

Enkele mogelijkheden van kwantitatieve evaluatie van screeningtests

J. M. A. HERMANNs en A. T. M. COOLS

Afdeling ontwikkelingspsychologie van het Instituut voor Pedagogische en Andragogische wetenschappen der R.U., Utrecht.

Samenvatting

Enkele manieren om de validiteit van screeningtests uit te drukken worden besproken. Aan de orde komen de parameters: copositiviteit, conegativiteit, overreferral ratio, underreferral ratio en totale screeningfout.

De afhankelijkheid van deze maten van de base-rate en het 'cutting-off-point' worden besproken. Ook de begrippen 'valse positieven' en 'valse negatieven' worden hieraan gerelateerd. Een en ander wordt toegelicht aan de hand van onderzoek naar de Denver Developmental Screening Test.

Tenslotte worden enkele onderzoeken naar psychologische screeningtests besproken en in bovenstaand kader geplaatst.

1. Inleiding

In de ontwikkelingspsychologische en sociaal-medische literatuur tref men de laatste jaren een aantal publikaties aan over 'screening'-tests (bijv. Frankenburg e.a. 1971; Schlesinger en van Wieringen 1972; v. Wiechen 1972). In een screeningprocedure wordt een populatie in twee groepen gesplitst: één groep - meestal de grootste - die een bepaald kenmerk met een grote waarschijnlijkheid niet bezit, en een andere - meestal kleinere - groep die het kenmerk met een grote waarschijnlijkheid wél bezit.

Over het algemeen zijn deze tests snel en eenvoudig af te nemen. Kenmerken waarop gescreend wordt zijn bijvoorbeeld: leerstoornissen; problemen in de auditieve en visuele perceptie; spasticiteit; phenylketonurie. De individuen die volgens de test het kenmerk bezitten zullen vervolgens een uitgebreid onderzoek moeten ondergaan. Daarin tracht men tot de vaststelling te komen dat het bedoelde kenmerk in feite al dan niet aanwezig is. Over het algemeen neemt men aan dat naarmate detectie vroeger plaats vindt de daarop volgende behandeling c.q. begeleiding meer succes heeft (Gresnigt en Gresnigt, 1973; Hoejenbos, 1971; Kalverboer 1973).

Eén van de onderwerpen, die binnen dit kader momenteel in de belangstelling staat, is het vroegtijdig opsporen van ontwikkelingsstoornissen bij (zeer) jonge kinderen, bijv. mentale retardatie, spasticiteit, culturele deprivatie.

Zowel binnen de testresearch als in organisaties (met name de gezondheidszorg) waarin de efficiëntie van screeningprocedures geëvalueerd moet worden doet zich de vraag voor naar kwantitatieve maten om de waarde van een screeningtest uit te drukken. In de research betreft het de vraag naar de validiteit van de test.

In organisatorische kaders betreft het de kosten-baten problematiek.

In het onderstaande zullen enkele van deze maten besproken worden. Verder zal ingegaan worden op de wijze van interpretatie en gebruiksmogelijkheden hiervan. Dit zal mede gebeuren aan de hand van een validiteitsonderzoek van Frankenburg en anderen betreffende de Denver Developmental Screening Test (Frankenburg, Goldstein en Camp 1971).

2. De validiteit van een screeningtest

2.1. Omschrijving validiteit

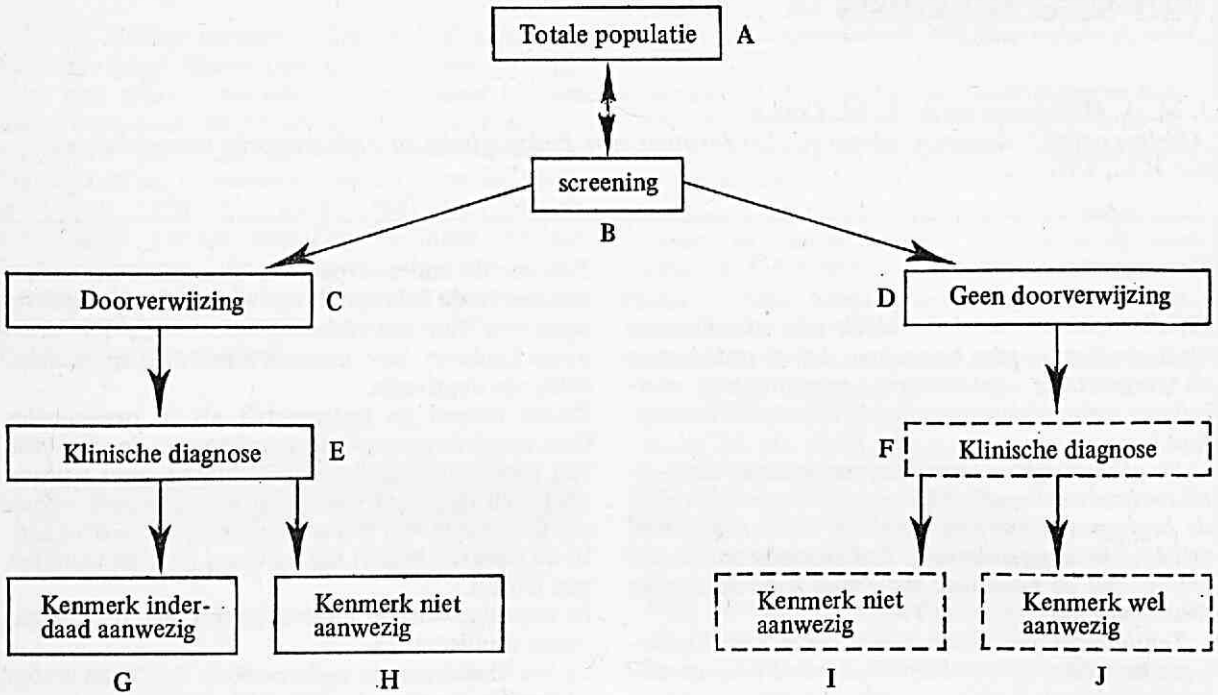
De vraag naar de validiteit van een test wordt hier als volgt opgevat: kan men door de test bereiken wat men zich ten doel gesteld heeft, nl. het efficiënt verdelen van een populatie in twee subpopulaties. Het gaat hier niet om de begripsvaliditeit, maar om de predictieve validiteit. Predictie wordt opgevat als voorspelling betreffende scores of uitslagen die met behulp van andere beoordelingsmiddelen verkregen zijn (Drenth 1966).

