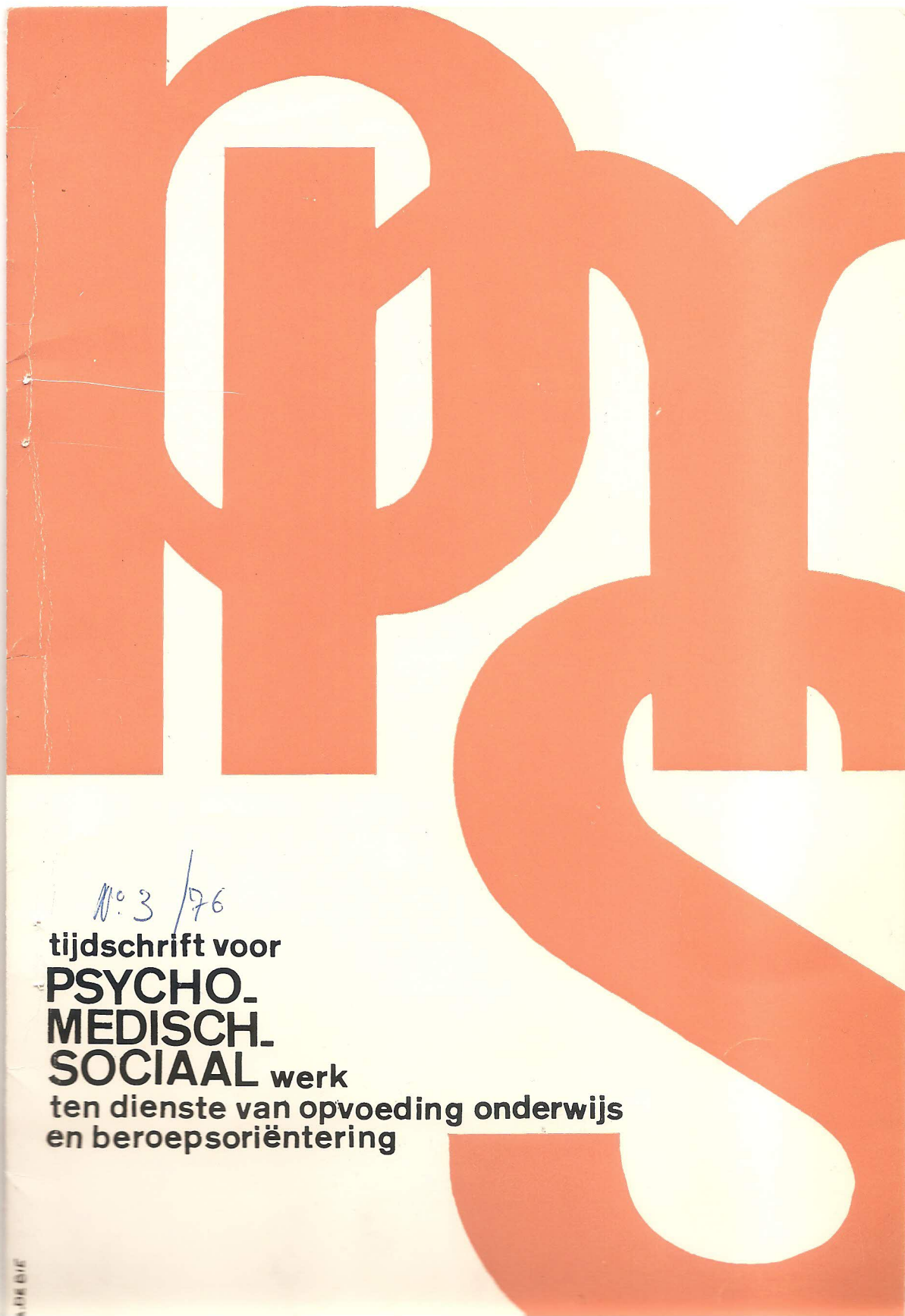


Constructie van meerkeuze-vragen

Een synthetisch overzicht van formele kenmerken

door A. VERKENS
Psychopedagoog



№ 3 / 76

tijdschrift voor

**PSYCHO-
MEDISCH-
SOCIAAL**

werk

ten dienste van opvoeding onderwijs
en beroepsoriëntering

Constructie van meerkeuze-vragen

Een synthetisch overzicht van formele kenmerken

door A. VERKENS
Psychopedagoog

INLEIDING

De meerkeuze-vraagvorm kent een ruime toepassing in schoolvorderingentests, objectieve evaluatieproeven en studietoetsen. In de literatuur over itemconstructie worden m.b.t. deze vraagvorm tal van richtlijnen opgesomd. De toepassing van deze richtlijnen bij de itemredactie maakt het mogelijk om begripsvalide items te realiseren. In dit artikel beperken we ons tot de richtlijnen betreffende de formele aspecten van de meerkeuze-vragen en in het bijzonder van de alternatieven. We beschouwen als formele aspecten die kenmerken van de meerkeuze-vraag die inherent zijn aan de itemvorm en relatief onafhankelijk van de inhoud (bijv. de rangschikking van de alternatieven, het aantal alternatieven).

De regels of richtlijnen die momenteel veel in gebruik zijn, hebben een geschiedenis van meer dan dertig jaar en zijn bijzonder verspreid geworden door Ebel (1). De meeste richtlijnen steunen op intuïtie, zijn veelal a priori opgesteld en kenden in de loop der jaren weinig wijzigingen. Bij de studie van deze richtlijnen hebben we een dubbele beperking ingevoerd. Vooreerst steunt onze samenstelling van richtlijnen op twintig handboeken, die als representatief kunnen beschouwd worden op het gebied van „educational measurement”. In tweede instantie hebben we ons voornamelijk beperkt tot de studie van de alternatieven, omdat hun constructie de meeste moeilijkheden oplevert voor de itemschrijvers en aldus ook gevoelige invloed heeft op de begripsvaliditeit van de items.

