

Appendix

Eenvoudige statistische begrippen

In deze appendix bespreken we enkele statistische begrippen, die voor een goed begrip van diverse technische passages in dit boek nuttig zijn. Overigens pretenderen we geen volledigheid, noch een korte ‘cursus’ statistiek voor de sociale en de gedragswetenschappen. Deze appendix kan hooguit het geheugen opfrissen.

We maken in dit boek veelal geen onderscheid tussen steekproef en populatie, althans niet in de behandelde psychometrische begrippen. De reden is om de eenvoud in de toch al vaak lastige materie zo veel mogelijk te bewaren.

Het aantal items in een test geven we aan met k . Items geven we aan met indices g en h , waarbij $g, h, = 1, \dots, k$. Scores op items geven we aan met toevalsvariabelen X_g en X_h . Dichotome items hebben doorgaans scores $X_g = 0, 1$ en polytome items scores $X_g = 0, \dots, m$, zodat het aantal geordende antwoordcategorieën van die items met $m + 1$ wordt aangegeven. De notatie m komt overigens niet veel voor in dit boek.

De ruwe testscore of gewoon de ruwe score geven we aan met X . Dit is de ongewogen som van de k itemscores in de test:

$$X = \sum_{g=1}^k X_g. \quad [\text{A.1}]$$

Personen geven we aan met index i ; we gaan er steeds van uit dat we de beschikking hebben over de scores van n personen, zodat $i = 1, \dots, n$.

Dit is meestal een steekproef. De scores van persoon i op item g geven we aan met X_{ig} en de ruwe score met X_i .

Centrale tendentie

Het gemiddelde van een variabele, zoals de ruwe score, is

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad [\text{A.2}]$$

Dit gemiddelde geeft de centrale tendentie van een verdeling van ruwe scores aan, dus een soort van geneigdheid van de onderzochte groep. De ene groep is geneigd hoger op de test te scoren – beter te presteren – dan de andere, en dat wordt weerspiegeld in de groepsgemiddelden. Een tweede maat voor centrale tendentie is de mediaan. Indien van alle n personen de ruwe scores worden geordend van klein naar groot, dan is de mediaan de middelste van deze scores. Bijvoorbeeld: de mediaan is de 45e score van 89 (n) geordende scores – er komen in de ordening 44 scores voor en 44 scores na de mediaan. Bij een even aantal scores, bijvoorbeeld $n = 90$, nemen we het gemiddelde van de twee middelste scores, dus van de 45e en de 46e. De derde maat is de modus. Dit is de waarde van X die het meeste voorkomt in de groep. In een kleine steekproef komen diverse scores meestal in dezelfde, geringe mate voor. De modus is dan niet zinvol, omdat er dan diverse scores met dezelfde frequentie zijn en bovendien de frequentie per score onnauwkeurig is. De modus kan derhalve het beste alleen worden bepaald voor grote steekproeven.

De mediaan en de modus zijn niet gevoelig voor uitschieters. Het gemiddelde is dat veel meer. De mediaan wordt wel gebruikt om de centrale tendentie aan te geven wanneer een verdeling scheef is. Bij normaalverdelingen vallen gemiddelde, mediaan en modus samen. In dit boek gebruiken we bijna uitsluitend het gemiddelde. De mediaan en de modus treft men vooral aan bij de normering en de interpretatie van testprestaties.

Dichotome variabelen – met scores 0 en 1 – hebben een gemiddelde dat heel eenvoudig is. Bedenk eerst dat we de som moeten uitrekenen van n nullen en enen. Dit levert het aantal enen in de steekproef, aangeduid met n_1 , waarna deling door n de proportie enen oplevert. Voor een denkbeeldige steekproef van enen en nullen hebben we dan:

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} (1 + 1 + 1 + 0 + 1 + 0 + 0 + \dots + 1) = \\ &= \frac{n_1}{n} = p_g.\end{aligned}\quad [\text{A.3}]$$

In de testtheorie staat dit itemgemiddelde bekend als de p -waarde van een item. De proportie nullen wordt dan aangegeven met q , en de relatie met de p -waarde is

$$q_g = 1 - p_g. \quad [\text{A.4}]$$

Spreiding

Vaak is het nuttig te weten hoe ver een score verwijderd ligt van het gemiddelde van de verdeling. Ten eerste kan dit worden uitgedrukt in de afwijkingsscore of de deviatiescore, aangegeven met x ; die is gedefinieerd als

$$x = X - \bar{X}.$$

Deze afwijkingsscore kan zowel negatieve als positieve waarden aannemen. Het gemiddelde van de afwijkingsscores in de groep waarvoor ze berekend zijn, dus met gebruikmaking van het groepsgemiddelde, is per definitie 0. Door gebruik te maken van de eigenschap van somtekens, dat $\sum (X + Y) = \sum X + \sum Y$, volgt dit uit

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n \bar{X} = \bar{X} - \bar{X} = 0.$$

De spreiding in n testcores wordt gegeven door de variantie – hier de steekproefvariantie, aangeduid met $S^2(X)$ en soms met S_X^2 . Dit is de gemiddelde, gekwadrateerde afwijkingsscore in de groep (vraag: waarom niet de gemiddelde afwijkingsscore?):

$$S^2(X) = \frac{1}{n} \sum_{i=1}^n x_i^2. \quad [\text{A.5}]$$

Voordat we verder gaan, toch even een uitstapje naar de populatie. Veel mensen raken in de war van de zogenaamde n -weging en $(n - 1)$ -weging van de variantie. De eerste gebruiken we om de spreiding van de scores in een specifieke groep te beschrijven, de tweede als we de groep zien als een representatieve steekproef van een nader omschreven populatie, waarvan we de variantie willen schatten. De $(n - 1)$ -weging is een wiskundige truc die ervoor zorgt dat de steekproef-

variantie een zuivere schatter is van de populatievariantie; dus, dat het gemiddelde van heel veel van dergelijke schattingen precies op de populatievariantie uitkomt. De n -weging zou dan een onderschatting opleveren.

Omdat de variantie iets zegt over spreiding in termen van gekwadrateerde eenheden en dit lastig te interpreteren is, neemt men voor de interpretatie vaak de standaarddeviatie:

$$S(X) = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}. \quad [\text{A.6}]$$

Voor dichotome items krijgen we weer sterk vereenvoudigde formules:

$$S^2(X_g) = p_g q_g, \quad [\text{A.7}]$$

en

$$S(X_g) = \sqrt{p_g q_g}. \quad [\text{A.8}]$$

Transformaties

Twee transformaties van de ruwe scores zijn in het bijzonder interessant, de afwijkingsscore of deviatiescore, en de z -score of standaard-score. We herhalen de definitie van de afwijkingsscore:

$$x = X - \bar{X}. \quad [\text{A.9}]$$

De twee belangrijkste eigenschappen zijn: $\bar{x} = 0$ en $S_x^2 = S_X^2$. De eerste eigenschap kwamen we al tegen, de tweede zegt dat de variantie van n afwijkingsscores gelijk is aan die van de oorspronkelijke n ruwe scores waaruit ze zijn berekend. We vermijden hier een overdaad aan wiskundige bewijsvoering. Wel is zonder wiskunde in te zien dat als van alle scores het gemiddelde wordt afgetrokken, de verdeling alleen verschuift, maar dat de spreiding gelijk blijft.

Twee variabelen (X en Y) in afwijkingsscorevorm (x en y) hebben beide een gemiddelde gelijk aan 0 ($\bar{x} = \bar{y} = 0$), maar verschillende varianties ($S_x^2 \neq S_y^2$). Die varianties kunnen ook nog worden gelijkgesteld, door de afwijkingsscores te delen door hun standaarddeviaties. Conform de meeste statistiekboeken delen we echter door S_X en niet door S_x (hoewel ze gelijk zijn: $S_x = S_X$). Voor variabele X krijgen we dan

$$z_X = \frac{X - \bar{X}}{S_X}. \quad [\text{A.10}]$$

De z -score geeft de afwijking van score X van het groepsgemiddelde \bar{X} uitgedrukt in standaarddeviaties van de verdeling van X . De twee be-

langrijke eigenschappen van z-scores zijn: $\bar{z} = 0$ en $S_z^2 = 1$. Twee variabelen in z-scorevorm hebben dus beide gemiddelde 0 en variantie 1. Verschillen en overeenkomsten in de interpretatie van afwijkingsscores en z-scores zijn de volgende. Stel dat $x_{Jan} = -7$, dan weten we dat de ruwe score van Jan 7 punten onder het groepsgemiddelde ligt. We weten niet of dit veel is, want daarvoor zouden we moeten weten wat de spreiding van de ruwe scores in de hele groep is. Stel, de standaarddeviatie is 3.5, dan is $z_{Jan} = -2$. Als de ruwe scores normaal verdeeld zijn, dan is in een tabel voor de normaalverdeling te vinden dat ongeveer 97.5% van de groep een hogere score heeft dan Jan, en dat is veel. Als de standaarddeviatie echter 14 zou zijn, dan was $z_{Jan} = -0.5$, en dus had ongeveer 69% een hogere score. De score van Jan was dan nog steeds aan de lage kant, maar niet meer dramatisch. Afwijkingsscores en z-scores laten beide zien aan welke kant van het groepsgemiddelde een score ligt, maar de laatste geeft de afstand in standaarddeviaties en is dus informatiever.

Samenhang

De samenhang tussen twee variabelen, X en Y , wordt uitgedrukt in hun covariantie, de mate waarin zij samen variëren. De covariantie is het gemiddelde product van de afwijkingsscores in de groep:

$$S(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i. \quad [A.11]$$

De covariantie geeft de richting van het verband, maar niet de sterkte. Een positieve covariantie betekent dat positieve scores op x en y in het algemeen met elkaar samengaan en evenzo negatieve scores. Een negatieve covariantie betekent dat er een tendens is dat positieve scores op x samengaan met negatieve scores op y , en negatieve scores op x met positieve op y . Een covariantie van 4.3 tussen twee variabelen geeft echter niet per se een sterker positief verband aan dan een covariantie van .9 tussen twee andere variabelen. Dit komt doordat de variantie van X en Y doorspeelt in hun covariantie: een grote variantie van een of beide variabelen betekent veel grote afwijkingsscores en dus grote producten xy in de formule voor de covariantie. De invloed van de spreiding van de variabelen kan worden uitgeschakeld door de variabelen te standaardiseren en de covariantie te berekenen tussen hun z-scoreversies:

$$S(z_X, z_Y) = \frac{1}{n} \sum_{i=1}^n z_{X_i} z_{Y_i}.$$

Aangetoond kan worden dat dit precies de product-momentcorrelatie is, want:

$$\frac{1}{n} \sum_{i=1}^n z_{X_i} z_{Y_i} = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{S_X} \frac{y_i}{S_Y},$$

en omdat de constante waarden $1/S_X$ en $1/S_Y$ voor het somteken mogen worden gezet, krijgen we

$$\frac{1}{S_X S_Y} \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{S_{XY}}{S_X S_Y} = r_{XY}.$$

De correlatie is dus de covariantie tussen de z-scoreversies van twee variabelen. De correlatie is ongevoelig voor de variantie van de variabelen, en heeft vaste onder- en bovengrenzen:

$$-1 \leq r_{XY} \leq 1. \quad [\text{A.12}]$$

Nu weten we dus dat een correlatie van .8 een sterk positief verband aangeeft en een correlatie van $-.2$ een zwak negatief verband. Overigens kunnen die extreme waarden alleen worden gehaald als de verdelingen van X en Y precies identiek zijn. Als bijvoorbeeld de ene variabele normaal verdeeld is en de andere scheef, dan worden die extreme waarden niet meer gehaald. We laten dit onderwerp verder rusten. Heel belangrijk om te weten is dat de correlatie de sterkte van het lineaire verband tussen twee variabelen aangeeft, dus de mate waarin het verband kan worden beschreven door een lineaire functie. Wie hierover meer wil weten, kan bijvoorbeeld terecht in de hoofdstukken waar de regressieanalyse wordt behandeld (zie de hoofdstukken 6 en 8). Ook kunnen twee variabelen een perfect verband hebben, maar als dit bijvoorbeeld een kwadratisch verband is, is de correlatie toch 0.

Lineaire transformaties

In de testtheorie worden ruwe scores vaak omgezet in andere scores. We zagen al afwijkingsscores en z-scores, en een ander voorbeeld zijn de IQ-scores (hoofdstuk 5). Dit zijn alle transformaties van het type

$$Y = a + bX \quad [\text{A.13}]$$

waarin a en b constante getallen zijn. Deze omzetting van X in Y is

lineair; we noemen dit een lineaire transformatie. Een grafiek van deze functie geeft, bij specifieke keuzes voor a en b , daadwerkelijk een rechte lijn te zien. De lezer kan zelf nagaan dat de omzetting van de ruwe score in afwijkingsscores gebeurt via een lineaire transformatie met $a = -\bar{X}$ en $b = 1$. De omzetting van X naar z_X gaat via $a = -\bar{X}/S_X$ en $b = 1/S_X$.