2.2. De screeningprocedure

Men kan het gebruik van een screeningtest weer-geven in een schema (zie figuur 1).

Doorgaans zal in deze procedure de rechterarm van

Figuur 1: Schematische weergave screeningprocedure en onderzoek van zijn predictieve validiteit.



het schema bij D ophouden: er vindt geen doorverwijzing plaats. Wanneer men de groep in D niet-temin toch diagnostiseert zal dit resulteren in een verdeling van de totale populatie over de elementen G, H, I en J.

Aan de hand van deze verdeling kan men de validiteit of de efficiëntie van de screeningstest bepalen. Naarmate er meer leden van de populatie in G en I en minder in H en J aanwezig zijn, zal de test meer valide genoemd kunnen worden.

2.3. Correlatiecoëfficiënten als validiteitsmaten

De predictieve validiteit van een psychologische test wordt veelal uitgedrukt in een correlatiecoëfficiënt (bijv. Pearson Product Moment, Punt Biseriële of Phi-coëfficiënt). Deze maten geven aan in hoeverre twee tests onderling samenhangen. Een dergelijke maat geeft echter nooit informatie over het aantal foutieve beslissingen dat genomen wordt in een ja-nee beslissingssituatie (Taylor en Russell 1939), zoals deze zich bij een screening voordoet. Bij het schatten van de predictieve validiteit van een screeningstest in termen van het schema in figuur 1, geven

de correlatiecoëfficiënten derhalve slechts globale, weinig bruikbare informatie.

2.4. Andere validiteitsschattingen

De resultaten van de screeningprocedure zoals hierboven beschreven kunnen in een 2×2 tabel beschreven worden (zie tabel 1).

Tabel 1: Twee bij twee tabel om een gescreende populatie onder te verdelen.

screening	Diagnose		
	Kenmerk aanwezig	Kenmerk afwezig	
Positief*	a	b	a + b
Negatief**	c	d	c + d
	a + c	b + d	a + b + c + d

* positief = door de screeningstest doorverwezen.

** negatief = door de screeningstest niet doorverwezen.

Tot deze verdeling kan men ook komen middels verschillende schattingsprocedures. Deze worden beschreven door Grant (1974).

De tabel kan als volgt beschreven worden: In kwadrant a bevinden zich de gevallen die volgens de screeningtest én diagnose het gezochte kenmerk bezitten (de werkelijke positieven).

In kwadrant d bevinden zich de gevallen die zowel volgens de screeningtest als volgens de diagnose het gezochte kenmerk niet bezitten (de werkelijke negatieven). In kwadrant c bevinden zich de gevallen die volgens de screeningtest het kenmerk niet bezitten, echter volgens de diagnose wél (de valse negatieven). In kwadrant b bevinden zich de gevallen die volgens de screeningtest het kenmerk bezitten, echter volgens de diagnose niet (de valse positieven).

Met behulp van bovenstaande tabel is het mogelijk de predictieve validiteit van een screeningtest te bepalen. Hiertoe kan de vraag naar de overeenstemming tussen screeningtest en diagnose in de volgende onderdelen ontleed worden:

- I. Welk deel van de gevallen die volgens de diagnose het kenmerk bezitten wordt door de screeningtest positief genoemd?
- II. Welke deel van de gevallen die volgens de diagnose het kenmerk niet bezitten wordt door de screeningtest negatief genoemd?
- III. Hoeveel van de door de screeningtest opgespoorde gevallen zijn volgens de diagnose ten onrechte opgespoord?
- IV. Hoeveel van de door de screeningtest niet opgespoorde gevallen zijn volgens de diagnose ten onrechte niet opgespoord?
- V. Welk gedeelte van het totale aantal gescreende gevallen is door de screeningtest fout beoordeeld?

Vraag I wordt beantwoord door de ratio $\frac{a}{a + c}$.

Deze noemt men de *copositiviteit*(1) van de screeningtest. Vraag II wordt beantwoord door de ratio $\frac{d}{b + d}$: de *conegativiteit*(1) van de screeningtest.

Vraag III wordt beantwoord door de ratio $\frac{b}{a + b}$.

Deze wordt de *overreferral ratio* genoemd. Vraag

IV wordt beantwoord door de ratio $\frac{c}{c + d}$ Deze

wordt de *underreferral ratio* genoemd. Vraag V wordt beantwoord door de ratio $\frac{b + c}{a + b + c + d}$.

Wij spreken hier van de totale *screeningfout*.

Deze en andere mogelijke validiteitsschattingen van screeningtest zullen verder worden besproken.

3. Onderzoek Frankenburg en anderen naar de D.D.S.T. (1971)

3.1. De Denver Developmental Screening Test (D.D.S.T.)

In de V.S. is sinds enige jaren de D.D.S.T. een veelgebruikte test. De D.D.S.T. is een screeningtest voor kinderen van twee weken tot 6 jaar. De afname (5 tot 20 minuten) resulteert in een uitspraak dat het testresultaat 'abnormal', 'questionable' of 'normal' is.

Bij testgebruik in de praktijk worden de kinderen met een 'abnormal' of 'questionable' testresultaat voor verder onderzoek doorverwezen.

De D.D.S.T. is bewerkt en genormeerd voor de Nederlandse populatie door de schrijvers van dit artikel (Cools en Hermanns, in press).

Als het gaat om het opsporen van 'ontwikkelingsstoornissen', wat het doel van deze test is, zijn veelal meerdere criteria nodig om de diagnose E en F (zie fig. 1) te kunnen stellen.

In het onderstaande zullen echter als voorbeeld slechts scores op een ontwikkelingstest of intelligentie-test, ongeveer gelijktijdig of later dan de screeningtest afgenomen, gehanteerd worden.

3.2. Resultaten van het onderzoek

Frankenburg en anderen hebben de resultaten van de D.D.S.T. afgezet tegen de scores op de Revised Bayley Infant Scale of de Stanford-Binet form L.M. (Frankenburg e.a. 1971).