De meeste auteurs geven een opsomming van richtlijnen (bijv. Ebel, Wesman), meestal geïllustreerd met enkele voorbeelden van goede en mislukte items (2). Zonder daarom het belang van een dergelijke benadering te betwijfelen, wordt ze hier niet gevolgd. In dit synthetisch overzicht worden de regels geordend rond

knelpunten bij het item-schrijven en dit met de bedoeling een gestructureerd en bruikbaar overzicht van richtlijnen op te stellen. Tevens beperken we ons niet tot de a priori geformuleerde regels maar we betrekken in dit overzicht ook de empirische onderzoeken voor zover deze voorhanden zijn in de literatuur. Systematisch onderzoek is op dit domein immers fragmentarisch en weinig uitgesproken (3).

Een eerste groep richtlijnen betreft het verband tussen itemstam en alternatieven en vormt het onderwerp van paragraaf 1. Een tweede groep richtlijnen heeft uitsluitend betrekking op de alternatieven en wordt behandeld in paragraaf 2. Er is sprake van richtlijnen voor het schrijven van alternatieven, omdat deze richtlijnen zowel op de afleiders als op de juiste oplossing betrekking hebben.

Om de tekst niet te overladen met voetnoten en geciteerde auteurs, is op het einde van dit artikel een lijst opgenomen van de geraadpleegde literatuur bij het samenstellen van de richtlijnen.

§ 1. Overgang van stam naar alternatieven

1.1 Grammaticale structuur van het item

1.1.1 *Grammaticale overgang van stam naar alternatieven*

Bij het bedenken van afleiders kan het de itemschrijver overkomen, dat itemstam en afleiders niet grammaticaal juist op elkaar volgen. Op deze wijze induceert men een indicator tot de juiste oplossing wanneer juiste oplossing en itemstam wel grammaticaal bij elkaar aansluiten. Om de homogeniteit van de alternatieven te bevorderen zal men trachten afleiders te construeren met eenzelfde grammaticale functie als de juiste oplossing.

Het juist aansluiten van itemstam en alternatieven is het duidelijkst te controleren als de itemstam een onvolledige zin is. Alle alternatieven moeten in dit geval als vervollediging van de itemstam kunnen worden ingeschakeld.

De itemstam kan ook in de vragende vorm geformuleerd zijn zoals bijv. „Wat is Antarctica?” Als deze vraag in de bevestigende vorm wordt getransformeerd, dan doet zich vorige situatie opnieuw voor, nl. Antarctica is...

Dunn en Goldstein stelden in dit verband vast dat de moeilijkheidsgraad van meerkeuze-vragen daalt als er een verschil was tussen de grammaticale overgang van itemstam naar het juiste antwoord en de overgang van itemstam naar de afleiders (4). M.a.w. het is wenselijk dat itemstam en alternatieven grammaticaal juist op elkaar aansluiten.

1.1.2 *Itemstam als onvolledige zin versus itemstam in de vragende vorm*

Een andere variabele, m.n. de zinsvorm van de itemstam, werd eveneens in het onderzoek van Dunn en Goldstein opgenomen. Zij stelden vast dat het formuleren van de itemstam als onvolledige

zin of als een vraag, geen invloed heeft op de moeilijkheidsgraad van het item. Een dergelijke conclusie sluit wel in dat voorgaande regel betreffende de grammaticale overgang van stam naar alternatieven is gerealiseerd (5). Voor de beginnende itemschrijver is het beter om een directe vraag als itemstam te formuleren. Een directe vraag verplicht de itemschrijver tot meer duidelijkheid dan een onvolledige zin.

1.1.3 Negatief geformuleerde itemstam

Een item waarop meerdere juiste antwoorden mogelijk zijn, kan men negatief benaderen. De itemstam is dan een negatieve vraag. In dit geval moet als juiste oplossing het alternatief aangestreept worden dat niet overeenkomt met de opgave van de itemstam.

Voorbeeld :

Welke van de volgende vier woorden is *niet* juist gespeld ?

- a) bureau
- b) directeur
- c) onmiddelijk
- d) weiland

Sommige auteurs menen dat het aanwenden van een negatief geformuleerde stam alleen dan moet gebeuren als de leerdoelstelling dit vereist (bijv. wanneer men de leerlingen hun taalvaardigheid wenst te onderzoeken om negatief geformuleerde vragen te beantwoorden). Verder kan men argumenteren dat een negatieve vraag meestal moeilijker te begrijpen is voor de leerlingen van het basisonderwijs. De meeste problemen kunnen in positieve termen geformuleerd worden. Deze formulering is niet alleen verbaal gemakkelijker maar voorkomt tevens dat de leerlingen over de woorden „niet, geen, minstens” zouden overkijken.

1.2 Herhalen van woorden uit de itemstam in de alternatieven

Het letterlijk of als bekend synoniem herhalen van een sleutelwoord uit de itemstam in het juiste alternatief moet vermeden worden, omdat hierdoor de leerling een indicator krijgt voor het juiste antwoord. Volgens Dunn en Goldstein geeft het opnemen van een dergelijke indicator aanleiding tot een daling van de moeilijkheidsgraad van de items.

