

# TOETSING VAN KENNIS BIJ TECHNISCHE OPLEIDINGEN

L. VAN GENDEREN EN H. F. MULDER

## 1. INLEIDING

Per jaar neemt Hoogovens (aanduiding voor Koninklijke Nederlandsche Hoogovens en Staalfabrieken en aangesloten bedrijven), 120 à 200 leerling-vaklieden in opleiding voor bankwerker, instrumentenvakman, elektriciën en dergelijke. Deze worden in 2 of 3 jaar (afhankelijk van het vak), opgeleid voor diploma's van de landelijke leerlingstelsels van Bemetel en V.E.V. In totaal omvat het leerlingstelsel jaarlijks 350—400 cursisten. De opleiding vindt grotendeels plaats aan de Bedrijfsschool, deels via roulatie in het bedrijf. Naast deze en vele andere opleidingsactiviteiten worden aan het Opleidingscentrum van Hoogovens avondcursussen gegeven op uiteenlopende vakgebieden. Hierbij zijn ruim duizend cursisten betrokken. Het onderstaande heeft betrekking op cursisten van het leerlingstelsel en van het avondvakonderwijs.

In de volgende paragrafen komt een onderwerp aan de orde waarmee bijna elke opleiding te maken heeft, n.l. de noodzaak vast te stellen welke kennis een cursist heeft en welke niet.

Deze behoefte heeft men op verschillende tijdstippen en met verschillende oogmerken:

1. aan het begin van een opleiding, want men moet aansluiten bij het bestaande kennispakket;
2. gedurende de opleiding, teneinde na te gaan wat wel en wat niet is begrepen;
3. het nemen van beslissingen: overgaan naar volgende klas, wel of niet toekennen van diploma.

Ten behoeve van het verkrijgen van inzicht in verworven kennis werd door de meeste leraren van de Bedrijfsschool van Hoogovens huiswerk opgegeven. Daarnaast zijn proefwerken in gebruik.

Leiding en leraren van de Bedrijfsschool zagen en zien het thuis maken van opgaven door de leerlingen als een belangrijk onderdeel van en hulpmiddel bij het leerproces.

Men beoogt er mee:

- dat de leerling zich rustig en in eigen tempo zal bezinnen op de behandelde leerstof;

— dat de leraar zich een beeld kan vormen van de mate waarin de leerling zich de gewenste kennis heeft eigen gemaakt.

Uitvoerige discussies binnen een groep leraren over het thuis maken van sommen e.d., alsmede een enquête onder leerlingen, bracht het volgende aan het licht:

- van rustige bezinning door de jongens komt over het algemeen weinig terecht. Het werk wordt vaak op 't laatste moment afgerafeld, soms ook overgenomen van anderen. De jongens kijken naar het cijfer of „hoeveel goed?” en niet naar de bijgeschreven kritiek of de aard van de gemaakte fouten.
- Het wel of niet maken van huiswerk of de kwaliteit van het gemaakte werk, had en heeft in de regel weinig consequenties.
- De correctie van het gemaakte werk en de beoordeling ervan gebeurt door de leraren niet op uniforme wijze. Bijv. door de een steeds, door de ander steekproefsgewijs. Bij de een per les, door de ander per serie lessen. De een geeft cijfers, de ander aantallen goed en fout.  
De een laat het werk door de klas nazien, de ander doet het zelf. Van de laatsten waren indertijd velen achter met hun correctiewerk.

Beslissingen t.a.v. bevorderingen en dergelijke werden genomen aan de hand van rapportcijfers, meestal samengesteld op grond van voor proefwerken en werkstukken verkregen cijfers.

In verband met de hele gang van zaken bij de huiswerkopdrachten en de correctie ervan, rees het vermoeden, dat de waarde van het huiswerk maken, de controle daarop, de correctie en de waardering ervan, in de bestaande vorm zeer betrekkelijk zou kunnen zijn.

Anderzijds bleef de mening, dat systematisch nagaan wat wel en niet is geleerd van behandelde lesstof, een essentieel onderdeel moet vormen van het leerproces.

De gesignaleerde bezwaren en moeilijkheden moeten worden gezien tegen de problemen die het maken van huiswerk en proefwerken in het algemeen oproepen, wanneer men vragen en opgaven gebruikt zonder deze te standaardizeren, normeren en ijken. Voorbeelden van de bedoelde moeilijkheden:

- men heeft tevoren geen maat voor de „zwaarte” van de opgaven;
- de cijferwaardering van het gemaakte werk achteraf, gebeurt meestal aan de hand van de „gemiddelde prestatie in de groep” en fluctueert daardoor van groep tot groep;

- het is tijdrovend voor de leerling — er gaat soms een hele lestijd mee heen;
- de correctie kost de leraar veel tijd, een nauwgezette correctie van één proefwerk van een hele klas kost uren;
- tussen afneming van de proef en terugrapportering van de resultaten aan de leerling ligt dikwijls een geruime tijd;
- de kans bestaat op subjectieve waardering van de prestatie.

Dit alles overwegende rees in een reeks gesprekken tussen de schrijvers van dit artikel over dit probleem, de gedachte een poging te ondernemen om de overdracht van lesstof te toetsen op een empirisch-statistische basis. Daarbij werd gestreefd naar aansluiting op ontwikkelingen in deze richting elders en gebruikmaking van ervaring, door de Psychologische Dienst opgedaan bij de constructie van tests.

Met enkele lesonderwerpen werd een begin gemaakt en elke — in overleg geconstrueerde toets — werd statistisch bewerkt. Op de constructiemethode wordt in par. 3 en 4 ingegaan. In par. 5 en 7 worden voorbeelden behandeld van twee soorten toetsen:

1. toets als hulpmiddel bij het leerproces (hierna genoemd lesstof-toetsen, afgekort LT).
2. toetsen voor het vaststellen van rapport- en examencijfers, op grond waarvan beslissingen worden genomen (hierna genoemd kern-toetsen, afgekort KT).