We zullen ons hier beperken tot de gegevens van het crossvalideringsonderzoek. In totaal werden 1.292 kinderen van 0-6 jaar gescreend. In de categorie 'abnormal' kwamen 20 kinderen terecht en in de categorie 'questionable' 54. Deze kinderen zullen we verder de 'positieven' noemen. Bij gebruik in de praktijk van de D.D.S.T. zou dus deze groep van 74 kinderen voor verder onderzoek worden doorverwezen. Van de overige 'negatieven' werd een willekeurig gekozen groep van 172 kinderen ook bij het onderzoek betrokken.

Er werden dus 246 kinderen met de Bayley (tot 2 jaar) en de Stanford Binet (vanaf 2 jaar) onderzocht. Indien een kind een DQ of IQ hoger dan of gelijk aan 80 haalde werd het 'negatief' genoemd. Als het DQ of IQ lager dan 80 was: 'positief'. De overeenstemming van deze beide tests met de D.D.S.T. kan in een tabel worden weergegeven (zie tabel 2). (2)

Tabel 2: Overeenstemming D.D.S.T.-resultaat en IQ/DQ (RBIS en Stanford-Binet) bij kinderen van 0 tot 6 jaar.
N = 246.

	IQ/DQ		tot.
	< 80	≥ 80	
abn. quest.	32	42	74
D.D.S.T. norm.	10	162	172
tot.	42	204	246

Daar echter het aantal valse negatieven en werkelijke negatieven berekend is op grond van een steekproef van de door de screeningstest negatief genoemde kinderen dient deze tabel gecorrigeerd te worden.

In kwadrant c komt te staan $\frac{10}{172} \times (1292 - 74) = 70$;

en in kwadrant d komt te staan $\frac{162}{172} \times (1292 - 74) = 1148$ (zie tabel 3).

Tabel 3: Geschatte overeenstemming D.D.S.T.-resultaat en IQ/DQ (RBIS en Stanford-Binet) bij kinderen van 0 tot 6 jaar.
N = 1292.

	IQ/DQ		tot.
	< 80	≥ 80	
abn. quest.	32	42	74
D.D.S.T. norm.	70	1148	1218
tot.	102	1190	1292

4. De validiteitsschattingen

4.1. Copositiviteit en conegativiteit

De copositiviteit van de D.D.S.T. in het voorbeeld van 3.2 is:

$$\frac{a}{a + c} = .31.$$

De conegativiteit is hier:

$$\frac{d}{d + b} = .96.$$

In de bovengenoemde validiteitstudie van Frankenburg e.a. wordt een copositiviteit van .92 genoemd. Gedeeltelijk is dit verschil te verklaren door het feit dat Frankenburg de 'questionable' gevallen niet bij zijn berekeningen betrokken heeft. Korrigiert men zijn gegevens hiervoor dan zou volgens zijn berekeningswijze de copositiviteit .76 zijn. Het overblijvende verschil (.76-.31) wordt verklaard door het niet corrigeren van tabel 2. Dit had moeten gebeuren omdat er een steekproef getrokken is uit het totale aantal door de screeningstest negatief genoemde kinderen. Ook t.a.v. de conegativiteit heeft Frankenburg geen correctie toegepast. Deze was zonder correctie .79.

4.2. Overreferral en underreferral ratio

De ratio van overreferral is in het voorbeeld

$$\frac{b}{a + b} = .57.$$

De ratio van underreferral is $\frac{c}{c + d} = .06$.

Een verklaring voor de hoge ratio van overreferral in dit voorbeeld (.57) zou kunnen zijn dat er slechts één criteriumtest gebruikt is i.p.v. multiple criteria.

De D.D.S.T. bleek namelijk ook kinderen op te sporen, die op 4-jarige leeftijd nog niet konden lopen terwijl ze wel een IQ > 80 haalden (Frankenburg, pers. comm.).

4.3. Totale screeningfout

De totale screeningfout is hier $\frac{b + c}{a + b + c + d} = .08$.

5. Wanneer is een screeningstest effectief?

5.1. Twee soorten beslissingsfouten.

Een screeningstest kan op twee manieren onnauwkeurig zijn:

er kunnen gevallen opgespoord worden die niet opgespoord hadden moeten worden en er kunnen gevallen niet opgespoord worden die wel opgespoord hadden moeten worden.

De eerste groep noemen we zoals gezien de valse positieven en de tweede groep de valse negatieven.

Men kan zich voorstellen dat het ernstiger is een ontwikkelingsstoornis niet op te sporen dan teveel kinderen voor verder onderzoek door te sturen. Men dient in dit geval een test te kiezen die een laag aantal valse negatieven oplevert. Van de andere kant is het ook mogelijk de grootst mogelijke efficiëntie van de instantie waarnaar doorverwezen wordt voorop te stellen. Men zal hier zo min mogelijk overbodig onderzoek willen verrichten en dus zorgdragen voor een zo gering mogelijk aantal valse positieven.

In de voorafgaande paragraaf zijn verschillende manieren aangegeven om de twee soorten fouten in getallen uit te drukken.

De genoemde maten zijn schattingen van de validiteit van de gebruikte test.

Het absolute aantal fouten dat in een bepaalde situatie gemaakt wordt is echter niet alleen afhankelijk van de validiteit van de screeningtest maar ook van de 'base rate'.

5.2. De 'base rate' en de totale screeningfout

De 'base rate', 'Grundwahrscheinlichkeit', is de frequentie van het verschijnsel in de populatie ten aanzien waarvan men voorspellingen wenst te doen. Zo wordt de base rate van niet-schoolbekwaam zijn rond de .10 geschat, hetgeen betekent dat in de huidige situatie 1 op de 10 kinderen niet schoolbekwaam wordt genoemd (Tiedeman, 1974).

Om een screeningprocedure te kunnen evalueren is het noodzakelijk te weten welke de base rate is van het op te sporen kenmerk. De base rate van ontwikkelingsstoornissen wordt wel geschat op .15 (Rutter e.a., 1970).

Als er 1.000 kinderen met de D.D.S.T. gescreend worden, mag men dus op grond van deze base rate verwachten dat er 850 echte negatieven zijn (d.w.z. geen ontwikkelingsstoornissen; geen doorverwijzing voor verder onderzoek) en 150 echte positieven (d.w.z. wel ontwikkelingsstoornissen; moeten worden doorverwezen). Van de 850 echte negatieven zullen 816 kinderen door de D.D.S.T. negatief (2) genoemd worden (de conegativiteit is .96). Van de 150 echte positieven zullen er 47 ook door de D.D.S.T. als positief worden benoemd (copolitiviteit = .31).