In het algemeen gelden de volgende formules voor de lineaire transformatie, $Y = a + bX$:

Gemiddelde:

$$\bar{Y} = a + b\bar{X}. \quad [\text{A.14}]$$

Dus: het gemiddelde van X ondergaat dezelfde transformatie als de afzonderlijke scores.

Variantie:

$$S^2(Y) = b^2 S^2(X). \quad [\text{A.15}]$$

De variantie van Y verschilt een factor b^2 van de variantie van X . De constante a heeft geen invloed op de variantie van Y .

Standaarddeviatie:

$$S(Y) = |b|S(X). \quad [\text{A.16}]$$

De standaarddeviatie van Y verschilt een factor $|b|$ van de standaarddeviatie van X .

Covariantie: Laat $Y_1 = a + bX_1$ en $Y_2 = c + dX_2$ (Y_1 en X_1 zijn variabelen, en c en d zijn constanten); dan is

$$S(Y_1, Y_2) = bdS(X_1, X_2). \quad [\text{A.17}]$$

Net als bij de variantie spelen de optelconstanten, a en c , geen rol. Verder kan worden opgemerkt dat als bijvoorbeeld b positief is en d negatief, de covariantie dan van teken verandert.

Correlatie:

$$\begin{aligned} r(Y_1, Y_2) &= r(X_1, X_2) \text{ als } bd > 0, \text{ en} \\ r(Y_1, Y_2) &= -r(X_1, X_2) \text{ als } bd < 0. \end{aligned} \quad [\text{A.18}]$$

Dus: alleen het teken van het product bd heeft invloed op de correlatie, namelijk op het teken ervan.

Lineaire combinaties

Een lineaire combinatie is een som van variabelen, al dan niet gewogen. Laat de som aangeduid worden met X en de te sommeren variabelen met X_g , $g = 1, \dots, k$. In onze notatie is X de ruwe score en

zijn de andere variabelen de k itemscores. Met name in hoofdstuk 6 over de betrouwbaarheid en aanverwante onderwerpen, spelen lineaire combinaties een belangrijke rol. Het is derhalve belangrijk te weten hoe, in dit geval, het gemiddelde en de variantie van de somscore en de te sommeren variabelen zich tot elkaar verhouden. We behandelen alleen het geval van de ongewogen variabelen of, in ons geval, itemscores – dus itemscores met alle het gewicht 1.

Gemiddelde:

$$\bar{X} = \sum_{g=1}^k \bar{X}_g. \quad [\text{A.19}]$$

Dus: het gemiddelde van een somvariabele is de som van de gemiddelden van de afzonderlijke, gesommeerde variabelen.

Variantie: we gebruiken de dubbele som $\sum \sum_{g \neq h}$ om aan te geven dat we optellen over alle paren van verschillende variabelen, X_g en X_h , zodat $g \neq h$; dus, paren met twee keer dezelfde variabele doen niet mee. Wel de paren (1,2), (1,3), . . . , (1,k); (2,1), (2,3), . . . , (2,k); . . . ; (k,1), (k,2), . . . , (k-1,k). Dit zijn in totaal $k(k-1)$ paren. De variantie van X schrijven we dan als:

$$S^2(X) = \sum_{g=1}^k S^2(X_g) + \sum \sum_{g \neq h} S(X_g, X_h). \quad [\text{A.20}]$$

De variantie van een somvariabele is de som van de varianties van de afzonderlijke variabelen in de som plus de som van alle covarianties tussen deze variabelen.

Covariantie: tot slot geven we de covariantie van twee somvariabelen.

We kiezen voor het gemak twee eenvoudige somvariabelen,

$X = X_1 + X_2$ en $Y = Y_1 + Y_2$. De covariantie van X en Y is

$$S(X, Y) = S(X_1, Y_1) + S(X_1, Y_2) + S(X_2, Y_1) + S(X_2, Y_2). \quad [\text{A.21}]$$

Dus, de covariantie van twee somvariabelen is gelijk aan de som van de covarianties van de variabelen waaruit de twee sommen zijn opgebouwd.

Opgdrachten

- 1 Gegeven zijn de testcores, X , van vijf personen: $X = 12, 10, 14, 11, 13$.
 - a Bereken voor elke persoon de afwijkingsscore, x , en de standaardscore, z_X .

- b Bereken de gemiddelden van x en z_X in deze groep van vijf personen.
 - c Bereken de standaarddeviaties van x en z_X in deze groep.
- 2 Stel, een tentamen statistiek bestaat uit 40 meerkeuze-items van het goed/fout-type.
 - a Wat is de variatiebreedte van de ruwe score?
 - b Als de cesuur (zak/slaag-grens) 28 is, heeft iemand met een tentamencijfer 7 dan een ruwe score lager dan 28?
- 3 Beantwoord de volgende vragen.
 - a Laat in een figuur zien dat de frequentieverdeling van de afwijkingsscores ten opzichte van de verdeling van de ruwe scores is verschoven.
 - b Over welke afstand is de verdeling verschoven?
 - c Laat tevens in de figuur zien dat het gemiddelde van de verdeling der afwijkingsscores 0 is.
- 4 Lever het bewijs voor de volgende wiskundige stellingen.
 - a Het gemiddelde van een variabele in afwijkingsscorevorm is 0.
 - b De variantie van een variabele in afwijkingsscorevorm is gelijk aan de variantie van de ruwe score.
- 5 Beantwoord de volgende vragen.
 - a Als ik weet dat iemands ruwe score gelijk is aan 37, wat weet ik dan over zijn positie in de groep?
 - b Als iemands afwijkingsscore gelijk is aan 14, wat weet ik dan over zijn positie in de groep?
- 6 Laat in een figuur zien wat er gebeurt met de frequentieverdeling van de ruwe scores als er standaardscores van gemaakt worden.
- 7 Lever wiskundige bewijzen van de volgende stellingen.
 - a Standaardscores hebben een gemiddelde 0.
 - b Standaardscores hebben een variantie 1.
- 8 Stel, iemand heeft op een test een z-score van -1.3 .
 - a Welke informatie biedt dit over zijn positie in de groep?
 - b Als de scores in de onderzochte groep normaal verdeeld zijn, hoeveel procent van de respondenten heeft dan een hogere testscore dan deze persoon? (hint: maak gebruik van uw statistiekboek).

- c Als de scores niet normaal verdeeld zijn (en we weten verder niet hoe de verdeling er wel uit ziet), wat kan men dan zeggen over het percentage respondenten met een score hoger dan $z = 0$?
- 9 Welke score is het meest informatief over iemands positie in de groep, de percentielscore of de standaardscore? Beargumenteer uw antwoord.
- 10 In de onderstaande tabel is de temperatuur in zes koelkasten van verschillende merken weergegeven in respectievelijk graden Fahrenheit (X) en graden Celsius (Y). Gegeven is:

$$Y = \frac{5}{9}(X - 32).$$

koelkast	X	x	Y	y
1	55.4	14.4	13.0	
2	30.2	-10.8	-1.0	
3	35.6	-5.4	2.0	
4	33.8	-7.2	1.0	
5	50.0	9.0	10.0	
6	41.0	0.0	5.0	

- a Bereken het gemiddelde en de standaarddeviatie van X en Y .
- b Vul de afwijkingsscores, y , in de tabel in.
- c Ziet u een snelle manier om gemiddelde en standaarddeviatie van Y te berekenen als u die van X reeds kent?
- 11 Gegeven zijn de scores van zes personen op drie tests:

persoon	X_1	X_2	X_3
1	1	3	5
2	0	8	10
3	1	9	10
4	0	11	8
5	1	14	10
6	0	15	11

- a Bereken de gemiddelden en de standaarddeviaties van de drie variabelen.
 - b Bereken de z-scores voor de tweede en de derde variabele.
 - c Bereken de correlatie tussen X_2 en X_3 .
 - d Bereken gemiddelde en variantie van de getransformeerde variabele $Y = 3 + 2X_2$.
 - e Ziet u bij opgave d een eenvoudiger oplossing die veel rekenwerk bespaart? Welke oplossing is dat?
- 12 Als van een dichotoom gescoord item met 0/1-scoring de p-waarde bekend is, is daarmee ook de variantie bekend.
- a Als $p = 0,3$, wat is dan de variantie van het item?
 - b Als de variantie van het item 0.24 is, wat is dan de p-waarde?
 - c Laat algemeen zien (dus, niet met getallen, maar met een formule) dat als we de variantie kennen, daarmee het gemiddelde – dus, de p-waarde – niet uniek vastligt.
- 13 Gegeven zijn de scores van tien personen op vier dichotome items:

persoon	X_1	X_2	X_3	X_4
1	1	1	0	0
2	0	1	1	0
3	1	1	1	1
4	1	1	1	1
5	0	0	0	0
6	0	0	1	0
7	1	1	1	0
8	1	0	1	0
9	0	0	1	0
10	0	0	0	0

- a Bereken de p-waarden van de items.
- b Bereken de standaarddeviaties van de items.
- c Bereken alle afwijkingsscores en standaardscores.
- d Bereken de correlaties tussen de items 1 en 2, 1 en 3, en 1 en 4. Gebruik telkens twee formules.
- e Bereken de correlatie tussen item 1 en de totaalscore op de andere drie items, die we aangeven met $R_{(-1)}$, en tevens de correlatie van item 4 met de restscore, $R_{(-4)}$.

14 Gegeven is een tabel met de scores van zes personen op drie tests:

persoon	X_1	X_2	X_3
1	10	4	10
2	8	10	1
3	8	10	7
4	4	-2	1
5	4	4	-11
6	2	-2	-2

- Bereken de afwijkingsscoreversies van de drie variabelen.
 - Bereken alle varianties en covarianties van de drie variabelen, en zet ze in een tabel met drie rijen en drie kolommen, die respectievelijk met de eerste, de tweede en de derde variabele corresponderen. Deze tabel is de variantie-covariantiematrix.
 - Bereken de somvariabele, $X = X_1 + X_2 + X_3$, en bereken op twee manieren het gemiddelde en de variantie van X .
 - Bereken direct uit de variantie-covariantiematrix: $S^2(X_1 + X_2)$, $S^2(X_1 + X_3)$ en $S^2(X_2 + X_3)$.
- 15 Wat geldt voor de covarianties tussen de items als iedereen op de test dezelfde ruwe score (aantal goed) heeft?

Literatuur

- Abrahams, N.M., & Ralf, E.F. (1972). Pratfalls in moderator research. *Journal of Applied Psychology*, 56, 245-251.
- Adkins, D. (1961). *Test construction*. Columbus, OH: Merrill Books.
- Adorno, T.W., Frenkel-Brunswik, E., Levinson, D.J., & Sanford, R.M. (1950). *The authoritarian personality*. New York: Harper.
- Aiken, L.R. (1971). *Psychological and educational testing*. Boston, MA: Allyn & Bacon.
- Algera, J.A., Jansen, P.W.G., Roe, R.A., & Vijn, P. (1984). Validity generalization: Some critical remarks on the Schmidt-Hunter procedure. *Journal of Occupational Psychology*, 57, 197-210.
- Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth.
- Altink-van den Berg, W.M.M. (1988). *Selectie voor hoger onderwijs in ontwikkelingslanden*. Dissertatie. Amsterdam, Vrije Universiteit.
- Altink, W.M.M., & Thijs, G.D. (1984). The issue of equity: research on selection processes for educational programmes in developing countries. *Institute of Development Studies Bulletin*, 15, 74-80.
- American Psychological Association (APA) (1954). Technical recommendations for psychological tests and diagnostic techniques. *Suppl. Psychological Bulletin*, 51.
- American Psychological Association (APA) (1969). Job testing and the disadvantaged. *American Psychologist*, 24, 637-650.
- Amrine, M. (1965). The 1965 congressional inquiry into testing: A commentary. *American Psychologist*, 20, 858.
- Anastasi, A. (1961, 1988). *Psychological testing*. New York: MacMillan.
- Anastasi, A. (1967). Psychology, psychologists, and psychological testing. *American Psychologist*, 22, 297-306.
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Anderson, N., & Herriot, P. (1997). *International handbook of selection and assessment*. Chichester: Wiley.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.
- Angoff, W.H. (1988). Validity: An evolving concept. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 19-32). Hillsdale, NJ: Erlbaum.