Tot heden bestaan de toetsen uit een serie vragen. Bij elke vraag worden 4 mogelijke antwoorden gepresenteerd. Eén van deze antwoorden is het juiste. De cursist moet door het plaatsen van een kruisje op een met de vragen corresponderend antwoordblaadje aangeven, welk antwoord volgens hem van toepassing is. Voorbeeld:

Een boor met een veranderende spoedhoek wordt gebruikt omdat:

- a. de spanen hier gemakkelijker breken;
- b. de spanen niet zo gemakkelijk zullen breken;
- c. het koelmiddel sneller de boorpunt bereikt;
- d. de boor sterker is.

De drie antwoordmogelijkheden naast het enige juiste antwoord noemt men wel „afleiders”.

Vraag plus het juiste antwoord plus afleiders noemt men een item. De 4-antwoorden-vorm is niet principieel. Ook andere vormen zijn mogelijk en zullen in de zich daartoe lenende gevallen stellig worden toe-

gepast. Te denken valt aan vragen waarop slechts één getal of woord het juiste antwoord is. Dit wordt dan opengelaten en op die plaats door de cursist ingevuld, of er worden uitspraken gepresenteerd waarachter staat: juist-onjuist. De leerling moet dan doorstrepen wat volgens hem niet van toepassing is.

Voorlopig concentreren wij ons nog op de meervoudige antwoordvorm ten einde met één vorm grondige ervaring op te doen. Deze uniforme vorm vergemakkelijkt en bespoedigt de scoring, telling en bewerking van de verkregen gegevens.

Momenteel krijgen de LT's de meeste aandacht omdat zij van het grootste belang worden geacht als hulpmiddel bij het leerproces. Met goed geconstrueerde LT's kunnen leerling én leraar snel geconfronteerd worden met slagen en falen bij het overdragen, begrijpen en beklijven van leerstof.

Meting met LT, of bepaalde kenniselementen wel of niet zijn „overgekomen”, kost weinig tijd, zodat aan uitleg, herhaling, bijbrengen van nieuwe elementen meer tijd kan worden besteed, dan bij klassieke toetsingsmethoden (huiswerk, overhoren, proefwerken) mogelijk is.

De mogelijkheid om, puttend uit een omvangrijk item-bestand, toetsen te vervaardigen voor het nemen van beslissingen (wel-niet voortzetten van opleidingen, bijv.), zien wij als een welkom bijproduct van onze werkwijze, doch dit was niet ons hoofddoel, toen wij begonnen te experimenteren met LT. Zoals gezegd, ging het in eerste instantie om de efficiency en effectiviteit van het leerproces.

## 2. VOORWAARDEN WAARAAN TOETSEN MOETEN VOLDOEN

1. De vragen moeten de vereiste kennis in een *bepaald vakgebied bestrijken*. Bijvoorbeeld: wil men de kennis van de elementaire elektriciteitsleer onderzoeken, dan mogen vragen over de wet van Ohm niet ontbreken.
2. De toetsen moeten *valide* zijn, dat wil zeggen, zij moeten meten wat zij pretenderen te meten. Een mechanicoets die in wezen een rekenoets is, is als mechanicoets niet valide.
3. Kennis en de daarmee samenhangende toetsen moeten *per tijdsperiode* (bijv. tussen twee rapporten) *doeltreffend gedoseerd* zijn. Op een bepaald moment  $t$  in de opleiding, moet de toets niet minder vragen dan aan kennis is (of kan worden) overgebracht dan tot moment  $t$  het geval is (of kan zijn), maar ook niet meer.
4. De toetsen moeten *betrouwbaar* zijn, dat wil zeggen, er moet weinig invloed zijn van toevalefouten of toevalligheden. De kans op

toevalstreffers moet kleiner zijn dan volgens kansrekening mogelijk is.

(Zie over de begrippen validiteit en betrouwbaarheid o.a. De Groot, lit. 5).

5. De toetsen moeten het vertrouwen hebben van leraren en leerlingen.
6. De toetsen moeten snel en objectief scorebaar zijn door ondeskundigen. (Bij grote aantallen zal men tot machinale scoring overgaan of een computer inschakelen).
7. De toetsen moeten snel te maken zijn door de cursisten. Voor LT rekenen wij op ongeveer 10 minuten, voor KT op één à anderhalf lesuur.

Deze voorwaarden brengen mee, dat aan de constructie veel zorg moet worden besteed. De door ons gevolgde methode van constructie wordt voor de LT behandeld in par. 3, voor de KT in par. 7. De analyse gericht op voldoen aan eis no. 4 komt in par. 4 ter sprake.

### 3. LEESSTOFTOETSEN • CONSTRUCTIEMETHODE

Toetsing van de overdracht van lesstof vraagt om een adequate samenstelling van die stof. Deze moet zijn afgestemd op de praktijk. Tussen wat men in de les brengt en het met het leren nagestreefde doel moet een verband bestaan. Daarbij doet het er niet veel toe of de lesstof bedoeld is als hulpmiddel voor andere lessen (meetkunde en gonio als hulpmiddel voor werktuigkunde, bijv.) of voor het leren vervullen van een bedrijfsfunctie (storingsbankwerker).

Ideaal zou zijn: lesstof gebaseerd op een functieanalyse die zo ver in onderdelen is uitgesplitst, dat daaruit de vereiste te leren kennis en vaardigheden kunnen worden afgeleid.