Men kan een en ander in een contingentietabel uitdrukken (zie tabel 4).

Tabel 4: Geschatte efficiëntie van de D.D.S.T. in het opsporen van ontwikkelingsstoornissen bij een base rate van .15 als $N = 1000$.

	Ontwikkelingsstoornissen		
	ja	nee	tot.
abn. quest.	47	34	81
D.D.S.T. norm.	103	816	919
tot.	150	850	1000

Uit de tabel kan men aflezen dat 103 kinderen niet doorverwezen zijn, die wel doorverwezen hadden moeten worden en dat 34 kinderen ten onrechte zijn doorverwezen.

5.3. Het veranderen van het 'cutting-off-point'

In bepaalde situaties is een screeningtest onbruikbaar omdat hij teveel valse positieven opspoot of omdat er teveel valse negatieven zijn. Nu bestaat soms de mogelijkheid om een test zo te veranderen dat er minder valse positieven opgespoord worden, of dat er minder valse negatieven zijn. Men kan een test minder valse positieven laten opsporen door de test gemakkelijker passeerbaar te maken. Men noemt dit het verlagen van het cutting-off-point.

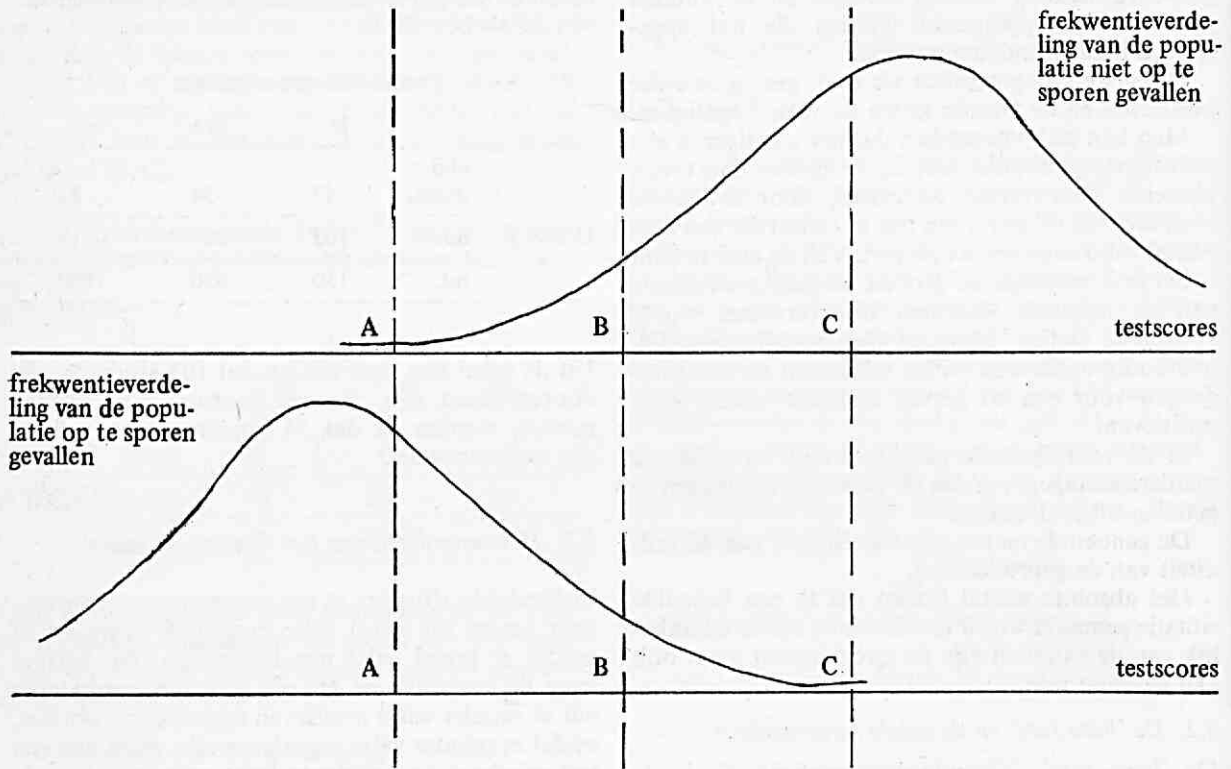
Onder cutting-off-point wordt een bepaalde test-score verstaan (zie fig. 2.). Personen die lager scoren worden positief genoemd. Alle hogere scores zijn dan negatief. Als het cutting-off-point verhoogd wordt waardoor de test moeilijker passeerbaar wordt, zal het aantal valse negatieven afnemen. De aantallen valse positieven en valse negatieven zijn onderling afhankelijk: neemt het aantal valse positieven toe dan neemt het aantal valse negatieven af, en omgekeerd. De afhankelijke relatie tussen valse negatieven en valse positieven is in figuur 2 afgebeeld.

In fig. 2 kan men zien dat bij een cutting-off-point A er vrijwel geen gevallen uit de niet-op-te-sporen populatie positief worden genoemd. Het aantal valse positieven is miniem. Daarentegen worden in dat geval wel veel gevallen uit de op te sporen populatie negatief genoemd. Het aantal valse negatieven is relatief groot.

Bij doorverwijzen van gevallen met een testscore lager dan B zullen er ongeveer evenveel valse positieven als valse negatieven zijn.

Als testscore C als cutting-off-point gehanteerd

Figuur 2: De relatie tussen valse positieven en valse negatieven en verschillende cutting-off-points.



A, B, en C zijn cutting-off-points; gevallen met een testscore die lager is dan (in de figuur links ligt van) het cutting-off-point worden positief genoemd; scores hoger dan (rechts van) dit punt worden negatief genoemd.

wordt zullen er relatief veel valse positieven en weinig valse negatieven zijn.

In ons voorbeeld met de D.D.S.T. is gekozen voor zo weinig mogelijk kans op valse negatieven. Men kan echter ook voor een lager cutting-off-point kiezen. Als men op het continuum 'abnormal-questionable-normal' het cutting-off-point niet tus-

sen questionable en normal, maar tussen abnormal en questionable plaatst ziet de twee bij twee tabel van tabel 3 er als volgt uit (zie tabel 5).