- Ariel, A. (2005). *Contributions to test-item bank design and management*. Dissertatie. Enschede, Universiteit Twente.
- Astin, A.W. (1964). Criterion-oriented research. *Educational and Psychological Measurement*, 24, 807-829.
- Baker, F.B. (1984). Technology and testing: State of the art and trends for the future. *Journal of Educational Measurement*, 21, 399-406.
- Baker, F.B. (1992). *Item response theory. Parameter estimation techniques*. New York: Marcel Dekker (nieuwe editie: F.B. Baker & S.-H. Kim, 2004).
- Bakx, A.W.E.A., Sijtsma, K., Sanden, J.M.M. van der, & Taconis, R. (2002). Development and evaluation of a student-centered multimedia self-assessment instrument for social-communicative competence. *Instructional Science*, 30, 335-359.
- Bali, S.K., Drenth, P.J.D., Flier, H. van der, & Young, W.C.E. (1984). Contribution of aptitude tests to the prediction of school performance in Kenya: a longitudinal study. Lisse: Swets & Zeitlinger.
- Ballard, P.B. (1920). *Mental tests*. London: Hodder & Stoughton.
- Ballard, P.B. (1924). *The new examiner*. London: University of London Press.
- Barendregt, J.T. (1961). *Research in psychodiagnostics*. Den Haag: Mouton.
- Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology*, 57, 221-237.
- Baumgarten, F. (1928). *Die Berufsprüfungsprüfungen: Theorie und Praxis*. München: Barth Verlag.
- Bender, L. (1938). A visual motor gestalt test and its clinical use. Res. Mon. No. 3. New York: American Orthopsychiatric Association.
- Bennett, C.C. (1967). What price privacy. *American Psychologist*, 22, 371-376.
- Berg, R.H. van den (2001). *Psychologisch onderzoek in een multiculturele samenleving*. Dissertatie. Amsterdam, Vrije Universiteit.
- Berge, J.M.F. ten, Sniijders, T.A.B., & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika*, 46, 201-213.
- Berge, J.M.F. ten, & Zegers, F.E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika*, 43, 575-579.
- Berry, J.W., Poortinga, Y.H., Segal, M.H., & Dasen, P.R. (1992). *Cross-cultural psychology: Research and applications*. New York: Cambridge University Press.
- Binet, A., & Simon, Th.A. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11, 191-244.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (1968). *Statistical theories of mental test scores* (pp. 396-479). Reading, MA: Addison-Wesley.
- Bishop, G.F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220-232.
- Biskin, B.H., & Kolotkin, R.L. (1977). Effects of computerized administration on scores on the Minnesota Multiphasic Personality Inventory. *Applied Psychological Measurement*, 1, 543-549.
- Bleichrodt, N. (1989). *Intelligentiemeting ten behoeve van het onderwijs in Indonesië*. Dissertatie. Amsterdam, Vrije Universiteit.
- Bleichrodt, N., Drenth, P.J.D., Resing, W.C.M., & Zaal, J.N. (1984). *Revisie Amsterdamse Kinder Intelligentie Test. Instructie, normen, psychometrische gegevens*. Lisse: Swets & Zeitlinger.

- Bleichrodt, N., Drenth, P.J.D., Resing, W.C.M., & Zaal, J.N. (1987). *Revisie Amsterdamsse Kinder Intelligentie Test*. Handleiding. Lisse: Swets & Zeitlinger.
- Bleichrodt, N., & Vijver, A.J.R. van de (2001). *Diagnostiek bij allochtonen; mogelijkheden en beperkingen van psychologische tests*. Lisse: Swets & Zeitlinger.
- Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield, IL: Charles C. Thomas.
- Bloom, B.S. (1956). *Taxonomy of educational objectives*. New York: McKay.
- Bobertag, O. (1911). *Ueber Intelligenzprüfungen nach der Methode von Binet und Simon*. Leipzig: Barth.
- Bock, R.D. (1997). The nominal categories model. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 33-49). New York: Springer.
- Boeck, P. de, & Wilson, M. (2004). *Explanatory item response models. A generalized linear and nonlinear approach*. New York: Springer.
- Boeke, P.E. (1962). *Psychodiagnostische problemen van de epilepsie*. Assen: Van Gorcum.
- Boekkooi-Timminga, E. (1987). Simultaneous test construction by zero-one programming. *Methodika*, 1, 101-112.
- Boomsma, A., Duijn, M.A.J. van, & Snijders, T.A.B. (2001). *Essays on item response theory*. New York: Springer.
- Boring, E.G. (1945). The use of operational definitions in science. *Psychological Review*, 52, 243-245.
- Borsboom, D. (2005). *Measuring the mind. Conceptual issues in contemporary psychometrics*. Cambridge UK: Cambridge University Press.
- Borsboom, D., Mellenbergh, G.J., & Heerden, J. van (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203-219.
- Borsboom, D., Mellenbergh, G.J., & Heerden, J. van (2004). The concept of validity. *Psychological Review*, 111, 1061-1071.
- Bos, D.J. (1974). *Schoolkeuze adviezen: Resultaten controle van 5 jaar*. Dissertatie. Amsterdam, Universiteit van Amsterdam. Den Haag: Mouton.
- Both, F.J. (1967). *Manuele vaardigheid: een oriëntatie*. Groningen: Wolters.
- Bouma, J.M. (1988). *Perceptual asymmetries and hemispheric specialization*. Dissertatie. Amsterdam: Free University Press.
- Bouwmeester, S. (2005). *Latent variable modeling of cognitive processes in transitive reasoning*. Dissertatie. Tilburg, Universiteit van Tilburg.
- Bouwmeester, S., & Sijtsma, K. (2004). Measuring the ability of transitive reasoning, using product and strategy information. *Psychometrika*, 69, 123-146.
- Bouwmeester, S., Sijtsma, K., & Vermunt, J.K. (2004). Latent class regression analysis for describing cognitive developmental phenomena: An application to transitive reasoning. *European Journal of Developmental Psychology*, 1, 67-86.
- Boxtel, H.W. van, Snijders, J.Th., & Welten, V.J. (1980). *Voorlopige verantwoording en handleiding. Vorm III ISI-Reeks*. Groningen/Lisse: Wolters-Noordhoff/Swets & Zeitlinger.
- Brainerd, C.J., & Kingma, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. *Developmental Review*, 4, 311-377.
- Brainerd, C.J., & Reyna, V.F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10, 3-47.
- Brennan, R.L. (2001). *Generalizability theory*. New York: Springer.
- Brim, O.G. (1965). American attitudes towards attitude tests. *American Psychologist*, 20, 125-130.

- Brink, W.P. van den, & Mellenbergh, G.J. (1998). *Testleer en testconstructie*. Amsterdam: Boom.
- Brogden, H.E. (1949). A new coefficient: application to biserial correlation and to estimation of selective efficiency. *Psychometrika*, 14, 169-182.
- Bühler, Ch., & Hetzer, H. (1932). *Kleinkindertests*. Leipzig: Barth.
- Buiten, B. (1970). *Biografische vragenlijsten*. Dissertatie. Groningen, Rijksuniversiteit Groningen.
- Burke, H.R. (1958). Raven's progressive matrices: A review and critical evaluation. *Journal of Genetical Psychology*, 93, 199-228.
- Burt, C. (1921). *Mental and scholastic tests*. London: London University Press.
- Campbell, D.T. (1957). A typology of tests, projective and otherwise. *Journal of Consulting Psychology*, 21, 207-217.
- Campbell, D.T. (1960). Recommendations for APA tests regarding construct, trait or discriminant validity. *American Psychologist*, 15, 546-553.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campbell, J.P., & Knapp, D.J. (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Erlbaum.
- Cattell, J. McKeen, J. (1890). Mental tests and measurements. *Mind*, 15, 373-380.
- Cattell, R.B. (1956-1957). *Sixteen personality factors questionnaire*. Yonkers-on-Hudson, NY: World Book.
- Choppin, B.H. (1976). Recent developments in item banking: a review. In D.N.M. de Gruijter & L.J.Th. van der Kamp (red.), *Advances in psychological and educational measurement* (pp. 233-245). New York: Wiley.
- CITOGroep (2004). *Thuis in de Citotoets. Een goede kennismaking voor thuis* (uitgave op dvd). Arnhem: CITOGroep.
- Claparède, E. (1924). *Comment diagnostiquer les aptitudes chez les écoliers*. Paris: Presses Universitaires de France.
- Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Cole, N.S. (1973). Bias in selection. *Journal of Educational Measurement*, 10, 237-255.
- Cole, N.S. (1981). Bias in testing. *American Psychologist*, 36, 1067-1077.
- Conrad, H.S. (1967). Clearance of questionnaires with respect to invasion of privacy, public sensitiveness, ethical standards, etc. *American Psychologist*, 22, 356-359.
- Coombs, C.H. (1964). *A theory of data*. Ann Arbor, MI: Mathesis Press.
- Costa, P.T., Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-RTM) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L.J. (1961, 1984). *Essentials of psychological testing*. New York: Harper & Row.
- Cronbach, L.J. (1964). *Educational psychology*. New York: Harcourt, Brace & World.
- Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (red.), *Educational Measurement* (pp. 443-507). Washington DC: American Council on Education.
- Cronbach, L.J. (1988). Five perspectives on the validity argument. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.

- Cronbach, L.J., & Gleser, G.C. (1953). Assessing similarity profiles. *Psychological Bulletin*, 50, 456-473.
- Cronbach, L.J., & Gleser, G.C. (1957, 1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dague, P. (1972). Development, application and interpretation of tests for use in French-speaking black Africa and Madagascar. In L.J. Cronbach & P.J.D. Drenth (red.), *Mental tests and cultural adaptation* (pp. 63-74). Den Haag: Mouton.
- Dam, K. van (1996). *Dansende beren. Beoordelingsprocessen bij personeelsselectie*. Dissertatie. Amsterdam, Universiteit van Amsterdam.
- Darlington, R.B. (1976). A defence of rational personnel selection, and two new methods. *Journal of Educational Measurement*, 13, 43-52.
- Dayton, C.M. (1998). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.
- Defares, P.B., Kema, G.N., & Werff, J.J. van der (1962). Intelligentie, tempo en nauwkeurigheid in de GALO-test. *Paedagogische Studiën*, 39, 1-17.
- Dehue, T. (1990). *De regels van het vak. Nederlandse psychologen en hun methodologie 1900-1985*. Amsterdam: Van Gennep.
- Dekker, M.J.A. (1987). *Intelligentie van visueel gehandicapte kinderen in de leeftijd van 6 tot 15 jaar*. Dissertatie. Amsterdam: vU Uitgeverij.
- Dekker, M.J.A., Drenth, P.J.D., & Zaal, J.N. (1989). *Handleiding Intelligentietest voor Visueel Gehandicapte Kinderen in de leeftijd van 6 tot 15 jaar (drie delen)*. Zeist: Vereniging Bartiméus.
- Delft, D. van (2005). *Heike Kamerlingh Onnes. Een biografie*. Amsterdam: Uitgeverij Bert Bakker.
- DiBello, L.V., Stout, W.F., & Roussos, L.A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P.D. Nichols, S.F. Chipman & R.L. Brennan (red.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Erlbaum.
- Doornen, L.J.P. van (1988). *Physiological stress reactivity*. Dissertatie. Amsterdam, Vrije Universiteit.
- Draney, K.L., Pirolli, P., & Wilson, M. (1995). A measurement model for a complex cognitive skill. In P.D. Nichols, S.F. Chipman & R.L. Brennan (red.), *Cognitively diagnostic assessment* (pp. 103-125). Hillsdale, NJ: Erlbaum.
- Drenth, P.J.D. (1965a). *De psychologische test*. Arnhem: Van Loghum Slaterus.
- Drenth, P.J.D. (1965b). *Test voor niet-verbale abstractie*. Amsterdam: Swets & Zeitlinger.
- Drenth, P.J.D. (1967). *Protesten contra testen*. Amsterdam: Swets & Zeitlinger.
- Drenth, P.J.D. (1972). Implications of testing for individual and society. In L.J. Cronbach & P.J.D. Drenth (red.), *Mental tests and cultural adaptation* (pp. 23-35). Den Haag: Mouton.
- Drenth, P.J.D. (1975). Tests for developing countries; rationale and objectives. *Nederlands Tijdschrift voor de Psychologie*, 30, 5-22.
- Drenth, P.J.D. (1977). The use of intelligence tests in developing countries. In Y.H. Poortinga (red.), *Basic problems in cross cultural psychology* (pp. 249-258). Amsterdam: Swets & Zeitlinger.

