

ENKELE INDRUKKEN VAN HET CONGRES
„POSSIBILITIES AND LIMITATIONS OF
EDUCATIONAL TESTING”

(mei 1967)

G. J. MELLEBERGH en R. F. VAN NAERSSSEN

1 Inleiding

Van 16-24 mei werd in Berlijn het Congres „Possibilities and Limitations of Educational Testing” gehouden. De beide auteurs van dit artikel werden door de Universiteit van Amsterdam in staat gesteld aan dit Congres deel te nemen. Zij leverden de volgende bijdragen:

R. F. van Naerssen: A signal/noise ratio index used for item selection in teacher-made tests.

G. J. Mellenbergh en A. D. de Groot: Some experiences with an educational testing program at the end of the sixth grade.

Er waren ruim 110 deelnemers aan het Congres, waaronder vele met beroemde namen op het gebied van „Educational Testing”, zoals Ebel, Lindquist, Messick, Henrysson en Chauncey. De thema's die tijdens het Congres ter sprake gebracht werden, zijn: toepassing van schooltests in verschillende landen – paedagogische doeleinden van testgebruik – itemtypen voor de meting van schoolvorderingen – nieuwe itemtypen voor het meten van vaardigheden – tests voor belangstelling en sociaal gedrag – itemconstructie, analyse en geautomatiseerde informatieverwerking – tests voor lager, middelbaar en hoger onderwijs – „teacher made” tests – invloed van testtype op het leren – presentatie van testresultaten aan docenten en interpretatie van deze resultaten – het gebruik van tests voor evaluatie en verbetering van de onderwijsmethoden – validiteit en homoskedasticiteit – belang van tests voor onderzoek van het leerplan en de onderwijsdoelstellingen. In dit artikel wordt geen systematisch verslag gegeven van al deze onderwerpen; er is echter getracht een indruk te geven van de belangrijkste punten die ter sprake gekomen zijn.

2 Testen in verschillende delen van de wereld

Studietoetsen worden op grote schaal gebruikt in Amerika, Zweden, Japan en Australië. Iowa State University test bijvoorbeeld 10 miljoen leerlingen per jaar met studietoetsen; een Instituut in Japan 4 miljoen kinderen per jaar. Ook in Engeland wordt op dit gebied vrij

veel werk gedaan. Spectaculair zijn de vorderingen in sommige ontwikkelingslanden (India, Maleisië, Chili, Nigeria, Mali en Kenia). In de meeste landen van Europa wordt nog bijzonder weinig op dit gebied gedaan. In dit opzicht onderscheidt Nederland zich niet van de overige Europese landen, hoewel er de laatste jaren in sommige onderwijskringen meer belangstelling komt voor studietoetsen.

3 *Vergelijking tussen „essay” tests en studietoetsen*

Gardner vergeleek studietoetsen met de conventionele proefwerken en tentamina (zogenaamde „essay” tests). Bij studietoetsen moet de leerling het goede antwoord kiezen uit een aantal alternatieven, terwijl de leerling bij de „essay” test zelf zijn antwoord moet formuleren. *Gardner* maakte een vergelijking tussen deze beide vormen van testen:

- a. Studietoetsen kunnen objectief en snel nagekeken worden; de „essay” tests moeten door de docenten beoordeeld worden, wat een zeer tijdrovende bezigheid is.
- b. Studietoetsen zijn betrouwbaarder. Uit vele onderzoeken is gebleken dat de betrouwbaarheid van studietoetsen hoger is dan van „essay” tests.
- c. Studietoetsen bestaan uit meer items dan „essay” tests.
Een studietoets (meestal 40 à 100 vragen) kan zo samengesteld worden dat de toets representatief is voor de gehele stof. Een „essay” test (meestal 3 à 8 vragen) is in het algemeen niet representatief voor de gehele stof vanwege het geringe aantal vragen.
- d. Bij „essay” tests zijn leerlingen die „vlot” kunnen schrijven in het voordeel.
- e. De constructie van studietoetsen vereist veel meer tijd dan de samenstelling van „essay” tests.
- f. Bij studietoetsen kunnen de leerlingen het goede antwoord raden; er zijn echter correcties voor raden mogelijk.
- g. Bij „essay” tests moeten de leerlingen de stof kunnen integreren en organiseren.
- h. Studietoetsen en „essay” tests zouden wellicht een verschillende invloed kunnen hebben op de wijze van bestuderen van de stof door de leerlingen.