§ 2. Kenmerken van de alternatieven

2.1. Lengte van de alternatieven

In het algemeen beveelt men aan om de alternatieven zo kort mogelijk te houden. Maakt men de alternatieven *zo kort mogelijk*, dan krijgt de leerling beter overzicht en wordt de aandacht gericht op de essentiële verschillen tussen de alternatieven. Vooral het herhalen van woorden en zinnen in ieder van de alternatieven wordt afgeraden. Met korte alternatieven bedoelt men dat de alternatieven toch meestal uit meer dan één woord moeten bestaan. Een onderzoek

van Owens bevestigt de geldigheid van deze algemene richtlijn. Hij berekende de hertestingsbetrouwbaarheid van de items afkomstig uit een test over geschiedenis en stelde vast dat items met hoge hertestingsbetrouwbaarheid zich o.a. onderscheiden van minder betrouwbare items naar de bondigheid van de alternatieven. M.a.w. het bondig formuleren van de alternatieven levert een bijdrage tot de betrouwbaarheid van de items.

Wanneer men wil nagaan of een leerling een bepaalde definitie kent, dan kan men het best het te definiëren woord in de stam van het item opnemen. In dit geval wordt de algemene regel, m.n. het formuleren van korte alternatieven, niet toegepast. Plaatst men immers omgekeerd de omschrijving in de stam, de te definiëren term als juist antwoord en de andere termen als afleiders, dan vergroot men de kans dat een leerling het item oplost door verbale associatie.

Voorbeeld :

Welke van de volgende uitspraken vertegenwoordigt de beste definitie van de term „protoplasma“?

- a) Een complex colloïdaal systeem bestaande uit water, proteïnen en vetten.
- b) Een substantie die in staat is te groeien door een regelmatige opeenvolging van veranderingen tot een meer complexe eenheid.
- c) Een complex mengsel van proteïnen, vetten en koolhydraten, die op veranderingen in de omgeving reageren.
- d) Een complex colloïdaal systeem van proteïnen, vetten, koolhydraten, anorganische zouten en enzymen die leven vertonen.

De gegeven definities variëren in juistheid en volledigheid. Wie de omschrijving van het begrip heeft „van buiten“ geleerd (verbale associatie tussen de term protoplasma en de definitie ervan), wordt hier in verwarring gebracht wegens de gelijkennis tussen de verschillende omschrijvingen. Deze conditie is moeilijker te realiseren als de definitie in de itemstam wordt opgenomen.

Opdat de lengte van de alternatieven niet als indicator zou fungeren, zal men trachten om afleiders te construeren van eenzelfde lengte als de juiste oplossing. *Dunn en Goldstein* toonden aan, dat de moeilijkheidsgraad van items daalde, wanneer het juiste alternatief korter of langer was dan de vorige alternatieven.

2.2 Aantal alternatieven

Volgens *Travers* is het optimaal aantal alternatieven in een item afhankelijk van de aard van het gestelde probleem en van het ontwikkelingsniveau van de leerlingen, voor wie het item bestemd is. Andere auteurs streven ernaar om items te construeren met vier of vijf alternatieven. Zij aanvaardden echter minder alternatieven, indien het onmogelijk is om drie of vier afleiders te construeren zonder absurditeiten te introduceren. De kwaliteit van het item wordt immers meer gewaarborgd door de kwaliteit van de afleiders dan door het aantal afleiders. In verband met de rol van het ontwikkelingsniveau bij de keuze van het aantal alternatieven, meent *Wesman* dat men voor 6- à 7-jarige kinderen het aantal alternatieven moet beperken tot drie.

De voorkeur voor items met vier of vijf alternatieven steunt niet op a priori overwegingen. Volgende twee argumenten worden hierbij veelal aangehaald.

2.2.1 Betrouwbaarheid

Theoretisch kan worden aangetoond, dat een test aan betrouwbaarheid wint, wanneer de items uit meerdere alternatieven zijn samengesteld. Onderstaande tabel illustreert duidelijk, dat de betrouwbaarheid van een test verhoogt, als de items meerdere alternatieven *bevatten.*

TABEL 1.

Verwachte betrouwbaarheid van een test bestaande uit 100 items als functie van het aantal alternatieven per item (9)

Aantal alternatieven per item	Verwachte betrouwbaarheid
2	.74
3	.83
4	.86
5	.87
6	.88

Onder leiding van Remmers werden verschillende onderzoeken uitgevoerd naar de relatie tussen de betrouwbaarheid van een test en het aantal alternatieven per item. Volgend onderzoek is illustratief voor zijn werk. Hijzelf en House berekenden de betrouwbaarheid van vier vormen van een rekentest (10). De basistest bestond uit items met vijf alternatieven. Telkens werd op verantwoorde wijze één afleider geëlimineerd, zodat uiteindelijk vier tests bekomen werden met respectievelijk vijf, vier, drie en twee alternatieven. Er werden vier equivalenten groepen proefpersonen geselecteerd en elke groep kreeg één testvorm aangeboden. Op alle testuitslagen werd de correctieformule voor raden toegepast. Uit een vergelijking van de halveringsbetrouwbaarheidscoëfficiënten van de verschillende tests mochten de auteurs besluiten, dat de betrouwbaarheid stijgt als het aantal alternatieven toeneemt.

Remmers houdt in zijn onderzoeken geen rekening met de tijdsduur waarbinnen b.v. een test, bestaande uit items met vijf alternatieven, wordt opgelost t.o.v. een test, bestaande uit items met twee alternatieven. In vele gevallen zal het oplossen van tweekeuze-items, in vergelijking met vijfkeuze-items, minder tijd in beslag nemen. Wanneer de tijdslimiet constant wordt gehouden, kan een groter aantal tweekeuze-items worden aangeboden dan vijfkeuze-items. Aangezien het verlengen van een test gepaard gaat met een verhoging van de betrouwbaarheid van de test, kan men besluiten dat uitsluitend omwille van de betrouwbaarheid van een test het niet noodzakelijk is vijfkeuze-items te construeren. De overwegingen van Travers, betreffende de aard van het gestelde probleem en het ontwikkelingsniveau van de leerlingen, zullen dus meer doorslaggevend zijn bij de keuze van het aantal alternatieven.