- Drenth, P.J.D. (1983). *Cultuur en intelligentie: een psychometrische optiek*. Amsterdam: KNAW/Noord Hollandse Uitgeversmaatschappij.
- Drenth, P.J.D. (1987). Intelligence tests in education, evaluation and selection. In C. Kagicibasi (red.), *Growth and progress in cross-cultural psychology* (pp. 293-301). Lisse: Swets & Zeitlinger.
- Drenth, P.J.D. (1988a). Psychologische selectie en discriminatie. *GEDRAG en Organisatie*, 1, 18-26.
- Drenth, P.J.D. (1988b). Intelligentie-onderzoek in het onderwijs: ervaringen in ontwikkelingslanden. In W.K.B. Hofstee & G.A. Lang (red.), *Nederlands psychologisch onderzoek, deel II* (pp. 9-24). Lisse: Swets & Zeitlinger.
- Drenth, P.J.D. (1989). Personeelsbeoordeling. In P.J.D. Drenth, Hk. Thierry & Ch.J. de Wolff (red.), *Nieuw handboek arbeids- en organisatiepsychologie* (Stud. Ed., pp. 353-401). Deventer: Van Loghum Slaterus.
- Drenth, P.J.D., & Flier, H. van der (1985). Beoordelen en testen in de psychologie. In J.F. Orlebeke, P.J.D. Drenth, R.H.C. Jansen & C. Sanders (red.), *Compendium van de psychologie (deel 5)* (pp. 255-293). Muiderberg: Coutinho.
- Drenth, P.J.D., Flier, H. van der, & Omari, I.M. (1983). Educational selection in Tanzania. *Evaluation in Education*, 7, 93-217.
- Drenth, P.J.D., & Hoolwerf, G. (1970). *Numerieke Aanleg Test (NAT '70)*. Amsterdam: Swets & Zeitlinger.
- Drenth, P.J.D., & Hoolwerf, G. (1977). *Hernormering Testserie Hoger Niveau*. Amsterdam: Swets & Zeitlinger.
- Drenth, P.J.D., & Kranendonk, L.A. (1973). *Schaal voor interpersoonlijke waarden*. Amsterdam: Swets & Zeitlinger.
- Drenth, P.J.D., & Wieringen, P.C.W. van (1969). *Verbale Aanleg Testserie 1969 (VAT '69)*. Amsterdam: Swets & Zeitlinger.
- Drenth, P.J.D., Wieringen, P.C.W. van, & Hoolwerf, G. (2001). *Drenth Testserie Hoog Niveau (DITH)*. Lisse: Swets Test Publishers.
- Drever, J., & Collins, M. (1936). *Performance tests of intelligence: A series of non-linguistic tests for deaf and normal children*. Edinburgh: Oliver and Boyd.
- Du Bois, Ph.H. (1970). *A history of psychological testing*. Boston, MA: Allyn & Bacon.
- Dunnette, M.D. (1966). *Personnel selection and placement*. London: Tavistock.
- Ebel, R.L. (1961). Must all tests be valid? *American Psychologist*, 16, 640-646.
- Ebel, R.L. (1963). The social consequences of educational testing. *Proceedings of the 1963 Invitational Conference on Testing Problems* (pp. 130-143). Princeton, NJ: Educational Testing Service.
- Ebel, R.L. (1965). *Measuring educational achievement*. Englewood Cliffs, NJ: Prentice-Hall.
- Edwards, A.L. (1957). *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts.
- Eggen, T.J.H.M. (2004). *Contributions to the theory and practice of computerized adaptive testing*. Dissertatie. Enschede, Universiteit Twente.
- Eggen, T.J.H.M., & Kelderman, H. (1987). Rapportage van parameters uit het Raschmodel. *Tijdschrift voor Onderwijsresearch*, 12, 121-132.
- Eggen, T.J.H.M., & Sanders, P.F. (1993). *Psychometrie in de praktijk*. Arnhem: CITO-groep.
- Embretson, S.E. (red.) (1985). *Test design*. Orlando, FL: Academic Press.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

- Emons, W.H.M., Sijtsma, K., & Meijer, R.R. (2004). Testing hypotheses about the person-response function in person-fit analysis. *Multivariate Behavioral Research*, 39, 1-35.
- Emons, W.H.M., Sijtsma, K., & Meijer, R.R. (2005). Global, local and graphical person-fit analysis using person response functions. *Psychological Methods*, 10, 101-119.
- Engel, J.F., Blackwell, R.D., & Miniard P. (1990). *Consumer behavior*. Orlando, FL: The Dryden Press.
- Engelen, R.J.H. (1989). *Parameter estimation in the logistic item response model*. Dissertatie. Enschede, Universiteit Twente.
- Esbroeck, R. van (1982). *Analyse van 'faking-gedrag' van adolescenten in beroepenvragenlijsten*. Brussel, Vrije Universiteit.
- Evers, A. (1992). *Handleiding Amsterdamse Beroepen Interesses Vragenlijst, ABIV 92*. Lisse: Swets & Zeitlinger.
- Evers, A., Anderson, N., & Voskuil, O. (2005). *The Blackwell handbook of personnel selection*. Oxford: Blackwell Publishing.
- Evers, A., & Lucassen, W. (1991). *Handleiding DAT '83. Differentiële Aanleg Testserie*. Lisse: Swets & Zeitlinger.
- Evers, A., Lucassen, W., & Wiegersma, S. (1999). *Beroepen Interesses Test BIT, versie 1997*. Handleiding. Lisse: Swets & Zeitlinger.
- Evers, A., Vliet-Mulder, J.C. van, & Groot, C.J. (2000a). *Documentatie van tests en testresearch in Nederland. Deel I: Testbeschrijvingen*. Assen: Van Gorcum.
- Evers, A., Vliet-Mulder, J.C. van, & Groot, C.J. (2000b). *Documentatie van tests en testresearch in Nederland. Deel II: Testresearch*. Assen: Van Gorcum.
- Evers, A., Vliet-Mulder, J.C. van, & Laak, J. ter (1992). *Documentatie van tests en testresearch in Nederland*. Assen: Van Gorcum.
- Feij, J.A. (1974). An investigation into the meaning of the achievement motivation test: I: questionnaire correlates. *Nederlands Tijdschrift voor de Psychologie*, 29, 171-190.
- Feij, J.A. (1978). *Temperament: onderzoek naar de betekenis van extraversie, emotionaleiteit, impulsiviteit en spanningsbehoefte*. Dissertatie. Amsterdam, Vrije Universiteit.
- Finkle, R.B. (1976). Managerial assessment centers. In M.D. Dunnette (red.), *Handbook of industrial and organization psychology* (pp. 861-888). Chicago, IL: Rand McNally.
- Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern (CH): Huber.
- Fischer, G.H., & Formann, A.K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement*, 6, 397-416.
- Fischer, G.H., & Molenaar, I.W. (1995). *Rasch models. Foundations, recent developments, and applications*. New York: Springer.
- Fischer, G.H., & Pendl, P. (1980). Individual testing on the basis of the dichotomous Rasch model. In L.J.Th. van der Kamp, W.F. Langerak & D.N.M. de Gruijter (red.), *Psychometrics for educational debates* (pp. 171-188). New York: Wiley.
- Fiske, D.W. (1965). The subject looks at psychological tests. *Proceedings of the 1964 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service.
- Fleishman, E.A. (1953). Testing for psychomotor abilities by means of apparatus tests. *Psychological Bulletin*, 50, 241-262.
- Fleishman, E.A. (1954). Dimensional analysis of psychomotor abilities. *Journal of Experimental Psychology*, 48, 437-454.

- Fleishman, E.A. (1956). Psychomotor selection tests: research and application in the UAF. *Personnel Psychology*, 9, 449-468.
- Flier, H. van der (1972). Evaluating environmental influences on test scores. In L.J. Cronbach & P.J.D. Drenth (red.), *Mental tests and cultural adaptation* (pp. 447-452). Den Haag: Mouton.
- Flier, H. van der (1980). *Vergelijkbaarheid van individuele testprestaties*. Lisse: Swets & Zeitlinger.
- Flier, H. van der, & Boomsma-Suerink, J.L. (1994). *Handboek GATB B 1002-B* (2e druk). Separate uitgave Normen GATB B 1002-B. Utrecht: Stichting GATB Research.
- Flier, H. van der, Keers, W., & Drenth, P.J.D. (1976). *Applikatie Programmeurs Test*. Amsterdam: Swets & Zeitlinger.
- Flier, H. van der, Mellenbergh, G.J., Adèr, H.J., & Wijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement*, 21, 131-145.
- Fokkema, S.D., & Dirkzwager, A. (1962). De vertaling en bewerking van de DAT voor Nederland. *Nederlands Tijdschrift voor de Psychologie*, 17, 121-148.
- Fokkema, S.D., & Dirkzwager, A. (1968). *Handleiding voor de Differentiële Aanleg Test*. Lisse: Swets & Zeitlinger.
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.
- Frederiksen, N., & Melville, S.D. (1954). Differential predictability in the use of test scores. *Educational and Psychological Measurement*, 14, 647-656.
- French, J.W. (1954). *Kit of selected tests for reference aptitude and achievement factors*. Princeton, NJ: University of Princeton.
- Frijda, N.H. (1967). *Gelaat en karakter*. Haarlem: De Toorts.
- GALO Handleiding (herziene versie) (1981). Groningen: Noordelijk Instituut voor Toegepaste Psychologie en Bedrijfswetenschappen.
- Geer, J.P. van de (1961). *De mening van de psycholoog*. Haarlem: De Toorts.
- Ghiselli, E.E. (1955, 1966). *The validity of occupational aptitude tests*. New York: Wiley.
- Ghiselli, E.E. (1956a). Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology*, 40, 374-377.
- Ghiselli, E.E. (1956b). The placement of workers: a concept and problems. *Personnel Psychology*, 9, 1-16.
- Ghiselli, E.E. (1960). The prediction of predictability. *Educational and Psychological Measurement*, 20, 3-8.
- Giessen, R.W. van der (1957). *Enkele aspecten van het probleem der predictie in de psychologie, speciaal met het oog op de selectie van militair personeel*. Amsterdam: Swets & Zeitlinger.
- Glas, C.A.W. (1989). *Estimating and testing Rasch models*. Dissertatie. Enschede, Universiteit Twente.
- Goldberg, L.R. (1990). An alternative 'description of personality': The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.
- Goldstein, K., & Scheerer, M. (1945). *Goldstein-Scheerer Tests*. New York: Psychological Corporation.
- Gorsuch, R.L. (1983). *Factor analysis* (2nd edition). Hillsdale, NJ: Erlbaum.
- Goslinga, C.J. (1933). *Het boek der Richteren: Korte verklaring der Heilige Schrift*. Kampen: Kok.
- Grayson, D.A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika*, 53, 383-392.

- Green, B.F. (1988). Construct validity of computer-based tests. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 77-86). Hillsdale, NJ: Erlbaum.
- Green, B.F., Bock, R.D., Humphreys, L.G., Linn, R.L., & Reckase, M.D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347-360.
- Greenleaf, E.A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351.
- Greuter, M.A.M. (1988). *Personeelsselectie in perspectief*. Haarlem: Thesis.
- Groot, A.D. de (1961). *Methodologie*. Den Haag: Mouton.
- Groot, A.D. de (1966). *Vijven en zessen*. Groningen: Wolters.
- Groot, A.D. de, e.a. (1967). *Amsterdamsche schooltoetsen*. Groningen: Wolters.
- Groot, A.D. de (1970). Some badly needed nonstatistical concepts in applied psychometrics. *Nederlands Tijdschrift voor de Psychologie*, 25, 360-376.
- Groot, A.D. de, & Naerssen, R.F. van (1977). *Studietoetsen. Construeren. Afnemen. Analyseren. Deel I en II*. Den Haag: Mouton.
- Gross, M.L. (1962). *The brain watchers*. New York: Random House.
- Gross, A.L., & Su Wen-Huey (1975). Defining a 'fair' or 'unbiased' selection model: a question of utilities. *Journal of Applied Psychology*, 60, 345-351.
- Gruijter, D.N.M. de (1982). *Tentamineren en beslissen*. 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs svo, svo reeks 63.
- Gruijter, D.N.M. de, & Hambleton, R.K. (1984). On problems encountered using decision theory to set cutoff scores. *Applied Psychological Measurement*, 8, 1-8.
- Gruijter, D.N.M. de, & Kamp, L.J.Th. van der (1984). *Statistical models in psychological and educational testing*. Lisse: Swets & Zeitlinger.
- Guilford, J.P. (1936, 1954). *Psychometric methods*. New York: McGraw-Hill.
- Guilford, J.P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-437.
- Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guilford, J.P., & Zimmerman, W.S. (1949). *The Guilford-Zimmerman Temperament Survey*. Beverly Hills CA: Sheridan Psychological Services.
- Guion, R.M. (1965). *Personnel testing*. New York: McGraw-Hill.
- Guion, R.M. (1980). On trinitarian doctrines of validity. *Professional Psychology*, 11, 385-398.
- Guion, R.M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Erlbaum.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star & J.A. Clausen (red.), *Measurement and prediction* (pp. 60-90). Princeton, NJ: Princeton University Press.
- Haassen, P.P. van, Bruyn, E.E.J. de, Pijl, Y.J., Poortinga, Y.J., Lutje-Spelberg, H.C., Steene, G. van der, Coetsier, P., Spoelders-Claes, R., & Stinissen, J. (1986). *WISC-R, Nederlandse uitgave; scoring en normen*. Lisse: Swets & Zeitlinger.
- Hagenaars, J.A., & McCutcheon, A.L. (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.
- Haladyna, T.M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Erlbaum.
- Hambleton, R.K., & Cook, L.L. (1977). Latent trait models and their use in