Wat betreft punt g. kunnen we opmerken dat bij studietoetsen de leerlingen weliswaar niet zelf een verhaal moeten samenstellen, maar het is vaak mogelijk om vragen voor studietoetsen te construeren, die inzicht en begrip meten. In vele gevallen is gebleken dat goed gecon-

strueerde studietoetsen dezelfde faktor kunnen meten als „essay” tests waarin de leerlingen zelf hun antwoord moeten formuleren. Er zijn echter onderwerpen die absoluut niet gemeten kunnen worden met studietoetsen, bijvoorbeeld: onderzoeken of een student zelfstandig een empirisch onderzoek kan uitvoeren, kan men alleen door de student zelf een onderzoek te laten doen. Zo zal dus ook volgens *Gardner* de „essay” test een rol blijven vervullen naast de (objectieve) studietoets. Van de mogelijkheden van de laatste is men echter in het onderwijs als geheel (ook in Amerika) nog onvoldoende op de hoogte.

Op punt h. werd ingegaan door *French*. Zijn paper vermeldde een onderzoek over de invloed van het testtype op de wijze van voorbereiding van de leerstof. Eén groep studenten werd verteld dat zij getest zouden worden met een „essay” test en een andere groep studenten dat zij een studietoets zouden krijgen. Nadat de studenten de stof bestudeerd hadden, kregen beide groepen beide soorten tests. Er bleken geen significante verschillen te bestaan tussen de twee groepen op de twee tests. Verder werd de studenten gevraagd hoe zij zich voorbereid hadden op de tests. Slechts 14% van de studenten zei een speciale wijze van voorbereiding toegepast te hebben afhankelijk van het type test dat zij verwachtten te krijgen. De conclusie luidt dat het soort es dat gebruikt wordt, praktisch geen invloed heeft op de wijze van voorbereiding door de leerlingen.

4 *Mogelijkheden van studietoets*

Studietoetsen zijn vooral bedoeld om de studieprestaties van leerlingen te beoordelen. Het is echter gebleken dat studietoetsen veel meer mogelijkheden bieden. Tijdens dit Congres is duidelijk geworden dat de nadruk de laatste tijd meer op deze andere mogelijkheden komt te liggen dan op de beoordeling van de studieprestaties.

In de eerste plaats zijn studietoetsen operationalisaties van onderwijsdoelstellingen. Bij de constructie van een studietoets moet de docent zich telkens afvragen wat de doelstellingen van zijn onderwijs zijn. De docent wordt voor vragen gesteld als „Moeten de leerlingen bepaalde dingen alleen uit het hoofd kennen of moeten ze met de geleerde begrippen kunnen werken?”, „Welke delen van de stof zijn belangrijk en welke zijn minder belangrijk?”. Een boek waarover veel gediscussieerd werd was de „*Taxonomy of Educational Objectives*” van *Bloom*. *Bloom* verdeelt de onderwijsdoelstellingen in 6 hoofdcategorieën: kennis, begrip, toepassing, analyse, synthese en evaluatie. Vooral *Kulkarni* en *Dave* (India) bleken sterk beïnvloed te zijn door

