

Mogelijkheden en beperkingen van clusteranalyse

Illustratie met de WISC-R bij kinderen met leerproblemen

Joris Cracco¹ en Evert Thiery²

Samenvatting

Dit artikel is het eerste deel van een tweeluik. In het tweede deel zal de factoranalyse aan bod komen. Dit tweeluik heeft vooral een educatieve betekenis. Het is gericht naar de clinicus die in de literatuur steeds meer met deze technieken wordt geconfronteerd en die wil leren wat de techniek inhoudt en wat de mogelijkheden en beperkingen zijn, zonder al te veel rekening te moeten houden met de statistische aspecten.

In dit artikel beschrijven we wat clusteranalyse is. We onderzoeken tevens de mogelijkheden en beperkingen van deze groep van technieken. We leggen er de nadruk op dat de clusters steeds moeten worden onderzocht op hun betrouwbaarheid en hun validiteit. Deze theoretische overwegingen worden geïllustreerd in het tweede gedeelte van het artikel. Op basis van de WISC-R-profielen van 59 kinderen met leerproblemen proberen we een betrouwbare en valide typologie te creëren. De gevonden clusters voldoen echter niet aan deze voorwaarden. We onderzoeken een aantal mogelijkheden die dit kunnen verklaren en trekken hieruit conclusies voor verder onderzoek.

1 Joris Cracco is orthopedagoog en werkzaam als directeur in het Centrum voor Ambulante Revalidatie De Klinker te Knokke-Heist.

2 Prof. dr. Evert Thiery is neuropsychiater en ere gewoon hoogleraar aan de Universiteit Gent. Hij is tevens werkzaam in het Consultatiecentrum Neuropsychologie te Gent.

Mogelijkheden en beperkingen van clusteranalyse

Illustratie met de WISC-R bij kinderen met leerproblemen

Joris Cracco en Evert Thiery

1 Inleiding

Bij neuropsychologisch gericht onderzoek wordt steeds meer gebruik gemaakt van multivariate statistische technieken. Hierbij denken we meer specifiek aan de recente en veelbelovende evolutie waarbij men verschillende subtypes van leerstoornissen poogt te onderscheiden (Feagans & Appelbaum, 1986; Fisk & Rourke, 1979; Fuerst, Fisk & Rourke, 1990; Korhonen, 1991a, 1991b; Lyon, Stewart & Freedman, 1982; Lyon & Watson, 1981; Lyon, Watson, Reitta, Porch & Rhodes, 1981; McKinney, 1984; Morris, 1988; Morris, Blashfield & Satz, 1981, 1986; Petrauskas & Rourke, 1979; Rourke, 1985, 1989, 1991; Spreen & Haaf, 1986; Watson & Goldgar, 1983). Nochtans lijken veel psychologen en (ortho)pedagogen een afkeer te hebben ontwikkeld voor de statistiek. Zonder twijfel is dit voor een belangrijk deel te wijten aan het misbruik dat men vaak van deze hulpwetenschap maakt en aan de complexiteit van sommige statistische technieken. We mogen echter het kind niet met het badwater weggooien. De hierboven vermelde evolutie is te belangrijk om zomaar te negeren. Anderzijds moeten we een gezond wetenschappelijk scepticisme aan de dag leggen en iedere studie evalueren op haar sterke en zwakke punten. Om dit te kunnen, is het noodzakelijk om een rudimentair begrip te hebben van de toegepaste methodes en op de hoogte te zijn van hun mogelijkheden en beperkingen.

Dit artikel heeft in de eerste plaats een educatieve opzet. Het is de bedoeling om de lezer die er minder vertrouwd mee is, kennis te laten maken met een veel gebruikte multivariate statistische methode, nl. de clusteranalyse. Hierbij zullen we zo weinig mogelijk gebruik maken van formules en pogen om de bedoeling, werkwijze, mogelijkheden en beperkingen van deze techniek op een begrijpelijke wijze naar voor te brengen. We richten ons dus niet naar de statisticus, die hier enkel een beknopte bespreking van deze techniek mag verwachten. Ter illustratie geven we de resultaten van een eigen onderzoek waarbij werd gepoogd om op basis van WISC-R-subtest-scores een valide en betrouwbare typologie van kinderen met leerproblemen te creëren.

In een tweede artikel zal later de factoranalyse aan bod komen. Hierbij zal ter illustratie de factorstructuur van de WISC-R worden onderzocht.

2 Situering

Stel dat men van 100 proefpersonen 10 tests heeft afgenomen. We krijgen dus $10 \times 100 = 1000$ waarden die onze proefgroep beschrijven. In tabel 1 tonen we hoe de datamatrix er dan uitziet:

Tabel 1: Theoretische datamatrix van 100 proefpersonen op 10 tests

| | TEST 1 | TEST 2 | ... | TEST 10 |
|--------|-------------|-------------|-----|--------------|
| pp. 1 | X_{1-1} | X_{1-2} | ... | X_{1-10} |
| pp. 2 | X_{2-1} | X_{2-2} | ... | X_{2-10} |
| . | . | . | ... | . |
| . | . | . | ... | . |
| . | . | . | ... | . |
| pp.100 | X_{100-1} | X_{100-2} | ... | X_{100-10} |

Het probleem dat zich nu stelt, is het indelen van de proefpersonen in groepen. We willen dit op zodanige wijze doen dat proefpersonen *binnen* een groep zo veel mogelijk op elkaar gelijken (wat betreft hun testcores), terwijl ze tegelijkertijd zo veel mogelijk verschillen van diegenen die in een andere groep werden ingedeeld. Dit is precies wat clusteranalyse beoogt. Hierbij komen echter drie belangrijke problemen kijken (Morris e.a., 1981): (1) in welk opzicht moeten de personen op elkaar gelijken of van elkaar verschillen? (2) hoeveel groepen moeten we maken? en (3) hoe gaan we de groepen samenstellen?. Voor geen enkel van deze problemen bestaat er een algemeen aanvaarde oplossing (Fletcher & Satz, 1985). Het is dus aan de onderzoeker om - naargelang het soort onderzoek dat hij doet en naargelang de aard van zijn data - een *verantwoorde* keuze te maken. In de volgende paragrafen lichten we deze drie problemen toe.

3 Clusteranalyse

3.1 Welke methode?

Zelfs de meest geavanceerde computer kan onmogelijk alle mogelijke indelingen die onze datamatrix biedt, uitrekenen. Er bestaan verschillende rekenmethodes om op een zo efficiënt mogelijke wijze tot een "redelijke" oplossing te komen. De twee belangrijkste groepen van deze algoritmen zijn de hiërarchische agglomeratieve methodes en de iteratieve verdelingsmethodes (Morris e.a., 1981).

3.1.1 Hiërarchische agglomeratieve methodes

De hiërarchische methodes beginnen met de paren proefpersonen die het meest op elkaar lijken tot één cluster samen te voegen. Bij de volgende stap worden deze clusters (die bestaan uit een of twee proefpersonen) opnieuw in paren samengevoegd. Men gaat zo verder tot er niets meer valt samen te voegen, m.a.w. totdat we nog één cluster over hebben. Naargelang de wijze waarop men clusters verenigt, kunnen we verschillende hiërarchische methodes onderscheiden (Aldenderfer & Blashfield, 1978; Blashfield, 1976; Edelbrock, 1979; Morris e.a., 1981). Hieronder geven we een kort overzicht van enkele van de meest gebruikte methodes.

De enkelvoudige verbinding (single linkage)

Deze techniek wordt ook wel "minimum linkage" of "nearest neighbor" genoemd. De gelijkenis tussen twee clusters wordt bij deze methode bepaald op basis van de gelijkenis tussen de dichtstbijzijnde leden van de twee clusters. Een individu wordt bij een cluster gevoegd indien dit individu meer lijkt op ten minste één lid van die cluster dan op gelijk welk ander lid van alle andere clusters. Deze werkwijze heeft het grote nadeel dat ze de neiging heeft een ketenachtige clusterstructuur te vormen (men spreekt in zo'n geval vaak van een ziggurat). Figuur 4 toont een voorbeeld van een dergelijke clusterstructuur.

De volledige verbinding (complete linkage)

Deze methode vormt de tegenhanger van de vorige en wordt ook wel "maximum linkage" of "furthest neighbor" genoemd. Hier lijkt ieder lid van een cluster meer op het meest verschillende lid van die cluster dan op het meest verschillende lid van gelijk welke andere cluster. De kans dat een bepaalde cluster een nieuw lid krijgt, daalt naargelang de cluster groter wordt. Hierdoor heeft dit procédé de neiging om clusters van gelijke grootte te vormen.

De gemiddelde verbinding (average linkage)

Op deze techniek bestaan verschillende varianten, naargelang de wijze waarop men het gemiddelde berekent en naargelang men toestaat om de clusters een verschillend gewicht toe te kennen. Bij de meest geciteerde methode (de "unweighted pair-group method using arithmetic averages" afgekort als UPGMA) wordt de afstand tussen twee clusters berekend als de gemiddelde afstand van alle paren individuen, waarbij één lid van het paar tot de eerste en het andere lid tot de tweede cluster behoort. Evenals de enkelvoudige verbinding heeft dit procédé de neiging om uitbijters op te leveren. Deze kans neemt toe naargelang de clusters groter worden. Een voordeel ten opzichte van de twee vorige technieken is dat alle leden van een cluster een bijdrage leveren tot het bepalen van de gelijkenis en niet enkel het dichtstbijzijnde of het verst afgelegen lid.

De minimumvariantiemethode (minimum variance)

Deze techniek is de enige van de hier vermelde die een objectief statistisch criterium hanteert om te bepalen hoe clusters moeten worden samengevoegd. Men berekent alle mogelijke combinaties van clusters. Hieruit kiest men die oplossing die de toename van de variantie binnen de clusters zo klein mogelijk houdt. Deze techniek wordt meestal aangeduid als *Wards* methode. Een belangrijke kritiek is dat deze techniek geen rekening houdt met de covariantiestructuur van de datamatrix.

3.1.2 Iteratieve methodes

De hiërarchische methodes hebben het grote nadeel dat wanneer de computer eenmaal een bepaalde proefpersoon heeft ingedeeld, deze niet meer kan worden verplaatst naar een andere cluster die meer geschikt lijkt. Dit kan wel met de iteratieve methodes (Blashfield & Aldenderfer, 1978). Deze beginnen met een vooropgesteld aantal clusters te maken en met het berekenen van het centrum van iedere groepering. Bij de volgende stap wordt iedere proefpersoon aan de cluster met het dichtstbijzijnde centrum toegewezen. Daarna worden de nieuwe centra van de clusters berekend en worden de proefpersonen opnieuw in groepen ingedeeld op basis van deze nieuwe centropunten. Dit opnieuw indelen gaat door totdat geen enkele proefpersoon van cluster verandert.