In de praktijk komen deze ideale functie-analyses weinig voor, zodat gesteund moet worden op bestaande lessen, boeken, proefwerken en „in leraren opgeslagen ervaring”. Wat hier als een moeilijkheid naar voren wordt gebracht geldt evenzeer de klassieke proefwerkmethode. In de regel komt de zwakte van de relatie tussen proefwerk en „wat de praktijk vraagt” niet vaak aan het licht.

Meestal maken de leraren zelf hun proefwerken, afgestemd op wat zij aan lesstof noodzakelijk achten. Maakt men op een aangepaste wijze gebruik van toetsen, dan is er kans op, dat verschillende leraren in verschillende jaren met dezelfde toets gaan werken. En eventueel verschil in behandeling van de lesstof zou zich dan kunnen wreken. Vandaar dat ten behoeve van lesstoftoetsen veel aandacht moet worden besteed

aan de opstelling van het over te dragen kennispakket. De hierna beschreven werkwijze heeft dan ook het volgende bijproduct opgeleverd. De lesstof wordt nog eens kritischer bekeken op behoefte en nut, dan in het algemeen het geval is bij het opstellen van proefwerken. Men wordt gedwongen zich te voren rekenschap te geven van de te brengen leerstof.

Scherper dan tevoren vraagt men zich af:  
welk kennisonderdeel kan in de praktijk nodig zijn, wat kan worden gemist, hoe moet het kenniselement worden gepresenteerd en overgedragen?

Na deze bezinning op de te behandelen lesstof en op de leselementen waarvan men de kennisoverdracht wil toetsen, worden voor de samenstelling van de toetsen de volgende stappen gezet.

1. Begonnen wordt met een splitsing van het geheel van het te doceren vak in onderdelen en onderwerpen. Dit gebeurt door een leraar die in een bepaald vak les geeft.
2. Daarna wordt de per les te geven stof omgezet in vragen en antwoorden. Voor een gedeelte doet dit een leraar aan de hand van de vragen die hij in het algemeen gedurende de lessen bij huiswerk of proefwerken stelt.

Voor een deel vindt dit op de volgende wijze plaats.

Een leraar woont lessen van een collega bij. Hij noteert wat op het bord wordt getekend en de kennis-elementen waarop de lesgevende leraar in zijn les de nadruk legt. Uit deze observatie haalt hij vragen, de juiste antwoorden en de afleiders.

In beide gevallen gaat het proces verder met de volgende stappen.

3. De in stap 2 gevonden vragen worden besproken in een groepje leraren (3 à 4), die dezelfde stof geven aan vergelijkbare groepen leerlingen. De groep leraren wordt bijeen geroepen en voorgezeten door de leraar die de vragen opstelt.

Een enkele maal wordt een leraar van een hogere jaargroep erbij gehaald. Deze krijgt zodoende informatie over het peil waarop hij de leerlingen „over” krijgt en hij kan wensen t.a.v. het volgens hem vereiste niveau kenbaar maken.

In het algemeen wordt men het in deze discussie snel eens over de aantallen vragen en de formulering ervan.

4. Per vraag worden door de groep mogelijke antwoorden op de vragen opgesteld. Daarover wordt veelal breedvoerig gesproken. De

voorzittende leraar moet soms knopen doorhakken. Maar dit gebeurt pas nadat *ieder* zijn mening heeft gegeven.

*Opm.*: de onder 3 en 4 genoemde groepsprocedure heeft de volgende voordelen:

- grotere ideeënproductie;
  - gelijktrekken van lesstofbehandeling;
  - zekerheid dat deze leraren in de praktijk met de door hen zelf geproduceerde toetsen gaan werken. Legt men deze wetenschap van buiten af aan hen op, dan komt van een effectief gebruik n.l. weinig terecht.
5. De op deze wijze samengestelde toets wordt getypt en vermenigvuldigd. Het voorlopige karakter van de toets komt tot uiting in de gebezigde reproductietechniek.
  6. De voorlopige toets wordt door de organiserende leraar geprobeerd in een van zijn lessen. Daarbij wordt gelet op de reacties van de leerlingen, teneinde zwakke kanten van de toets te ontdekken en eventuele suggesties voor verbetering op te vangen.
  7. De uitkomsten van de toets worden statistisch bewerkt (zie par. 4). Aan de hand van de in de voorgaande stap opgedane ervaringen worden noodzakelijk geachte wijzigingen aangebracht.
  8. Andere leraren proberen de toets op andere — vergelijkbare — groepen.
  9. De onder 8. verkregen resultaten worden statistisch bewerkt.
  10. De toets wordt herzien aan de hand van de uitkomsten van stap no. 9. Deze toets wordt gedurende 1 jaar gebruikt in de verschillende klassen.
  11. Statistische bewerking van de via stap 10 verkregen resultaten. De toets krijgt aan de hand hiervan een „definitieve” vorm. „Definitief” tussen aanhalingstekens omdat voortgezet gebruik tot verandering in items en groepering van de items kan voeren.
  12. De resultaten, per item verkregen met de toets, worden op een kaart t.b.v. de „Item bibliotheek” vastgelegd.

#### 4. ITEMANALYSE

Elke proefset items wordt onderzocht voor de volgende gezichtspunten:

- 4.1 De relatie tussen de prestatie bij een item en de totale prestatie bij de toets, uitgedrukt in een correlatiecoëfficiënt  $r$ .



4.2 De frequentie van de keuze van de juiste antwoorden; deze levert de p-waarde op.

4.3 Hiervoor zijn de volgende bewerkingen nodig:

verdeel de door de cursisten behaalde scores in 2 groepen met gelijke aantallen, een groep met hoge scores en een evengrote groep met lage scores.