In dit geval is het aantal valse positieven 6 (copositiviteit = .14) en het aantal valse negatieven 88 (conegativiteit = .99).

Tabel 5: Geschatte overeenstemming D.D.S.T.-resultaat en IQ/DQ scores bij een laag D.D.S.T.-cutting-off-point.

		DQ/IQ		tot.
		< 80	≥ 80	
D.D.S.T.	abn. +	14	6	20
	norm. quest. —	88	1184	1272
	tot.	102	1190	1292

Tabel 6: Aantallen valse positieven en valse negatieven bij een hoog en een laag D.D.S.T.-cutting-off-point.

	D.D.S.T.-cutting-off-point	
	laag*	hoog**
aantal valse posit.	6	42
aantal valse negat.	88	70

* zie tabel 5
** zie tabel 3

In tabel 6 staan de twee soorten fouten bij de twee cutting-off-points aangegeven.

Uit tabel 6 is af te lezen dat Frankenburg c.s. door het nemen van het hoge cutting-off-point i.p.v. het lage, bij een winst van 18 echte positieven, 36 valse positieven méér op de koop toe hebben genomen.

5.4. Schatting van het totale aantal beslissingsfouten

De totale screeningfout is gedefinieerd als

$$\frac{b + c}{a + b + c + d}$$

Als weer een IQ/DQ score kleiner

dan 80 als operationalisatie van ontwikkelingsstoornissen wordt genomen is de totale screeningfout .09 (112 kinderen).

Als er geen screeningtest gebruikt zou worden en alle kinderen als negatief beschouwd zouden worden, zou ten aanzien van .08 (102 kinderen) een foutieve beslissing worden genomen. Door het gebruik van de test neemt het aantal foutieve beslissingen met ongeveer .01 (11 kinderen) toe!

Toch blijkt uit de cijfers betreffende valse positieven en valse negatieven dat het verantwoord kan zijn om de test toe te passen. Het aantal valse negatieven neemt namelijk door het testgebruik af van 102 tot 70. Als men er vanuit gaat dat het acceptabel is 42 kinderen ten onrechte voor verder onderzoek door te verwijzen om daarnaast 32 kinderen door te sturen die anders niet gedetekteerd zouden zijn, is screening met dit instrument alleszins te rechtvaardigen.

In sommige situaties is het echter zo dat fouten van beide soorten even ernstig kunnen zijn. Dit zou het geval kunnen zijn bij het gebruiken van een (schoolbekwaamheids) test als selectiemiddel voor het al dan niet naar de basisschool gaan.

Het al dan niet zinnig zijn van het gebruik van een test hangt ook hier weer af van de base rate. Als deze op 10% (Tiedeman 1974) gesteld wordt zal de totale screeningfout niet hoger mogen zijn dan deze 10%. In alle andere gevallen maakt men mét testgebruik meer beslissingsfouten dan zónder gebruik. Nu zijn van de meeste schoolbekwaamheidstests geen validiteitsmaten in termen van het onder 4.1. tot 4.5. beschrevene aanwezig. Veelal zijn wel andere validiteitscoëfficiënten aanwezig. Taylor en Russel (1939) geven tabellen aan de hand waarvan men, gegeven een bepaalde base rate, kan bepalen hoe hoog de validiteitscoëfficiënt moet zijn om meer goede beslissingen mét dan zonder testgebruik te kunnen nemen. Hieruit blijkt dat een test een validiteitscoëfficiënt moet hebben van ongeveer .70 wil, gege-

ven een base rate van 10%, testgebruik minder onder- en/of overreferrals opleveren dan geen testgebruik (d.w.z. alle gevallen negatief noemen). Meehl en Rosen (1955) tonen aan dat testgebruik het meest efficiënt is bij een base rate van .50.

Hier leveren ook tests met een vrij matige validiteit een bijdrage aan het aantal juiste beslissingen. Is de base rate hoger of lager, dan moeten naarmate de base rate verder van 50% afwijkt hogere eisen gesteld worden aan de validiteit van de test.

6. Enkele voorbeelden van onderzoek naar psychologische screeningtests

6.1. Terreinbepaling

Duidelijk moet gesteld worden dat beoordelingen van de validiteit van tests zoals hier beschreven, alleen van toepassing zijn waar tests classificierend gebruikt worden. Met name gaat het hier om dichotome classificaties.

6.2. Testen van zuigelingen in verband met adoptie

Illingworth (1971) onderzocht 230 adoptiekinderen van 6 tot 10 maanden oud. Bij dezelfde kinderen nam hij een IQ test af toen ze 7 jaar waren. Hij concludeerde dat zijn onderzoek 'provides evidence that the developmental assessment in infancy has some predictive value'. Hij deelde de baby's in 4 groepen in n.l. superior, possibly above average (1), average (2), doubtful (3) en inferior (4).

Op 7-jarige leeftijd bleken de gemiddelde IQ's van deze groepen te zijn: groep 1, 111; groep 2, 108; groep 3, 98; groep 4, 76.

We hebben zijn gegevens in een contingentietabel geplaatst, ervan uitgaande dat zijn groepen 3 en 4 (doubtful en inferior) latere IQ's van 84 of lager willen voorspellen (zie tabel 7).

Tabel 7: Overeenstemming test Illingworth bij kinderen van 6 maanden tot 10 maanden en IQ scores op 7-jarige leeftijd.
N = 230 (Illingworth 1971).

	IQ		tot.	
	< 84	> 84		
categorieën	3,4	19 (a)	50 (b)	69
onderzoek	1,2	6 (c)	155 (d)	161
Illingworth	tot.	25	205	230

De copositiviteit is .76; de conegativiteit is .76; de overreferral ratio is .72; de underreferral ratio is .04 en de totale screeningfout is .24. Kinderen die later een relatief laag IQ hebben kunnen dus met deze test redelijk accuraat geïdentificeerd worden blijkens de copositiviteit. In totaal werd echter bijna één kwart van de kinderen foutief gecategoriseerd. In ongeveer drie kwart van de gevallen waarin de ontwikkeling van de baby dubtful of inferior werd genoemd, bleek het IQ met 7 jaar 84 of hoger te zijn (in 37 gevallen d.w.z. ongeveer de helft, zelfs hoger dan 95). Op een werkelijke bijdrage van de test in de beslissingssituatie mag derhalve niet gerekend worden.