Deze methodes hebben als bijkomend voordeel dat ze grotere databestanden aankunnen (Hintze, 1987). Een nadeel is dat de uiteindelijke oplossing sterk afhangt van de eerste indeling. Deze eerste indeling kan op basis van het toeval worden gekozen of de onderzoeker kan zelf een schatting maken. Een veel gebruikte techniek is om eerst een hiërarchische methode toe te passen en de oplossing daarvan te gebruiken als eerste indeling van een iteratieve methode (DeLuca, Adams & Rourke, 1991). Soms worden de proefpersonen niet ingedeeld in een nieuwe cluster op basis van het dichtstbijzijnde centropunt maar op basis van de mate waarin een bepaald statistisch criterium wordt geoptimaliseerd (Blashfield e.a., 1978). Dit criterium is meestal het verminderen van de mate van variantie binnen iedere cluster of een combinatie van het verminderen van de binnenclustervariantie en het vergroten van de tussenclustervariantie.

3.2 In welk opzicht moeten de personen worden vergeleken?

De tweede cruciale vraag die we ons moeten stellen, is welke soort informatie we in onze datamatrix het belangrijkst vinden. Op basis van zijn of haar testcores moet iedere proefpersoon een plaats toegewezen krijgen die rekening houdt met alle dimensies die in de testcores bevat liggen (Adams, 1985). Deze dimensies zijn (1) de hoogte (2) de spreiding van de scores en (3) de vorm van het testprofiel (Skinner, 1978). De eerste dimensie is eenvoudig te begrijpen: iemand met hoge scores op de gemeten variabelen wordt onderscheiden van iemand met lage scores. De tweede dimensie geeft aan in welke mate de resultaten afwijken van het gemiddelde. De vorm zegt iets over het patroon van hoge en lage scores (het testprofiel).

We hebben in de vorige paragraaf gezien dat een clusteralgoritme poogt om personen die goed op elkaar gelijkken in dezelfde groep onder te brengen. Bij het operationaliseren van deze "gelijkenis" zal er meer of minder belang worden gehecht aan een van deze drie dimensies. De "gelijkheidsmaat" speelt dus een belangrijke rol. Wanneer men bijvoorbeeld de correlatie berekent tussen de individuen (niet tussen de variabelen) en op basis daarvan de gelijkenis bepaalt, krijgt men uitsluitend informatie over het patroon van de testresultaten (de vorm). Informatie over de spreiding en de hoogte van de scores wordt genegeerd. Een tweede vaak gebruikte gelijkheidsmaat is het kwadraat van de euclidische afstand. De gelijkenis tussen twee personen wordt hier bepaald door van ieder paar variabelen het kwadraat van het verschil te berekenen en deze kwadraten op te tellen. Dit blijkt uit de volgende formule:

Formule 1: Het kwadraat van de euclidische afstand

$$D_{xy}^2 = \sum_i (X_i - Y_i)^2$$

Deze afstandsmaat combineert informatie over de hoogte, de spreiding en een interactie tussen spreiding en vorm (Skinner, 1978).

3.3 Hoeveel clusters vormen de optimale oplossing?

Wanneer men te veel clusters kiest, worden de gegevens moeilijk te interpreteren. Verder zal ieder computeralgoritme net zo veel clusters maken als de onderzoeker vraagt, ook wanneer dit betekent dat een groep proefpersonen die maar minimaal van elkaar verschillen in twee moet worden gesplitst. Anderzijds kan men triviale informatie krijgen wanneer men te weinig clusters vraagt en heeft men de kans dat de computer proefpersonen die sterk van elkaar verschillen tot "cohabitatie" samen dwingt.

Ook voor dit probleem bestaat er geen vast omliggende oplossing. Alhoewel een aantal richtlijnen kunnen worden verstrekt, blijft het de (subjectieve) keuze van de onderzoeker om het aantal te bepalen (Morris e.a., 1981). Men kan bijvoorbeeld de voorkeur geven aan de klinische interpreteerbaarheid van een gegeven oplossing. Er bestaan ook clustercoëfficiënten die bijvoorbeeld aangeven hoeveel variantie er wordt verklaard door het toevoegen of weglaten van een cluster, of hoe groot de afstand is tussen twee clusters die worden samengevoegd. Wanneer er van de ene stap naar de andere een groot verschil optreedt in deze verklaarde variantie of wanneer de afstand tussen twee samengevoegde clusters veel groter is dan de afstand bij de vorige stap, betekent dit vaak dat twee groepen werden samengevoegd die toch wel aanzienlijk van elkaar verschillen.

3.4 Nadelen van clusteranalytische technieken

Clusteranalyses geven een statisch beeld van de realiteit. Ze zijn zeer geschikt als beschrijvende techniek maar minder om rekening te houden met de ontwikkelingsaspecten (Adams, 1985). Er zijn bijvoorbeeld aanwijzingen dat de kenmerken van de verschillende subtypes van kinderen met leerproblemen veranderen met de leeftijd van deze kinderen (Spreeen e.a., 1986). Clusteranalyse is in essentie een cross-sectionele techniek en veel minder geschikt voor longitudinale studies.

Een groot gevaar dat inherent is aan het gebruik van clusteranalyses, is dat ze de schijn geven van een objectiviteit die in de realiteit niet aanwezig is. Uit het voorgaande mag blijken dat bij ieder onderzoek waarbij deze technieken worden toegepast, heel wat keuzes moeten worden gemaakt. Er bestaan hierbij enkele vuistregels maar geen algemeen aanvaarde criteria. Het maken van een of andere keuze kan de uiteindelijke oplossing grondig beïnvloeden. Bij clusteranalyses kan men geen significantietoetsen toepassen (Adams, 1985) en er bestaat geen "confirmatorische clusteranalyse" waarbij men een hypothetisch model kan toetsen aan alternatieve modellen.

In verband staand met het voorgaande is het risico dat men via clusteranalyse vervalt in een soort "naïef empirisme", namelijk dat men op gelijk welk gegevensbestand analyses gaat uitvoeren zonder rekening te houden met de theorievorming.

Ten slotte zal iedere clusteranalyse wel met een oplossing voor de dag komen, ook wanneer de datamatrix zuiver op basis van het toeval tot stand kwam (Fletcher e.a., 1985) of wanneer de proefgroep homogeen is. Het uitvoeren van een clusteranalyse is dan ook maar een begin (DeLuca e.a., 1991). Nadien moet de gevonden oplossing grondig worden gevalideerd. In de volgende paragraaf hebben we het over de verschillende technieken die bij het valideren en bij het controleren van de betrouwbaarheid kunnen worden toegepast.

3.5 Validering van clusteroplossingen

3.5.1 Externe validiteit

De externe validiteit "...involves evidence of prognostic usefulness, descriptive validity, clinical meaningfulness, and generalizability across samples" (DeLuca e.a., 1991). De meest gebruikte werkwijze om de externe validiteit te beoordelen, is nagaan of de gevonden clusters significant van elkaar verschillen op grond van variabelen die niet in de oorspronkelijke clusteranalyse werden betrokken. Zo voerde Korhonen (1991a, 1991b) een clusteranalyse uit op een proefgroep van 82 leergestoorde en 84 controlekinderen, op basis van acht neuropsychologische testvariabelen. Hij vond zes clusters die significant van elkaar te onderscheiden waren op basis van het geslacht van de kinderen, hun leeftijd, sociaal-economische status van de ouders, lezen en spelling, neurologische soft signs en gedragsbeoordelingen door leerkracht en ouders. Spijtig genoeg stopt de externe validering te vaak bij deze stap. Het moeizame proces dat gepaard gaat

met het aantonen van de ecologische, klinische en predictieve validiteit wordt vaak achterwege gelaten. Ook het onderzoek naar de interactie tussen subtype en behandeling komt zelden aan bod (zie echter Fiedorowicz & Trites, 1991; Lyon & Flynn, 1991).

3.5.2 Betrouwbaarheid

Morris e.a. (1981) beschrijven verschillende methodes om de betrouwbaarheid van een clusteroplossing na te gaan. Het meest gebruikte procédé bestaat erin dat men de data op een zodanige wijze manipuleert dat men de veralgemeenbaarheid van de oplossing kan aantonen. Zo kan men bijvoorbeeld de groep proefpersonen op toevallige basis in twee delen opsplitsen en dan op beide subgroepen de clusteranalyse opnieuw toepassen. Wanneer men bij beide subgroepen een oplossing krijgt die een grote overeenkomst vertoont met het aanvankelijke resultaat, heeft men reden om de clusters als betrouwbaar te beschouwen. Deze methode wordt aangeduid met de "split-sample"-techniek. Een variant hierop is dat men proefpersonen aan de oorspronkelijke datamatrix toevoegt of proefpersonen weglaat.

Soms gebruikt men ook zogenaamde "Monte Carlo"-technieken om een hypothetische controlegroep samen te stellen. Over het gebruik van dergelijke technieken bestaat heel wat controverse, waarop we hier echter niet zullen ingaan. Eén techniek bestaat erin dat men een hypothetische proefgroep creëert. De testcores van deze hypothetische groep zijn toevallig gekozen, maar met die beperking dat ze hetzelfde gemiddelde en dezelfde standaarddeviatie en covariantiematrix hebben als de oorspronkelijke proefgroep. Als een clusteranalyse op deze hypothetische groep een oplossing oplevert die sterk lijkt op die van de oorspronkelijke, is de kans groot dat de gevonden oplossing op toeval berust. Men kan ook een reeks testgroepen maken door uit de oorspronkelijke proefgroep verschillende steekproeven met teruglegging te nemen.

Een derde mogelijkheid is dat men op dezelfde proefgroep verschillende soorten gelijkaardige clusteranalyses toepast. Wanneer deze met gelijkvormige oplossingen voor de dag komen, is dit eveneens een indicatie van een goede betrouwbaarheid. Hierbij is het wel belangrijk dat men gebruik maakt van clusteranalyses die van dezelfde assumpties vertrekken. De hiërarchische technieken vertrekken bv. van de assumptie van de ultrametrische ongelijkheid. D.w.z. dat voor elk drietal observaties geldt dat er twee zijn die dicht bij elkaar liggen en even ver van de derde. Wanneer de gebruikte clusteralgoritmes te veel van elkaar verschillen, is het niet meer dan normaal dat men een totaal verschillende oplossing bekomt.

Een vierde techniek bestaat erin dat men een iteratieve analyse laat volgen op een hiërarchische en dan nagaat hoeveel proefpersonen door de iteratieve analyse aan een andere cluster werden toegewezen. DeLuca e.a. (1991) stellen echter de bruikbaarheid van deze techniek in vraag. Ze voerden verschillende soorten clusteranalyses uit op twee datamatrices waarvan de clusteroplossing op voorhand reeds was gekend. Op deze wijze konden ze nagaan in hoeverre de verschillende technieken de proefpersonen correct hadden geclassificeerd. De correlatie tussen

het aantal correct geclassificeerde waarnemingen en het percentage opnieuw toegewezen waarnemingen na een iteratieve methode was nihil of zelfs licht negatief.