Per item worden de volgende aantallen geteld:

j. goede antwoorden in de groep met hoge scores

k. foutieve „ „ „ „ „ „ „ „

l. goede „ „ „ „ „ „ lage „

m. foutieve „ „ „ „ „ „ „ „

4.4 Berekening van  $r$ . Neemt men hiervoor de tetrachorische correlatiecoëfficiënt  $r_{tet}$  (bepaald via een cosinus-pi-coëfficiënt) dan is weinig rekenwerk vereist bij gebruik van de daarvoor bestaande tabellen (zie bijvoorbeeld Guilford, lit. 8).

De uitkomst van  $(j \times m) : (k \times l)$  wordt in de tabel opgezocht en geeft meteen de bijbehorende correlatiecoëfficiënt. De gevonden  $r_{tet}$  waarden zijn weliswaar benaderend, doch voor ons doel betrouwbaar.

Guilford (lit. 8.) e.a. geven voorts tabellen voor het aflezen van de produkt-moment-correlatiecoëfficiënt, de zgn abac's. De ingangen van deze tabellen zijn: de omvang van het aantal juiste antwoorden bij 27 % van de hoogste en 27 % van de laagste scores.

Abac's zijn formeel slechts bij grote aantallen proefpersonen bruikbaar, doch kunnen eventueel worden benut bij groepen van ongeveer 100 personen, wanneer men voor een eerste benadering van de betreffende uitkomsten aan de betrouwbaarheid geen hoge eisen stelt.

In lit. 15 wordt de coëfficiënt van Long behandeld als een verkorte rekenwijze voor de benadering van  $r$  in de item analyse.

4.5 Vaststelling van de lengte van de toets, ofwel het aantal items. De correlatie tussen score op een item en de totaalscore hangt mede af van de lengte van de toets. De formule van Spearman-Brown biedt gelegenheid om na te gaan hoeveel maal de toets verlengd moet worden om de gewenste betrouwbaarheid te verkrijgen.

Deze formule luidt:

$$r_{nN} = \frac{nr + I}{1 + (n-1)r}$$

II

Men kan er voorts de nieuwe betrouwbaarheid mee vaststellen als men



een toets met een bepaald aantal items verlengt. (Zie hiervoor o.a. Drenth lit. 3. p. 179).

#### 4.6 Berekening van $p$

$$p + (j + 1) : (j + k + 1 + m).$$

(Voor andere technieken voor itemanalyse zie Lindquist, lit. 11).

#### 4.7 Voorbeelden van bewerking van itemgegevens

Tabel 1.

Bepaling van gegevens over een denkbeeldig item.

Keuze van de antwoordmoge- lijkheden in de groepen met	aantal perso- nen	frequentie van de keuze van de ant- woorden no.:					onbe- ant- woord	coëff.
		1	2	3	4	5		
Hoge scores	100	4	18	55	1	21	1	$r = .52$
Lage scores	100	20	25	22	0	31	2	$p = .38$
				↑				
				juiste antwoord				

#### Berekening van $r_{tet}$

$$jm : kl = 55 \times 76 : 44 \times 22 = 4.3$$

Bovengenoemde tabel geeft voor  $jm : kl = 4.3$  een  $r_{tet}$  van 0.52.

#### Berekening van $p$ .

$$p = 77 : 200 = 0.38 \text{ (afgerond op 2 decimalen).}$$

*Opn.*: zou in bovenstaand voorbeeld het juiste antwoord (dit is no. 3 van tabel 1) in de groep met de hoge scores bijv.  $25 \times$  zijn gekozen en in de groep met de lage scores  $30 \times$ , dan zou dit aanleiding geven het item te herzien.

Dit zou ook moeten gebeuren wanneer de groep met hoge scores bij een bepaald item vaker foutieve afleiders kiest dan het goede antwoord. Derde mogelijkheid: de frequentie van de keuzen is voor alle antwoordmogelijkheden nagenoeg gelijk.

In het algemeen kunnen bij de itemanalyse dit soort ernstige fouten in de toetsconstructie snel aan het licht komen.

Vatten we het bovenstaande nog eens samen in een overzicht van de met itemanalyse nagestreefde doelen, dan krijgen we het volgende:

- zonder een validatie onderzoek te verrichten kan men toch zien of een item *onbruikbaar* is;

- aanwijzingen verkrijgen voor de definitieve *vormgeving* van de vragen en de antwoorden. Een nooit gekozen afleider of een te frequent gekozen afleider vervangt men door een ander. Sommige afleiders zullen te moeilijk blijken, andere te gemakkelijk. Deze zullen veranderd moeten worden.
- Items die laag correleren met de totale score op de toets zullen in hun geheel moeten worden herzien of moeten vervallen.
- Vaststelling van de moeilijkheidsgraad. Items met een hoge correlatie met de totaal score, doch met een weinig frequente keuze van het goede antwoord, zijn moeilijker dan die met een hoge correlatie, doch een zeer frequente goede antwoordkeuze. De moeilijkheidsgraad van items kan een richtsnoer zijn voor de bepaling van de volgorde van de items in de toetsen.  
(Voor constructie van toetsen zie lit. 1,2,3,4,9,11).

Hieronder is een LT over een gedeelte van het vakonderwerp boren afgedrukt. Links staan de vragen, rechts de antwoorden en afleiders. De letter voor het juiste antwoord is cursief gedrukt.