2.2.2. Raden

Als tweede argument ten voordele van items met meerdere alternatieven, wordt aangehaald dat bij dergelijke items de mogelijkheid om lukraak het juiste antwoord aan te strepen, eerder klein is. Wanneer men op toeval een antwoord aanduidt tussen de alternatieven van een meerkeuze-vraag, dan zal men voor enkele items de juiste oplossing selecteren. Gemiddeld genomen, zal dit met een frequentie van k/n gebeuren; k is het aantal items door raden opgelost en n is het aantal alternatieven per item. Door een groter aantal alternatieven per item te kiezen kan men de winst, verbonden aan het raden, verminderen, aangezien de verhouding k/n kleiner wordt. Het heeft evenwel weinig zin afleiders op te nemen die geen enkele aantrekking uitoefenen. In dit geval kan men beter items opnemen met 1 à 2 afleiders (11).

Het lukraak aanstrepen van een antwoord wordt aangeduid met de term raden in enge zin. De term „raden” wordt ook als algemeen begrip gebruikt om een domein van gedragingen aan te duiden, die optreden wanneer een proefpersoon antwoordt op een vraag in keuze-antwoordvorm aangeboden, waarvoor hij het antwoord niet kent. De problematiek van het raden is verbonden aan de keuze-antwoordvorm en heeft aanleiding gegeven tot het toepassen van correcties voor raden. Volgens E b e l heeft men de bedoeling om met dergelijke correcties een vermindering na te streven van de hogere uitslag, die proefpersonen zouden bekomen door lukraak naar de juiste oplossing te raden (12).

Volgende correctieformule is veel verspreid in de literatuur (13).

$$S = R - \frac{W}{n-1}$$

S: gecorrigeerde score

R: aantal juist opgeloste items

W: aantal foutief opgeloste items

n: aantal alternatieven per item

De problematiek, betreffende de correctie voor raden, heeft in de literatuur over „educational measurement” veel aandacht gekregen. Volgens M a y o is het gebruik van de correctieformule reeds afgenomen in vergelijking met de zestiger jaren, maar het onderzoek erover blijft doorgaan zonder innovaties (14). In functie van het onderwerp, dat we hier bespreken, m.n. het item-schrijven, kan een korte samenvatting volstaan van de literatuurstudie, die we over de correctie voor raden hebben ondernomen. Wanneer een correctie voor raden wordt toegepast, is het wenselijk om vooraf per item de alternatieven dusdanig te kiezen, dat ze een gelijke plausibiliteit bezitten (15). Zowel de motivatie van de leerlingen als de instructies kunnen het lukraak raden al dan niet bevorderen. B.v. leerlingen die weinig gemotiveerd zijn om een proef goed op te lossen, zullen gemakkelijker hun toevlucht zoeken in het lukraak raden, indien ze een opgave niet kunnen oplossen. Het lukraak raden neemt eveneens

toe als de instructies nadruk leggen op snelheid en accuraatheid (16). Daarom stelt Cureton voor om de correctieformule alleen dan toe te passen wanneer het tests betreft die onder tijdsdruk worden afgenomen (17). Dergelijke tests zouden moeten aanvangen met gemakkelijke items en geleidelijk overgaan naar moeilijke items. Voor tests, die niet onder tijdsdruk worden afgenomen, stellen Davis en Hendrickson voor om gewichten toe te kennen aan de alternatieven. Naar gelang een alternatief meer gelijkenis vertoont met de juiste oplossing, wordt een hoger gewicht toegekend. Op deze wijze wordt de einduitslag op een test afhankelijk gesteld van de kwaliteit van de afzonderlijke antwoorden en niet van het aantal juiste en verkeerde antwoorden (18).

We kunnen besluiten dat de bevindingen van Cureton en Davis een genuanceerd en tevens wetenschappelijk verantwoord gebruik van de correctieformule toelaten.

2.3 Rangschikking van de alternatieven

In de literatuur over het item-schrijven wordt aan de rangschikking van de alternatieven aandacht geschonken omwille van volgende argumenten.

Ten eerste vertonen beginnende item-schrijvers spontaan een tendens om het juiste alternatief meer frequent in het midden van de lijst alternatieven te plaatsen, eerder dan aan het begin of op het einde. Dergelijke handelwijze komt overeen met wat we een indicator hebben genoemd. Immers, indien leerlingen vaststellen dat de juiste oplossing b.v. meestal overeenkomt met het derde alternatief, dan zullen ze, voor items die ze niet kunnen oplossen, een goede kans maken om de juiste oplossing te selecteren door het derde alternatief aan te strepen.

Ten tweede stelt men vast, dat bepaalde leerlingen, wanneer ze de oplossing van een item niet kennen, de voorkeur geven aan een alternatief dat een bepaalde plaats inneemt in de rij alternatieven (19). Dit fenomeen, aangeduid als „positional response set”, staat onafhankelijk van de itemschrijver en moet onderscheiden worden van vorige werkwijze, waar de leerlingen weloverwogen inspelen op een indicator die hen werd bezorgd door de itemschrijver. De „positional response set” wordt omschreven als een antwoordtendens of een neiging die maakt dat een persoon standvastig andere antwoorden geeft op bepaalde items dan hij zou geven als deze items onder een andere vorm werden aangeboden. De meeste onderzoeken wijzen uit dat antwoordtendensen de validiteit van een test verlagen. Andere antwoordtendensen dan de „positional response set” zijn b.v. de instemmingstendens, de ontkennende antwoordtendens (20). Het al of niet optreden van een antwoordtendens bij een subject vindt zijn verklaring in motivationele factoren en attituden. Een andere persoonlijkheidsvariabele, die bepalend is voor de antwoorden van een subject, wordt aangeduid als cognitieve stijl. Dit is een voor het individu kenmerkende relatief constante wijze van functioneren met

betrekking tot de intellectuele activiteiten en de waarneming (b.v. de „field dependence” of veldafhankelijkheid). In dit verband meent Thorndike dat de invloed van het onderzoek over de „field dependence” op de constructie van pedagogische tests („educational tests”) in de toekomst nog zal toenemen (21).

