abilities of man. Can.J.Psychol. 10 (1956) 121-31
- FEUERSTEIN, R. A dynamic approach to the causation, prevention, and alleviation of retarded performance. In: H.C. Haywood (ed.). Socio-cultural aspects of mental retardation. New York, Appleton-Century-Crofts, 1970. pp. 341-77
- FEUERSTEIN, R. Cognitive assessment of the socioculturally deprived child and adolescent. In: L.J. Cronbach & P.J.D. Drenth (eds). Mental tests and cultural adaptation. The Hague, Mouton, 1972a. pp. 265-75
- FEUERSTEIN, R. Studies in cognitive modifiability; the dynamic assessment of retarded performance; vol. 1. Clinical LPAD battery. Jerusalem, Hadassah WIZO Canada Research Institute, 1972b
- FEUERSTEIN, R., & D. KRASILOWSKY. Interventional strategies for the significant modification of cognitive functioning in the disadvantaged adolescent. J.Amer.Acad.Child Psychiatry 11 (1972) 572-82
- FIENBERG, S.E. The analysis of multi-dimensional contingency tables. Ecology 51 (1970) 419-33
- FIENBERG, S.E. The analysis of cross-classified categorical data. Cambridge, Mass., MIT Press, 1977
- FLOWER, R. The evaluation of auditory abilities in the appraisal of children with reading problems. In: A. Figurell (ed.). Perception and reading. Newark, I.R.A., 1968
- FRENCH, J.W., R.B. EKSTROM & L.A. PRICE. Manual for Kit of Reference Tests for Cognitive Factors; rev. ed. Princeton, N.J., Educational Testing Service, 1963
- FROSTIG, M. Frostig Visual Development Program. Chicago, Follett Publ., 1964

- COSTER, W. DE, I. MERVIELDE, R. PARMENTIER & A. VANDIERENDONCK. Cognition, cognitive science and psychology. In: M. de Mey, R. Pinxten, M. Poriau & F. Vandamme (eds). International workshop on the cognitive viewpoint. Gent, University of Ghent, 1977. pp. 229-38
- CRONBACH, L.J. Essentials of psychological testing; 3rd ed. New York etc., Harper & Row, 1970
- CRONBACH, L.J. Five decades of public controversy over mental testing. Amer.Psychol. 30 (1975a) 1-14
- CRONBACH, L.J. Beyond the two disciplines of scientific psychology. Amer. Psychol. 30 (1975b) 116-27
- CRONBACH, L.J. On the design of educational measures. In: D.N.M. de Gruyter & L.J.Th. van der Kamp (eds). Advances in psychological and educational measurement. London etc., Wiley, 1976. pp. 199-208
- CRONBACH, L.J., & L. FURBY. How we should measure "change" - or should we? Psychol.Bull. 74 (1970) 68-80
- CRONBACH, L.J., & G.C. GLESER. Psychological tests and personnel decisions; 2nd rev. ed. Urbana, Ill., University of Illinois Press, 1965
- CRONBACH, L.J., G.C. GLESER, N. HARINDER & N. RAJARATNAM. The dependability of behavioral measurements. New York, Wiley, 1972
- CRONBACH, L.J., & R.E. SNOW. Aptitudes and instructional methods; a handbook for research on interactions. New York, Irvington, 1977
- DAVIE, R. Children at increased educational risk; some results and some reservations. In: K. Wedell & E.C. Raybould (eds). The identification of educationally 'at risk' children. Birmingham, School of Education, University of Birmingham, 1976. pp. 30-5
- DELACATO, C.H. The diagnosis and treatment of speech and reading problems. Springfield, Ill., Thomas, 1963
- DIXON, W.J., & F.J. MASSEY. Introduction to statistical analysis; 2nd ed. New York etc., McGraw-Hill, 1957
- DRENTH, P.J.D. Protesten contra testen. Amsterdam, Swets & Zeitlinger, 1967.
Inaugurele rede Amsterdam VU
- DRENTH, P.J.D. Inleiding in de testtheorie. Deventer, Van Loghum Slaterus, 1975
- DRENTH, P.J.D., J.F. PETRIE & N. BLEICHRODT. Handleiding bij de Amsterdamse Kinder Intelligentie Test (AKIT). Amsterdam, Swets & Zeitlinger, 1968