- the analysis of educational test data. *Journal of Educational Measurement*, 14, 75-96.
- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer/Nijhoff Publishing.
- Hamers, J.H.M., & Resing, W.C.M. (1993). Learning potential assessment: Introduction. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsseenaars (red.), *Learning potential assessment. Theoretical, methodological and practical issues* (pp. 23-41). Amsterdam/Lisse: Swets & Zeitlinger.
- Hamers, J.H.M., Sijtsma, K., & Ruijsseenaars, A.J.J.M. (1993). *Learning potential assessment. Theoretical, methodological and practical issues*. Amsterdam/Lisse: Swets & Zeitlinger.
- Harman, H.H. (1976). *Modern factor analysis*. Chicago, IL: University of Chicago Press.
- Hathaway, S.R. (1964). MMPI professional use by professional people. *American Psychologist*, 19, 204-210.
- Hathaway, S.R., & McKinley, J.C. (1951). *The Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation.
- Heck, G.L.M. van (1981). *Anxiety: the profile of a trait*. Dissertatie. Tilburg, Katholieke Hogeschool.
- Heijden, P.G.M. van der, & Sijtsma, K. (1996). Fifty years of measurement and scaling in the Dutch social sciences. *Statistica Neerlandica*, 50, 111-135.
- Heinen, T. (1996). *Latent class and discrete latent trait models. Similarities and differences*. Thousand Oaks, CA: Sage.
- Hemker, B.T., Ark, L.A. van der, & Sijtsma, K. (2001). On measurement properties of continuation ratio models. *Psychometrika*, 66, 487-506.
- Hemker, B.T., Sijtsma, K., Molenaar, I.W., & Junker, B.W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, 62, 331-347.
- Herk, H. van, Poortinga, Y.H., & Verhallen, T.M.M. (2004). Response styles in rating scales. Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, 35, 346-360.
- Hermans, H.J.M. (1976). *Handleiding bij de Prestatie Motivatie Test, PMT* (herziene uitgave). Amsterdam: Swets & Zeitlinger.
- Hessen, D.J. (2003). *Differential item functioning. Types of DIF and observed score based detection methods*. Dissertatie. Amsterdam, Universiteit van Amsterdam.
- Hettema, P.J. (1966). *Stijlkenmerken in de waarneming*. Dissertatie. Nijmegen, Katholieke Universiteit Nijmegen.
- Hetzer, H. (1937). *Psychologische Untersuchung der Konstitution des Kindes*. Leipzig: Barth.
- Heus, P. de, Leeden, R. van der, & Gazendam, B. (1995). *Toegepaste data-analyse. Technieken voor niet-experimenteel onderzoek in de sociale wetenschappen*. Utrecht: Lemma.
- Hoekstra, H.A., Ormel, J., & Fruyt, F. (1996). *NEO PI-R – NEO FFI. Big Five Persoonlijkheidsvragenlijsten. Handleiding*. Lisse: Swets & Zeitlinger.
- Hofer, P.J., & Green, B.F. (1985). The challenge of competence and creativity in computerized psychological testing. *Journal of Consulting and Clinical Psychology*, 53, 826-838.
- Hoffman, K.I., & Lundberg, G.D. (1976). A comparison of computer-monitored group tests with paper-and-pencil tests. *Educational and Psychological Measurement*, 36, 791-809.

- Hofstee, W.K.B. (1966). [Boekbespreking van: P.J.D. Drenth, *De psychologische test*]. *Nederlands Tijdschrift voor de Psychologie*, 21, 462-468.
- Hofstee, W.K.B. (1969). Individuele verschillen en averechtse toepassing. *Nederlands Tijdschrift voor de Psychologie*, 24, 482-493.
- Hofstee, W.K.B. (1982). De methodische deskundigheid van de psycholoog. *De Psycholoog*, 17, 697-707.
- Hofstee, W.K.B. (1983a). Validiteitsgeneralisatie en -specificiteit: antwoord op Roe e.a. *De Psycholoog*, 18, 402-409.
- Hofstee, W.K.B. (1983b). *Selectie*. Utrecht: Het Spectrum.
- Hofstee, W.K.B. (1985). Liever klinisch? Grenzen aan het objectiviteitsbeginsel bij beoordeling en selectie. *Nederlands Tijdschrift voor de Psychologie*, 40, 459-473.
- Hofstee, W.K.B. (1990) Toepasbaarheid van psychologische tests bij allochtonen. *De Psycholoog*, 25, 291-294.
- Hofstee, W.K.B., Campbell, W.H., Eppink, A., Evers, A., Joe, R.C., Koppel, J.M.H. van de, Zweers, H., Choenni, C.E.S., & Zwan, T.J. van der (1990). *Toepasbaarheid van psychologische tests bij allochtonen* (LBR-reeks, nr. 11). Rotterdam: Landelijk Bureau Racismebestrijding.
- Hofstee, W.K.B., Raad, B. de, & Goldberg, L.R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63, 146-163.
- Hoijtink, H. (1990). PARELLA, *Measurement of latent traits by proximity items*. Leiden: DSWO Press.
- Holland, P.W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, 55, 577-601.
- Holland, P.W., & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Holland, P.W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
- Holtzman, W.H. (1958). *The Inkblot Test*. Austin, TX: University of Texas.
- Holzinger, K.J., & Harman, H.H. (1941). *Factor analysis*. Chicago, IL: University of Chicago Press.
- Hoorn, W. van, Meijer, J., & Oostdam, R. (2003). *Nederlandse Differentiatie Testserie. Handleiding*. London: Harcourt.
- Horst, P. (1962). The logic of personnel selection and classification. In R.M. Gagné (red.), *Psychological principles in system development* (pp. 231-271). New York: Holt, Rinehart & Winston.
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Erlbaum.
- Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response theory*. Homewood, IL: Dow-Jones-Irwin.
- Hunt, E., & Pellegrino, J. (1985). Using interactive computing to expand intelligence testing: A critique and prospectus. *Intelligence*, 9, 207-236.
- Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Jackson, P.H., & Agunwamba, C.C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: algebraic lower bounds. *Psychometrika*, 42, 567-578.
- Jackson, D.N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55, 243-252.

- Janda, L.H. (1998). *Psychological testing: theory and applications*. Needham Heights, MA: Allyn & Bacon.
- Jansen, A. (1963). *De toetsing van grafologische uitspraken*. Dissertatie. Amsterdam, Universiteit van Amsterdam.
- Jansen, B.R.J., & Maas, H.L.J. van der (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review*, 17, 321-357.
- Jansen, P.G.W. (1983). *Rasch analysis of attitudinal data*. Dissertatie. Nijmegen, Katholieke Universiteit Nijmegen.
- Jansen, P.G.W. (1984). Een nieuwe correctie voor 'restriction of range'. *Tijdschrift voor Onderwijsresearch*, 9, 180-182.
- Jansen, P.G.W. (1991). *Het beoordelen van managers. Effectiviteit van assessment center methoden bij selectie en ontwikkeling van managers*. Baarn: Uitgeverij H. Nelissen.
- Jansen, P.G.W., Roe, R.A., Vijn, P., & Algera, J.A. (1986). *Validity generalization revisited*. Delft: Delft University Press.
- Jessor, R., & Hammond, K.R. (1957). Construct validity and the Taylor Manifest Anxiety Scale. *Psychological Bulletin*, 54, 161-170.
- Junker, B.W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M.A.J. van Duijn & T.A.B. Sijnders (red.), *Essays on item response theory* (pp. 247-276). New York: Springer.
- Junker, B.W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Kadis, A.L. (1950). Finger painting as a projective technique. In L.E. Abt & L. Bellak (red.), *Projective psychology* (pp. 403-431). New York: Grove Press.
- Kampen, D. van (1977). Eysenck's psychoticisme construct gemeten volgens de vragenlijstmethode. *Nederlands Tijdschrift voor de Psychologie*, 32, 111-122.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16, 277-298.
- Katzell, R.A. (1966). *Involuntary psychological testing in employment settings*. APA Meeting, New York.
- Kelderman, H. (1987). *Quasi-loglinear models for test and item analysis*. Dissertatie. Enschede, Universiteit Twente.
- Kelderman, H., & Eggen, T.J.H.M. (1986). Eenvoudige aanschouwelijke interpretaties van het Raschmodel. In G.R. Buning, T.J.H.M. Eggen, H. Kelderman & W.J. van der Linden (red.), *Het gebruik van het Raschmodel voor een decentraal toetservicesysteem (Rapport 86-3)*. Enschede, Universiteit Twente, Toegepaste Onderwijskunde.
- Kelly, G.A. (1963). *A theory of personality*. New York: Norton.
- Kelly, T.L. (1923). *Statistical methods*. New York: MacMillan.
- Kelly, T.L. (1928). *Crossroads in the mind of man: a study of differentiable mental abilities*. Stanford, CA: Stanford University Press.
- Kleinmuntz, B. (1970). Clinical information processing by computer. In Th.N. Newcomb (red.), *New directions in psychology* (pp. 123-210). New York: Holt, Rinehart & Winston.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.
- Kloot, W.A. van der (1997). *Meerdimensionele schaaltechnieken voor gelijkennis- en keuze-data. Ruimtelijke modellen voor psychologie, marktonderzoek en andere wetenschappen*. Utrecht: Lemma.

- Knippenberg, A. van, & Siero, F.W. (1994). *Multivariate analyse. Beknopte inleiding en toepassingen*. Houten: Bohn Stafleu Van Loghum.
- Koch, K. (1949). *Der Baumtest*. Bern: Huber.
- Kok, F.G. (1988). *Vraagpartijdigheid*. Dissertatie. Amsterdam, Universiteit van Amsterdam.
- Kolen, M.J., & Brennan, R.L. (1995). *Test equating. Methods and practices*. New York: Springer.
- Koning, E. de, Sijtsma, K., & Hamers, J.H.M. (2002). Comparing four IRT models when analyzing two tests for inductive reasoning. *Applied Psychological Measurement*, 26, 302-320.
- Koning, R.J. (1983). Automatisering van psychologisch onderzoek. *De Psycholoog*, 18, 225-237.
- Kouwer, B.J. (1955). *Gewetensproblemen van de toegepaste psychologie*. Groningen: Wolters.
- Kouwer, B.J. (1957). *Tests in de psychologische praktijk*. Utrecht: Bijleveld.
- Kouwer, B.J. (1963). *Het spel van de persoonlijkheid*. Utrecht: Bijleveld.
- Kouwer, B.J., & Kema, G.N. (1957-1958). *Groninger Afsluitingsonderzoek Lager Onderwijs*. Groningen: Wolters.
- Krantz, D.H., Luce, R.D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, vol. 1: Additive and polynomial representations*. New York: Academic Press (in totaal 3 delen).
- Kreft, I., & Leeuw, J. de (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Kubinger, K.D. (red.) (1988). *Moderne Testtheorie*. Weinheim: Beltz Verlag.
- Kuder, G.F., & Richardson, M.W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2, 151-160.
- Lawshe, C.H. (1952). Employee selection. *Personnel Selection*, 5, 31-34.
- Lawshe, C.H., & Balma, M.J. (1948, 1966). *Principles of personnel testing*. New York: McGraw-Hill.
- Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star & J.A. Clausen (red.), *Measurement and prediction* (pp. 362-412). Princeton, NJ: Princeton University Press.
- Leest, P. van (1997). *Persoonlijkheidsmeting bij alloctonen*. Lisse: Swets & Zeitlinger.
- Lenep, D.J. van (1948). *Four Picture Test*. Utrecht: Nederlandse Stichting voor Psychotechniek.
- Lenep, J.E. van (1958). *Beleving en verbeelding in het tekenen*. Amsterdam: Swets & Zeitlinger.
- Levy, S. (1950) Figure drawing as a projective test. In L.E. Abt & L. Bellak (red.), *Projective psychology* (pp. 257-297). New York: Grove Press.
- Licht, R. (1988). *Event-related potential asymmetries and word reading in children*. Dissertatie. Amsterdam, Vrije Universiteit.
- Lienert, G.A. (1961). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lievens, S. (1975). *Thematische apperceptie test voor de pubescentie*. Lisse: Swets & Zeitlinger.
- Likert, R.A. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 149-158.
- Linden, W.J. van der (1980). Decision models for use with criterion-referenced tests. *Applied Psychological Measurement*, 4, 469-492.