Ten slotte bestaan er nog verschillende statistische maateenheden die worden voorgesteld om de betrouwbaarheid te controleren. Morris e.a. (1981) wijzen er echter op dat deze coëfficiënten meestal voor iets anders werden ontwikkeld en dat de adequaatheid ervan nog ongekend is.

Soms spreekt men over "interne validiteit" in plaats van over betrouwbaarheid. Adams (1985) en DeLuca e.a. (1991) vechten de eerste term aan op basis van het argument dat interne validiteit te maken heeft met de fase van de theorievorming. Deze fase moet reeds voor het gebruik van statistische technieken zijn afgerond en behelst o.a. consideraties omtrent het kiezen van geschikte variabelen. Dit aspect komt in de volgende paragraaf aan bod.

3.5.3 Theorievorming

Men moet steeds voor ogen houden dat clusteranalyse een exploratieve techniek is en *op zichzelf* slechts een geringe of zelfs geen bewijskracht heeft. De beslissing dat men gebruik zal maken van een clusteranalyse en de keuze tussen de verschillende alternatieven (keuze van clustertechniek, gelijkheidsmaat, variabelen...) kunnen slechts op een zinvolle wijze gebeuren indien men werkt met een theoretisch kader. Skinner (1981) drukt dit als volgt uit: "The essential point is that choice among alternative classification models and corresponding computer algorithms must be consistent with theory".

Een hiërarchische clusteranalyse is bijvoorbeeld enkel een geschikte classificatietechniek indien men een systeem wil opbouwen van discrete groepen. Hierbij kan een persoon niet terzelfder tijd tot meer dan één groep behoren. Wanneer men meer belang hecht aan het beschrijven van relaties tussen individuen, die worden voorgesteld als punten in een multidimensionele ruimte, kan men beter gebruik maken van een Q-factoranalyse (dit is een factoranalyse waarbij men vertrekt van de correlaties tussen individuen i.p.v. de correlaties tussen variabelen). Deze beide technieken werden reeds gebruikt bij studies die waren gericht op het zoeken naar subtypes bij kinderen met leerstoornissen. Vreemd genoeg gaat de discussie over de voor- en nadelen van de ene techniek t.o.v. de andere vrijwel altijd over het vermogen om betrouwbare en valide classificaties te creëren, maar zelden over het soort classificatie waarin men *a priori* is geïnteresseerd.

Het spreekt vanzelf dat ook de keuze van een afstandsmaat in eerste instantie zal worden gedicteerd door theoretische overwegingen. Indien men enkel informatie wil over testprofielen (ongeacht de hoogte van de testresultaten), dan heeft het geen zin om de euclidische afstand te gebruiken, daar deze gelijkheidsmaat wordt beïnvloed door zowel de vorm, de hoogte als de spreiding van de resultaten.

4 Illustratie

4.1 Vraagstelling en hypotheses

In wat volgt, geven we een illustratie van een onderzoek waarbij clusteranalyse werd toegepast. We vertrekken hierbij van 2 vraagstellingen waarvan de tweede afhangt van de eerste:

- (1) Kan men - aan de hand van WISC-R-profielen - klinisch zinvolle, betrouwbare en valide subtypes van kinderen met leerproblemen onderscheiden?
- (2) Indien we dergelijke groepen kunnen onderscheiden, zullen kinderen met leerstoornissen en kinderen met leermoeilijkheden dan in verschillende subtypes worden geclassificeerd?

Deze vragen kunnen we herformuleren in de twee volgende nulhypotheses:

- (1) Het is niet mogelijk om - door middel van clusteranalyse met WISC-R-profielen - een betrouwbare en valide typologie van kinderen met leerproblemen te creëren.*
- (2) Indien het mogelijk is om aan de hand van WISC-R-profielen een valide en betrouwbare typologie te creëren, dan zullen kinderen met leerstoornissen en kinderen met leermoeilijkheden in elk subtype proportioneel even veel voorkomen.*

4.1.1 Eerste hypothese

Uit vorig onderzoek naar de factorstructuur van de WISC-R weten we dat deze intelligentietest kan worden ingedeeld in twee stabiele en één minder stabiele factoren (Kaufman, 1975, 1979). De twee belangrijkste factoren zijn een verbale (subtests Informatie, Overeenkomsten, Woordenschat en Begrijpen) en een visuospatiële (Onvolledige Tekeningen, Plaatjes Ordenen, Blokpatronen, Figuur Leggen en Doolhoven). Deze dichotomie sluit aan bij Luria's (1973) links/rechts-dimensie van de organisatie van de hersenen. Verder toont onderzoek bij volwassenen aan dat linkshemisferische laesies vaak gepaard gaan met een depressie van het VIQ, terwijl het omgekeerde (relatief laag PIQ) geldt bij een rechtshemisferische beschadiging (Kolb & Whishaw, 1990). Het is ook opmerkelijk dat de bestaande typologieën van leergestoorde kinderen vaak (impliciet of expliciet) steunen op deze links/rechts-dimensie. Een goed voorbeeld hiervan zijn de P-type en L-type dyslexie van Bakker (1992).

Het is trouwens niet zo verwonderlijk dat de boven/onder- en de voor/achter-dimensies van Luria minder tot uiting komen in testresultaten van leergestoorde kinderen. De boven/onder-dimensie betreft nl. vooral het regelen van de aandacht. Stoornissen op dit niveau zullen doorgaans een invloed hebben op *alle* testresultaten (ongeacht de inhoud en de afnamemodaliteiten) en dus minder tot uiting komen in het testprofiel. Daar de frontale hersenkwabben slechts volledig worden ontwikkeld tijdens de puberteit, is het niet verwonderlijk dat ook de voor/achter-dimensie bij kinderen niet of minder tot uiting komt.

We verwachten dat we op z'n minst twee valide en betrouwbare WISC-R-profielen zullen vinden, namelijk een met een relatief laag VIQ en een met een relatief laag PIQ. Voortgaand op de hypothese van Rourke en collega's omtrent de invloed van neurologische deficiënties op de psychosociale problemen van leergestoorde kinderen (Fuerst e.a., 1989, 1990; Rourke, 1988, 1991), kunnen we ook verwachten dat de kinderen met een relatief laag PIQ meer psychosociale problemen zullen vertonen. Rourke stelt namelijk dat deficiënties op het gebied van de niet-verbale conceptvorming, de communicatie en de probleemoplossingsvaardigheden bij ten minste één subtype van kinderen met leerstoornissen leiden tot psychosociale problemen. Voor een meer uitgebreid overzicht hiervan verwijzen we naar Cracco (1993) en Cracco & Thiery (1993).

4.1.2 Tweede hypothese

In de meeste definities van leerstoornissen vindt men een aantal kenmerken terug:

- er moet een discrepantie zijn tussen de intelligentie van het kind en zijn/haar schoolvorderingen;
- de leerproblemen kunnen niet worden verklaard door sensorische, sociaal-educatieve of emotionele problemen (in dit geval spreekt men in Vlaanderen meestal van leermoeilijkheden, een term die wij hier ook verder zullen gebruiken).

Deze definities komen de laatste tijd onder zware druk te staan. Een veel gehoorde kritiek is dat ze vooral zeggen wat een leerstoornis *niet* is. Men stelt zich ook steeds meer de vraag of er inderdaad wel zo'n essentieel verschil is tussen kinderen die een discrepantie vertonen tussen IQ en schoolvorderingen en diegenen die dit niet doen (Siegel, 1989, 1992; Shaywitz, Fletcher, Holahan & Shaywitz, 1992). We verwachten dan ook dat we - indien er betrouwbare en valide subtypes worden gevonden - in iedere gevonden categorie een even grote proportie kinderen met leerstoornissen als kinderen met leerproblemen zullen vinden.

4.2 Proefgroep

De proefgroep bestaat uit 59 van de 62 kinderen die tussen juni 1992 en mei 1993 bij een van de auteurs werden aangemeld met leerproblemen. Van de drie kinderen die niet in de proefgroep werden opgenomen, hadden er twee een TIQ < 70 en was er één kind met onvolledige data. De gemiddelde leeftijd van de kinderen is 105,15 maanden, met een verhouding van 42 jongens tegenover 17 meisjes.

4.3 Testvariabelen

De meeste van deze kinderen werden met een uitgebreide batterij tests onderzocht. In dit onderzoek zullen enkel de twaalf subtests van de WISC-R (Van Haasen, De Bruyn, Pijl, Poortinga, Spelberg, Vander Steene, Coetsier, Spoelders-Claes & Stinissen, 1986), de vier hoofdschalen van de Achenbach Child Behavior Checklist (Verhulst, Koot, Akkerhuis & Veerman, 1990) en de schoolvorderingen voor technisch lezen, spelling en rekenen worden gebruikt.

Voor de schoolvorderingen werd ernaar gestreefd om voor alle leeftijden zo veel mogelijk dezelfde soort toets te gebruiken. Tabel 2 geeft een overzicht van de gebruikte tests. De ruwe scores werden omgezet in z-scores. Bij de CITO-rekentoetsen komen er enkele items voor die niet te vertalen zijn voor een Vlaamse populatie. In dergelijke gevallen werd het item niet meegerekend en werd de totale score geschat met de regel van drie. Indien een kind werd getest op een maand waarop de toets niet was genormeerd, werden het gemiddelde en de standaarddeviatie geschat d.m.v. extra- of intrapolatie. Deze methode veronderstelt een lineaire progressie in de scores, hetgeen niet noodzakelijk waar hoeft te zijn. Het gebruik van de normen die het dichtst bij het ogenblik liggen waarop het kind werd getest, kan anderzijds leiden tot grote onder- of overschatting.

Tabel 2: Overzicht van de gebruikte schoolvorderingstests

| Technisch lezen | Spelling | Rekenen |
|--|--|---|
| SBO 1'30 test A vorm (1ste leerjaar) | SBO dictee F vorm M1° (jan 1ste lj t.e.m. maart 1ste lj) | SBO rekenen F vorm M1° (jan 1ste lj t.e.m. maart 1ste lj) |
| 1 minuut test (vanaf sept 2de lj) | SBO dictee E1°-M2° (apr 1ste lj t.e.m. okt 2de lj) | SBO rekenen T vorm (apr 1ste lj t.e.m. dec 2de lj) |
| | PDO praxis (vanaf nov 2de lj) | SBO wiskundig denken (jan 2de lj t.e.m. maart 2de lj) |
| | | CITO rekenproeven (vanaf apr 2de lj) |

4.4 Gebruikte procedure

4.4.1 Indeling in groepen

De proefgroep werd ingedeeld in een groep met leerstoornissen (groep LS) en een groep met leermoeilijkheden (groep LM). Een kind werd als leergestoord beschouwd op basis van de volgende criteria:

(1) Afwezigheid van

- meertalige opvoeding
- verminderd gehoor (meer dan -30dB)
- indicaties van sociaal-culturele deprivatie
- mentale handicap (TIQ < 70)

Wegens financiële problemen kon de visus niet van alle kinderen worden onderzocht. Van de 33 kinderen van wie de visus wel werd onderzocht, werd er geen enkel kind gedetecteerd met ongecorrigeerde visusproblemen.