- GROOT, A.D. DE. Vijven en zessen; cijfers en beslissingen: het selectie-proces in ons onderwijs. Groningen, Wolters, 1967
- GRUYTER, D.N.M. DE, & L.J.Th. VAN DER KAMP (eds). Advances in psychological and educational measurement. London etc., Wiley, 1976
- GUILFORD, J.P. The nature of human intelligence. New York etc., McGraw-Hill, 1967
- GULLIKSEN, H. Looking back and ahead in psychometrics. Amer.Psychol. 29 (1974) 251-61
- GUTHRIE, J.T., & H.K. GOLDBERG. Visual sequential memory in reading disability. J.Learn.Disabil. 5 (1972) 41-6
- GUTTMAN, L. A new approach to factor analysis: the radex. In: P. Lazarsfeld (ed.). Mathematical thinking in the social sciences. Glencoe, Ill., The Free Press, 1954
- HALBERSTADT-FREUD, H.C. Psychodiagnostiek met kinderen. Amsterdam, Swets & Zeitlinger, 1973
- HARMAN, H.H. Modern factor analysis; 2nd rev.ed. Chicago etc., University of Chicago Press, 1967
- HARRIS, C.W. (ed.). Problems in measuring change. Madison, University of Wisconsin Press, 1963
- HAYWOOD, H.C., J.W. FILLER, M.A. SHIFMAN & G. CHATELANAT. Behavioral assessment in mental retardation. In: P. McReynolds (ed.). Advances in psychological measurement; vol. 3. San Francisco, Jossey-Bass, 1975. pp. 96-136
- HEESEN, H., D. STRELITSKI & A. VAN DER WISSEL. Handleiding voor de Schiedamse Rekentest (S.R.T.). Groningen, Wolters-Noordhoff, 1969
- HEIDEN, L. VAN DER. De Verhaaltjestest; het gebruik van vereenvoudigingsmiddelen in een standaardsituatie. Amsterdam, Vrije Universiteit, 1976.
Doktoraal skriptie VU
- HERMAN, D.O. A study of sex differences in the Wechsler Preschool and Primary Scale of Intelligence. Proceedings of the 76th Annual Convention of the American Psychological Association 3 (1968) 455-6
- HISKEY, M.S. Manual for the Hiskey-Nebraska Test of Learning Aptitude (HNTLA). Lincoln, Nebr., Union College Press, 1966
- HOOGSTRATEN, J., & G.J. MELLENBORG. Relevante variabelen bij het doorverwijzen na de lagere school; een experiment. T.Onderwijs-research 3 (1978) 161-72
- HUNT, E. What kind of computer is man? Cognitive Psychol. 2 (1971) 57-98
- JENSEN, A.R. How much can we boost IQ and scholastic achievement? Harvard Educ.Rev. 39 (1969) 1-123

- JENSEN, A.R. Educability and group differences. New York, Harper & Row, 1973
- JOHN, E.R., et al. Neurometrics. Science 196 (1977) 1393-1410
- JONES, M.B. A two-process theory of individual differences in motor learning. Psychol.Rev. 77 (1970) 353-60
- KALVERBOER, A.F. A neurobehavioral study in preschool children. London, Heinemann, 1975
- KALVERBOER, A.F., & H.J. SANDERS-ALLERSMA. Een Profieltest voor de ruimtelijk-constructieve ontwikkeling. In: Klinische psychologie in Nederland; dl. 2. Deventer, Van Loghum Slaterus, 1973. pp. 65-98
- KENDLER, H.H., & T.S. KENDLER. Vertical and horizontal processes in problem solving. Psychol.Rev. 69 (1962) 1-16
- KEOGH, B.K. Hyperactivity and learning disorders; review and speculation. Except.Child. 38 (1971) 101-9
- KEOGH, B.K., & L.D. BECKER. Early detection of learning problems; questions, cautions and guidelines. Los Angeles, University of California, 1972. (Technical Report)
- KEOGH, B.K., R.J. HALL & L.D. BECKER. Early identification of exceptional children for educational programming. Los Angeles, Graduate School of Education, University of California, 1974. (Technical Report SERP 1974-A6)
- KEPHART, N.C. The slow learner in the classroom. Englewood Cliffs, N.J., Merrill, 1960
- KERLINGER, F.N., & E.J. PEDHAZUR. Multiple regression in behavioral research. New York etc., Holt, Rinehart & Winston, 1973
- KESSLER, K. Health and educational screening of school-age children; definition and objectives. Amer.J.publ.Hlth 62 (1972) 191-8
- KING, J.D., & R.A. SMITH. Abbreviated forms of the Wechsler Preschool and Primary Scale of Intelligence for a kindergarten population. Psychol.Reports 30 (1972) 539-42
- KIRK, R.E. Experimental design; procedures for the behavioral sciences. Belmont, Calif., Brooks/Cole Publ., 1968
- KIRK, S.A. The diagnosis and remediation of psycholinguistic disabilities. Urbana, Ill., Institute for Research on Exceptional Children, University of Illinois, 1966
- KIRK, S.A., J.J. MCCARTHY & W.D. KIRK. Examiner's manual; Illinois Test of Psycholinguistic Abilities (ITPA); rev. ed. Urbana, Ill., University of Illinois Press, 1968
- KLOPFER, B., & D.M. KELLEY. The Rorschach technique. New York, Yonkers World Book Company, 1942

- KLUEVER, R. Mental abilities and disorders of learning. In: H.R. Myklebust (ed.). Progress in learning disabilities; vol. 2. New York etc., Grune & Stratton, 1971. pp. 196-212
- KOHNSTAMM, G.A. Over de AKIT, de WPPSI, en de UTANT, drie nieuwe tests voor 4- tot 7-jarigen. Ned.T.Psychol. 23 (1968) 269-90
- KOHNSTAMM, G.A. Geslachtsverschillen in prestaties op schoolvordering - toetsen en enkele tests aan het einde van de basisschool. Ned.T.Psychol. 28 (1973) 351-67
- KOHNSTAMM, G.A., A.P. MESSER & A.K. DE VRIES. De Utrechtse Taalniveau Test voor 4-7 jarigen (UTANT 4-7). Amsterdam, Swets & Zeitlinger, 1971
- KORNMAN, R. Diagnose der Testbedingungen - ein Ansatz zur Erklärung schwacher Testleistungen und zur Indikation von Förderungsmaßnahmen; lezing voor het Congres Ortho '78 van de Ver. voor Ortho-Agogische Activiteiten O. & A., 1978. In: L. Stevens et al. Congresboek Ortho '78. Groningen, Wolters-Noordhoff (wordt gepubliceerd)
- KRAPP, A., & H. MANDL. Schulreifetest und Schulerfolg. München, Oldenbourg, 1971
- KRATOCHWILL, T.R. The movement of psychological extras into ability assessment. J.spec.Educ. 11 (1977) 299-311
- KWAAITAAL, T., & E. ROSKAM. Analysis of variance psytab varian/02. Nijmegen, Nijmegen University, 1968
- LAGAE, C., & L. COETSLER. Wechsler Intelligence Scale for Children (WISC); handleiding bij de aanpassing voor Vlaamse kinderen. Gent, Rijksuniversiteit Gent, 1963
- LAWTON, D. Social class, language and education. London, Routledge & Kegan Paul, 1969
- LEZAK, M.D. Neuropsychological assessment. New York, Oxford University Press, 1976
- LIENERT, G.A. Testaufbau und Testanalyse; 2e Aufl. Weinheim, Beltz, 1961
- LINDEN, F.J. VAN DER. Intelligentie en schoolprestaties van jonge schoolkinderen; faktoranalyse van testprestaties van leerlingen uit de eerste klas van het basisonderwijs. Ned.T.Psychol. 25 (1970) 278-301
- LURIA, A.R. The role of speech in the regulation of normal and abnormal behavior; ed. by J. Tizard. New York, Liveright, 1961
- LIJF, R.T.P. VAN. Beschrijving van Nederlandstalige toetsen voor kleuteren basisonderwijs. Arnhem, CITO, 1971
- MACCOBY, E.E., & C.N. JACKLIN. The psychology of sex differences. Stanford, Calif., Stanford University Press; London, Oxford University Press, 1975