Om de „positional response set” of het lokalisatie-effect op te vangen, kan men de juiste oplossing een gelijk aantal keren elk van de mogelijke posities laten innemen. Verschillende methodes zijn bruikbaar om de plaats van de juiste oplossing tussen de afleiders te bepalen.

Wanneer elk alternatief uit slechts één woord bestaat, dan kan men de alternatieven rangschikken naar hun beginletter (alfabetisch). Doet men dit consequent, dan kunnen de leerlingen inzien welk systeem is gevolgd en zullen ze in de volgorde van de alternatieven geen sleutel voor het juiste antwoord zoeken. In dit verband is het wenselijk om in de instructies aan de leerlingen een toelichting te geven betreffende de rangschikking van de alternatieven (22). Als de alternatieven uit getallen bestaan, kan men ze ordenen van klein naar groot. Ook wanneer de alternatieven op een kwalitatieve schaal geplaatst kunnen worden, kan men ze plaatsen in volgorde van grootte.

Zijn de alternatieven niet volgens grootte of beginletter te rangschikken, dan bevelen de meeste auteurs aan om een aselechte keuze te maken uit de mogelijke volgorden. Dit kan verwezenlijkt worden door middel van een tabel van aselekt gekozen getallen, zoals b.v. onderstaande tabel van Mosier en Price.

TABEL 2.

Overzicht van de mogelijke permutaties met vijf cijfers

35124	13425	12453	24531	54231	42153
34215	43512	54132	32541	32154	25143
12354	43251	15324	52314	21453	24135
25134	51243	25314	42315	54321	52143
14523	13254	15423	51342	12543	34152
24513	43152	54312	42531	23451	52134
52431	31245	13524	51234	21534	15342
31452	42513	12534	14532	35412	23514
54213	25341	43521	41325	53421	34251
53412	34521	21435	32451	51432	32145
43125	13542	35241	41352	52341	51324
23541	13245	45132	31542	54123	43215
45231	41235	24351	31425	23415	41523
21354	14235	23145	52413	35214	24153
15243	53124	21543	31524	25431	41532
42135	21345	25413	12435	35142	34125
31254	45213	32415	12345	51423	23154
14253	45321	53214	35421	14352	24315
53142	13452	42351	45123	32451	45312
15432	41253	14325	53241	15234	34512

Deze tabel bevat de 120 mogelijke permutaties met vijf cijfers. De permutaties zijn op toevallige wijze gerangschikt (23).

Elke permutatie stemt overeen met een item, bestaande uit vijf alternatieven, i.c. één juiste oplossing en vier afleiders. Om het nummer van de juiste oplossing te bepalen, wordt uit vijf kaartjes, genummerd van 1 tot 5, op toevallige wijze één kaartje getrokken. Vervolgens wordt dit nummer gebruikt om binnen elke permutatie (of item) de plaats van dit nummer (of juiste oplossing) vast te stellen. Als men b.v. het nummer 2 trekt, dan komt de juiste oplossing in het eerste item (zie de eerste permutatie in de tabel) als vierde alternatief aan bod. De andere nummers uit de permutatie worden vervangen door de afleiders. In het tweede item komt de juiste oplossing in dit geval op de derde plaats te staan.

2.4 Eén alternatief is het juiste of beste antwoord

De meest verspreide meerkeuze-vragen bestaan uit een itemstam en een aantal alternatieven, waarvan één het juiste of beste antwoord is. In het geval van een meerkeuze-vraag, waarop het *juiste* antwoord gegeven wordt, is één alternatief absoluut juist en zijn de andere alternatieven absoluut fout. Aangezien deze eis niet voor alle vragen realiseerbaar is, acht Ebel volledige juistheid van het goede alternatief dan ook niet steeds noodzakelijk. Men kan een leerling ook opdracht geven om het beste antwoord van de gegeven alternatieven aan te duiden. Het verschil tussen items met het juiste antwoord en met het beste antwoord is niet zozeer een kwestie van vorm, maar hangt vooral af van de soort leerstof en de mentale operaties die men met het item wenst te onderzoeken.

2.5 Meerdere alternatieven zijn juiste antwoorden

Voor sommige opgaven of opdrachten, b.v. deze die het creatief denken van de leerling willen onderzoeken, kan men de voorkeur geven aan items met meer dan één juist antwoord. Men voorziet dan meerdere alternatieven die een juist antwoord zijn op de itemstam. Bij dergelijke items zal men de leerlingen duidelijke instructies geven dat zij *alle juiste* antwoorden moeten aanduiden. Gronlund en Remmers menen, dat meerkeuze-vragen met meerdere juiste antwoorden zo weinig mogelijk gebruikt moeten worden. Volgens hen kunnen deze vragen in de meeste gevallen beter worden aangeboden onder vorm van waar/vals-vragen. Halleux-Hendrick deelt deze mening niet. Zij construeerde 75 meerkeuze-vragen over wiskunde, die werden aangeboden aan 72 leerlingen van het eerste jaar A6A2. Ongeveer de helft van de aangeboden meerkeuze-vragen bevatte één goede oplossing. De andere helft items werd geconstrueerd volgens het type meerkeuze-vraag met meerdere juiste oplossingen. Ter illustratie geven we volgend voorbeeld van dit itemtype (24).