---

## VAKTHEORIETOETS

### *Boren I*

- |  |  |
|--|--|
| 1. De boren welke in de werkplaats worden gebruikt zijn gemaakt van: | <ul style="list-style-type: none"> <li>a. <i>gelegeerd koolstofstaal</i></li> <li>b. <i>sneldraaistaal</i></li> <li>c. hoogwaardig staal</li> <li>d. hardmetaal.</li> </ul>  |
| 2. De schacht van een boor is:                                       | <ul style="list-style-type: none"> <li>a. <i>gehard om slijtage te verminderen</i></li> <li>b. <i>gehard om vreten tegen te gaan</i></li> <li>c. ongehard om breuk te vermijden</li> <li>d. ongehard om de boor goedkoper te maken.</li> </ul>   |
| 3. Een spiraalboor bezit geleidingsranden:                           | <ul style="list-style-type: none"> <li>a. <i>dit zijn tevens snijkanten en helpen hierdoor een gat te maken in staal</i></li> <li>b. dienen voor geleiding van de boorpunt in ondiepe gaten</li> <li>c. dienen voor geleiding van de boorpunt in diepe gaten</li> <li>d. helpen het gat een betere rondheid te geven.</li> </ul> |

4. De boorpunt van een spiraalboor geschikt voor het boren van staal is:
- 90°
  - 100°
  - 120°
  - 130°
5. De spiraalhoek (spoedhoek) van de spaangroeven voor een boor geschikt om staal te boren is:
- 15°
  - 25°
  - 35°
  - 45°
6. Als de boorpuntshoek en de vrijloophoek van een boor voor staal goed geslepen zijn, dan zijn de:
- hoofdsnijkkanten bol
  - hoofdsnijkkanten hol
  - hoofdsnijkkanten recht en niet evenwijdig
  - hoofdsnijkkanten recht en evenwijdig.
7. De lip aan de conische schacht van een spiraalboor dient:
- alleen om de boor uit de boor-spil te drijven zonder kans op beschadigingen
  - om de eerste aanzetkracht op te vangen en de boor uit te drijven
  - alleen om slippen van de boor te voorkomen
  - voor het opvangen van de aanzetkracht.
8. Als de dwarssnijkant van een spiraalboor niet in het midden zit, dan zijn:
- de hoofdsnijkkanten even lang
  - de hartlijn van de boor de boorpuntshoek middendoor
  - zijn de hoofdsnijkkanten niet even lang en deelt de hartlijn de boorpuntshoek niet middendoor
  - de vrijloophoeken niet gelijk
9. De spaangroeven in een spiraalboor dienen:
- voor het vormen van de hoofdsnijkkanten
  - om wrijving te verkleinen
  - om ronde gaten te krijgen
  - om de wand van het boorgat gladder te maken
10. De dwarssnijkkanten verspanen niet omdat:
- de wighoek te groot is
  - de boorpuntshoek te groot is
  - de vrijloophoek te klein is
  - de spaanhoek negatief is.

---

De eerste bewerking van de gegevens, verkregen met deze LT, is verzameld in tabel 2.

TABEL 2 — Eerste bewerking van gegevens verkregen met lesstoftoets boren I

		Frequentie van de keuze van de antwoorden van item no.																								
SCORE GROEP		1					2					3					4					5				
		ANTWOORD					ANTWOORD					ANTWOORD					ANTWOORD					ANTWOORD				
		a	b <sup>1)</sup>	c	d		a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d				
Hoog	1	12	21½ <sup>3)</sup>	2	1	5	4	26½	1	8	3	18½	7	0	0	36½	0	10	21	4½	1					
Laag	1	22	7½	3	3	9	3	22½	2	14	6	7½	9	0	1	35½	0	11	21	3½	1					
Som	1	34	29	5	4	14	7	49	3	22	9	26	16	0	1	72	0	21	42	8	2					
P		.40					.67					.36					.99					.58				
r		.60					.20					.51					.03 <sup>2)</sup>					.00				

		Frequentie van de keuze van de antwoorden van item no.																								
SCORE GROEP		6					7					8					9					10				
		ANTWOORD					ANTWOORD					ANTWOORD					ANTWOORD					ANTWOORD				
		a	b	c	d		a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d				
Hoog	1	1	0	1	34½	14½	21	1	0	0	0	34½	2	34½	2	0	0	5	6½	4	21					
Laag	1	2	1	6	27½	28½	5	1	2	1	0	29½	6	32½	3	0	1	5	10½	11	10					
Som	1	3	1	7	62	43	26	2	2	1	0	64	8	67	5	0	1	10	17	15	31					
P		.85					.36					.88					.92					.42				
r		.09 <sup>2)</sup>					.34 <sup>2)</sup>					.07 <sup>2)</sup>					.01 <sup>2)</sup>					0.47				

Aantal proefpersonen 73

<sup>1</sup> de correcte antwoorden zijn cursief gedrukt.

<sup>2</sup> Wegens de kleine aantallen in enkele cellen moeten deze correlaties met veel voorzichtigheid worden gehanteerd.

<sup>3</sup> De middelste waarneming in de serie van 73 is voor de helft bij de hoge en bij de lage scores geteld, daar komen de waarden 1/2 vandaan. I.v.m. besparing van rekenwerk kan men bij een oneven serie waarnemingen, de middelste eventueel ook weglaten, i.p.v. te verdelen over hoog en laag, zoals in dit voorbeeld is gedaan.

De verkregen gegevens kunnen als volgt aangeduid worden:

*Bruikbaar* zijn de items 1, 3 en 10. Want het juiste antwoord wordt dan door de hoog scorende groep significant vaker gekozen dan door de laag-scorende. Zie ook de r-waarden. De verdeling van de frequenties is redelijk, zij het dat de afleiders c en d bij item 1 wat weinig gekozen worden. Verbetering van deze afleiders is gewenst.

*Een of meer afleiders herzien*: item 2, 6, 7, 8. Want bij deze items worden sommige afleiders te weinig gekozen. Vraag en correcte antwoord kunnen gehandhaafd blijven, want er is een duidelijk verschil in de keuze van het juiste antwoord tussen hoog en laag scorenden.