(2) Ten minste één schoolvorderingsscore wijkt significant af ten opzichte van het TIQ. Het significantieniveau werd vastgesteld op 5 procent ($p = 0.05$) en werd berekend volgens formule 2 (Reynolds, 1981).

$$z_{(D)} = \frac{|Z_x - Z_y|}{\sqrt{(1 - r_{xx}) - (1 - r_{yy})}}$$

Waarbij :

Z_x = z -score op test x
 Z_y = z -score op test y
 r_{xx} = betrouwbaarheid voor test x
 r_{yy} = betrouwbaarheid voor test y

Formule 2 z-score voor het berekenen van significante verschillen tussen twee testcores

Een kind werd ingedeeld in de categorie leermoeilijkheden indien aan de volgende voorwaarden werd voldaan:

- (1) Geen mentale handicap (TIQ > 70)
- (2) Geen leerstoornis (cf. supra)
- (3) Het kind haalt op ten minste één schoolvorderingstoets een z-score lager dan -1.

Kinderen die noch in de eerste groep noch in de tweede groep konden worden geclassificeerd, werden als een restgroep beschouwd. Het betreft hier vier kinderen met slechts lichte leerproblemen en zes kinderen die hun schooljaar dubbelen. Deze laatste groep heeft wel degelijk een schoolse achterstand, maar hun schoolvorderingen worden door het dubbelen kunstmatig omhoog getrokken, zodat het niet valt uit te maken of ze tot de groep leerstoornissen of de groep leermoeilijkheden behoren.

Tabel 3: Testscores van de totale proefgroep, de groep leerstoornissen (LS), de groep leerproblemen (LM) en de restgroep (rest).

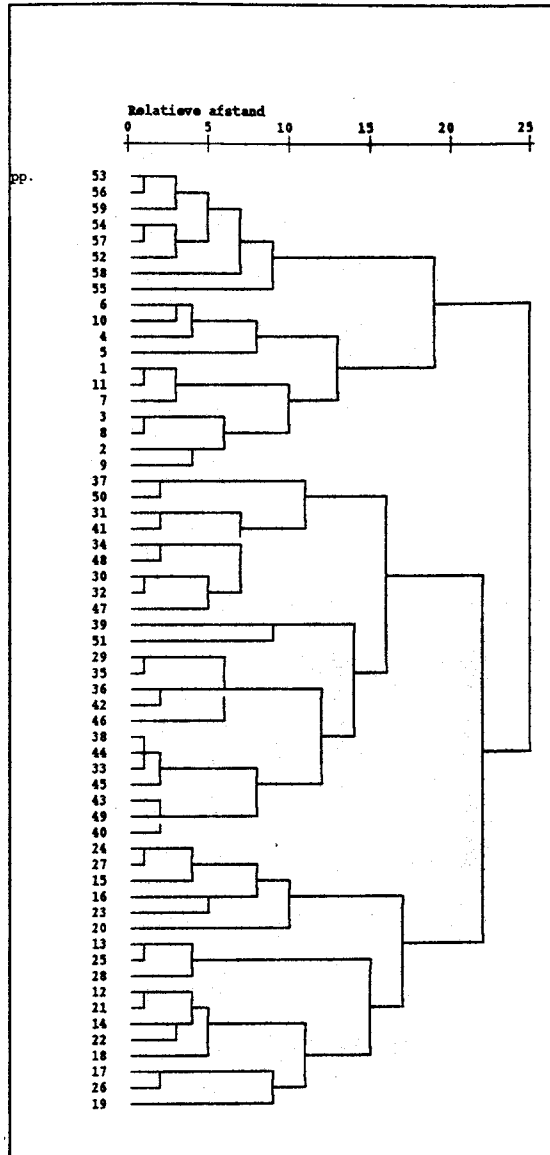
| M (SD) Range | LS N = 31 | LM N = 18 | Rest N = 10 | Totaal N = 59 |
|-------------------------|-------------------------|------------------------|-------------------------|--------------------------|
| TIQ | 96.48 (12.07) 73 122 | 83.33 (7.34) 70 98 | 86.90 (9.60) 71 104 | 90.85 (11.95) 70 122 |
| TVIQ | 96.13 (10.85) 78 122 | 83.39 (8.67) 71 108 | 91.90 (10.25) 68 107 | 91.53 (11.45) 68 122 |
| TPIQ | 97.61 (16.30) 67 134 | 86.44 (8.28) 71 101 | 84.20 (10.71) 73 99 | 91.93 (14.57) 67 134 |
| F1IQ | 99.94 (12.24) 80 130 | 85.28 (9.95) 69 110 | 95.70 (11.00) 71 110 | 94.75 (12.95) 69 130 |
| F2IQ | 98.68 (17.69) 69 137 | 85.39 (9.27) 67 102 | 84.10 (12.98) 71 104 | 92.15 (16.16) 67 137 |
| F3IQ | 90.32 (10.35) 72 112 | 87.22 (7.37) 74 100 | 87.10 (12.11) 72 104 | 88.83 (9.83) 72 112 |
| I | 8.55 (2.81) 4 17 | 6.17 (2.23) 3 11 | 7.90 (2.13) 5 11 | 7.71 (2.72) 3 17 |
| O | 10.35 (2.59) 7 19 | 7.89 (2.05) 5 12 | 10.00 (3.20) 2 13 | 9.54 (2.74) 2 19 |
| R | 8.58 (2.53) 4 15 | 7.00 (1.75) 3 10 | 6.90 (2.60) 4 11 | 7.81 (2.43) 3 15 |
| W | 9.87 (3.21) 4 18 | 7.67 (2.63) 3 14 | 8.80 (2.70) 2 11 | 9.02 (3.08) 2 18 |
| B | 11.29 (2.71) 5 16 | 8.78 (2.67) 1 12 | 10.60 (3.31) 6 15 | 10.41 (2.97) 1 16 |
| C | 8.16 (1.98) 5 11 | 7.67 (2.40) 4 12 | 8.70 (2.50) 4 12 | 8.10 (2.19) 4 12 |
| OT | 10.52 (3.53) 5 16 | 7.89 (2.37) 2 12 | 8.60 (3.20) 3 14 | 9.39 (3.34) 2 16 |
| PO | 10.19 (2.76) 6 18 | 8.00 (2.35) 3 11 | 9.00 (2.71) 6 14 | 9.32 (2.77) 3 18 |

Tabel 3: Testscores van de totale proefgroep, de groep leerstoornissen (LS), de groep leerproblemen (LM) en de restgroep (rest).

| M (SD) Range | LS N = 31 | LM N = 18 | Rest N = 10 | Totaal N = 59 |
|-----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| BP | 9.48 (3.70) 3 17 | 7.67 (2.40) 5 15 | 6.50 (2.76) 2 11 | 8.42 (3.37) 2 17 |
| FL | 8.45 (4.02) 1 17 | 7.11 (2.40) 3 12 | 5.60 (3.66) 1 11 | 7.56 (3.64) 1 17 |
| SU | 9.06 (2.98) 5 18 | 9.72 (1.96) 5 12 | 8.70 (2.87) 5 13 | 9.20 (2.67) 5 18 |
| DH | 9.97 (2.50) 4 15 | 8.28 (1.96) 4 11 | 8.20 (2.39) 5 13 | 9.15 (2.45) 4 15 |
| CBC TOT T score | 44.55 (9.08) 28 70 | 45.44 (10.52) 28 59 | 39.00 (6.34) 25 47 | 43.88 (9.30) 25 70 |
| CBC INT T score | 44.94 (9.89) 28 66 | 44.94 (10.12) 25 60 | 41.90 (5.49) 32 49 | 44.42 (9.30) 25 66 |
| CBC EXT T score | 45.45 (8.75) 26 66 | 44.44 (9.79) 28 58 | 38.30 (7.66) 21 47 | 43.93 (9.14) 21 66 |
| CBC COM T score | 43.97 (10.90) 30 80 | 35.83 (4.38) 24 45 | 43.30 (12.09) 21 63 | 41.37 (10.17) 21 80 |
| Technisch lezen z-score | -0.84 (0.86) -2.19 0.76 | -0.71 (0.70) -1.54 0.62 | -0.02 (0.92) -0.87-1.93 | -0.66 (0.86) -2.19 1.93 |
| Spelling z-score | -0.87 (1.20) -3.06 2.47 | -0.53 (1.00) -2.32 1.28 | 0.25 (0.80) -0.69 1.59 | -0.57 (1.14) -3.06 2.47 |
| Rekenen z-score | -1.24 (1.56) -5.49 1.23 | -1.02 (1.32) -4.19 0.59 | -0.10 (0.55) -0.84 0.95 | -0.98 (1.41) -5.49 1.23 |
| Aantal maanden onderwijs | 27.90 (16.94) 4 61 | 25.67 (14.74) 4 54 | 28.80 (18.92) 5 62 | 27.37 (16.40) 4 62 |
| Leeftijd in Maanden | 106.13 (20.43) 79 147 | 102.83 (18.36) 76 137 | 106.30 (21.89) 77 143 | 105.15 (19.79) 76 147 |

Noot: TIQ = Totaal IQ, VIQ = Verbaal IQ, PIQ = Performaal IQ, F1IQ = Factor 1 IQ, F2IQ = Factor 2 IQ, F3IQ = Factor 3 IQ, I = Informatie, O = Overeenkomsten, R = Rekenen, W = Woordenschat, B = Begrijpen, C = Cijferreeksen, OT = Onvolledige Tekeningen, PO = Plaatjes Ordenen, BP = Blokpatronen, FL = Figuur Leggen, SU = Substitutie, DH = Doolhoven, CBC = Child Behavior Checklist, INT = Internaliseren, EXT = Externaliseren, COM = Competentie.

In tabel 3 geven we van deze drie groepen en van de totale proefgroep de gemiddelde score, standaarddeviatie en range (laagste/hogste score) van de verschillende testvariabelen, de chronologische leeftijd en het aantal volledige maanden genoten onderwijs. De Achenbach T-scores werden zo omgezet dat lagere scores steeds wijzen op meer problemen. Gezien de schoolvorderings- en de IQ-scores werden gebruikt om de verschillende groepen te definiëren, heeft het weinig zin om significante verschillen te zoeken tussen deze variabelen. Om eventuele significante verschillen tussen de andere variabelen te ontdekken, werd een Kruskal Wallis-test uitgevoerd (een ANOVA was hier niet bruikbaar, gezien de drie groepen in aantal proefpersonen en spreiding van de testresultaten te veel van elkaar verschillen). We vonden enkel een significant verschil bij de competentiescore van de Achenbach ($\chi^2 = 8,97$, $p = 0.01$) en een tendens tot significantie bij de subschaal externaliseren ($\chi^2 = 4.97$, $p = 0.08$). Deze tendens zou kunnen worden verklaard door het feit dat de kinderen in de restgroep met slechts lichte of geen leerproblemen eigenlijk gedragsproblemen hadden. De lagere competentiescore bij de leermoeilijke kinderen is waarschijnlijk een weerspiegeling van hun lagere begaafdheid.