- Linden, W.J. van der (1983). *Van standaardtest naar itembank*. Oratie. Enschede, Universiteit Twente.
- Linden, W.J. van der (1984). Some thoughts on the use of decision theory to set cutoff scores: comment on De Gruijter and Hambleton. *Applied Psychological Measurement*, 8, 9-17.
- Linden, W.J. van der (1985). Het besliskundig gebruik van test scores. *Nederlands Tijdschrift voor de Psychologie*, 40, 400-411.
- Linden, W.J. van der (1994). Fundamental measurement and the fundamentals of Rasch measurement. In M. Wilson (red.), *Objective measurement: Theory into practice* (Vol. 2, pp. 3-24). Norwood, NJ: Ablex.
- Linden, W.J. van der (1998, red.). Optimal test assembly. *Applied Psychological Measurement*, 22, 195-302 (speciaal nummer).
- Linden, W.J. van der (2005). *Linear models for optimal test design*. New York: Springer.
- Linden, W.J. van der, & Glas, C.A.W. (2000). *Computerized adaptive testing. Theory and practice*. Dordrecht: Kluwer Academic Publishers.
- Linden, W.J. van der, & Hambleton, R.K. (1997a). *Handbook of modern item response theory*. New York: Springer.
- Linden, W.J. van der, & Hambleton, R.K. (1997b). Item response theory: Brief history, common models, and extensions. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 1-28). New York: Springer.
- Linden, W.J. van der, & Mellenbergh, G.J. (1977). Optimal cutting scores using a linear loss function. *Applied Psychological Measurement*, 1, 593-599.
- Lindquist, E.F. (red.) (1951). *Educational measurement, Part 2*. Washington, DC: American Council on Education.
- Linschoten, J.S. (1964). *Idolen van de psycholoog*. Utrecht: Bijleveld.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694.
- Lord, F.M. (1952). A theory of test scores. *Psychometric Monograph no. 7*, Psychometric Society.
- Lord, F.M. (1970). Item characteristic curves estimated without knowledge of their mathematical form. A confrontation of Birnbaum's logistic model. *Psychometrika*, 35, 43-50.
- Lord, F.M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14, 117-138.
- Lord, F.M. (1980a). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F.M. (1980b). Some how and which for practical tailored testing. In L.J.Th. van der Kamp, W.F. Langerak & D.N.M. de Gruijter (red.), *Psychometrics for educational debates* (pp. 189-205). New York: Wiley.
- Lord, F.M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, 48, 233-245.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lumsden, J. (1976). Test theory. *Annual Review of Psychology*, 27, 251-280.
- Lüscher, M. (1949). *Psychologie des Farben*. Basel: Test Verlag.
- Luteijn, F., & Ploeg, F.A.E. van der (1983). *Handleiding GIT*. Lisse: Swets & Zeitlinger.
- Maanen, L. van, Been, P.H., & Sijtsma, K. (1989). Problem solving strategies and

- the Linear Logistic Test Model. In E.E.Ch.I. Roskam (red.), *Mathematical psychology in progress* (pp. 267-287). New York/Berlin: Springer.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Masters, G.N., & Wright, B.D. (1997). The partial credit model. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 101-121). New York: Springer.
- McArthur, D.L., & Choppin, B.H. (1984). Computerized diagnostic testing. *Journal of Educational Measurement*, 21, 391-397.
- McGee, R.K. (1967). Response set in relation to personality: an orientation. In I.A. Berg (red.), *Response set in personality assessment* (pp. 1-31). Chicago, IL: Aldine.
- Mead, A.D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.
- Meehl, P.E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.
- Meehl, P.E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns and cutting scores. *Psychological Bulletin*, 52, 194-216.
- Meerling (1988). *Methoden en technieken van psychologisch onderzoek. Deel 2 Data-analyse en psychometrie*. Meppel: Boom.
- Meier, N.Ch. (1929-1963). *Meier Art Tests*. New York: Psychological Corporation.
- Meijer, R.R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods*, 8, 72-87.
- Meijer, R.R., & Baneke, J.J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods*, 9, 354-368.
- Meijer, R.R., & Nering, M.L. (1999). Computerized adaptive testing. *Applied Psychological Measurement*, 23, 187-278 (speciaal nummer).
- Meijer, R.R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25, 107-135.
- Meijer, R.R., Sijtsma, K., & Smid, N.G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 14, 283-298.
- Meili, R. (1951). *Lehrbuch der psychologischen Diagnostik*. Bern: Huber.
- Mellenbergh, G.J. (1971). *Studies in studietoetsen*. Dissertatie. Amsterdam: Psychologisch Laboratorium, Universiteit van Amsterdam.
- Mellenbergh, G.J. (1985). Vraagzuiverheid: definitie, detectie en onderzoek. *Nederlands Tijdschrift voor de Psychologie*, 40, 425-435.
- Mellenbergh, G.J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, 19, 91-100.
- Mellenbergh, G.J., & Linden, W.J. van der (1979). The internal and external optimality of decisions based on tests. *Applied Psychological Measurement*, 3, 257-273.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 33-45). Hillsdale, NJ: Erlbaum.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.

- Millman, J., & Arter, J.A. (1984). Issues in item banking. *Journal of Educational Measurement*, 21, 315-330.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mokken, R.J. (1971). *A theory and procedure of scale analysis*. Den Haag/Berlin: Mouton/De Gruyter.
- Mokken, R.J. (1997). Nonparametric models for dichotomous responses. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 351-367). New York: Springer.
- Mokken, R.J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 417-430.
- Molenaar, I.W. (1983). Rasch, Mokken en schoolbeleving. In S. Lindenberg & F.N. Stokman (red.), *Modellen in de sociologie* (pp. 195-213). Deventer: Van Loghum Slaterus.
- Molenaar, I.W. (1997). Nonparametric models for polytomous items. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 369-380). New York: Springer.
- Morgan, C.D., & Murray, H.A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry*, 34, 289-306.
- Mulaik, S.A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Mulder, J.L. (1997). *Het meten van verbale lange termijn geheugenstoornissen bij neurologische en psychiatrische patiënten*. Dissertatie. Amsterdam, Universiteit van Amsterdam.
- Münsterberg, H. (1914). *Grundzüge der Psychotechnik*. Leipzig: Barth.
- Murphy, K.R., & Davidshofer, C.O. (1998). *Psychological testing. Principles and applications*. Upper Saddle River, NJ: Prentice Hall.
- Murray, H.A. (1935, 1943). *Thematic Apperception Test*. Cambridge, MA: Harvard University Press.
- Naerssen, R.F. van (1962). *Selectie van chauffeurs: onderzoeken ten behoeve van de selectie van chauffeurs bij de Koninklijke Landmacht*. Groningen: Wolters.
- Naerssen, R.F. van (1965). Application of the decision-theoretical approach to the selection of drivers. In L.J. Cronbach & G.C. Gleser (red.), *Psychological tests and personnel decisions* (pp. 273-290). Urbana, IL: University of Illinois Press.
- Niemöller, K., & Schuur, W.H. van (1983). Stochastic models for unidimensional scaling: Mokken and Rasch. In D. McKay, N. Schofield & P. Whiteley (red.), *Data analysis and the social sciences* (pp. 120-170). London: Francis Pinter Publ.
- Nijenhuis, J. te (1997). *Comparability of test scores for immigrants and majority group members in The Netherlands*. Dissertatie. Amsterdam, Vrije Universiteit.
- NIP (1976). *Beroepsethiek voor psychologen*. Amsterdam: Nederlands Instituut van Psychologen.
- Nitko, A.J., & Hsu, T.C. (1984). A comprehensive microcomputer system for classroom testing. *Journal of Educational Measurement*, 21, 377-390.
- Novick, M.R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1-18.
- Novick, M.R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13.
- Nunnally, J.C. (1972). *Educational measurement and evaluation*. New York: McGraw-Hill.
- Nunnally, J.C. (1978)/Nunnally, J.C. & Bernstein, I.R. (1994). *Psychometric theory*. New York: McGraw-Hill.

- Oppenheim, A.N. (1966). *Questionnaire design and attitude measurement*. London: Heinemann.
- Ord, J.G. (1968). *The Pacific Design Construction Test and Manual*. Melbourne: Australian Council on Educational Research.
- Orlebeke, J.F. (1988). Motivatie. In J.F. Orlebeke, P.J.D. Drenth, R.H.C. Janssen & C. Sanders (red.), *Compendium van de psychologie* (deel 4) (pp. 131-149). Muiderberg: Coutinho.
- Ortar, G. (1972). Some principles for adaptation of psychological tests. In L.J. Cronbach & P.J.D. Drenth (red.), *Mental tests and cultural adaptation* (pp. 111-120). Den Haag: Mouton.
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana-Champaign, IL: University of Illinois Press.
- Owers, W.H. (1976). Background data. In M.D. Dunnette (red.), *Handbook of industrial and organization psychology* (pp. 13-49). Chicago, IL: Rand McNally.
- Parshall, C.G., Davey, T., & Pashley, P.J. (2000). Innovative item types for computerized testing. In W.J. van der Linden & C.A.W. Glas (red.), *Computerized adaptive testing. Theory and practice* (pp. 129-148). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Parshall, C.G., Spray, J.A., Kalohn, J.C., Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Petersen, N.S., & Novick, M.R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 13, 3-29.
- Pfister, M. (1951). *Farbpyramidentest*. Bern: Huber.
- Piaget, J., Inhelder, B., & Szeminska, A. (1948). *La géométrie spontanée de l'enfant*. Paris: Presses Universitaires de France.
- Pieters, J.P.M., & Ven, A.H.G.S. van der (1982). Precision, speed and distraction. *Applied Psychological Measurement*, 6, 93-109.
- Pintner, R., & Paterson, D.G. (1917, 1927). *Pintner-Paterson Scale of Performance Tests*. New York: Appleton.
- Poortinga, Y.H. (1982). Cross-culturele psychologie en minderhedenonderzoek. *De Psycholoog*, 17, 708-720.
- Porteus, S.D. (1933). *The Maze Test and Mental Differences*. Vineland, NJ: Smith.
- Post, W.J. (1992). *Nonparametric unfolding models. A latent structure approach*. Leiden: DSWO Press.
- Raad, B. de, & Perugini, M. (red.) (2002). *Big Five assessment*. Seattle, WA: Hogrefe & Huber Publishers.
- Raijmakers, M.E.J., Jansen, B.R.J., & Maas, H.L.J van der (2004). Rules and development in triad classification task performance. *Developmental Review*, 24, 289-321.
- Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 495-502.
- Ramsay, J.O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 611-630.
- Ramsay, J.O. (1997). A functional approach to modeling test data. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 381-394). New York: Springer.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.

- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 18, 49-57.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Raven, J.C. (1938). *Progressive matrices*. London UK: H.K. Lewis.
- Raven, J.C., Court, J.H., & Raven, J. (1979). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section I General Overview*. London: H.K. Lewis.
- Resing, W.C.M. (1990). *Intelligentie en leerpotentieel; een onderzoek naar leerpotentieel van jonge kinderen uit het basis- en speciaal onderwijs*. Lisse: Swets & Zeitlinger.
- Resing, W.C.M., Bleichrodt, N., & Drenth, P.J.D. (1986). Het gebruik van de RAKIT bij allochtoon etnische groepen. *Nederlands Tijdschrift voor de Psychologie*, 41, 179-188.
- Resing, W.C.M., & Drenth, P.J.D. (2001). *Intelligentie: Weten en meten*. Amsterdam: Uitgeverij Nieuwezijds.
- Roe, R.A. (1983). *Grondslagen der personeelsselectie*. Assen: Van Gorcum.
- Roe, R.A., Algera, J.A., Jansen, P.G.W., & Vijn, P. (1983a). Ernst met methodologische deskundigheid. *De Psycholoog*, 18, 202-206.
- Roe, R.A., Algera, J.A., Jansen, P.G.W., & Vijn, P. (1983b). De olifant en de nieuwe kleren van de keizer: een antwoord aan Hofstee. *De Psycholoog*, 18, 503-512.
- Roid, G.H., & Haladyna, T.M. (1982). *A technology for test-item writing*. New York: Academic Press.
- Rorschach, H. (1921). *Psychodiagnostik*. Bern: Huber.
- Rosenzweig, S. (1945). The picture association method and its application in a study of reactions to frustrations. *Journal of Personality*, 14, 3-23.
- Rost, J. (1988). *Quantitative und qualitative probabilistische Testtheorie*. Bern: Huber.
- Rost, J. (1996). *Testtheorie. Testkonstruktion*. Bern: Huber.
- Rotter, J.B. (1951). Word association and sentence completion methods. In H.H. Anderson & G.L. Anderson (red.), *An introduction to projective techniques* (pp. 279-311). New York: Prentice-Hall.
- Roussos, L., & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20, 355-371.
- Rudinger, G., Chaselon, F., Zimmermann, E.J., & Henning, H.J. (1985). *Qualitative Daten*. München: Urban & Schwarzenberg.
- Rulon, P.J., Tiedeman, D.V., Tatsuoka, M.M., & Langmuir, C.R. (1967). *Multivariate statistics for personnel classification*. New York: Wiley.
- Sackett, P.R., Schmitt, N., Tenopyr, M.L., Kehoe, J., & Zedeck, S. (1985). Commentary on forty questions about validity generalization and meta-analysis. *Personnel Selection*, 38, 697-798.
- Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometric Monograph*, no. 17, Psychometric Society.
- Samejima, F. (1977). Weakly parallel tests in latent trait theory with some criticisms of classical test theory. *Psychometrika*, 42, 193-198.
- Samejima, F. (1997). Graded response model. In W.J. van der Linden & R.K. Hambleton (red.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.
- Sandbergen, G. (1973). *Zekerheidsaanduiding in het meten van studieprestaties*. Dissertatie. Amsterdam, Universiteit van Amsterdam.
- Saunders, S.B. (1956). *The clinical interaction, with special reference to the Rorschach*. New York: Harper.