- McNEMAR, Q. The revision of the Stanford-Binet Scale. Boston, Houghton-Mifflin, 1942
- MADAUS, G.F., E.M. WOODS & R.L. NUTTALL. A causal model analysis of Bloom's taxonomy. *Amer.Educ.Research J.* 10 (1973) 253-62
- MANN, L. Psychometric phrenology and the new faculty psychology: the case against ability assessment and training. *J.spec.Educ.* 5 (1971) 3-14
- MARK, H.J. Elementary thinking and the classification of behavior. *Science* 135 (1962) 75-87
- MARK, H.J. Psychodiagnostics in patients with suspected minimal brain dysfunctions (MBD). In: *Minimal brain dysfunction in children*. Bethesda, Md., US. Dept. of Health, Education and Welfare, 1969. (PHS publication no. 2015, N and SDCP Monograph)
- MARK, H.J. Psychodiagnostics in brain dysfunctions; a seminar report; ed. by J.J.F. Schrooters & S.D. Hecht-Ponger. Leiden, NIPG/TNO, 1972
- MARK, H.J., & S.A. MARK. Neuropsychology examinations in young children. In: J. Hellmuth (ed.). *Exceptional infant; vol. 2. Studies in abnormalities*. New York, Brunner/Mazel, 1971
- MASSEY, J.O. WPPSI test profile. Palo Alto, Calif., Consulting Psychologists Press, 1968
- MATHENEY, A.P., A.M. BROWN & R.S. WILSON. Rotations on WPPSI geometric and block design subtests at four, five and six years. *J. Psychol.* 82 (1972) 217-21
- MELLENBERGH, G.J. Bekend, maar onbemand; dl. 1. Amsterdam, Universiteit van Amsterdam, 1976
- MELLENBERGH, G.J., & J. PANNEKOEK. The moderator effect reconsidered. Amsterdam, Universiteit van Amsterdam, 1979
- MEY, M. DE. The cognitive viewpoint: its development and its scope. In: M. de Mey, R. Pinxten, M. Poriau & F. Vandamme (eds). *International workshop on the cognitive viewpoint*. Gent, University of Ghent, 1977. pp. XVI-XXXII
- MICHON, J.A., E.G.J. EIJKMAN & L.F.W. DE KLERK. Inleiding. In: J.A. Michon, E.G.J. Eijkmán & L.F.W. de Klerk (red.). *Handboek der psychonomie*. Deventer, Van Loosdrecht, 1976. pp. 3-13
- MILLIREN, A.P., & T.E. NEWLAND. Statistically significant differences between subtest scaled scores for the WPPSI. *J.School Psychol.* 7 (1968/1969) 16-9
- MITTLER, P. The psychological assessment of mental and physical handicaps. London, Methuen, 1970