het werk van *Bloom*. Voor elke cursus moet vastgesteld worden hoe de verhouding is tussen deze 6 doelstellingen. De studietoets over de cursus moet dan samengesteld worden overeenkomstig de vastgestelde verhouding tussen de doelstellingen. Bijvoorbeeld bij een cursus Geschiedenis stelt men van tevoren vast dat het vooral gaat om inzicht in bepaalde historische samenhangen. De docent bepaalt bijvoorbeeld dat in de toets over deze cursus 10% kennisvragen staan en 30% begrip, 30% analyse, 20% synthese en 10% toepassing. *Ebel* en *Wong* verzetten zich tegen een al te formeel vasthouden aan deze doelstellingen. Deze sprekers waren van mening dat de door *Bloom* genoemde doelstellingen een soort „kaart” vormen in het hoofd van de docent. De docent houdt bij de constructie van de studietoets wel rekening met de doelstellingen, maar het is weinig zinvol om precies van tevoren vast te leggen hoeveel items men van elke categorie van *Bloom* moet maken.

In de tweede plaats kunnen studietoetsen op grote schaal gebruikt worden doordat automatische verwerking van het materiaal mogelijk is. Voor de verwerking van grote hoeveelheden materiaal kunnen „optical readers” benut worden, die antwoordbladen van leerlingen lezen en de antwoorden overbrengen op ponskaarten. De ponskaarten kunnen dan weer verwerkt worden met de computer. *Lindquist* vertelde dat men in Iowa State University apparatuur ontwikkeld heeft die in staat is om 60.000 tests (elk bestaande uit ongeveer 60 vragen) per uur te verwerken. Deze deskundige gaf ook een toekomstbeeld van het gebruik van de computer bij studietoetsen. De computer is veel beter in staat dan welke persoon ook om fouten gemaakt bij de invulling van de antwoordbladen op te sporen en te corrigeren. Ook is het mogelijk de invullingen in de testboekjes zelf door de „optical reader” te laten verwerken waardoor geen antwoordbladen meer nodig zijn. Voorts zal men spoedig alle mogelijke itemvormen en scoringswijzen kunnen gebruiken, waardoor men nog beter precies datgene kan meten wat men wil meten.

Ten derde kan een studietoets statistisch geanalyseerd worden. Voor grote hoeveelheden materiaal heeft men een computer nodig; voor kleine hoeveelheden materiaal kan de docent zelf (met behulp van benaderingsformules) zeer snel een analyse van de test maken. Een statistische analyse van de studietoets geeft de docent „feedback” over het onderwijs dat door hem gegeven is. Hij kan bijvoorbeeld zien welke onderdelen van de stof door de leerlingen niet begrepen zijn en wat veel voorkomende misvattingen van de leerlingen zijn. *Schrader* benadrukte vooral de communicatie tussen de psychometrici en de docenten: de psy-

chometrici moeten de resultaten van de testanalyse in een dergelijke vorm gieten dat de docenten (meestal zonder kennis van de psychometrica) het kunnen begrijpen en zelf iets met de informatie kunnen doen. Verscheidene papers hadden betrekking op min of meer handige item-indices, die door de docenten geïnterpreteerd konden worden.

In de vierde plaats kan een studietoets een hulpmiddel zijn om het leerproces van de leerlingen te bevorderen. *Wantman* gaf hier een fraai voorbeeld van. Hij construeerde voor een cursus Statistiek, die door hem gegeven werd, twee studietoetsen. Beide studietoetsen bestreken de gehele stof van de cursus. Eén van de studietoetsen werd aan de studenten mee naar huis gegeven. De studenten maakten de studietoets thuis en vervolgens besprak *Wantman* de fouten die de studenten gemaakt hadden. Elke student kon aan de hand van de studietoets vaststellen welke onderdelen van de stof hij nog niet voldoende beheerste en wat hij nog niet begrepen had. De onderdelen die voor de student nog niet duidelijk waren trachtte de docent in een gesprek te verhelderen. Een tweede voorbeeld van dit gebruik van studietoetsen werd gegeven door *Wong*. Zij verdeelde een cursus Wiskunde in 55 verschillende eenheden en voor elke eenheid werd een studietoets geconstrueerd. De leerlingen begonnen met het bestuderen van de stof van de 1e eenheid; daarna maakten zij de studietoets over deze 1e eenheid. Als een leerling 90% van de vragen van de studietoets goed had, dan kon hij overgaan naar de 2e eenheid; als hij minder dan 90% van de vragen goed had, dan werden hem door de docent die onderdelen, die hij niet goed begrepen had, uitgelegd.