*Onbruikbaar*: items 4, 5, 9. Want: geen verschil in keuze van het juiste antwoord tussen hoog en laag scorenden en/of weinig keuze van

afleiders. Deze items zullen of moeten worden vervangen, of antwoord en afleiders zullen moeten worden herzien.

Waarschijnlijk voldoet t.a.v. vraag 4 de volgende vraagstelling beter:

4. De boorpuntshoek van een spiraalboor, geschikt voor het boren van staal is: (grootte van de hoek invullen).

En voor vraag 5:

5. De spiraalhoek (spoedhoek) van de spaangroeven voor een boor geschikt om staal te boren is  $25^\circ$ , dit is: juist — onjuist.

Voor vraag 9:

9. De spaangroeven in een spiraalboor dienen voor het vormen van de hoofdsnijkant. Dit is juist — onjuist.

Als voorbeeld hoever men kan gaan met ingewikkelde opgaven waar kan worden volstaan met vragen, waarbij het juiste antwoord bestaat uit één woord of getal, of doorstrepen wel — niet en dergelijke, zie men A. J. Derksen, Leerboek Elektronica, De Muiderkring N.V. Bussum.

#### 6. PROEFWERKEN IN DE VORM VAN TOETSEN - ALGEMEEN

Voor het vaststellen van rapportcijfers worden bij de Bedrijfsschool proefwerkcijfers gehanteerd. Daartoe worden de antwoorden bij proefwerken, bestaande uit vragen en opgaven, door de leraren beoordeeld en in de vorm van een proefwerkcijfer gewaardeerd.

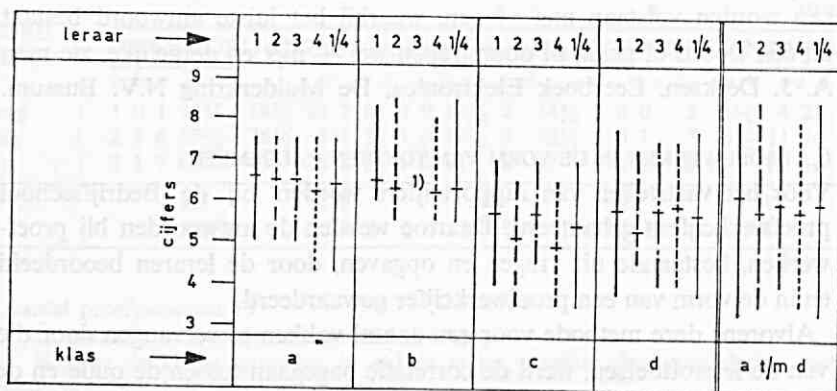
Alvorens deze methode voor een aantal vakken te vervangen door die van de lesstoftoetsen, werd de correlatie nagegaan tussen de oude en de nieuwe methode en tussen de cijfers die leraren onafhankelijk van elkaar voor eenzelfde proefwerk gaven.

Van de door de verschillende leraren gegeven cijfers werden rekenkundig gemiddelden bepaald en de op deze wijze verkregen „gemiddelde leraren cijfers” werden gecorreleerd met de uitkomsten van lesstoftoetsen.

Uiteraard kwamen in deze toetsen dezelfde onderwerpen aan de orde als in de proefwerken het geval was. Als voorbeeld van de resultaten, verkregen bij de oude en nieuwe methode worden hier table 3 en figuur 1 gegeven.

Tabel 3. Proefwerkcijfers vergeleken met toetsresultaten.  
Lesonderwerpen: boren, zagen, beitels, vijlen.

Klas	n	gemiddelde leraren- cijfers	correlatie leraren- gemiddelde en toets	p voor de correlatie
a	21	6.4	.46	< 0.05
b	22	6.3	.36	0.10
c	22	5.4	.23	> 0.10
d	22	5.7	.51	< 0.02
a t/m d	87	6.0	.37	< 0.02



— = gemiddelde

1) = niet beoordeeld

Figuur 1 brengt in beeld: de variatiebreedte van de cijferwaardering en het gemiddelde per klas en per leraar, en de cijferwaarderingen per leraar over de 4 klassen samen.

Uit figuur 1 en tabel 3 komt het volgende naar voren:

— een *zelfde proefwerk* wordt door verschillende leraren *uiteenlopend* gewaardeerd, zie bijv. hoe de gemiddelden variëren.

In 19 van de 87 gevallen werd door een of meer leraren een onvoldoende (< 5.5) toegekend, terwijl de overige leraren voor het *zelfde werk* een voldoende gaven.

— De vier *klassen* worden verschillend gewaardeerd. De rangorde is: a, b, c, d.

Leraren zijn al gauw geneigd deze verschillen aan de kwaliteit

van de groepen toe te schrijven, men spreekt van goede en slechte, ijverige en luie klassen, klassen van hoog en laag niveau, enz.

Het laatste is in ons geval geen afdoende verklaring. Want de intelligentiescores liggen voor de 4 klassen op een zelfde gemiddelde en de vooropleiding van de leerlingen is nagenoeg gelijk.

— de hoogte van de waardering verschilt per leraar.

De volgorde van de hoogte van de cijferwaardering van de leraren is van hoog tot laag: 1, 3, 2, 4.

— het gebruik van de cijferschaal loopt per leraar uiteen. De variatiebreedte ( $w$ ) is bij de een groter dan bij de ander.

— de uitersten van de cijferschaal werden niet gebruikt.