- MÖNKES, F.J., H. ROST & N.H. COFFIE. Nijmeegse Schoolbekwaamheidstest (NST); monografie en handleiding. Nijmegen, Nederlandse Universiteitspers, 1969
- MORIARTY, A. Coping patterns of preschool children in response to intelligence test demands. *Genetic Psychol. Monographs* 64 (1961) 3-128
- MYKLEBUST, H.R. Learning disabilities; definition and overview. In: H.R. Myklebust (ed.). *Progress in learning disabilities*; vol. 1. New York, Grune & Stratton, 1968
- MYKLEBUST, H.R., & B. BOSCHES. Minimal brain damage in children; final report. Washington, D.C., Department of Health, Education and Welfare, 1969
- NARROL, H., & D.G. BACHOR. An introduction to Feuerstein's approach to assessing and developing cognitive potential. *Interchange* 6 (1975) 2-16
- NATIONAL ADVISORY COMMITTEE ON HANDICAPPED CHILDREN. Washington, D.C., U.S. Office of Education, 1967
- NELISSEN, J., N. VERLOOP & M. ZWARTS. Intelligentie en rekenen; pleidooi voor een meer procesmatige benadering van het intelligentiebegrip. *Pedagog. Studiën* 55 (1978) 413-26
- NEWELL, A. Production systems: models of control structures. In: W.G. Chase (ed.). *Visual information processing*. New York etc., Academic Press, 1973. pp. 463-526
- NEWELL, A., J.C. SHAW & H.A. SIMON. Elements of a theory of human problem solving. *Psychol. Rev.* 65 (1958) 151-66
- NIE, N.H., C.H. HULL, J.G. JENKINS, K. STEINBRENNER & D.H. BRENT. Statistical package for the social sciences (SPSS); 2nd ed. New York etc., McGraw-Hill, 1975
- NOTA SPECIAAL ONDERWIJS. Den Haag, Staatsuitgeverij, 1977
- NUNNALLY, J.C. Psychometric theory. New York etc., McGraw-Hill, 1967
- ORPET, R.E., & C.E. MEYERS. Six structure-of-intellect hypotheses in six-year-old children. *J. educ. Psychol.* 57 (1966) 341-6
- PIAGET, J. The psychology of intelligence. Paterson, N.J., Littlefield Adams, 1960
- PHS, PUBLIC HEALTH SERVICE. Minimal brain dysfunction in children; terminology and identification; phase one of a three-phase project. Washington, D.C., U.S. Government Printing Office, 1966.
(PHS publication no. 1415, NINDB Monograph; 3)
- PHS, PUBLIC HEALTH SERVICE. Minimal brain dysfunction in children. Educational, medical and health related services; phase two of a three-phase project. Washington, D.C., U.S. Government Printing Office, 1969.
(PHS publication no. 2015, N and SDCP Monograph)

- PRINSEN, J.C.C., & J.A. KROPMAN. De nederlandse sportvisser. Nijmegen, Instituut voor Toegepaste Sociologie, 1975
- RIGSBY, M.B. Rates of reinforcement in testing and classroom settings. Clemson, South Carolina, Clemson University, 1972.
Unpublished master's thesis
- RISPENS, J. Auditieve aspecten van leesmoeilijkheden; een onderzoek naar de relatie tussen auditieve diskriminatie, auditieve analyse, auditieve synthese en leesmoeilijkheden. Utrecht, Elinkwijk, 1974
Proefschrift Utrecht
- RISPENS, J. Functietrainingsprogramma's: ondanks alles ... toch maar doen? Enkele opmerkingen over effectiviteit en uitgangspunten van functietrainingsprogramma's. Pedagog.Studiën 54 (1977) 98-109
- ROSENTHAL, R. Teacher's expectancies: determinants of pupils IQ gains. Psychol.Reports 19 (1966) 115-8
- RUPP, J.C.C. Opvoeding tot schoolweerbaarheid; een Utrechtse kinderpsychologische studie. Utrecht, Wolters-Noordhoff, 1969.
Proefschrift
- RUSCHVAL, M.L., & J.G. WAY. The WPPSI and the Stanford-Binet: a validity and reliability study using gifted preschool children. J. Consult.clin.Psychol. 37 (1971) 163
- SANDBERGEN, S. "Testslimheid", de invloed van oefening en coaching op test scores. Ned.T.Psychol. 23 (1968) 502-29
- SATTLER, J.M. Assessment of children's intelligence; rev. repr. Philadelphia etc., Saunders, 1974.
(1st ed. 1972)
- SCHROOTS, J.J.F. Prediktie, inferentie en decisie; een evaluatie van de neuropsychologische test. Ned.T.Psychol. 27 (1972) 309-40
- SCHROOTS, J.J.F. Handleiding voor het standaard- en grenzentoetsend onderzoek van schoolverlatende kleuters (experimentele versie). Leiden, NIPG/TNO, 1973.
Interim-rapport 3
- SCHROOTS, J.J.F. Some issues in the early detection of learning disabilities. In: K. Wedell & E.C. Raybould (eds). The early identification of educationally 'at risk' children. Birmingham, School of Education, University of Birmingham, 1976.
(Educ.Rev. no. 6)
- SCHROOTS, J.J.F. De bijdrage van de psychologische test aan de diagnostiek van leerstoornissen. In: L. Stevens et al. Congresboek Ortho '78. Groningen, Wolters-Noordhoff (wordt gepubliceerd)
- SCHROOTS, J.J.F., & R.J. VAN ALPHEN DE VEER. Leidse Diagnostische Test; konstruktie en psychometrische verkenning. Leiden, NIPG/TNO, 1974.
Interim-rapport 8