- Scheiblechner, H. (1972). Das Lernen und Lösen komplexer Denkaufgaben. *Zeitschrift für experimentelle und angewandte Psychologie*, 19, 476-506.
- Schlien, J.M. (1958). Mental testing and modern society. *The Humanist*, 18, 356-364.
- Schmidt, F.L. (1988). Validity generalization and the future of criterion-related validity. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 173-189). Hillsdale, NJ: Erlbaum.
- Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt, F.L., Hunter, J.E., & Caplan, J.R. (1981). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, 66, 261-273.
- Schmidt, F.L., Hunter, J.E., Pearlman, K., & Hirsh, H.R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Selection*, 38, 697-798.
- Schmitt, N., & Chan, D. (1998). *Personnel selection. A theoretical approach*. Thousand Oaks, CA: Sage.
- Schuur, W.H. van (2003). Mokken scale analysis: between the Guttman scale and parametric item response theory. *Political Analysis*, 11, 139-163.
- Sechrest, L. (1963). Incremental validity: a recommendation. *Educational and Psychological Measurement*, 23, 153-158.
- Shavelson, R.J., & Webb, N.M. (1991). *Generalizability theory. A primer*. Thousand Oaks, CA: Sage.
- Shye, S., Elizur, D., & Hoffman, M. (1994). *Introduction to facet theory. Content design and intrinsic data analysis in behavioral research*. Thousand Oaks, CA: Sage.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. Tokyo: McGraw-Hill Kogakusha, LTD (new version by S. Siegel & N.J. Castellan Jr., 1988).
- Sijtsma, K. (1983). Rasch-homogeniteit empirisch onderzocht. *Tijdschrift voor Onderwijsresearch*, 8, 104-121.
- Sijtsma, K. (1993). Kaf en koren onder Nederlandse tests. Recensie van 'Documentatie van tests en testresearch in Nederland'. *De Psycholoog*, 28, 502-503.
- Sijtsma, K. (1998). Methodology Review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement*, 22, 3-31.
- Sijtsma, K., Algera, J.A., & Altink, W.M.M. (1987). Een stap terug? *Nederlands Tijdschrift voor de Psychologie*, 42, 250-253.
- Sijtsma, K., & Meijer, R.R. (2001). The person response function as a tool in person-fit research. *Psychometrika*, 66, 191-207.
- Sijtsma, K., & Molenaar, I.W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika*, 52, 79-97.
- Sijtsma, K., & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.
- Sijtsma, K., & Verweij, A.C. (1999). Knowledge of solution strategies and IRT modeling of items for transitive reasoning. *Applied Psychological Measurement*, 23, 55-68.
- Smith, P.B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural psychology*, 35, 50-61.
- Smith, M., & Smith, P. (2005). *Testing people at work: Competencies in psychometric testing*. Oxford, UK: BPS Blackwell.
- Snijders, J.Th., & Snijders-Oomen, N. (1958). *Niet-verbaal intelligentieonderzoek van horenden en doofstommen: Snijders-Oomen niet-verbale intelligentieschaal SON*. Groningen: Wolters.

- Snijders, J.Th., Tellegen, P.J., & Laros, J.A. (1988). *Snijders-Oomen niet-verbale intelligentietest SON-R 5½-17. Verantwoording en handleiding*. Groningen: Wolters-Noordhoff.
- Snijders, J.Th., & Verhage, F. (1962). *Voorlopige handleiding bij de Groninger Intelligentie Test*. Amsterdam: Swets & Zeitlinger.
- Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Snijders-Oomen, A.W.M. (1943). *Intelligentie onderzoek van doofstomme kinderen*. Nijmegen: Berkhout.
- Social Science Research Division, Standard Oil Comp NJ (1961). *The early identification of management potential*. New York: SONJ (internal report).
- Spada, H. (1976). *Modelle des Denkens und Lernens*. Bern: Huber.
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-195.
- Spearman, C.E. (1927). *The abilities of man, their nature and measurement*. New York: Macmillan.
- Spitz, J.C. (1968). *Statistiek voor psychologen, pedagogen, sociologen*. Amsterdam: Noord-Hollandsche uitgevers mij.
- Staabs, G. von (1951). *Der Scenotest*. Zürich: Hirzel.
- Stern, W. (1911). *Die differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Stevens, S.S. (1951). *Handbook of experimental psychology*. New York: Wiley.
- Steyer, R., & Eid, M. (2001). *Messen und Testen*. Berlin: Springer
- Stigler, S.M. (1986). *The history of statistics. The measurement of uncertainty before 1900*. Cambridge MA: Harvard University Press.
- Stinissen, J., Willems, P.J., Coetsier, P., & Hulsman, W.L.L. (1970). *WAIS, Nederlandstalige bewerking. Handleiding*. Amsterdam: Swets & Zeitlinger.
- Stokman, F.N., & Schuur, W.H. van (1980). Basic scaling. *Quality & Quantity*, 14, 5-30.
- Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F., Star, S.A., & Clausen, J.A. (red.) (1950). *Measurement and prediction*. Princeton, NJ: Princeton University Press.
- Stout, W.F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55, 293-325.
- Stout, W.F. (2002). Psychometrics: From practice to theory and back. *Psychometrika*, 67, 485-518
- Stout, W.F., Habing, B., Douglas, J., Kim, H., Roussos, L., & Zhang, J. (1996). Conditional covariance based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, 20, 331-354.
- Strien, J.W. van (1988). *Handedness and hemispheric laterality*. Dissertatie. Amsterdam, Vrije Universiteit.
- Strien, P.J. van (1966). *Kennis en communicatie*. Utrecht: Bijleveld.
- Strien, P.J. van (1976). *Personeelsselectie in discussie*. Meppel: Boom.

- Strong, E.K. (1927-1959). *Strong Vocational Interest Blank for Men and Women*. Stanford University: Stanford University Press.
- Sudman, S., & Bradburn, N.M. (1982). *Asking questions. A practical guide to questionnaire design*. San Francisco, CA: Jossey-Bass.
- Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush & E. Galanter (red.), *Handbook of Mathematical Psychology* (pp. 1-76). New York: Wiley.
- Szondi, L. (1947). *Experimentelle Triebdiagnostik*. Bern: Huber.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics (4th edition)*. Boston, MA: Allyn and Bacon.
- Tatsuoka, K.K., (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P.D. Nichols, S.F. Chipman, & R.L. Brennan (red.), *Cognitively diagnostic assessment* (pp. 327-359). Hillsdale, NJ: Erlbaum.
- Tatsuoka, K.K., & Tatsuoka, M.M. (1982). Detection of aberrant response patterns and their effect on dimensionality. *Journal of Educational Statistics*, 7, 215-131.
- Tatsuoka, K.K., & Tatsuoka, M.M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement*, 20, 221-230.
- Tatsuoka, M.M. (1971). *Multivariate analysis*. New York: Wiley.
- Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. Discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Tellegen, P.J., Zegers, F.E., & Liebrand, W.B.G. (1999). *Testserie Hoger Niveau Adaptieve Computerversie*. Lisse: Swets & Zeitlinger.
- Terman, L.M. (1916). *The measurement of intelligence*. Boston, MA: Houghton Mifflin.
- Terman, L.M. (1960). *The Stanford-Binet Intelligence Scale*. Boston, MA: Houghton-Mifflin.
- Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika*, 50, 411-420.
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, 567-577.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H.I. Braun (red.), *Test validity* (pp. 147-169). Hillsdale, NJ: Erlbaum.
- Thomson, G.H. (1938, 1951). *The factorial analysis of human ability*. Boston, MA: Houghton Mifflin.
- Thorndike, R.L. (1949). *Personnel selection*. New York: Wiley.
- Thorndike, R.L. (red.) (1971a). *Educational measurement*. Washington DC: American Council on Education.
- Thorndike, R.L. (1971b). Concepts of culture-fairness. *Journal of Educational Measurement*, 8, 63-70.
- Thorndike, R.L., & Hagen, E. (1959). *Ten thousand careers*. New York: Wiley.
- Thurstone, L.L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406-427.
- Thurstone, L.L. (1935). *The vectors of mind*. Chicago, IL: University of Chicago Press.
- Thurstone, L.L. (1938). Primary mental abilities. *Psychometric Monograph*, 1.
- Top, W. (1973). Naar de bliksem met de psychotechniek. *De Vakbeweging*, 64, 1.
- Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Trabasso, T., Riley, C.A., & Wilson, E.G. (1975). The representation of linear order

- and spatial strategies in reasoning: A developmental study. In R.J. Falmagne (red.), *Reasoning: representation and process in children and adults* (pp. 201-229). Hillsdale, NJ: Erlbaum.
- Traub, R.E., & Lam, Y.R. (1985). Latent structure and item sampling models for testing. *Annual Review of Psychology*, 36, 19-48.
- Turnbull, W.W. (1966). Testing for guidance and selection. In C.I. Chase & H. Glenn Ludlov (red.), *Readings in educational and psychological measurement* (pp. 299-311). Boston, MA: Houghton Mifflin.
- Ven, A.H.G.S. van der (1971). Time-limit tests. A critical evaluation. *Nederlands Tijdschrift voor de Psychologie*, 26, 580-591.
- Ven, A.H.G.S. van der (1976). An error score model for time-limit tests. *Tijdschrift voor Onderwijsresearch*, 1, 215-226.
- Vernon, Ph.E. (1950). *The structure of human abilities*. New York: Wiley.
- Vernon, Ph.E. (1963). *Personality assessment: A critical survey*. London: Methuen.
- Vernon, Ph.E. (1969). *Intelligence and cultural environment*. London: Methuen.
- Verouden, G., Ross, F., Stef, A., & Scheele, J. (1987). *Psychologische selectie van etnische minderheden*. Amsterdam: GG & GD.
- Verweij, A.C. (1994). *Scaling transitive inference in 7-12 year old children*. Dissertatie. Amsterdam, Vrije Universiteit.
- Vijn, P., & Molenaar, I.W. (1981). Robustness regions for dichotomous decisions. *Journal of Educational Statistics*, 6, 205-235.
- Vijver, F.J.R. van de (1987). Het gebruik van computer-ondersteunde tests in de diagnostische praktijk. *De Psycholoog*, 22, 10-15.
- Vijver, F.J.R. van de (1988). Systematizing the item content in test design. In R. Langeheine & J. Rost (red.), *Latent trait and latent class models* (pp. 291-307). New York: Plenum Press.
- Vijver, F.J.R. van de, & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology*, 79, 852-859.
- Visser, R.S.H. (1973). Een nieuwe systematische indeling van tests. *Gedrag*, 1, 225-239.
- Vrijhof, B.J., Mellenbergh, G.J., & Brink, W.P. van den (1983). Assessing and studying utility functions in psychometric decision theory. *Applied Psychological Measurement*, 7, 341-357.
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Erlbaum.
- Wartegg, E. (1953). *Schichtdiagnostik. Der Zeichentest (wzt)*. Göttingen: Verlag für Psychologie.
- Watson, G., & Glaser, E.M. (1964). *Critical thinking appraisal*. New York: Psychological Corporation.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.
- Wechsler, D. (1955a). *Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
- Wechsler, D. (1955b). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.
- Weeren, P. van (1965). *De Nederlandse Onderwijs-Differentiatie Test*. Groningen: Wolters.
- Weiss, D.J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53, 774-789.

- Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.
- Wiegersma, S. (1963). *Selectieproblemen*. Amsterdam: Swets & Zeitlinger/NIPG.
- Wiggins, J.S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wilde, G.J.S. (1962). *Neurotische labiliteit gemeten volgens de vragenlijstmethode*. Amsterdam: Van Rossen.
- Willems, P.J. (1964). Een proefschrift beproefd. *Nederlands Tijdschrift voor de Psychologie*, 19, 124-135.
- Willink, B. (1998). *De tweede Gouden Eeuw. Nederland en de Nobelprijzen voor natuurwetenschappen 1870-1940*. Amsterdam: Uitgeverij Bert Bakker.
- Wit, J. de, & Compaan, E. (2005). *Differentiële Aanleg Test DAT NL, handleiding A*. London: Harcourt.
- Witkin, H.A. (1949). *Perception and personality*. New York: Harper.
- Wolff, Ch.J. de, & Bosch, G. van den (1980). Personeelsaanname. In P.J.D. Drenth, Hk. Thierry, P.J. Willems & Ch.J. de Wolff (red.), *Handboek arbeids- en organisatiepsychologie* (pp. 2.7.1-2.7.25). Deventer: Van Loghum Slaterus.
- Wollenberg, A.L. van den (1979). *The Rasch model and time-limit tests*. Dissertatie. Nijmegen, Katholieke Universiteit Nijmegen.
- Wollenberg, A.L. van den (1983). Measuring subjects on a joint scale by means of time-limit tests. *Tijdschrift voor Onderwijsresearch*, 8, 145-156.
- Wollenberg, A.L. van den (1985). Speed and precision in intelligence tests: Facts or artifacts? *Tijdschrift voor Onderwijsresearch*, 10, 69-81.
- Wright, B.D., & Bell, S.R. (1984). Item banks: What, why, how. *Journal of Educational Measurement*, 21, 331-345.
- Yamamoto, K. (1966). Psychological testing: invasion of privacy? *Educational Leadership*, 23, 363-368.
- Young, F.A. (1966). Involuntary psychological testing in basic research. Symposium APA Involuntary Psychological Testing, New York.
- Zaal, J.N. (1998). Assessment center methods. In P.J.D. Drenth, Hk. Thierry, & Ch.J. de Wolff (red.), *Handbook of work and organizational psychology* (pp. 89-122) (2nd edition). Hove: Psychology Press.
- Zeeuw, J. de (1971). *Algemene psychodiagnostiek I*. Amsterdam: Swets & Zeitlinger.
- Zeeuw, J. de, Dekker, R., & Resing, W.C.M. (2004). *Algemene psychodiagnostiek I. Testmethoden*. Leiden: PITS.
- Zegers, F.E., & Berge, J.M.F. ten (1982). Necessary and sufficient conditions for parallelism of tests in classical test theory. *Tijdschrift voor Onderwijsresearch*, 7, 76-79.
- Zwinderman, A.H. (1991). *Studies of estimating and testing Rasch models*. Dissertatie. Nijmegen, Katholieke Universiteit Nijmegen.