Voorbeeld :

Meerkeuze-vraag met meer dan één juiste oplossing

Plaats een x bij de juiste oplossing(en).

Maak volgend produkt.

$$\begin{aligned} E &= (x, y, z) \\ E &= (t) \\ EF &= (xt, yt, zt) \\ &= (x, t, y, z) \\ &= (tx, ty, tz) \\ &= (xt, zt, yt) \\ &= (yt, xt, zt) \end{aligned}$$

De leerlingen waren voor beide groepen items totaal onwetend over het aantal correcte alternatieven (25). Uit een vergelijking van beide groepen items bleek duidelijk dat alle leerlingen, zonder uitzondering, een groter procent meerkeuze-vragen met één juiste oplossing correct beantwoordden in vergelijking met de meerkeuze-vragen met meerdere juiste oplossingen (26).

Volgens Halleux-Hendrick heeft de meerkeuze-vraagvorm met meerdere juiste oplossingen een bijzondere pedagogische betekenis, nl. de leerlingen krijgen de kans om maximaal hun kennis en vaardigheden te tonen (27). Het oplossen van dergelijke vragen neemt echter meer tijd in beslag, aangezien alle alternatieven in overweging moeten worden genomen. Het gebruik van deze vraagvorm moet niet veralgemeend worden en de gepaste aanwending moet gefundeerd zijn op de aard van de leerdoelen die men beoogt te onderzoeken (b.v. creatief denken).

2.6 „Geen van deze” als alternatief

In sommige proeven of tests gebruikt men bij meerdere of bij alle items als laatste alternatief: „het juiste antwoord is niet gegeven” of nog „geen van deze”. Deze alternatieven kunnen volgens Ebel en Nunally alleen dan als alternatief in een item worden toegepast, wanneer gevraagd wordt naar het juiste antwoord en niet naar het beste antwoord. In dit laatste geval doet zich de mogelijkheid voor dat b.v. van een aantal onjuiste alternatieven één alternatief nog wel als beste antwoord kan worden beschouwd, nl. het minst onjuiste.

Het gebruik van „geen van deze” kan als volgt worden gemotiveerd. Bij meerkeuze-vragen zijn vele leerlingen geneigd steeds één van de alternatieven aan te strepen, ook al kennen ze het juiste antwoord niet. Zij rekenen er op, dat het juiste antwoord noodzakelijk tussen de alternatieven staat. Een leerling, die de overeenkomstige doelstelling van een item niet heeft bereikt, kan evenwel een foutief antwoord hebben gevonden, dat niet tussen de keuzen staat. Deze leerling moet dan zijn oordeel of zijn berekening herzien en wordt dus bij het beantwoorden van het item onrechtstreeks geholpen door de alternatieven. Ofwel kan hij lukraak één van de alternatieven aanstrepen, hetgeen nadelig is voor de validiteit van de proef. Omdat de leerlingen het alternatief „geen van deze” als een mogelijk juist antwoord zouden beschouwen, kunnen volgens Ebel en Davis bij het begin van een proef of test een paar gemakkelijke items worden aangeboden, waarin „geen van deze” het juiste alternatief is.

Wat betreft de toepassing van „geen van deze” als alternatief, zijn items over rekenen en spelling het meest aangewezen. Hierbij is een antwoord helemaal juist of zonder fout. ^{meer}

Empirisch onderzoek over het gebruik van het alternatief „geen van deze” laat geen eenduidige uitspraak toe. *Wesman* en *Bennett* gebruiken „geen van deze” in een woordenschat- en in een wiskundetest. Elke test werd aangeboden in twee vormen; de ene vorm met vijf specifieke alternatieven en de andere vorm met vier specifieke alternatieven en „geen van deze” als vijfde alternatief. Er werd tussen beide vormen geen verschil vastgesteld wat betreft gemiddelde uitslag, moeilijkheidsgraad, validiteit en betrouwbaarheid (28). In een studie van *Boyn-ton* met items over spelling werd een verschil gevonden tussen items met „geen van deze” als alternatief en items zonder dit alternatief. Zij stelde vast, dat items met „geen van deze” moeilijker waren (29).

Als onderdeel van een onderzoeksproject, opgezet door *Rimland*, werd een studie gemaakt naar de invloed van items met het alternatief „juist antwoord niet gegeven” op de moeilijkheidsgraad en de validiteit van een test (30). Voor dit onderzoek werd een subtest uit de *U.S. NAVY Arithmetic Test (ARI)* gebruikt, m.n. de „reasoning subtest”. Elk item uit de oorspronkelijke subtest bestond uit vijf numerische keuze-antwoorden. In de experimentele vorm werd voor elk van de dertig items het minst frequent gekozen alternatief vervangen door „juist antwoord niet gegeven”. Elke vorm werd aangeboden aan 2.000 rekruten. De beide groepen waren vergelijkbaar wat betreft intelligentie en rekenvaardigheid. *Rimland* stelde vast dat de validiteit van de „reasoning subtest” niet veranderde door opname in de items van het alternatief „juist antwoord niet gegeven”. Het gebruik van dit alternatief gaf wel aanleiding tot een kleine, maar statistisch significante verhoging van de moeilijkheidsgraad van de subtest. Ondanks de bereikte statistische significantie, meent *Rimland*, dat de vastgestelde verschillen van geringe betekenis zijn. In zijn conclusies sluit hij aan bij het onderzoek van *Wesman* en *Bennett*, die vonden dat het gebruik van „geen van deze” weinig invloed heeft op de itemkarakteristieken (31).