- SCHROOTS, J.J.F., & R.J. VAN ALPHEN DE VEER. Leidse Diagnostische Test (LDT). Dl. 1: Handleiding. Amsterdam, Swets & Zeitlinger, 1976
- SCHROOTS, J.J.F., A.E. AKKERMAN & A. DE GROOT. Leidse Diagnostische Test (LDT). Dl. 2: Handleiding voor het stroomdiagram; van klinisch oordeel tot psychologisch rapport. Lisse, Swets & Zeitlinger, 1978a
- SCHROOTS, J.J.F., A.E. AKKERMAN & A. DE GROOT. Leidse Diagnostische Test (LDT). Dl. 3: Stroomdiagram. Lisse, Swets & Zeitlinger, 1978b
- SCHROOTS, J.J.F., A.E. AKKERMAN & A. DE GROOT. Leidse Diagnostische Test (LDT). Dl. 4: Uitspraken. Lisse, Swets & Zeitlinger, 1978c
- SCHROOTS, J.J.F., D.J. BAKKER, R.J. VAN ALPHEN DE VEER & J.H.A. GROENENDAAL. Temporele orde en het lees-leerproces. Ned.T.Psychol. 30 (1975) 337-61
- SCHUCMAN, H. Evaluating the educability of the severely mentally retarded child. Psychol.Monographs 74 (1960) 1-32
- SEWELL, T.E., & R.A. SEVERSON. Learning ability and intelligence as cognitive predictors of achievement in first-grade black children. J.educ.Psychol. 66 (1974) 948-55
- SHAPIRO, M.B. An experimental approach to diagnostic psychological testing. J.ment.Sci. 97 (1951) 748-64
- SIGEL, I.E., A. SECRIST & G. FORMAN. Psycho-educational intervention beginning at age two: reflections and outcomes. In: J.C. Stanley (ed.). Compensatory education for children, ages 2 to 8. Baltimore, Johns Hopkins University Press, 1973
- SILVERSTEIN, A.B. A short form of Wechsler's scales for screening purposes. Psychol.Reports 21 (1967) 842
- SIXMA, J. Leesvoorwaardentoets. Groningen, Wolters-Noordhoff, 1971
- SKINNER, B.F. The technology of teaching. New York, Appleton-Century-Crofts, 1968
- SKINNER, B.F. Cumulative record: a selection of papers; 3rd ed. New York, Appleton-Century-Crofts, 1972
- SMEAD, V.S. Ability training and task analysis in diagnostic/prescriptive teaching. J.spec.Educ. 11 (1977) 113-25
- SNIJDERS, J.Th., & N. SNIJDERS-OOMEN. Snijders-Oomen Niet-Verbale Intelligentieschaal (S.O.N.). Groningen, Wolters, 1958
- SNIJDERS, J.Th., & N. SNIJDERS-OOMEN. Snijders-Oomen Niet-Verbale Intelligentieschaal S.O.N. 2½-7. Groningen, Tjeenk Willink, 1975
- SOUDIJN, K.A. Cognitie: notities over een Gentse conferentie. De Psycholoog 12 (1977) 423-8

- SOUDIJN, K.A. Heeft de psychotherapeut een achilleshiel? De Psycholoog 14 (1979) 117-27
- SPPD HILVERSUM. Hilversumse Begrippentoets Rekenen; Instructie. Hilversum, Stichting Schoolpedagogische en -Psychologische Dienst Hilversum, 1973
- STARREN, J.C.M. De ontwikkeling van een nieuwe versie van de S.O.N. voor 7-17 jarigen. Groningen, Rijksuniversiteit, 1973. Proefschrift
- STEVENS, L.M. Curriculum Schoolrijpheid III; een evaluatie-onderzoek. 's-Hertogenbosch, Malmberg, 1975
- STEVENS, L.M. The infant teacher and developmental assessment. J.Assoc. educ.Psychol. 4 (1978) 43-9
- STEVENSON, H.W., Th. PARKER, A. WILKINSON, A. HEGION & E. FISH. Longitudinal study of individual differences in cognitive development and scholastic achievement. J.educ.Psychol. 68 (1976) 377-400.
- STINISSEN, J., P.J. WILLEMS, L. COETSIER & W.L.L. HULSMAN. Handleiding Nederlandse bewerking Wechsler Adult Intelligence Scale (WAIS). Amsterdam, Swets & Zeitlinger, 1969
- STOTT, L.H., & R.S. BALL. Infant and preschool mental tests; review and evaluation. Monographs Society for Research in Child Develop. 30 (1965) no. 3
- STRIEN, P.J. VAN. Van psychotechniek tot maatschappijkritiek; drie decaden praktijkbeschouwing. De Psycholoog 4 (1969) 351-9
- SWANSON, L. Verbal encoding effects on the visual short-term memory of learning disabled and normal readers. J.educ.Psychol. 40 (1978) 539-44
- TAYLOR, E.M. Psychological appraisal of children with cerebral defects. Cambridge, Mass., Harvard University Press, 1959
- TERMAN, L.M., & M.A. MERRILL. Stanford-Binet Intelligence Scale; manual for the third revision Form L-M. Boston, Houghton Mifflin, 1960
- THOMASSEN, A.J.W.M., & G.A.M. KEMPEN. Geheugen. In: J.A. Michon, E.G.J. Eijkman & L.F.W. de Klerk (red.). Handboek der psychonomie. Deventer, Van Loghum Slaterus, 1976. pp. 354-87
- THURSTONE, L.L. Primary mental abilities. Psychometr.Monogr. no. 1, 1938
- THURSTONE, L.L., & T.G. THURSTONE. Technical supplement for the SRA Primary Mental Abilities for age 5 to 7. Chicago, Ill., University of Chicago Press, 1953a
- THURSTONE, L.L., & T.G. THURSTONE. Examiner manual for the SRA Primary Mental Abilities for age 5 to 7. Chicago, Ill., University of Chicago Press, 1953b. (vert. Pedagogisch Centrum, Enschede)