Register

A

absolute norm 175
actuarische en intuïtieve voorspelling 414
actuarische methode van voorspellen 27
adaptief testen 169, 302
aftestgrens 404
afwijkende patronen van itemscores 312
afwijkingsscore 464
alfacoëfficiënt 215
Algemene Standaard Testgebruik (AST-NIP) 439
Amsterdamse KinderIntelligentieTest (RAKIT) 96
antecedent probability 398, 405
antwoordtendenties 385
AST-NIP *zie* Algemene Standaard Testgebruik
attenuatiecorrectie 240
averechte diagnostiek 71

B

bandwijdte ('bandwidth') 430
base rate 398, 405
begripsvalidering 370
begripsvaliditeit 332
beoordeling
–, van groepen 70
–, van individuen 68
–, van invloed van situaties en methoden 71
–, van kwaliteit items in vooronderzoek 136
betekenisanalyse 338, 376
betrouwbaarheid 190, 328
–, acceptabele waarden 202

–, definitie 202
–, en spreiding van scores 243
–, en testlengte 235
–, en validiteit 238
–, kenmerk van test 50
–, van heterogene tests 244
–, van verschillen 241
betrouwbaarheidsindex 238
betrouwbare score 194, 290
bewerkte normen 172
bewerkte scores 172
bezwaren tegen testen
–, levensbeschouwelijke en menselijke 440 *zie* uniciteit van de mens
–, technische en methodologische 445
Big Five persoonlijkheidsstructuur 103
Binet-Simon-test 16, 21
biografische vragenlijsten 91
Birnbaum-model 273, 274, 285, 303
brede-bandtests 431

C

calibreren 296
Centraal Instituut voor ToetsOntwikkeling (CITO) 29, 34
–, Eindtoets Basisonderwijs 174
checklist 152
CITO *zie* Centraal Instituut voor ToetsOntwikkeling
Cohens kappa 48
collectieve algemene intelligentietest 80
Commissie Testaangelegenheden Nederland (COTAN) 33, 445
compensatorisch 415
computertests

- , adaptief testen 169
- , technologische bijdragen 162
- , wetenschappelijke bijdragen 166
- conceptueel criterium 344
- concurrent validity 334
- confirmerende validering 375
- confirmerende validiteit 391
- congruent validity 339
- conjunctief 415
- construct validity 337
- constructie van items 116
- contaminatie van het criterium 352
- content validity 335
- COTAN zie Commissie Testaangelegenheden Nederland
- covariantie van variabelen 466
- criterion-referenced measurement 175
- criteriumgedrag 344
- criteriummaat 344
- criteriumprestatie 344
- Cronbachs alfa 215
- cultuurvrije tests 110

D

- Deviatie-IQ 185
- deviatiescore 464
- diagnostisch toetsen per computer 164
- dichotoom item 133, 462
- differential item functioning 307
- differentiële validiteit 425
- directe tests 112
- discriminante validering 375
- discriminante validiteit 391
- discriminatie in het testgebruik 453
- discriminatieparameter 262
- discriminerende testinterpretatie 454
- doelinformatiefunctie 300
- drie-parameter logistische model 276, 285, 303

E

- educational measurement, invloed op testtheorie 31
- Educational Testing Service (ETS) 29
- een-parameter logistische model 264
- efficiëntie, kenmerk van test 41
- enkelvoudige algemene niveautests 78
- equivalering van scores en kenmerken van items 294
- essayvorm van items 119
- ethiek van het testen 437

- ETS zie Educational Testing Service
- experimentele psychologie, invloed op testtheorie 17

F

- face validity 339
- factoranalyse 378
- false positives 405
- fenomenologische of intuïtieve methode 27
- fysiologische methoden van testen 89

G

- gedrag in testsituatie
 - , van de proefleider 150
 - , van de proefpersoon 148
- gemiddelde van een variabele 463
- generaliseerbaarheid van metingen 245
- generaliseerbaarheidstheorie 246
- genetica, invloed op testtheorie 18
- genormaliseerde standaardscores 184
- geprecodeerde vraagvorm 119
- gesloten vraagvorm 119, 121
- gestratificeerde alfacoëfficiënt 245
- gewogen loting 457
- giscorrectie van scores 159
- giskans op goede antwoord 260
- groepsobservatietests 88
- groepstest 23, 106

H

- herhaalbaarheid van metingen 190
- heteroscedastische relatie 352
- historische ontwikkeling van het testen 15
- homoscedastische relatie 354
- hypothetische begrippen 64, 66

I

- identificatie van de te meten eigenschap 62
- incremental validity 340
- indirecte tests 112
- individualisme
 - , gecorrigeerd 458
 - , gekwalificeerd 457
- individuele beslissingen 400, 426
- individuele intelligentietest voor volwassenen 79
- individuele observatietests 88
- individuele ontwikkelingstest 79

individuele test 23, 106
 informatiefunctie 293, 300
 inhoudvaliditeit 335
 institutioneel beslissingsproces 420
 institutionele beslissingen 400
 intelligentiequotiënt (IQ) 176
 intelligentietest 16, 20, 78
 interbeoordelaarsbetrouwbaarheid 46, 49
 interestetests 90
 interne-consistentiemethode 215
 Interuniversitair Onderzoeksinstituut voor Psychometrie en Sociometrie (IOPS) 35
 intervalschaal 54, 275, 288
 IOPS zie Interuniversitair Onderzoeksinstituut voor Psychometrie en Sociometrie
 ipsatieve scores 69
 itembank 162, 294
 iteminformatiefuncties 293
 item-karakteristieke curve 258
 item-karakteristieke functie 258
 item-responsfunctie 258
 item-responstheorie 170, 253
 –, voor polytoom gescoorde items 318
 items
 –, voor prestatieniveautests 125
 –, voor tests voor gedragswijze 129
 itemscores 133

K

kans op een specifieke respons 257
 kanskapitalisatie 357
 klassieke testtheorie 170, 194
 KR20 222
 kritische testscore 404
 kruisvalidering 357
 kwalitatieve prestatietests 93
 kwantificering
 –, van antwoorden 131
 –, van reacties 64, 116

L

lambda₂-coëfficiënt 225
 latente-klassenanalyse 242
 leerpotentieeltests 82
 Likert-items 129
 lineaire combinaties van variabelen 468
 lineaire regressie 227

lineaire transformaties van variabelen 467
 lokale betrouwbaarheid 292

M

Maximalisering
 –, van doelmatigheid 455
 –, van gelijke kansen 457
 mediaan 463
 meerkeuzevraagvorm 119
 meetfout 194
 meetniveau 53
 mentale leeftijd 21
 meten
 –, bij fiat 254, 287
 –, bij implicatie 254, 287
 metrische schalen 288
 misbruik van tests
 –, binnendringen in het privéleven 451
 zie ook recht op privacy
 –, discriminatie 453
 –, misleiding 450
 –, schending van vertrouwen 449
 model
 –, van dubbele monotomie 283, 285
 –, van monotone homogeniteit 278, 285
 modellen volgens Mokken 278
 moderatorvariabelen 355
 modus 463
 moeilijkheidsparameter 261
 morfologische methoden van testen 89
 motivatie van de onderzochte 149
 mucoëfficiënten 226
 multipele of meervoudige regressie-analyse 356
 multiple-choicevraagvorm 119
 multi-stage testing 303
 multitrek-multimethodebenadering 389
 multivariate informatie 400

N

nadere bepaling van het criterium 343
 negatieve missers 405
 negatieve treffers 405
 NEO – meting van de Big-Five persoonlijkheidsstructuur 103
 NIP 446
 niveautest 107
 –, voor gedragswijze 94
 nominale schaal 54

nomologisch netwerk 66, 372
 nomologische validering 371
 normen
 –, gebaseerd op gemiddelde en spreiding 182
 –, gebaseerd op rangorde 179
 –, vergelijking met absolute standaard 175
 –, verhoudingsnormen 176
 normering 44
 –, kenmerk van test 43
 norm-referenced measurement 175

O

objectiviteit
 –, als onafhankelijkheid van de specifieke beoordelaar 46
 –, kenmerk van test 46
 observatietest 26, 87
 odds van persoon *i* op item *g* 289
 onderdelen van een test 38
 onderwijskundig meten, invloed op testtheorie 31
 ongekwalificeerd individualisme 456
 open vraag 120, 429
 open vraagvorm of opdracht 119
 operationalisering van de eigenschap 63, 64
 operationalisme 59
 opvattingen over meten 57, 59
 ordinale schaal 54, 281, 283

P

parallele tests 206
 parallelvormmethode 206
 percentielscores 179
 persoonlijkheidskenmerken 103
 persoonlijkheidstest 26
 persoonlijkheidsvragenlijsten 26, 91
 persoon-responsfunctie 315
 plaatsing van personen
 –, en kwalitatieve verschillen 423
 –, en niveaoverschillen 419
 plaatsingsbeslissingen 419
 plafondeffecten 295
 polytome items 134, 462
 populatie-onafhankelijkheid 275, 267, 278, 281, 283
 positieve missers 404
 positieve treffers 404
 predictieve validiteit 332, 341, 346

predictive validity 334
 product-momentcorrelatie 467
 projectietest 26, 95, 431
 pseudokansniveauparameter 261
 psychiatrie, invloed op testtheorie 16
 psychometrische kwaliteit van bestaande tests 445
 p-waarde van het item 138

Q

Q-technique of Q-sort 93

R

RAKIT zie Amsterdamse KinderIntelligentieTest 96
 rangorde, vaststellen van 44
 Rasch-model 263, 285, 303
 rating scale-item 122
 ratio scale 56
 ratioschaal 288
 recht op privacy 452 zie ook misbruik van tests, binnendringen in het privéleven 451
 relatie
 –, heteroscedastische 352
 –, homoscedastische 354
 relatieonderzoek van testscores 377
 representational measurement 59, 66
 Role Construct Repertoire 92
 ruwe score 173, 462

S

schaaltheorie, invloed op testtheorie 31
 schaaltransformaties, toegestaan 265
 schoolvorderingentests 28, 34
 scoring
 –, van antwoorden 151
 –, van reacties op gecodeerde items 153
 –, van reacties op items met openvraagvorm 152
 selectie- en diagnostische tests 28
 selectiepsychologie 30
 selectieratio 405
 semantische differentiaal 92
 sequentiële beslissingsmodellen 417
 snelheidstest 107
 sociale wenselijkheid 384
 somato-fysiologische methoden van testen 88
 soortgenootvaliditeit 339

Spearman-Brown-formule 213, 224, 235
 speciale niveaustests 83
 splitsingsmethode 212
 standaarddeviatie van een variabele 465
 standaardisatie
 –, kenmerk van test 42
 –, van het gedrag van de testleider 150
 standaardmeetfout 197, 231, 291
 –, van persoon i 197
 –, van testscore 203
 standaardschattingfout 232
 standaardscore 465
 –, of z-score 182
 standaardtest 170
 stanines 185
 stereotypen-bestendigend karakter van tests 448
 structuuronderzoek van testcores 377
 succeskans 258
 succesratio 405
 suppressorvariabele 354
 surplus-betekenis van hypothetische begrippen 64
 synthetische validiteit 338

T

testangst 149
 testbatterijen
 –, voor geschiktheden 82
 –, voor intelligentiefactoren 81
 testconstructie op basis van een item-bank 299
 testen
 –, op maat 302
 –, per computer 161 zie computertests
 testervaring 147
 testformulieren 39
 testhandleiding 39
 test-hertestmethode 210
 testinformatiefunctie 293
 testmateriaal 38
 test
 –, definitie 67
 –, voor gedragswijze 86
 –, voor prestatieniveau 78
 –, voor speciale geschiktheden 84
 –, voor speciale intelligentiefactoren 83
 –, voor speciale niet-intelligentiefactoren 84

testcores op intervalniveau 56
 test-wisness 147
 toegestane operaties op schaalwaarden 53
 toegestane schaaltransformaties 55, 265
 toevalscorrectie 156
 toevalskans 405
 trace line 258
 transparantheid 340
 trapsgewijze selectie 417
 trekvalidering 371
 T-scores 185
 twee-parameter logistische model 274
 two-stage testing 303

U

uniciteit van de mens 440 zie bezwaren tegen testen
 univariate informatie 400
 utiliteit van een testprocedure 411

V

validiteit 328, 408
 –, kenmerk van test 52
 validiteitsgeneralisatie 363
 variantie van een variabele 464
 veelvoudige algemene niveaustests 81
 veelvoudige minimumscore 414
 vergelijking klassieke testtheorie en item-responstheorie 320
 verhoudingsnormen 176
 verhoudingsschaal 56
 verschilschaal 266, 288
 verwachte beslissingsrendement 420
 verwachte rendement 427
 vloereffecten 295
 vorderingentests 85
 vraagonzuiverheid 306, 317
 vraagpartijdigheid 307

W

waarde- en attitudetests 91
 weging
 –, van items 135
 –, van itemscores 160

Z

zelfbeoordelingen 90
 z-score 465
 zuiverheid ('fidelity') 430