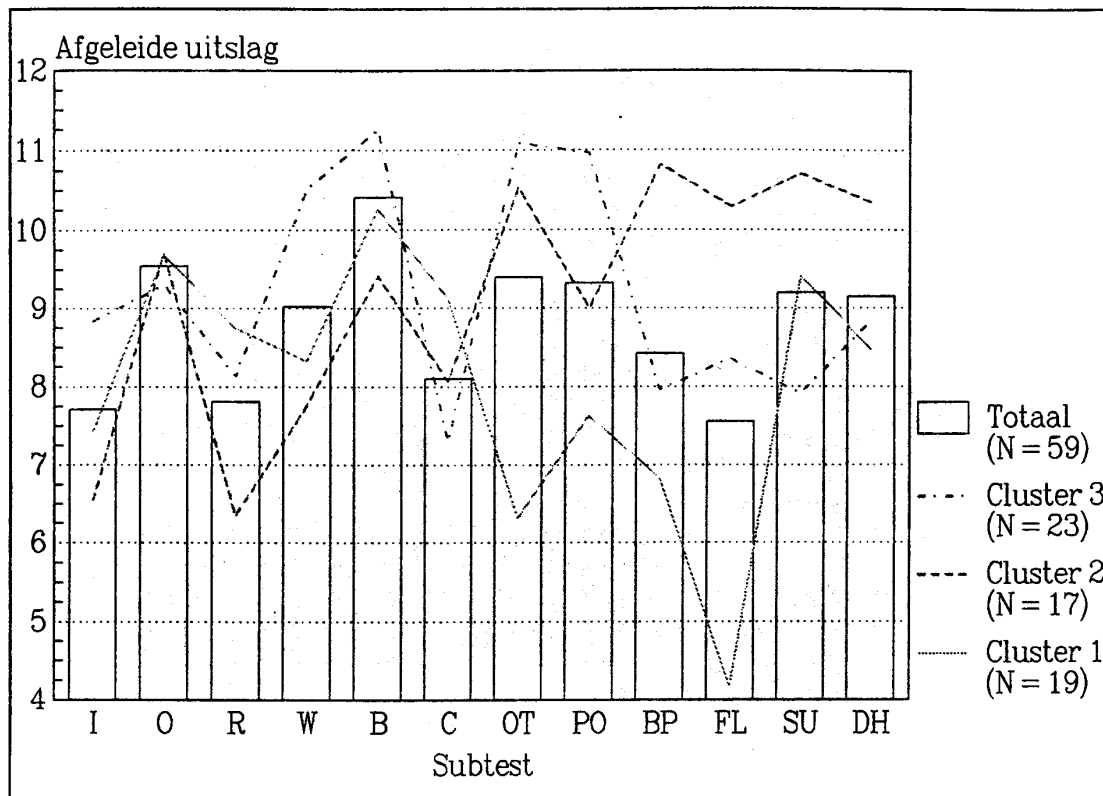


Figuur 1: Dendogram na clusteranalyse volgens Wards methode

4.4.2 Clusteranalyse

Om de eerste hypothese te onderzoeken, gingen we als volgt te werk. Eerst werd voor alle kinderen de gemiddelde WISC-R-standaardscore berekend, en per subtest het verschil tussen die subtest en het gemiddelde. Op deze nieuwe variabelen werd een clusteranalyse uitgevoerd volgens *Wards* methode, met het kwadraat van de euclidische afstand als gelijkheidsmaat. *Wards* methode werd verkozen omdat uit vorig onderzoek bleek dat dit algoritme vaak beter in staat blijkt te zijn om een clusterstructuur terug te vinden (Blashfield, 1976; Morey, Blashfield & Skinner, 1983) en omdat ze berust op een objectief statistisch criterium. Het kwadraat van de euclidische afstand is een afstandsmaat die zowel de hoogte, de spreiding als het profiel van de scores weerspiegelt. Door echter iedere subtestscore van het gemiddelde af te trekken, werd de informatie over de hoogte van de scores "uitgefilterd" (Skinner, 1978). Dit was nodig omdat we geïnteresseerd zijn in de profielen. In figuur 1 geven we het dendrogram dat uit deze analyse resulteerde. Bovenaan op deze figuur vindt men een schaal van 0 tot 25 die aangeeft hoe groot de afstand is tussen twee clusters die worden samengevoegd. Links vindt men de nummers van de proefpersonen.

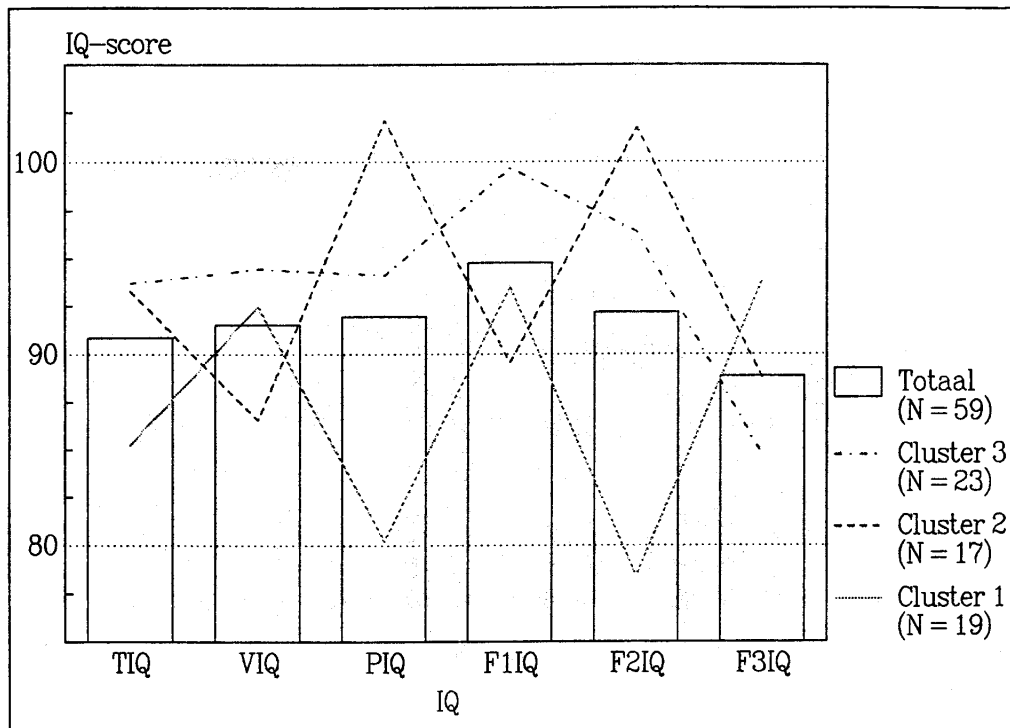
Om te bepalen hoeveel clusters er zouden worden weerhouden, gebruikten we de informatie van het dendrogram, de clustercoëfficiënten en de grafieken van een aantal verschillende mogelijke oplossingen (twee, drie, vier en zes clusters). Er werd gekozen voor een drieclusterstructuur, waarvan men in figuur 2 en 3 de grafieken kan zien. In de grafiek van de clustercoëfficiënten zien we namelijk een lichte knik bij de drieclusteroplossing. Deze drie clusters kunnen tevens goed worden geïnterpreteerd in functie van de driefactorenstructuur van de WISC-R. De eerste cluster scoort vooral zwak op subtests die een beroep doen op ruimtelijk inzicht, de tweede haalt zowel op de verbale factor als op de zogenaamde concentratiefactor lage resultaten en de derde cluster valt uitsluitend uit op de derde factor.



Figuur 2: Grafiek van de WISC-R-subtest scores van de drieclassering volgens Wards methode

Betrouwbaarheid van de clusteranalyse

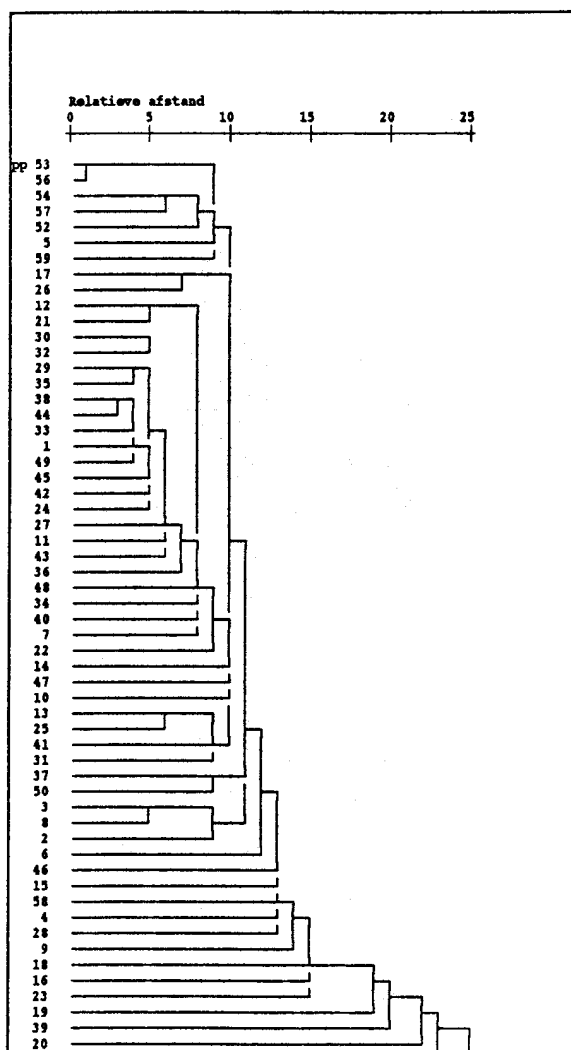
Om de betrouwbaarheid na te gaan, hebben we gebruik gemaakt van twee technieken. Eerst voerden we op dezelfde datamatrix een clusteranalyse uit met een aantal verschillende algoritmen. Daarna construeerden we een controlegroep en voerden op die groep opnieuw een clusteranalyse uit volgens *Wards* methode.