- TORGESON, J.K., Ch. BOWEN & Ch. IVEY. Task structure versus modality of presentation: a study of the construct validity of the visual-aural digit span test. *J.educ.Psychol.* 70 (1978) 451-6
- VAUGHAN, R.W., & L. HODGES. A statistical survey into a definition of learning disabilities; a search for acceptance. *J.Learn.Disabil.* 6 (1973) 68-74
- VENGER, L.A. De kwalitatieve benadering in de diagnostiek van de cognitieve ontwikkeling. *Pedagog.Studiën* 52 (1975) 406-11
- VERNON, P.E. Intelligence and cultural environment. London, Methuen, 1969
- VIGOTSKY, L.S. Thought and language. New York, Wiley, 1962
- VOLLE, F.O. A proposal for "testing the limits" with mental defectives for purposes of subtest analysis of the WISC verbal scale. *J.clin.Psychol.* 13 (1957) 64-7
- WATERINK, J. Theorie der opvoeding. Kampen, Kok, 1951
- WECHSLER, D. Manual for the Wechsler Intelligence Scale for Children (WISC). New York, Psychological Corporation, 1949
- WECHSLER, D. The measurement and appraisal of adult intelligence; 4th ed. Baltimore, Williams & Wilkins, 1958
- WECHSLER, D. Manual for the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). New York, Psychological Corporation, 1967
- WECHSLER, D. Manual for the Wechsler Intelligence Scale for Children-Revised (WISC-R). New York, Psychological Corporation, 1974
- WESTERLAAK, J.M. VAN, J.A. KROPAN & J.W.M. COLLARIS. Beroepenklapper. Nijmegen, Instituut voor Toegepaste Sociologie, 1975
- WHITE, S.H. Evidence for a hierarchical arrangement of learning processes. *Adv.Child Develop. & Behav.* 2 (1965) 187-220
- WHITELY, S.E., & R.V. DAWIS. Effects of cognitive intervention on latent ability measured from analogy items. *J.educ.Psychol.* 66 (1974) 710-7
- WHITELY, S.E., & R.V. DAWIS. A model for psychometrically distinguishing aptitude from ability. *Educ.Psychol.Meas.* 35 (1975) 51-66
- WOODROW, H. The relation between abilities and improvement with practice. *J.educ.Psychol.* 29 (1938) 215-30
- WOODROW, H. Factors in improvement with practice. *J.Psychol.* 7 (1939) 55-70
- YSSELDYKE, J.E., & J. SALVIA. Diagnostic-prescriptive teaching: two models. *Except.Child.* 41 (1974) 181-5
- YULE, W., M. BERGER, S. BUTLER, V. NEWHAM & J. TIZARD. The WPPSI: an empirical evaluation with a British sample. *Brit.J.educ.Psychol.* 39 (1969) 1-13

- ZAAL, C.T. De Verhaaltjestest; modifikaties en empirisch onderzoek. Amsterdam, Vrije Universiteit, 1976.
Skriptie
- ZIGLER, E., & E.C. BUTTERFIELD. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Develop. 39 (1968) 1-14
- ZWARTS, M.A. Een onderzoek naar de differentiatie van het aanvankelijk lezen binnen het kader van de schoolbegeleiding; interim-verslag 6: Instrumenten en technieken (dl. 1); Figuren, schema's, tabellen en bijlagen (dl. 2). Utrecht, Instituut voor Pedagogische en Andragogische Wetenschappen, Vakgroep Onderwijskunde, 1976.
(DAL-project)

BIJLAGE

	pag.
Tabel 12. Korrelaties tussen de tests	256
Tabel 15. Korrelaties tussen de tests, naar leeftijd	257
Tabel 17. Korrelaties tussen de tests, naar geslacht	258
Tabel 19. Korrelaties tussen de tests, naar sociaal- ekonomisch niveau	259
Tabel 21. Korrelaties tussen de LDT-E, FST en NST	260
Tabel 34. Vierweg kruistabel-gegevens voor Blokpatronen- Lezen/taal	262
Tabel 35. Vierweg kruistabel-gegevens voor Vouwblaadjes- Lezen/taal	263
Tabel 36. Vierweg kruistabel-gegevens voor Natikken- Lezen/taal	264
Tabel 37. Vierweg kruistabel-gegevens voor Woordenspan- Lezen/taal	265

Tabel 12. Korrelaties tussen de tests

N = 392

tests	Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span	Plaatjes aanwijzen	Zinnen nazeggen	Verhaaltje vragen
Vouwblaadjes	.30						
Natikken	.37	.31					
Woordenspan	.20	.15	.36				
Plaatjes aanwijzen	.29	.30	.46	.63			
Zinnen nazeggen	.24	.32	.39	.55	.55		
Verhaaltje vragen	.22	.26	.31	.27	.34	.54	
Begrip en inzicht	.15	.16	.29	.27	.34	.55	.45

Tabel 15. Korrelaties tussen de tests, naar leeftijd^a

N₁ = 219; N₂ = 173

tests		Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span	Plaatjes aanwijzen	Zinnen nazeggen	Verhaaltje vragen	Begrip en inzicht
Vouwblaadjes	1 ^a	.26							
	2	.29							
Natikken	1	.30	.25						
	2	.38	.33						
Woordenspan	1	.14	.10	.32					
	2	.24	.18	.37					
Plaatjes aanwijzen	1	.22	.25	.38	.57				
	2	.32	.30	.49	.67				
Zinnen nazeggen	1	.14	.32	.31	.53	.49			
	2	.35	.28	.45	.56	.59			
Verhaaltje vragen	1	.21	.34	.33	.26	.36	.56		
	2	.21	.14	.28	.26	.31	.51		
Begrip en inzicht	1	.04	.13	.22	.19	.25	.50	.39	
	2	.21	.12	.29	.34	.39	.58	.52	
gemiddelde (m)	1	15.90	7.97	5.42	6.86	4.03	43.90	15.28	15.62
	2	17.02	8.66	6.73	7.28	4.94	45.84	16.23	17.36
standaarddeviatie (s)	1	2.53	1.97	2.41	2.00	2.16	8.48	6.63	4.27
	2	2.08	1.90	2.47	2.04	2.43	8.52	6.90	4.31

^a 1 = jongste leeftijdsgroep (5;05 - 5;10); 2 = oudste leeftijdsgroep (5;11 - 6;04)