Anderson construeerde studietoetsen voor de 1e klas van de lagere school. Het nut van studietoetsen op deze jeugdige leeftijd is er in gelegen dat de onderwijzer reeds zeer vroeg kan vaststellen of een leerling een achterstand heeft in Rekenen of Taal. Op grond van deze informatie kan de onderwijzer de leerling reeds zeer snel trachten „bij te sturen”. De studietoets wordt hier dus gebruikt als een diagnostisch instrument.

5 Slotopmerkingen

Tijdens dit Congres is voornamelijk gesproken over studietoetsen. Er zijn echter ook lezingen gehouden over enkele andere tests, die van belang kunnen zijn voor het onderwijs. *Meis* rapporteerde over de constructie van een batterij van schoolrijpheidstests in Duitsland; *Seyfried* over de constructie van een batterij schoolrijpheidstests in Oostenrijk. *Malmquist* rapporteerde over een onderzoek in Zweden,

waarbij met behulp van tests kinderen met leesmoelijkheden opgespoord werden; 75% van deze kinderen kon met behulp van „remedial teaching” tot normale leesprestaties gebracht worden. *Irvine* rapporteerde over een onderzoek met „Raven's Progressive Matrices” in verschillende culturen. Een conclusie van dit onderzoek was dat ook deze intelligentietest niet „cultuur-vrij” is. *Torrance* demonstreerde een aantal creativiteitstests, die door hem en zijn medewerkers geconstrueerd waren. Ook *Kogan* rapporteerde over dergelijke tests.

Het is verbazingwekkend hoeveel onderzoek op dit netelige gebied reeds verricht is. Het bleek van belang om verschillende creativiteitsfactoren te onderscheiden, die weinig of niets met elkaar te maken hebben. Het zal echter nog wel even duren voordat hier nuttige toepassingen mogelijk zijn.

Ook vermelden we nog de interessante manier waarop sprekers van de Educational Testing Service (*Anderson, Rosner*) het mak en van de studietoetsen door onderwijzers propageren. Zij hebben hiervoor diacursussen gemaakt, die, met gesproken woord en muziek van een bandrecorder kunnen worden afgedraaid. We kregen de indruk dat geen woord te veel of te weinig op deze geprogrammeerde wijze werd overgebracht. Over het succes – soms wel eens mislukking – van deze cursussen op scholen werd vrij veel onderzoek verricht, dat dan weer diende om de cursussen te verbeteren. Deze sprekers, en ook *Dunn* (Australië) wezen er op dat vele onderwijzers het toetsen van de verworven kennis en vaardigheden nog te veel zien al siets wat onafhankelijk van het leerproces, desnoods als een noodzakelijk kwaad, gedaan moet worden, terwijl het in werkelijkheid een onmisbaar deel is van het leren: geen „curriculum” is effectief zonder toets.

Bijzonder opvallend was voor ons de achterstand van Nederland, samen met de meeste Europese landen, op het gebied van de studietoetsen. Dit Congres heeft het belang van studietoetsen bijzonder duidelijk gemaakt. Interessant was vooral de nadruk op andere mogelijkheden van studietoetsen naast het beoordelen van prestaties. In Nederland verkeren we nog in de situatie dat het noodzakelijk is om studietoetsen te construeren. Daarbij kunnen we een vruchtbaar gebruik maken van de vele mogelijkheden van studietoetsen, waarop dit Congres gewezen heeft.