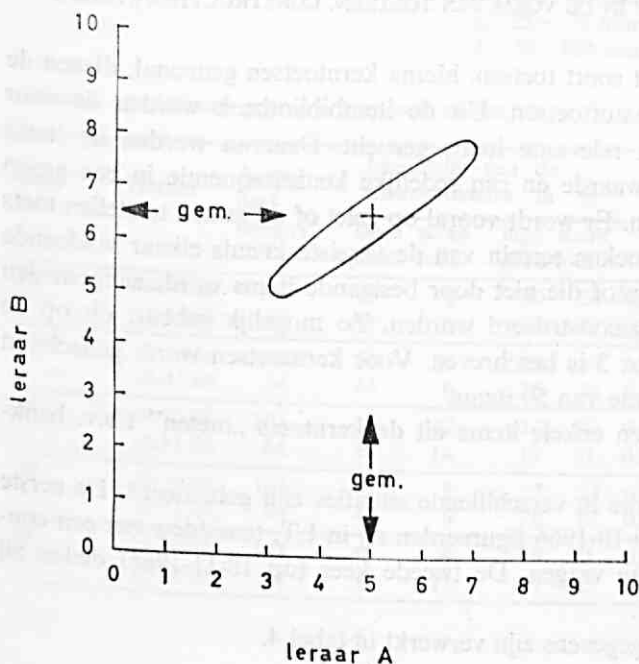
Enkele van de bij toetsen verkregen uitkomsten werden gecorrigeerd met intelligentietests.

Correlatiecoëff. kerntoets — G-factor GATB = 0.10

„ „ — „Gereedschappentest” = -0.03

(GATB = General Aptitude Test Battery).

De toetsen correleerden dus noch met de „algemene intelligentie”, noch met „kennis van gereedschappen”.





De indruk van sommige leraren, dat lesstoftoetsen intelligentie meten i.p.v. schoolkennis, werd niet bevestigd.

De correlaties tussen cijferwaarderingen van diverse leraren verschilden sterk per persoon en vak. Variaties van  $r = .40$  tot  $r = .90$  kwamen voor. De nu en dan hoge correlaties tussen de cijfers van de verschillende leraren werden nog al eens gehanteerd als een argument ten gunste van de klassieke waarderingswijze. Maar deze redenering bleek niet steekhoudend, omdat in deze gevallen de gemiddelde waardering per leraar sterk uiteen liep!

De een hanteerde bijv. een schaal van 5 tot 8 en kwam op een gemiddelde van 6.5, de ander werkte met cijfers van 3 tot 7 en kwam op een gemiddelde van 5. De correlatie was weliswaar hoog, doch in wezen was de waardering verschillend. Figuur 2 stelt zo'n voorbeeld grafisch voor: een hoge correlatie, maar verschillende gemiddelden.

Na een aantal experimenten groeide de overtuiging, dat de argumenten die voor de objectieve toetsmethoden worden aangevoerd ook gelden voor de stof en de cursisten van de Bedrijfsschool van Hoogovens. Alle bevindingen pleitten voor een vervanging van proefwerken door KT. Voor zover daarvoor voldoende itemmateriaal beschikbaar is, vindt deze vervanging daadwerkelijk plaats of is in bewerking.

#### 7. PROEFWERKEN IN DE VORM VAN TOETSEN. CONSTRUCTIEMETHODE EN VOORBEELD

Als basis voor dit soort toetsen, hierna kerntoetsen genoemd, dienen de items van de lesstoftoetsen. Uit de itembibliotheek worden de voor een bepaald vak relevante items gezocht. Daarvan worden de items met een hoge  $r$ -waarde en een redelijke keuzefrequentie in een groep leraren besproken. Er wordt vooral op gelet of de samen te stellen toets en het te onderzoeken terrein van de vereiste kennis elkaar voldoende dekken. Voor lesstof die niet door bestaande items wordt aangesneden moeten nieuwe geconstrueerd worden. Zo mogelijk gebeurt dit op de wijze zoals in par. 3 is beschreven. Voor kerntoetsen wordt gedacht in de orde van grootte van 50 items.

Hieronder volgen enkele items uit de kerntoets „meten” t.b.v. bankwerkers.

Het zijn items die in verschillende situaties zijn gehanteerd. De eerste maal, nml. op 19-10-1966 figureerden zij in LT, temidden van een context van verwante vragen. De tweede keer (op 18-11-1966) deden zij dienst in een KT.

De statistische gegevens zijn verwerkt in tabel 4.

20. Bij het meten moet rekening gehouden worden met de:
- hardheid van het materiaal;
  - de zachtheid van het materiaal;
  - oppervlaktestgesteldheid;
  - meetdruk van het meetinstrument;
42. De grootte van het meetgebied van een meetklok is waar te nemen aan:
- de verdeling van de grote wijzerplaat;
  - de verdeling van de kleine wijzerplaat;
  - de lengte van de meetstift;
  - het opschrift op de schaal.
44. Een meetklok wordt in het algemeen gebruikt voor:
- het meten van eindmaten;
  - het meten van bewegende onderdelen;
  - het meten van vlakheid;
  - het stellen van werkstukken op een draaibank.
45. Alle soorten winkelhaken behoren tot de groep:
- meetinstrumenten;
  - controle-instrumenten;
  - hulpgereedschappen;
  - hoekmeetgereedschappen.
46. Het meetgebied van een normale micrometer is:
- 0—25 mm;
  - 0—50 mm;
  - 25—75 mm;
  - 50—100 mm.

Tabel 4. Statistische gegevens kerntoets meten.