Samenvatting en conclusies

In de literatuur over het item-schrijven worden een aantal regels geformuleerd betreffende de formele kenmerken van meerkeuzevragen. Deze kenmerken zijn inherent aan het gebruik van de meerkeuze-vraagvorm en relatief onafhankelijk van de inhoud. Het realiseren van bepaalde formele kenmerken in de items draagt bij tot de homogeniteit van de alternatieven en deze eigenschap levert een bijdrage tot de begripsvaliditeit van het item.

In dit artikel werd een gestructureerd overzicht samengesteld van de belangrijkste regels en empirische onderzoeken over de formele kenmerken van de alternatieven. Een dergelijk overzicht heeft een dubbele bedoeling. Enerzijds moet het als uitgangspunt fungeren voor een handleiding over de redactie van items. Ander-

zijds kan het gebruikt worden bij de samenstelling van een vragenlijst om de kwaliteit van items te beoordelen. Statistische analyse van items bezorgt alleen informatie over het goed of slecht zijn van de items. Om items te verbeteren heeft de itemconstructeur echter behoefte aan informatie over de redenen waarom een item niet voldoet. Het belang van een vragenlijst naar de kwaliteit van de items moet gezien worden in deze optiek van het verbeteren van items (32).

De meeste regels betreffende de formele kenmerken van meerkeuze-vragen zijn vrij algemeen geformuleerd en weinig onderworpen aan empirisch onderzoek. Vandaar dat de meeste regels worden gekenmerkt door een voorlopige geldigheid. Om een gefundeerd oordeel te kunnen vormen over de geldigheid van de algemene regels, zou het empirisch onderzoek moeten worden uitgebreid en gediversifieerd naar: verschillende onderwijsniveaus; verschillende types van cognitieve operaties.

De onderzoekingen, die werden besproken, worden gekenmerkt door volgend opzet. Men creëert een onderzoeksconditie, waarin een bepaalde regel wordt toegepast op de items uit een test en een andere conditie met afwezigheid van de regel in kwestie. Vervolgens worden beide condities onderling vergeleken naar hun predictieve en gelijktijdige validiteitscoëfficiënten. In het perspectief van proeven voor didactische evaluatie behoeven de regels voor itemconstructie naar hun bijdrage tot de begripsvaliditeit te worden beoordeeld. Dit heeft voor gevolg dat de meeste onderzoekingen, die momenteel beschikbaar zijn in de literatuur, slechts een ondergeschikt belang hebben in de vernieuwde pedagogische testbeweging, omdat de begripsvaliditeit er te weinig aan bod komt of verwaarloosd is. Deze theoretische stellingname moet m.a.w. nog sterker toepassing vinden in het empirisch onderzoek. (Ontvangen, mei 1976)

Voetnoten

- (1) R.L. EBEL, Writing the test item. -In: E.F. LINDQUIST (Ed.), *Educational measurement*. Washington, ACE, 1951, p. 185-249.
- (2) *Ibid.*, p. 151; A.G. WESMAN, Writing the test item. -In: R.L. THORNDIKE (Ed.), *Educational measurement*. Washington, ACE, 1971 (2nd ed.), p. 113.
- (3) W.E. COFFMAN, Achievement tests. -In: R.L. EBEL (Ed.), *Encyclopedia of educational research*. New York, MacMillan, 1969 (4th ed.), p. 12-14.
- (4) T.F. DUNN & L.G. GOLDSTEIN, Test difficulty, validity and reliability as functions of selected multiple-choice item construction principles. *Educ. psychol. Measmt*, 1959 (19), 171-179.
- (5) *Ibid.*, p. 175.
- (6) *Ibid.*, p. 177.
- (7) W.A. OWENS, Retest consistency and the writing of life history items. *J. appl. Psychol.*, 1962 (46), 329-331.
- (8) T.F. DUNN & L.G. GOLDSTEIN, *O.c.*, p. 178-179.
- (9) R.L. EBEL, Expected reliability as a function of choices per item. *Educ. psychol. Measmt*, 1969 (29), 565-570.
- (10) H.H. REMMERS & J.M. HOUSE, Reliability of multiple-choice measuring instruments as a function of the Spearman-Brown formula IV. *J. educ. Psychol.*, 1941 (32), 372-376.
- (11) THORNDIKE, The problem of guessing. -In: R.L. THORNDIKE (Ed.), *O.c.*, p. 59-61.
- (12) In het volgend werk van Ebel is een bondige studie over de correctie voor raden te vinden, die tevens praktische richtlijnen bevat. Zie: R. L. EBEL