Tabel 17. Korrelaties tussen de tests, naar geslacht^aN_j = 199; N_m = 193

tests		Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span	Plaatjes aanwijzen	Zinnen nazeggen	Verhaaltje vragen	Begrip en inzicht
Vouwblaadjes	j ^a	.36							
	m	.26							
Natikken	j	.44	.31						
	m	.30	.33						
Woordenspan	j	.17	.12	.28					
	m	.22	.16	.43					
Plaatjes aanwijzen	j	.27	.29	.34	.62				
	m	.32	.29	.57	.63				
Zinnen nazeggen	j	.25	.32	.33	.58	.56			
	m	.23	.30	.44	.52	.53			
Verhaaltje vragen	j	.25	.31	.29	.30	.38	.54		
	m	.19	.21	.33	.23	.30	.54		
Begrip en inzicht	j	.13	.10	.16	.28	.33	.55	.41	
	m	.16	.28	.43	.28	.37	.57	.51	
gemiddelde (m)	j	16.44	7.85	5.99	6.95	4.25	44.11	15.58	16.85
	m	16.35	8.71	6.01	7.15	4.62	45.42	15.81	15.92
standaarddeviatie (s)	j	2.35	1.96	2.44	2.05	2.34	8.42	6.61	4.44
	m	2.47	1.88	2.60	2.00	2.31	8.63	6.92	4.26

^a j = jongen; m = meisje

Tabel 19. Korrelaties tussen de tests, naar sociaal-ekonomisch niveau^a

N₁ = 94; N₂ = 112; N₃ = 205

tests		Blok- patronen	Vouw- blaadjes	Natikken	Woorden- span	Plaatjes aanwijzen	Zinnen nazeggen	Verhaaltje vragen	Begrip en inzicht
Vouwblaadjes	1 ^a	.32							
	2	.30							
	3	.29							
Natikken	1	.28	.33						
	2	.39	.25						
	3	.41	.30						
Woordenspan	1	.13	.15	.25					
	2	.20	.17	.23					
	3	.21	.10	.39					
Plaatjes aanwijzen	1	.34	.32	.43	.64				
	2	.16	.26	.25	.57				
	3	.35	.29	.51	.63				
Zinnen nazeggen	1	.23	.18	.37	.59	.54			
	2	.18	.32	.20	.46	.48			
	3	.26	.30	.38	.55	.55			
Verhaaltje vragen	1	.15	.24	.36	.13	.26	.43		
	2	.05	.11	.21	.24	.28	.44		
	3	.33	.31	.32	.30	.39	.62		
Begrip en inzicht	1	.02	-.01	.14	.10	.22	.30	.21	
	2	.13	.17	.22	.28	.36	.50	.51	
	3	.19	.19	.27	.26	.30	.53	.48	
gemiddelde (m)	1	16.78	8.49	6.96	7.47	5.11	48.56	16.94	18.35
	2	16.25	8.43	6.16	7.30	4.68	46.15	15.53	16.67
	3	16.26	8.09	5.35	6.59	3.97	41.53	14.90	15.26
standaarddeviatie (s)	1	2.21	2.12	2.42	1.84	2.33	5.22	5.69	3.30
	2	2.75	1.81	2.21	2.00	2.33	7.45	6.84	4.45
	3	2.29	1.93	2.56	2.06	2.24	9.70	7.29	4.52

^a 1 = SEN-I (hoog); 2 = SEN-II (midden); 3 = SEN-III (laag)

Tabel 21. Korrelaties tussen de LDT-E, FST en NST
N = 340

(sub)tests	FST				NST										
	1	2	3	to- taal	1	2	3	4	5	6	7	8	9	10	to- taal
<i>LDT-E</i>															
Blokpatronen	.51	.33	.40	.52	.39	.50	.38	.23	.29	.26	.23	.33	.22	.26	.49
Vouwblaadjes	.38	.26	.35	.42	.30	.27	.35	.25	.19	.17	.21	.19	.19	.18	.37
Natikken	.42	.25	.40	.46	.34	.43	.48	.25	.31	.28	.24	.39	.25	.23	.51
Woordenspan	.27	.27	.36	.37	.30	.31	.38	.18	.22	.24	.16	.28	.24	.18	.39
Plaatjes aanwijzen	.43	.31	.45	.50	.34	.45	.51	.25	.27	.28	.25	.35	.33	.18	.52
Zinnen nazeggen	.44	.33	.43	.51	.41	.47	.54	.28	.31	.35	.31	.45	.39	.28	.61
Verhaaltje vragen	.33	.26	.29	.36	.37	.37	.39	.26	.29	.25	.29	.31	.35	.22	.50
Begrip en inzicht	.39	.19	.30	.39	.26	.32	.44	.32	.23	.27	.29	.38	.36	.09	.48
totaal	.60	.41	.57	.67	.52	.59	.67	.37	.41	.39	.40	.52	.44	.31	.74