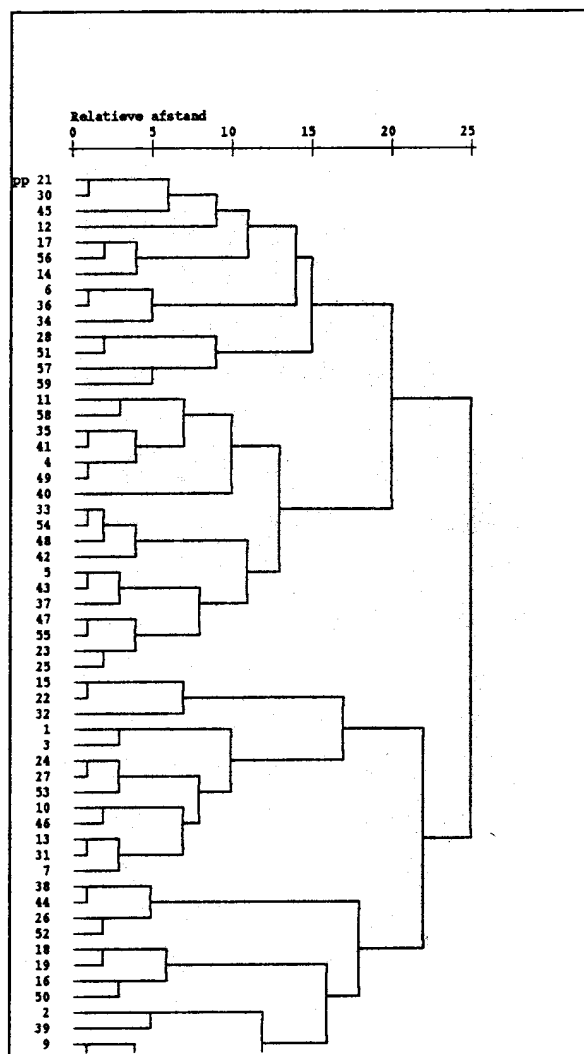
Figuur 3: Grafiek van de WISC-R-IQ-scores van de drieclassificatie volgens Wards methode

Indien de voorgaande classificatie betrouwbaar is, dan zou men steeds een gelijkaardige oplossing moeten vinden, ongeacht de gebruikte clustertechniek. Om dit na te gaan, voerden we op dezelfde datamatrix een analyse uit volgens de enkelvoudige, de volledige en de gemiddelde verbindingmethode. We geven in figuur 4 het dendrogram van de oplossing met de enkelvoudige verbindingmethode. Op dit dendrogram zien we duidelijk het typische keteneffect. Er zijn nog nauwelijks clusters te onderscheiden en deze oplossing werd dan ook niet verder onderzocht. Wanneer we de clusters volgens de minimumvariantiemethode en volgens de volledige en gemiddelde verbinding vergelijken, zien we een classificatie-overeenkomst van respectievelijk 54 procent en 64 procent. Dit wil bijvoorbeeld zeggen dat 38 op de 59 proefpersonen (64 %) door de gemiddelde verbindingmethode en *Wards* methode in een gelijkaardige cluster werden onderverdeeld. Deze procenten worden enigszins omlaag gedrukt door een aantal uitbijters, maar blijven niettemin te laag.

Een tweede mogelijkheid om de betrouwbaarheid na te gaan, bestaat erin dat men dezelfde clusteranalyse uitvoert op een gelijkaardige groep proefpersonen. We construeerden een controlegroep van kinderen die in de voorbije jaren waren aangemeld en waarvan we een volledig WISC-R-profiel hadden. We hanteerden dezelfde exclusiecriteria als die van de proefgroep. Uit een groep van 332 kinderen werden 59 paren gemaakt met de proefgroep. Dit volgens (in die volgorde) geslacht, TIQ en leeftijd. Door middel van een T-toets werd nagegaan of er significante verschillen waren tussen deze twee groepen voor de volgende variabelen: TIQ, TVIQ, TPIQ, F1IQ, F2IQ, F3IQ, de 12 WISC-R-subtestcores, de chronologische leeftijd in maanden en het geslacht. Op geen enkele van deze variabelen daalde het significantieniveau onder de 0,25.

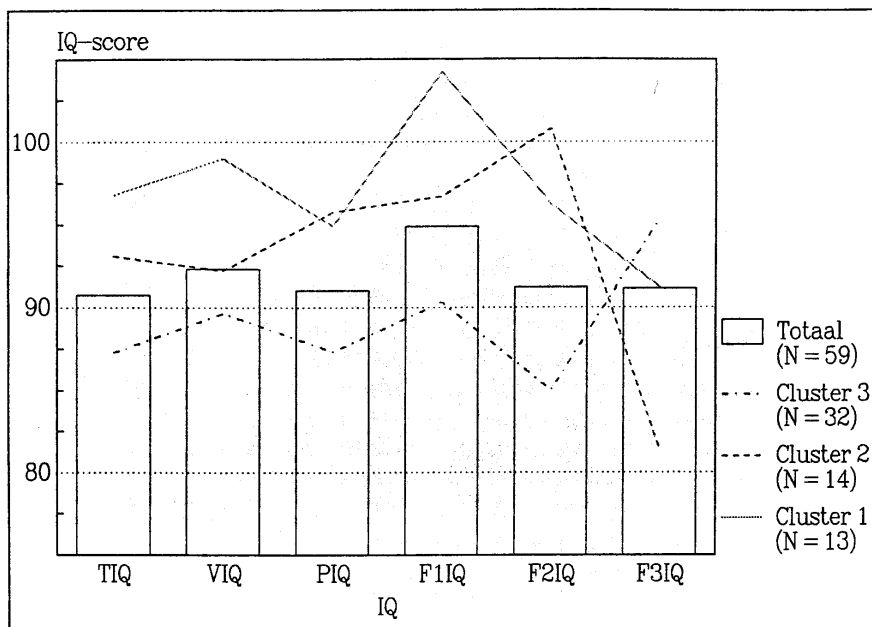


Figuur 4: Dendrogram na clusteranalyse volgens de enkelvoudige verbinding



Figuur 5: Dendrogram na clusteranalyse van de controlegroep volgens Ward's methode

In figuur 5 vindt de lezer het dendrogram en in figuur 6 de verschillende IQ-scores van deze clusteranalyse. Enkel de cluster met de lage scores op de derde factor wordt door de controlegroep duidelijk gerepliceerd. Cluster 1 van de proefgroep komt enigszins overeen met cluster 3 van de controlegroep, maar wel met dat verschil dat de verschillen tussen de IQ-scores opmerkelijk groter zijn bij de proefgroep.



Figuur 6: Grafiek van de WISC-R-scores volgens de clusteroplossing van de controlegroep

Samenvattend kunnen we dus stellen dat de betrouwbaarheid van onze clusteranalyse vrij laag uitvalt. Dit blijkt zowel wanneer we verschillende analyses op dezelfde proefgroep uitvoeren als wanneer we dezelfde analyse op een controlegroep verrichten.

Validiteit van de clusteranalyse

Meestal wordt de validiteit van een clusteranalyse gecontroleerd door na te gaan of de clusters van elkaar te onderscheiden zijn op basis van variabelen die niet in de clusteranalyse waren betrokken. Bij onze proefgroep werden twee multivariate variantie-analyses uitgevoerd (MANOVA). In beide analyses was het lidmaatschap tot een bepaalde cluster de onafhankelijke variabele.

In de eerste analyse waren de schoolvorderingsresultaten voor lezen, spellen en rekenen de afhankelijke variabelen. Het effect van de onafhankelijke variabele was niet significant (Hotellings T^2 0.08, p = 0.61). Bij de verschillende univariate tests werd ook voor de afzonderlijke schoolvorderingsresultaten geen significant effect gevonden. Er was wel een duidelijke tendens aanwezig tot lagere scores op de rekentest bij de cluster met het lage PIQ. Deze resultaten waren als volgt: technisch lezen $F(2,56) = 0,32$, p = 0,73; spelling $F(2,56) = 0,50$, p = 0,61; rekenen $F(2,56) = 1,96$, p = 0.15.

In de tweede analyse waren T-scores van de Achenbach-subschalen (totale problemen, internaliseren, externaliseren en competentie) de afhankelijke variabelen. Ook hier was het effect van de MANOVA en de verschillende ANOVA's niet significant. De resultaten voor de MANOVA

waren als volgt: Hotellings $T^2 = 0.21$, $p. = 0.21$. Voor de verschillende ANOVA's werden de volgende resultaten genoteerd: totale problemen $F(2,56) = 2,71$, $p. = 0,07$; internaliseren $F(2,6) = 1,21$, $p. = 0,31$; externaliseren $F(2,56) = 2,16$, $p. = 0,12$; competentie $F(2,56) = 1,33$, $p. = 0,27$. De tendens is hier dat kinderen met een relatief laag F3IQ minder probleemgedrag vertonen. Er is eveneens een tendens tot een lagere competentiescore bij de eerste cluster.

Samenvattend kunnen we stellen dat de uitgevoerde variantie-analyses onvoldoende bewijzen kunnen aanvoeren ter validering van de clusteranalyse. We zien echter wel een paar interessante tendensen. Zo zien we een tendens tot samenhang tussen een laag PIQ en rekenstoornissen. Deze tendens is in de literatuur herhaaldelijk beschreven, hoewel ook uit de klinische ervaring en de literatuur blijkt dat het hier een correlatie betreft en geen systematische bevinding (Rourke, 1985).

Een andere tendens die bij de analyse tot uiting komt, is dat we bij kinderen met een laag F3IQ ook minder probleemgedrag opmerken. Deze op het eerste gezicht paradoxale bevinding wordt echter duidelijk in het licht van het totale profiel. De kinderen met een laag F3IQ hebben een hoog PIQ en een laag VIQ en zijn in dit opzicht het spiegelbeeld van de andere groep, waar een hoog F3IQ geassocieerd is met een laag PIQ en een hoog VIQ. En het is andermaal uit ervaring en literatuur bekend dat kinderen met een discrepant IQ ten nadele van het PIQ, duidelijk meer kans hebben op probleemgedrag (Rourke, 1991).

Ten slotte kunnen we opmerken dat wanneer we *Wards* methode gebruiken, we een clusterstructuur zien verschijnen die klinisch zeer bekend overkomt: een cluster met vooral kinderen die problemen hebben met visuospatieel inzicht, een cluster met kinderen die sterk scoren op de tweede WISC-R-factor en een groep kinderen die uitvalt op de derde factor.

4.4.3 De tweede hypothese

Aangezien de gevonden clusters onvoldoende betrouwbaar en valide zijn, heeft het weinig zin om in deze populatie verder onderzoek naar de tweede hypothese te verrichten.

5 Discussie

5.1 Inleiding

Uit de illustratie die we gaven, blijkt duidelijk dat clusteranalyse niet zo'n evidente techniek is. Alhoewel een aantal resultaten zeer aannemelijk lijken en verwacht waren, konden ze onvoldoende worden gevalideerd en liet ook de betrouwbaarheid van de clusters te wensen over. We stellen ons eerst de vraag naar het waarom van deze resultaten. Daarna hebben we het meer algemeen over de noodzaak om typologieën te creëren en de problemen die hierbij komen kijken.

5.2 Resultaten van dit onderzoek

Uiteraard rijst de vraag waarom we geen betrouwbare en valide clusters konden vinden in onze proefgroep. In wat volgt overwegen we een aantal mogelijke verklaringen.