6.3. I.Q.-screenen in de praktijk van de kinderarts

Pless e.a. (1965) concluderen na een onderzoek met de Quick Test (een zeer korte test voor IQ schattingen), dat voor kinderen van 6 tot 15 jaar 'the Q.T. appears to be a rapid, reliable method for identifying those individuals whose (WISC) IQ is above or below 80'.

Op grond van hun gegevens is een twee-bij-twee tabel op te stellen die de relatie tussen de Quick Test en de WISC aangeeft.

Tabel 8: Overeenstemming tussen de Quick Test en de WISC bij kinderen van 6 tot 15 jaar. N = 50 (Pless e.a. 1965).

		WISC - IQ	
		< 80	≥ 80
Quick test.	< 80	9	4
	≥ 80	2	35

De copositiviteit is hier .82; de conegativiteit is .90; de overreferral ratio is .31; de underreferral ratio is .05; de totale screeningfout is .12. Pless e.a. adviseren de Quick Test voor gebruik in de praktijk van kinderartsen. Een snelle schatting van het ontwikkelingsniveau van de binnenkomende patiënten zou met deze test mogelijk zijn.

In de in dit onderzoek onderzochte populatie is de base rate ongeveer 20% (11 van de 50 kinderen hebben een WISC-IQ van minder dan 80). Zelfs bij deze relatief hoge base rate werd in 12% van de gevallen een foutief oordeel uitgesproken door de Q.T. (3). Dit ondanks een product-moment-correlatie van .84.

De Q.T. heeft echter een laag percentage underreferrals en een hoge conegativiteit. Omdat in de situatie waarin de kinderarts werkt follow-up onderzoek vaak relatief eenvoudig te realiseren is, zal het minder erg zijn teveel kinderen door te verwijzen dan kinderen met een IQ lager dan 80 over het hoofd te zien.

In deze situatie zal de Q.T. dus wellicht efficiënt kunnen functioneren, mits de base rate niet te laag is. Onderzoek met meer kinderen dan hier gebeurd is zal gedaan moeten worden (N = 50).

6.4. Vroegtijdige opsporing van leerstoornissen

Ferinden e.a. (1970) bespreken enkele tests die bij kleuters nauwkeurig mogelijke leerstoornissen in de 'elementary schools' zouden kunnen voorspellen. Het betreft hier de Wide Range Achievement Test (WRAT), de Evanston Early Identification Scale (EEIS), en de Metropolitan Reading Readiness Test form R (MRT). Deze tests werden afgenomen bij 'Kindergarten' kinderen. Vier maanden na het begin van de 'elementary school' werd een leestest afgenomen, een subtest van de WRAT.

De overeenstemming tussen de drie screeningtests en de leestest is in tabel 9 uitgedrukt.

Tabel 9: De overeenstemming van WRAT, EEIS en MRT scores van 'Kindergarten' Kinderen en leesscores na vier maanden leesonderwijs. N = 67.

		WRAT-leestest	
		beneden norm	boven norm
EEIS (N = 67)	high risk	15	1
	low-middle risk	26	25
WRAT (N = 66)	beneden norm	35	4
	boven norm	3	24
MRT (N = 67)	low-low normal	27	3
	average superior	12	25

De copositiviteit, de conegativiteit, de overreferral- en underreferral ratio van de drie tests staan in tabel 10 vermeld.

Tabel 10: De validiteitsmaten van de WRAT, de EEIS en de MRT.

	WRAT	EEIS	MRT
copositiviteit	.92	.36	.69
conegativiteit	.89	.96	.89
underr.-ratio	.11	.51	.32
overr.-ratio	.10	.06	.10
totale scr.fout	.11	.40	.22

Pas op grond van een tabel als deze kan een beargumenteerde beslissing genomen worden over het gebruik van een of meer van deze tests als screeningstests. In het artikel van Ferinden e.a. wordt slechts gesproken over correcte identificatie in termen van copositiviteit.

7. Discussie

In het bovenstaande werden enkele manieren aangegeven om de validiteit van een screeningtest aan te geven. Dezelfde maten kunnen gebruikt worden bij evaluatie van de effectiviteit van een screeningprocedure in bijvoorbeeld de gezondheidszorg.

Wij zouden willen voorstellen om copositiviteit te

definiëren als $\frac{a}{a+c}$, conegativiteit als $\frac{d}{b+d}$,

overreferral ratio als $\frac{b}{a+b}$, underreferral ratio

als $\frac{c}{c+d}$ en de totale screeningfout als

$\frac{b+c}{a+b+c+d}$.

In een aantal onderzoeken naar de validiteit van screeningstests die besproken werden bleken de volgende punten reden tot kritiek te geven.

a. De wijze van berekening van copositiviteit en conegativiteit was foutief, namelijk gebaseerd op een steekproef van de door de screeningtest negatief genoemde gevallen, zonder dat er voor de steekproeftrekking een correctie is toegepast (Frankenburg e.a. 1970).

- b. Er werd geen onderscheid gemaakt tussen twee mogelijke fouten bij het screenen, n.l. valse positieven en valse negatieven (Illingworth, 1971; Pless e.a., 1965; Ferinden e.a., 1970; Tiedeman 1974).
- c. Er werd niet gesproken over het rekening houden met de base rate in een bepaalde populatie (Illingworth, 1971; Pless e.a. 1965; Ferinden e.a. 1970, zie ook Tiedeman, 1974).

Noten

1. i.p.v. copositiviteit en conegativiteit wordt ook wel respectievelijk de term sensitiviteit en specificiteit gebruikt.
2. voor het gemak ervan uitgaand dat $IQ < 80$ identiek is aan ontwikkelingsstoornissen. De keuze van het IQ als criterium is gebaseerd op pragmatische gronden. Op inhoudelijke aspecten van deze keuze wordt niet ingegaan.
3. in 22% van alle gevallen was het verschil 9 punten of meer.