Tabel 21 (vervolg). Korrelaties tussen de LDT-E, FST en NST

N = 340

(sub)tests	FST				NST										
	1	2	3	to- taal	1	2	3	4	5	6	7	8	9	10	to- taal
<i>FST</i>															
1 Figuren natekenen															
2 Hekpatroon voortzetten	.50														
3 Ramen tekenen	.54	.41													
totaal	.89	.66	.84												
<i>NST</i>															
1 Vormwaarneming etc.	.43	.35	.41	.49											
2 Fijne motoriek	.69	.42	.45	.67	.43										
3 Begrip van grootte etc.	.51	.38	.44	.56	.48	.53									
4 Scherp waarnemen	.38	.24	.28	.38	.28	.34	.34								
5 Kritische waarneming	.34	.27	.33	.39	.40	.39	.38	.31							
6 Koncentratie	.31	.26	.24	.33	.28	.28	.32	.32	.26						
7 Geheugen	.27	.15	.27	.30	.23	.21	.34	.30	.15	.28					
8 Objektbegrip etc.	.38	.21	.31	.39	.38	.44	.49	.36	.36	.31	.36				
9 Weergeven verhaaltje	.29	.16	.29	.32	.30	.32	.37	.33	.26	.31	.38	.40			
10 Menstekening	.25	.24	.26	.30	.34	.34	.28	.11	.25	.14	.13	.14	.23		
totaal	.61	.42	.52	.66	.63	.69	.72	.61	.59	.59	.59	.67	.63	.46	
gemiddelde (m)	6.7	1.7	5.8	14.2	7.0	5.0	6.0	5.1	6.0	5.6	5.5	6.0	3.4	5.4	55.0
standaarddeviatie (s)	3.7	1.4	3.2	6.9	1.4	2.3	1.8	2.2	2.0	2.4	2.8	1.7	2.0	2.0	12.6

Tabel 34. Vierweg kruistabel-gegevens voor Blokpatronen - Lezen / taal^a

Blokpatronen			Lezen / taal			totaal
pretest	posttest	hulp	laag	midden	hoog	
laag	laag	met	2	2	2	6
laag	laag	zonder	0	0	2	2
laag	midden	met	13	10	9	32
laag	midden	zonder	4	3	1	8
laag	hoog	met	6	16	7	29
laag	hoog	zonder	2	1	3	6
midden	laag	met	0	0	0	0
midden	laag	zonder	0	0	0	0
midden	midden	met	6	5	2	13
midden	midden	zonder	1	1	3	5
midden	hoog	met	23	20	12	55
midden	hoog	zonder	11	11	16	38
hoog	laag	met	0	0	0	0
hoog	laag	zonder	0	0	0	0
hoog	midden	met	0	0	0	0
hoog	midden	zonder	0	0	0	0
hoog	hoog	met	12	8	7	27
hoog	hoog	zonder	14	15	34	63
totaal			94	92	98	284

^a voor toelichting: zie tekst van paragraaf 11.4

Tabel 35. Vierweg kruistabel-gegevens voor Vouwblaadjes - Lezen / taal^a

Vouwblaadjes			Lezen / taal			totaal
pretest	posttest	hulp	laag	midden	hoog	
laag	laag	met	5	4	1	10
laag	laag	zonder	0	0	0	0
laag	midden	met	8	12	2	22
laag	midden	zonder	0	2	2	4
laag	hoog	met	26	18	15	59
laag	hoog	zonder	4	26	2	32
midden	laag	met	0	0	0	0
midden	laag	zonder	0	0	0	0
midden	midden	met	1	0	0	1
midden	midden	zonder	2	1	1	4
midden	hoog	met	28	16	17	61
midden	hoog	zonder	13	13	14	40
hoog	laag	met	0	0	0	0
hoog	laag	zonder	0	0	0	0
hoog	midden	met	0	0	0	0
hoog	midden	zonder	0	0	0	0
hoog	hoog	met	7	12	19	38
hoog	hoog	zonder	4	15	25	44
totaal			98	119	98	315

^a voor toelichting: zie tekst van paragraaf 11.4

Tabel 36. Vierweg kruistabel-gegevens voor Natikken - Lezen / taal^a

Naticken			Lezen / taal			totaal
pretest	posttest	hulp	laag	midden	hoog	
laag	laag	met	7	5	0	12
laag	laag	zonder	5	2	1	8
laag	midden	met	21	5	4	30
laag	midden	zonder	3	5	1	9
laag	hoog	met	9	5	8	22
laag	hoog	zonder	2	5	3	10
midden	laag	met	0	0	0	0
midden	laag	zonder	0	0	0	0
midden	midden	met	2	7	2	11
midden	midden	zonder	0	1	1	2
midden	hoog	met	24	22	17	63
midden	hoog	zonder	10	21	11	42
hoog	laag	met	0	0	0	0
hoog	laag	zonder	0	0	0	0
hoog	midden	met	0	0	0	0
hoog	midden	zonder	0	0	0	0
hoog	hoog	met	10	12	30	52
hoog	hoog	zonder	6	11	26	43
totaal			99	101	104	304

^a voor toelichting: zie tekst van paragraaf 11.4

Tabel 37. Vierweg kruistabel-gegevens voor Woordenspan - Lezen / taal^a

Woordenspan			Lezen / taal			totaal
pretest	posttest	hulp	laag	midden	hoog	
laag	laag	met	19	9	3	31
laag	laag	zonder	14	6	1	21
laag	midden	met	10	6	5	21
laag	midden	zonder	4	7	4	15
laag	hoog	met	2	4	3	9
laag	hoog	zonder	1	5	2	8
midden	laag	met	0	0	0	0
midden	laag	zonder	0	0	0	0
midden	midden	met	6	5	5	16
midden	midden	zonder	8	5	4	17
midden	hoog	met	6	9	15	30
midden	hoog	zonder	10	12	22	44
hoog	laag	met	0	0	0	0
hoog	laag	zonder	0	0	0	0
hoog	midden	met	0	0	0	0
hoog	midden	zonder	0	0	0	0
hoog	hoog	met	10	11	16	37
hoog	hoog	zonder	9	20	23	52
totaal			99	99	103	301

^a voor toelichting: zie tekst van paragraaf 11.4