- Essentials of educational measurement*. Englewood Cliffs, Prentice-Hall, 1972, XII-622 pp.
- Zie ook: M. DEGRAEVE, Een correctie voor raden bij meerkeuze-vragen? *Tijdschr. Opvoedk.*, 1974-75 (20), 184-198.
- (13) R.L. THORNDIKE, *O.c.*, p. 59.
- (14) S.T. MAYO, The methodology and technology of educational and psychological testing. *Rev. educ. Res.*, 1968 (38), p. 92-101.
- (15) R.L. EBEL, *O.c.*, p. 223.
- (16) F. SWINEFORD & P.M. MILLER, Effects of directions regarding guessing on item statistics of a multiple-choice vocabulary test. *J. educ. Psychol.*, 1953 (44), 129-139.
- (17) E. CURETON, The correction for guessing. *J. exp. Educ.*, 1966 (34), 44-47.
- (18) F.B. DAVIS, A note on the correction for chance success. *J. exp. Educ.*, 1967 (35), 42-47.
- G. F. HENDRICKSON, Effect of differential option weighting on multiple-choice objective tests. *J. educ. Measmt*, 1971 (8), 291-296.
- (19) L.J. CRONBACH, *Essentials of psychological testing*. New York, Harper and Row, 1970 (3rd ed.), p. 148;
- L.J. CRONBACH, Response sets and test validity. *Educ. psychol. Measmt*, 1946 (6), 475-496.
- (20) Voor een overzicht over response sets, zie: J.C. NUNNALLY, *Psychometric theory*. New York, McGraw-Hill, 1967, p. 593-622.
- (21) R.L. THORNDIKE, Educational measurement for the seventies. -In: R.L. THORNDIKE (Ed.), *O.c.*, p. 9.
- (22) R.F. VAN NAERSSSEN, Meer gecompliceerde scoringswijzen. -In: A.D. DE GROOT, R.F. VAN NAERSSSEN & al., *Studietoetsen: construeren, afnemen, analyseren*. Den Haag, Mouton, 1969, p. 270.
- (23) C.I. MOSIER & H.G. PRICE, The arrangement of choices in multiple-choice questions and a scheme for randomizing choices. *Educ. psychol. Measmt*, 1945 (5), 379-382.
- Een voorstel, analoog aan dit van Mosier, werd geformuleerd door Anderson. Hij stelt voor om de rangschikking van de permutaties te baseren op „Tippit's random sampling numbers tabel”, daar waar Mosier zich baseert op logaritmefafels. Zie: S.B. ANDERSON, Sequence in multiple-choice item options. *J. educ. Psychol.*, 1952 (43), 364-368.
- (24) Jacqueline HALLEUX-HENDRICK, Construction de questions à choix multiple: une seule solution correcte? *Rev. belge Psychol. Pédag.*, 1969 (31), p. 122.
- (25) *Ibid.*, p. 117.
- (26) *Ibid.*, p. 121.
- (27) *Ibid.*, p. 122.
- (28) A.G. WESMAN & G.K. BENNETT, The use of „none of these” as an option in test construction. *J. educ. Psychol.*, 1946 (37), 541-554.
- (29) M. BOYNTON, Inclusion of „none of these” makes spelling items more difficult. *Educ. psychol. Measmt*, 1950 (10), 431-432.
- (30) B. RIMLAND, The effect of varying time limits and of using „right answer not given” in experimental forms of the U.S. Navy Arithmetic test. *Educ. psychol. Measmt*, 1960 (20), 533-539.
- (31) *Ibid.*, p. 538.
- (32) G.J. MELLEBERG, *Studies in studietoetsen*. (Ongepubliceerd doctoraatsproefschrift). Amsterdam, 1971, p. 98-99.

Literatuur

Bij het samenstellen van de regels betreffende de formele kenmerken van de alternatieven bestudeerden we de werken van volgende auteurs. Voor een aantal auteurs steunden we op de gegevens die Lans hierover samenbracht in haar doctoraatscriptie: Wilhelmina LANS, *Over itemconstructie*. (Ongepubliceerd Memorandum, nr AET-216). Amsterdam, 1967, 61 pp. Deze werken zijn met volgend teken aangeduid: *. De andere referenties verwijzen naar publikaties die we zelf hebben geraadpleegd.

ADAMS, G.S., *Measurement and evaluation in education, psychology and guidance*. New York, Rinehart and Winston, 1964, p. 428-437.

- ADKINS, D.C., *Construction and analysis of achievement tests*. Washington, Government Printing Office, 1947, p. 41-65.
- * ANSTEY, E., *Psychological tests*. London, 1966.
- BEAN, K.L., *Construction of educational and personnel tests*. New York, McGraw-Hill, 1953, p. 52-75.
- C.P.D.O., *Leidraad voor beoordeling en redactie van items*. (Project: Evaluatie van het pedagogisch-didactisch rendement van het onderwijs). Leuven, 1970, 20 pp.
- DAVIS, F.B., *Educational measurements and their interpretation*. Belmont, Wadsworth Publishing Company, 1964, IX-422 p.
- EBEL, R.L., Writing the test item. -In: E.F. LINDQUIST (Ed.), *Educational and psychology*, New York, Macmillan, 1968, XV-463 pp.
- EBEL, R.L., *Measuring educational achievement*. Englewood Cliffs, Prentice-Hall, 1965, p. 151-170.
- FURST, J., *Constructing evaluation instruments*. New York, David McKay, 1958, p. 250-275.
- * GERBERICH, N.E., GREENE, H.A. & JORGENSEN, A.N., *Measurement and evaluation in the modern school*. New York, 1964.
- GRONLUND, N.F., *Measurement and evaluation in teaching*. New York, Macmillan, 1965, p. 140-159.
- GRONLUND, N.E. (Ed.), *Readings in measurement and evaluation. Education and psychology*. New York, Macmillan, 1968, XV-463 pp.
- LANS, Wilhelmina & MELLENBERG, G.J., Constructie en beoordeling van items: formele aspecten. -In: A.D. DE GROOT, R.F. VAN NAERSEN & al., *Studietoetsen: construeren, afnemen en analyseren*. Den Haag, Mouton, 1969, p. 84-86.
- MARSHALL, J.C. & HALES, L.W., *Classroom test construction*. Massachusetts, Addison-Wesley, 1971, p. 92-112.
- * NUNNALLY, J.C., *Tests and measurements: assessment and prediction*. New York, MacGraw-Hill, 1959.
- NUNNALLY, J.C., *Educational measurement and evaluation*. New York, MacGraw-Hill, 1964, p. 121-128.
- REMMERS, H., RUMMEL, J. & GAGE, N., *A practical introduction to measurement and evaluation*. New York, Harper and Row, 1966 (2nd ed.), p. 209-253.
- TRAVERS, R., *Educational measurement*. New York, Macmillan, 1955, p. 184-187.
- * TRAVERS, R., *How to make achievement tests*. New York, 1950.
- WESMAN, A.G., Writing the test item. -In: R.L. THORNDIKE (Ed.) *Educational measurement*. Washington, ACE, 1971, p. 184-187.