5.2.1 Proefgroep

De eerste mogelijkheid zou kunnen zijn dat er in de proefgroep geen subtypes aanwezig waren, m.a.w. dat we te maken hadden met een homogene groep leerlingen. Dit lijkt ons echter zeer onwaarschijnlijk. We hebben reeds een groot aantal literatuurverwijzingen opgegeven waaruit blijkt dat kinderen met leerproblemen een heterogene groep vormen en in verschillende subtypes zijn in te delen. Ten tweede werden er bij de constructie van de proefgroep weinig exclusiecriteria gehanteerd, hetgeen de heterogeniteit in de hand zou moeten werken.

Het is wel mogelijk dat de proefgroep te klein was om alle subtypes tot hun recht te laten komen. Het kiezen van het aantal clusters is steeds een subjectieve zaak. Het spreekt voor zich dat kleine clusters minder tot uiting zullen komen in kleine proefgroepen. Het probleem van de grootte van de proefgroep wordt verder gecompliceerd door de grote verschillen in leeftijd van de onderzochte kinderen. Het is best mogelijk dat de WISC-R-profielen van kinderen met leerproblemen wijzigen met de leeftijd. Bij kleine proefgroepen moeten de gevonden verschillen groter zijn om een bepaald significantieniveau te bereiken. Het is dus niet uitgesloten dat de tendensen die bij de validering van de clusters tot uiting kwamen, wel significant zouden zijn geweest bij een grotere groep. Ten slotte moeten we er ook rekening mee houden dat de proefgroep zeker niet als representatief mag worden beschouwd voor de totale bevolking. Het betreft hier kinderen die werden aangemeld in één revalidatiecentrum en die binnen dit centrum nog eens naar één welbepaalde persoon werden doorverwezen. Er waren dus heel wat filters werkzaam voordat een kind in aanmerking kwam om in de groep te worden opgenomen.

5.2.2 Variabelen

Ten slotte is er het probleem van de variabelen. De WISC-R-subtestscores hebben het grote voordeel dat ze algemeen gekend zijn en algemeen worden gebruikt, en dat er reeds heel wat onderzoek is verricht naar de factorstructuur. Daartegenover staat dat de WISC-R geen neuropsychologische testbatterij is. De subtests van deze intelligentietest zijn vrijwel allemaal zeer complex van aard. Bijgevolg is het moeilijk om ze elk afzonderlijk eenduidig te interpreteren (dit is trouwens ook nooit de bedoeling geweest bij de constructie van deze test). De subtest Woordenschat van de WISC-R meet bv. niet enkel de productieve woordenschat. Doordat men meer punten geeft voor abstract geformuleerde woorden, doet men in een niet onbelangrijke mate ook een beroep op verbale vlotheid en verbale conceptvorming. Bij een neuropsychologische testbatterij zal men juist streven naar taken die zo eenduidig mogelijk een bepaalde vaardigheid meten en dan nog liefst een vaardigheid waar men een duidelijk verband kan leggen met de werking van een bepaalde regio in de hersenen. Natuurlijk zullen we nooit

beschikken over een test die 100 procent "zuiver" is, maar er is wel degelijk een verschil in gradatie mogelijk.

Verder speelt in een neuropsychologische testbatterij ook het geheel van de tests een zeer belangrijke rol. Men zal proberen om de batterij zodanig te construeren dat men de werking van de belangrijkste hersengebieden kan isoleren.

5.2.3 Implicaties voor verder onderzoek

Om rekening te houden met de hierboven geformuleerde bezwaren zouden we gebruik kunnen maken van een grotere proefgroep. Naast de testvariabelen die wij gebruikten, zouden er tevens een aantal neuropsychologische tests moeten worden afgenomen, zodat men zou kunnen nagaan of er op basis van deze variabelen een meer stabiele en valide clusteroplossing kan worden gevonden. Ten slotte zou het ook aan te raden zijn om het effect van de leeftijd onder controle te houden. Dit zou bijvoorbeeld kunnen door verschillende proefgroepen te construeren, waarbij per proefgroep het leeftijdsbereik zo beperkt mogelijk wordt gehouden.

5.3 Hanteren van typologieën

Tot slot willen we het nog even hebben over het gebruik van typologieën. Waarom is het belangrijk om kinderen in subtypes in te delen? Een dergelijke werkwijze is niet zonder gevaar. Het grootste risico is zonder twijfel dat de typologie rigide wordt gehanteerd. De ervaring leert dat mensen zich niet altijd eenduidig in een bepaalde categorie laten plaatsen. En dan ontstaat maar al te vaak de neiging om de persoon aan te passen aan de typologie (en niet omgekeerd). Dit gevaar is waarschijnlijk nog groter wanneer de subtypes tot stand kwamen met behulp van ingewikkelde en gesofisticeerde statistische technieken. Enerzijds is er minder controle mogelijk wanneer er minder mensen op de hoogte zijn van de techniek en anderzijds bestaat er nog steeds een soort naïef geloof dat ingewikkelde of mathematische methodologieën noodzakelijkerwijze de beste zijn.

Er zijn echter ook belangrijke voordelen verbonden aan het gebruik van typologieën. Ze laten ons toe om onze kennis en inzichten te ordenen en om patronen te ontdekken die we anders niet zouden zien. Op zijn beurt laat dit ons dan weer toe om (controleerbare) prognoses te maken. Het kan ook onze aandacht richten op bepaalde aspecten van het probleem die we anders gemakkelijk over het hoofd zouden zien. Bij het zogenaamde NLD-syndroom (Rourke, 1989) weten we bv. dat deze kinderen over een grote verbale vlotheid beschikken maar dat de taalpragmatiek gestoord is. Wanneer men dat niet bij voorbaat weet, bestaat er een zeer reële kans dat men dit over het hoofd zal zien.

De typologie moet uiteraard wel zijn bruikbaarheid en zijn effectiviteit kunnen bewijzen. Ze moet relevant zijn voor de gekozen problematiek, in staat zijn om controleerbare voorspellingen te formuleren, en gefundeerde aanwijzingen geven omtrent etiologie en behandeling. Zo kan men

bv. nagaan of de gevonden types van elkaar te differentiëren zijn door middel van neurofysiologische en beeldvormende technieken zoals uitgelokte potentialen en NMR-, PET- en CT-scan. Recente onderzoeken met uitgelokte potentialen hebben op dit gebied trouwens reeds beloftevolle resultaten opgeleverd (Bakker, Licht, Kok & Bouma, 1980; Bakker, Licht & van Strien, 1991; Dool, Stelmack & Rourke, 1993; Licht, Bakker, Kok & Bouma, 1992; Licht, Kok, Bakker & Bouma, 1986; van Strien, Bouma & Bakker, 1993).

Pas indien ook aan deze voorwaarden wordt voldaan, kan men spreken van een valide classificatie. Het gebruik van een gesofisticeerde technologie zoals de clusteranalyse *kan* een hulpmiddel zijn bij de eerste stappen in de constructie van de subtypes. Het is echter niet meer dan dat: een hulpmiddel bij het zetten van de eerste stappen, en er bestaat wel degelijk een reëel gevaar dat men meent dat deze eerste stappen een einddoel zijn.

Dankwoord

De auteurs danken prof. dr. André Vandierendonck van de Faculteit Psychologie en Pedagogische Wetenschappen, Vakgroep Experimentele Psychologie, Universiteit Gent. Zijn adviezen waren onmisbaar bij het tot stand komen van dit artikel.

Literatuur

- Aldenderfer, M.S., Blashfield, R.K. (1978). Computer programs for performing hierarchical cluster analysis. *Applied Psychological Measurement*, 2, 405-411.
- Adams, K.M. (1985). Theoretical, methodological, and statistical issues. In: B.P. Rourke (ed.), *Neuropsychology of learning disabilities: Essentials of subtype analysis*. New York: Guilford Press.
- Bakker, D.J. (1992). Neuropsychological classification and treatment of dyslexia. *Journal of Learning Disabilities*, 25, 102-109.
- Bakker, D.J., Licht, R., Kok, A., Bouma, A. (1980). Cortical responses to word reading by right- and left-eared normal and reading-disturbed children. *Journal of Clinical Neuropsychology*, 2, 1-12.
- Bakker, D.J., Licht, R., van Strien, J. (1991). Biopsychological validation of L- and P-type dyslexia. In: B.P. Rourke (ed.), *Neuropsychological validation of learning disability subtypes*. New York: Guilford Press.
- Blashfield, R.K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83, 377-388.
- Blashfield, R.K., Aldenderfer, M.S. (1978). Computer programs for performing iterative partitioning cluster analysis. *Applied Psychological Measurement*, 2, 533-541.
- Cracco, J. (1993). Niet-verbale leerproblemen; een overzicht van Rourke's model. *Tijdschrift voor Klinische Psychologie*, 23, 25-46.
- Cracco, J., Thiery, T. (1993). Neuropsychologische assessment en behandelingsplan van een meisje met niet-verbale leerproblemen: een gevalstudie. *Acta Ergotherapeutica Belgica*, 5, 103-111.

- DeLuca, J.W., Adams, K.A., Rourke, B.P. (1991). Methodological and statistical issues in cluster analysis. In: B.P. Rourke (ed.), *Neuropsychological validation of learning disability subtypes*. New York: Guilford Press.
- Dool, C.B., Stelmack, R.M., Rourke, B.P. (1993). Event-related potentials in children with learning disabilities. *Journal of Clinical Child Psychology*, 22, 387-398.
- Edelbrock, C. (1979). Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. *Multivariate Behavioral Research*, 14, 367-384.
- Feagans, L., Appelbaum, M. (1986). Validation of language subtypes in learning disabled children. *Journal of Educational Psychology*, 78, 358-364.
- Fiedorowicz, C., Trites, R. (1991). From theory to practice with subtypes of reading disabilities. In: B.P. Rourke (ed.), *Neuropsychological validation of learning disability subtypes*. New York: Guilford Press.
- Fisk, J.L., Rourke, B.P. (1979). Identification of subtypes of learning-disabled children at three age levels: A neuropsychological, multivariate approach. *Journal of Clinical Neuropsychology*, 1, 289-310.
- Fletcher, J.M., Satz, P. (1985). Theoretical, methodological, and statistical issues. In: B.P. Rourke (ed.), *Neuropsychology of learning disabilities: Essentials of subtype analysis*. New York: Guilford Press.
- Fuerst, D., Fisk, J., Rourke, B.P. (1989). Psychosocial functioning of learning-disabled children: Replicability of statistically derived subtypes. *Journal of Consulting and Clinical Psychology*, 57, 275-280.
- Fuerst, D., Fisk, J., Rourke, B.P. (1990). Psychosocial functioning of learning-disabled children: Relations between WISC Verbal IQ-Performance, IQ Discrepancies and Personality Subtypes. *Journal of Consulting and Clinical Psychology*, 58, 657-660.
- Hintze, J.L. (1987). *Number Cruncher Statistical System, Version 5.0*. Kaysville.
- Kaufman, A.S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley and sons.
- Kaufman, A.S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 and 16 Years. *Journal of Clinical and Consulting Psychology*, 43, 135-147.
- Kolb, B., Whishaw, I.Q. (1990). *Fundamentals of human neuropsychology*. New York: W.H. Freeman.
- Korhonen, T. (1991a). An empirical subgrouping of finnish learning-disabled children. *Journal of Clinical and Experimental Neuropsychology*, 13, 259-277.
- Korhonen, T. (1991b). Neuropsychological stability and prognosis of subgroups of children with learning disabilities. *Journal of Learning Disabilities*, 24, 48-57.
- Licht, R., Bakker, D J., Kok, A., Bouma, A. (1992). Grade-related changes in event-related potentials (ERPs) in primary school children: Differences between two reading tasks. *Journal of Clinical and Experimental Neuropsychology*, 14, 193-200.
- Licht, R., Kok, A., Bakker, D.J., Bouma, A. (1986). Hemispheric distribution of ERP components and word naming in preschool children. *Brain and Language*, 27, 101-116.
- Luria, A.R. (1973). *Grondslagen van de neuropsychologie*. Deventer: Van Loghum Slaterus.
- Lyon, G., Flynn, J. (1991). Educational validation studies with subtypes of learning-disabled readers. In: B.P. Rourke (ed.), *Neuropsychological validation of learning disability subtypes*. New York: Guilford Press.
- Lyon, R., Stewart, N., Freedman, D. (1982). Neuropsychological characteristics of empirically derived subgroups of learning disabled readers. *Journal of Clinical Neuropsychology*, 4, 343-365.