Literatuur

- Cools, A. en Hermanns, J., *Denver Ontwikkeling Screeningstest*. Amsterdam: Swets en Zeitlinger, in press.
- Drenth, P., *De Psychologische test*. Arnhem: van Loghem Slaterus, 1966.
- Ferinden, W. E., Jacobson, S. & Linden, N., Early identification of learning disabilities. *J. learning disabilities* 1970, 3, 48-52.
- Frankenburg, W. K., Goldstein, A. D. & Camp, B. W., The revised Denver Developmental Screening Test: its accuracy as a screening instrument. *Journal of Pediatrics* 1971, 79, 988-995.
- Grant, J. A., Quantitative evaluation of a screening program. *Am. J. public Health*, 1974, 64, 66-71.
- Gresnigt, H. A. A. & Gresnigt-Strengers, A. M. C., *Ouders en gezinnen met een diepzwakzinnig kind*. Amsterdam: Swets en Zeitlinger 1973.
- Hoejenbos, E., Vroegtijdige onderkenning van zwakzinnigheid. *N.O.Z. Kommunikatie Orgaan* 1971, 4, 3-22.
- Illingworth, R. S., The predictive value of developmental assessment in infancy. *Developm. Med. Child Neurol.* 1971, 13, 721-725.
- Kalverboer, A. F., Vroege ontwikkelingsdiagnostiek: enige feiten en vraagtekens. in: Wit, J. de, Bolle, H., Jessurun Cardozo - van Hoorn, R. (red.), *Psychologen over het kind, deel 3*. Groningen: Wolters-Noordhoff, 1973, 53-71.
- Meehl, P. en Rosen, A., Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psych. Bull.* 1955, 52, 194-216.

- Pless, B., Snider, M., Eaton, A. & Kearsley, R., A rapid screening test for intelligence in children. *Am. J. diseases of children* 1965, 109, 533-537.
- Rutter, M., Tizard, J. & Whitmore, K. (eds.), *Education, health and behavior*. London: Longman 1970.
- Schlesinger-Was, E. A. & Wieringen, J. C. van, Onderzoek 'psycho-motorische ontwikkeling van zuigelingen op 3 consultatiebureaux voor zuigelingen in de stad Utrecht'. Projekt 525a, N.I.P.G.-T.N.O., Leiden 1972.
- Taylor, H., en Russell, J. T., The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *J. appl. Psych.* 23, 1939, 565-578.
- Tiedemann, J., Die Problematik der Schuleignungsdiagnose unter entscheidungstheoretischem aspekt. *Z. Entw. Psycholog. Pädag. Psych.* 1974, 2, 124-132.
- Wiechen, H. J. van, Vroegtijdige opsporing van gehandicapte zuigelingen en peuters. *Huisarts en Wetenschap* 1972, 15, 35.

Curricula vitae

Drs. A. Cools (geb. 1939) en *drs. J. Hermanns* (geb. 1948) zijn leiders van een onderzoeksprojekt, waarin de mogelijkheid van vroegtijdige opsporing van een aantal problemen in de ontwikkeling van kinderen wordt nagegaan. Dit onderzoek heeft o.a. geresulteerd in de Denver Ontwikkeling Screeningtest (Swets en Zeitlinger, in druk).

Adres: Instituut voor Pedagogische en Andragogische Wetenschappen R.U., Utrecht, Maliesingel 23, Utrecht.

Bij het overlijden van Professor Dr. V. D'Espallier



Professor D'Espallier werd in 1904 te Antwerpen geboren, liep er lagere school en deed er zijn middelbare studies, heeft er heel zijn leven gewoond en stierf er op zondag 12 januari 1975.

Hij doorliep achtereenvolgens de lagere normaalschool (akte van onderwijzer) en de middelbare normaalschool (diploma van regent, d.i. onderwijsbevoegdheid in het lager secundair onderwijs).

Tijdens deze studietijd komt hij reeds in contact met twee belangrijke Vlaamse pedagogen: Leo Roels (later Inspecteur-Generaal van het Lager Onderwijs en samen met Jeunehomme auteur van het leerplan 1936) en Dr. F. De Hovre (wijsgerig pedagoog, stichter van het eerste Hoger Opvoedkundig Instituut te Gent en van het Vlaams Opvoedkundig Tijdschrift in 1929).

Alvorens in de praktijk te stappen gaat D'Espallier Filosofie en Pedagogiek studeren aan de Leuvense Universiteit. In 1927 is hij licentiaat in de Opvoedkunde. Twee jaar later verdedigt hij reeds zijn proefschrift 'Het beginsel der zelfwerkzaamheid op de lagere school theoretisch belicht' en promoveert hiermede tot Doctor in de Opvoedkunde. Onder zijn professoren noemen wij Prof. Dr. A. Michotte (experimenteel psycholoog), Prof. Dr. A. Fauville (pedagogische psychologie) en de promotor van zijn proefschrift Prof. Dr. R. Buyse (psycho-didacticus). In 1931 en '32 volgt hij als vacatiecursist colleges aan het Instituut J. J. Rousseau te Genève, waar hij persoonlijk Jean Piaget en diens publicaties leert kennen.

Van 1928 tot '37 is hij onderwijzer in de oefenschool van de normaalschool te Antwerpen. In 1927 verschijnt zijn eerste artikel. Sindsdien publiceert hij frequent in het Vlaams Opvoedkundig Tijdschrift.

Zijn bijdragen verwerven bekendheid door hun heldere en vlotte stijl, ook door de relevantie van de inhoud voor de opvoeding en in het bijzonder voor het onderwijs.

Van bij het ontstaan van het tweede Hoger Opvoedkundig Instituut in het Vlaamse land, nl. te Antwerpen in 1928, is D'Espallier en secretaris en docent. Hij is dit gebleven tot 1958. Onder zijn leerlingen komen tal van bekende inspecteurs van het lager onderwijs voor.

Hijzelf slaagt in 1933 in het gevreesde inspecteurs-examen, in werkelijkheid een zware selectieproef. In 1937 volgt zijn benoeming tot inspecteur.

Ondertussen verschijnen ook afzonderlijke publicaties van hem, boeken en brochures. Het moge hier volstaan er enkele te vermelden:

- Uit het werk van Piaget. 1932, 72 pp.
- Opvoedkundige vraagstukken van dezen tijd. 1933, 107 pp.
- Moderne schoolboeken. 1935, 90 pp.
- Mensenkennis en typologie. 1935, 66 pp.

Inmiddels is hij in Vlaanderen en Nederland bekend geworden als hoofdredacteur van het monumentale didactisch werk 'Nieuwe banen in het onderwijs', 2 volumens, 1933 (1ste editie) en 1938 (2de editie).

In 1936 wordt hij mede-hoofdredacteur van het Vlaamse Opvoedkundig Tijdschrift. (De andere twee zijn Dr. F. De Hovre en Dr. A. Decoene.)