- Lyon, R., Watson, B. (1981). Empirically derived subgroups of learning disabled readers: Diagnostic characteristics. *Journal of Learning Disabilities*, 14, 256-261.
- Lyon, R., Watson, B., Reitta, S., Porch, B., Rhodes, J. (1981). Selected linguistic and perceptual abilities of empirically derived subgroups of learning disabled readers. *Journal of School Psychology*, 19, 152-166.
- McKinney, J. (1984). The search for subtypes of specific learning disability. *Journal of Learning Disabilities*, 17, 43-50.
- Morey, L.C., Blashfield, R.K., Skinner, H.A. (1983). A comparison of cluster analysis techniques within a sequential validation framework. *Multivariate Behavioral Research*, 18, 309-329.
- Morris, R. (1988). Classification of learning disabilities: Old problems and new approaches. *Journal of Consulting and Clinical Psychology*, 56, 789-794.
- Morris, R., Blashfield, R., Satz, P. (1981). Neuropsychology and cluster analysis: Potentials and problems. *Journal of Clinical Neuropsychology*, 3, 79-99.
- Morris, R., Blashfield, R., Satz, P. (1986). Developmental classification of reading-disabled children. *Journal of Clinical and Experimental Neuropsychology*, 8, 371-392.
- Petrauskas, R.J., Rourke, B.P. (1979). Identification of subtypes of retarded readers: A neuropsychological, multivariate approach. *Journal of Clinical Neuropsychology*, 1, 17-37.
- Reynolds, C.R. (1981). The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. *The Journal of School Psychology*, 19, 350-358.
- Rourke, B.P. (ed.) (1985). *Neuropsychology of learning disabilities: Essentials of subtype analysis*. New York: Guilford Press.
- Rourke, B.P. (1988). Socioemotional disturbances of learning disabled children. *Journal of Consulting and Clinical Psychology*, 56, 801-810.
- Rourke, B.P. (1989). *Nonverbal Learning Disabilities: the syndrome and the model*. New York: Guilford Press.
- Rourke, B.P. (ed.) (1991). *Neuropsychological validation of learning disability subtypes*. New York: Guilford Press.
- Shaywitz, B.A., Fletcher, J.M., Holahan, J.M., Shaywitz, S.E. (1992). Discrepancy definitions compared to low achievement definitions of reading disability: Results from the Connecticut longitudinal study. *Journal of Learning Disabilities*, 25, 639-648.
- Siegel, L.S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities*, 22, 469-478.
- Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, 25, 618-629.
- Skinner, H.A. (1978). Differentiating the contribution of elevation, scatter and shape in profile similarity. *Educational and Psychological Measurement*, 38, 297-308.
- Skinner, H.A. (1981). Toward the integration of classification theory and models. *Journal of Abnormal Psychology*, 90, 68-87.
- Spreen, O., Haaf, R. (1986). Empirically derived learning disability subtypes: A replication attempt and longitudinal patterns over 15 years. *Journal of Learning Disabilities*, 19, 170-180.
- Van Haasen, P., De Bruyn, E., Pijl, Y., Poortinga, Y., Spelberg, H., Vander Steene, G., Coetsier, P., Spoelders-Claes, R., Stinissen, J. (1986). *WISC-R, scoring en normen*. Lisse: Swets & Zeitlinger.

- Van Strien, J.W., Bouma, A., Bakker, D J. (1993). Lexical decision performances in P-type dyslexic, L-type dyslexic, and normal reading boys. *Journal of Clinical and Experimental Neuropsychology*, 15, 516-524.
- Verhulst, F.C., Koot, J.M., Akkerhuis, G.W., Veerman, J.W. (1990). *Praktische handleiding voor de CBCL*. Assen: Van Gorum.
- Watson, B., Goldgar, D. (1983). Subtypes of reading disability. *Journal of Clinical Neuropsychology*, 5, 377-399.



Significant

- Elektronisch wetenschappelijk tijdschrift
- voor klinische research en reviews
- in revalidatie en psychosociale hulpverlening

Redactiesecretariaat **Significant**

Vormingsdienst SIG (Stichting Integratie Gehandicapten vzw)
Kerkham 1 bus 2, 9070 Destelbergen (België)
tel. +32 (0)9 238 31 25 - fax +32 (0)9 238 31 40
email: vormingsdienst.sig@skynet.be
<http://users.skynet.be/vsig>

Significant

Elektronisch wetenschappelijk
tijdschrift
voor klinische research en reviews
voor revalidatie en psychosociale
hulpverlening

Hoofdredactie

Jan Scheiris (SIG)

Kernredactie

Prof. Dr. C. Andries (VU Brussel)
Prof. Dr. H. Roeyers (U Gent)
Prof. Dr. E. Thiery (U Gent)
Prof. Dr. I. Zink (KU Leuven)
Dr. H. Hellemans (AKZ Antwerpen)
Joris Cracco (SIG)
Herman Van Hove (SIG)

Redactieraad

Prof. Dr. P.P. De Deyn (U Antwerpen)
Prof. Dr. J.P. Fryns (KU Leuven)
Prof. Dr. P. Ghesquière (KU Leuven)
Dr. J. Simons (KU Leuven)
Prof. Dr. H.J.M. Hermans (KU Nijmegen)
Prof. Dr. E. Vandenbussche (KU Leuven)
Prof. Dr. G. Van Hove (U Gent)
Prof. Dr. M. Vanryckeghem (U Central
Florida)
Drs. H. Van Waelvelde (KaHoG)
Prof. Dr. H. Vertommen (KU Leuven)
Prof. Dr. S. Westreich (VU Brussel)

Redactiesecretariaat

Vormingsdienst SIG
(Stichting Integratie Gehandicapten vzw)
Geert Andries
Kerkham 1 bus 2, 9070 Destelbergen (B)
tel. +32 (0)9 238 31 25 - fax 238 31 40
email: vormingsdienst.sig@skynet.be
<http://users.skynet.be/vsig>

Verantwoordelijke uitgever

Jan Scheiris
Stichting Integratie Gehandicapten vzw
Kerkham 1
9070 Destelbergen

Significant is een uitgave van
Vormingsdienst SIG (Stichting
Integratie Gehandicapten vzw) en
verschijnt vier keer per jaar op het
internet:

<http://users.skynet.be/vsig>

Abonneren of artikels los bestellen of
nabestellen is mogelijk:

- online via de website
- per post, fax of telefoon via het
redactiesecretariaat

Vormingsdienst SIG

Kerkham 1 bus 2, 9070 Destelbergen
tel. +32 (0)9 238 31 25
fax +32 (0)9 238 31 40

abonnement:

24,79 euro / 1000 frank voor 12
artikels

los bestellen:

2,48 euro / 100 frank per artikel

Significant

Elektronisch wetenschappelijk tijdschrift
voor klinische research en reviews
voor revalidatie en psychosociale hulpverlening

Nummer 1 Redactioneel

Jan Scheiris, hoofdredacteur

Significant 0, ons proefnummer, werd goed onthaald. De interesse, af te leiden uit het aantal bestelde artikels (en er zijn zelfs al abonnees), was voor de kernredactie overtuigend genoeg om ermee door te gaan.

Ook Significant 1 biedt u drie kwaliteitsartikels aan die voldoen aan onze hoofddoelstelling: informatie geven van een gedegen wetenschappelijk niveau, toegankelijk en snel beschikbaar en met een hoge klinische praktijkrelevantie.

Subtypering, classificatie, 'labeling' ... ze hebben tot doel orde te brengen in de vaak onoverzichtelijke hoeveelheid problematieken die zich in de revalidatie en psychosociale hulpverlening aandienen. Wie een heldere uitleg, inclusief een interessant voorbeeld, wil lezen over een van de methoden die wordt gebruikt om subtypes op het spoor te komen en te onderbouwen, vindt zijn gading in het artikel over **clusteranalyse** van *Joris Cracco* en *Evert Thiery*. Beiden bereiden trouwens een gelijkaardige bijdrage voor over een andere techniek, de factoranalyse, wat wellicht iets wordt voor Significant 2.

'Evidence based', wetenschappelijk onderbouwd werken ... het is al geruime tijd ook in ons werkveld een streven dat hoog in het vaandel wordt geschreven. *Johan Simons* heeft het over de wetenschappelijke fundering van de **psychomotorische diagnostiek en therapie**. Hij bespreekt o.a. reviews en meta-analyses over de effectiviteit van de psychomotorische discipline en gaat hierbij zowel in op de kwaliteit van de effectstudies als op wat ze leren over de waarde van deze aanpak in het licht van bepaalde doelstellingen en problematieken.

Wetenschappelijk onderzoek in het domein van de leerstoornissen werd lange tijd beheerst door dyslexie. Meer en meer groeit de aandacht voor kinderen die onverwachte en hardnekkige problemen hebben met die andere belangrijke schoolse vaardigheid, het rekenen. Wie op zoek is naar een duidelijke operationalisering van het begrip **rekenstoornis**, zijn afgrenzing van het begrip 'rekenprobleem' en de resultaten wil kennen van een onderzoek naar cognitieve profielen bij deze kinderen, kan terecht bij het artikel van *Annemie Desoete* en *Herbert Roeyers*.

Veel leesgenot! Uw reacties kunt u kwijt in de rubriek *Woord en wederwoord*.