Tevens is hij vóór de Tweede Wereldoorlog een der bedrijvigste voordrachtgevers van de Vlaamse Opvoedkundige Vereniging geweest.

Een tweede periode van deze uitermate actieve pedagoog begint in 1941. Hij behoort, samen met de Professoren Nuttin (psycholoog), Dellaert (psychiater) en een paar jaar later Kriekemans (ontwikkelingspsycholoog en pedagoog) tot de eerste groep die belast wordt met de vernederlandsing van de specifieke colleges aan de in 1924 opgerichte 'Ecole de pédagogie et de psychologie appliquée à l'éducation'.

Alhoewel dit voor D'Espallier een volledige leeropdracht betekende, heeft hij deze als lector en

later als buitengewoon hoogleraar uitgeoefend terwijl hij inspecteur bleef. Vanaf 1959 werd hij dan gewoon hoogleraar in de psychopedagogiek en de didactiek.

Meer dan 30 jaar heeft hij aan het Instituut voor Psychologie en Opvoedkunde (tot 1965) en verder aan de Faculteit der Psychologie en Opvoedkundige Wetenschappen van de Leuvense Universiteit gedoceerd en samen met zijn studenten-thesisten wetenschappelijk onderzoek verricht.

Onder de talrijke thema's vallen er enkele op door het aantal verhandelingen die eraan gewijd werden. Meestal is zo'n reeks studies uitgemond in een wetenschappelijke monografie. Hier kunnen de volgende titels vermeld worden:

- Bijdrage tot de psychologie van het enige kind. (Studia Psychologica.) Leuvense Universitaire Uitgaven, 1956.
- Het kind op de drempel van de lagere school. 3 volumens. (Studia Paedagogica.) Ibidem, 1957 (Toby-Jonckheereprijs voor experimentele pedagogiek 1959).
- Psycho-pedagogische studiën over de zittenblijver in het Middelbaar Onderwijs. (Studia Paedagogica.) Ibidem, 1968.

Als hoofdredacteur heeft D'Espallier niet geringe verdiensten. Reeds eerder werd zijn rol vermeld als één der hoofdredacteurs van het *Vlaamse Opvoedkundig Tijdschrift*, verschenen van 1919 tot 1953-54.

In 1954-55 vormt hij een Nederlands-Belgische redactie, die het werk van het *Vlaams Opvoedkundig Tijdschrift* (1919, 1ste jaargang) en van het Nederlandse *Tijdschrift van Zielkunde en Opvoedingsleer* (1968, 1ste jaargang) zal voortzetten in het *Tijdschrift voor Opvoedkunde*. Hij is hoofdredacteur gebleven tot en met jaargang 1969-1970. Dan heeft hij, overmand door de ziekte, het roer in handen gegeven van zijn leerling Professor Dr. C. De Keyser.

In de periode vlak voor deze fusie dirigeerde D'Espallier een omvangrijke Nederlands-Belgische redactieploeg, m.n. bij de voorbereiding van de *Katholieke Encyclopedie voor Opvoeding en Onderwijs*, 3 volumens, 1951-1954. In dit standaardwerk op pedagogisch gebied vinden wij een honderdtal korte en langere bijdragen waaronder zijn naam staat.

Ook in Nederland is zijn naam en zijn werk niet onbekend gebleven. Hij onderhoudt regelmatig contact met zijn collegae aan de Pedagogische Instituten en treedt er meermalen als gastspreker en als auteur in de belangstelling. In 1970 verschijnt van hem in *Pedagogische Studiën* een fundamenteel

artikel over 'Het probleem van de schoolrijpheid in een dynamisch perspectief.'

Meerdere bezoeken van Nederlandse studenten en professoren aan het Leuvense Instituut en omgekeerd heeft hij mede georganiseerd of er actief aan deelgenomen. Vele Nederlandse vrienden zullen hem wel herkennen met bijgaande foto, die veeleer de ietwat stroeve en gesloten kant laat zien van deze overigens joviale professor en collega.

Professor D'Espallier is op vele terreinen werkzaam geweest. Een van deze werd nog niet vermeld, nl. zijn grote humane en wetenschappelijke belangstelling voor de gehandicapte kinderen.

Reeds in 1932 publiceert hij samen met L. Boggaerts het boek 'Moeilijke kinderen op school'. In 1936 is hij medestichter met Prof. Dr. R. Dellaert en Juffrouw C. Arnou van de Medische-Psychologische Kinderkliniek te Antwerpen, de eerste Belgische Child Guidance Clinic. Hij is er tot aan zijn ziekte de actieve voorzitter van de Raad van Beheer gebleven. Zijn echte belangstelling ging echter naar de concrete hulp op diagnostisch en op pedotherapeutisch gebied aan de opgenomen kinderen.

Van 1960 tot aan het jaar van zijn ziekte is hij Voorzitter geweest van het Nationaal Verbond van het Katholiek Buitengewoon Onderwijs.

In 1965 kwam, dank zij zijn inspirerende medewerking en zijn stuwkracht de groep Orthopedagogiek tot stand als een der vier afstudeerrichtingen van de studies in de Pedagogische Wetenschappen aan de Universiteit te Leuven.

Van de meer dan honderd vijftig licentiaatsverhandelingen en zes doctoraatsproefschriften die onder zijn leiding te Leuven werden gemaakt, zijn er zes en veertig verhandelingen en twee proefschriften over een orthopedagogisch onderwerp gemaakt.

Samen met zijn familie, zijn collegae, medewerkers, vrienden en oud-studenten, was een groepje gehandicapte leerlingen aanwezig bij de stemmingsvolle uitvaartplechtigheid van deze actieve pedagogisch-didacticus en orthopedagoog.

K. SWINNEN

-
- R. VANDENBERGHE, Bibliografie van Professor Dr. V. D'Espallier. *Tijdschrift Opvoedkunde*, 1970-71 (16), nr 1, 39-45; nr 3, 166-171: A. Boeken; B. Artikels; C. Artikels in de Kathol. Encyclopedie voor opvoeding en onderwijs; D. Proefschriften en verhandelingen gemaakt onder zijn leiding.
C. DE KEYSER (Ed.), *Liber Amicorum Professor V. D'Espallier*. Leuven, in voorbereiding.