Item no.	Datum afn.	Aantal deelnemers	frequentie van de antwoordkeuze in %.				r	p
			hoge score		lage score			
			goed	fout	goed	fout		
20	19-10-66	103	45	5	30	20	0,61	0,75
	18-11-66	64	44	6	27	23	0,62	0,70
42	19-10-66	103	30	20	11	39	0,60	0,40
	18-11-66	64	44	6	16	34	0,80	0,59
44	19-10-66	103	28	22	13	37	0,49	0,41
	18-11-66	64	36	14	19	31	0,52	0,54
45	19-10-66	103	45	5	27	23	0,67	0,72
	18-11-66	64	45	5	28	27	0,67	0,73
46	19-10-66	103	46	4	29	21	0,69	0,75
	18-11-66	64	39	11	27	23	0,42	0,66

Bij de items 42 en 44 ziet men een frequenter keuze van het juiste antwoord in de KT-situatie.

Dit zou op een leereffect kunnen wijzen. Dwz: De cursisten zijn in de LT situatie met een bepaalde vraag geconfronteerd. Bij deze LT situatie behoort feed-back over goed- vs fout-beantwoording en een eventuele nieuwe behandeling van een leselement wanneer velen bij dit element bij de toets falen. Dit is bijv. bij vr. 42 en 44 het geval geweest.

Een maand later krijgt men dezelfde vragen weer, nu in de veel ruimere KT context. Wanneer de vraag dan door meer cursisten juist wordt beantwoord dan in de LT situatie van een maand tevoren, dan zou men van een leereffect kunnen spreken.

Ook bij andere experimenten kwamen wij regelmatig verbeteringen in de resultaten tegen, soms wel van 50 % of meer. Meestal betreft het dan items met een relatief hoge  $r$  en een  $p$ -waarde tussen .30 en .60.

Bij de items 20 en 45 trad geen verschil op en bij item 46 een verslechtering. Dit vindt vermoedelijk zijn oorzaak in een wat hoge  $p$  van deze items. Wellicht zullen deze alsnog worden herzien.

## 8. TOETSEN ALS EINDEXAMEN

Alvorens er toe over te gaan voor een bepaalde soort cursussen de klassieke manier van examineren deels te vervangen door toetsen, werden als proef 70 cursisten van de voortgezette avondopleiding machinebankwerker bij het eindexamen op de volgende manieren beoordeeld.

### 1. *Schriftelijk theorie-examen*

Er waren 10 vragen en opgaven. Daarvan moesten in  $1\frac{1}{2}$  uur 8 stuks (naar eigen keuze) worden gemaakt.

### 2. *Mondeling examen*

Gedurende een half uur werd door 2 examinatoren mondeling gevraagd over vaktechnische zaken, rekening houdend met de ervaring van de examinandus in het bedrijf. De examinatoren gaven onafhankelijk van elkaar hun cijfers.

### 3. *Toets*

Er werden 96 vragen gesteld, in  $1\frac{1}{2}$  uur te beantwoorden (oorspronkelijk waren er 100 vragen, doch 4 moesten worden afgekeurd wegens onduidelijkheden in de vraagstelling).

### *Resultaten:*

Op grond van de overeenstemming tussen de behaalde cijfers bij de combinatie mondeling en schriftelijk examen enerzijds en de toetsre-

sultaten anderzijds, werd besloten in 1967 het schriftelijk en mondeling examen voor enkele onderdelen van bepaalde avondcursussen te laten vervallen en te volstaan met een toets.

Voorlopig wordt (als overgangsmaatregel) kandidaten met toetsresultaten tussen vijf en zes, de mogelijkheid gegeven zich via een mondeling (her)examen opnieuw kansen te scheppen. Gemeend wordt hiermede zowel de objectiviteit en waarde van de toets als de belangen van de cursisten recht te doen wedervaren.

Of met herexamens bij twijfelgevallen zal worden doorgegaan is de vraag. Tot nu toe zakten bij het mondelinge herexamen alle cursisten die bij de toets een score tussen 5 en 6 behaalden.

#### *Bijkomend voordeel van de toetsen als examen:*

Leraren en bedrijfsfunctionarissen worden voor het afnemen van examens aan hun eigenlijke taak (= lesgeven) onttrokken en het betekent voor velen van hen een zware extra belasting. Voor het afnemen van een toets kan met een ondeskundige surveillant worden volstaan, terwijl de berekening van het examencijfer op kantoor kan gebeuren.

Wat men bespaart op het tijdrovende examineren kan beter aan lesgeven worden besteed.

#### *Constructie van toetsen als examen*

Toetsen t.b.v. een examen worden op dezelfde wijze samengesteld als kerntoetsen, het enige verschil is de omvang. Het is n.l. de bedoeling dat zij in het algemeen uit 100 items zullen bestaan, die met elkaar de gehele examenstof dekken.

### 9. DE CIJFERWAARDERING VAN TOETSEN

Bij de LT worden geen cijfers toegekend. Volstaan wordt met het in de klas bekend maken van de bij het maken van de toets naar voren gekomen knelpunten: waar is de gewenste kennis niet of in onvoldoende mate overgekomen.

Bij de KT zou in principe eveneens kunnen worden volstaan met bekendmaking van de totaalscores per cursist. In de praktijk blijken hieraan enkele bezwaren te kleven. Cursisten — en vooral hun ouders en relaties — weten niet goed raad met dit soort van scores. Als het om proefwerken en examens gaat, blijkt men zeer gehecht te zijn aan schoolcijfers. Een ernstiger bezwaar is het volgende. Voor een aantal vakken geldt nog steeds de klassieke proefwerkmethode en de daaraan verbonden waarderingscijfers.

Een vergelijking van deze schoolcijfers met KT-scores is voor niet-ingewijden bijzonder moeilijk. Het is daarom gewenst de KT-scores te

transformeren, zodat zij voor de „gebruikers” iets van schoolcijfers weg hebben.

Verder is een omzetting gemakkelijk om de scores van uiteenlopende soorten toetsen met elkaar te kunnen vergelijken. Overigens klemt deze noodzaak niet alleen wanneer men met objectief scorebare toetsen gaat werken. In feite zou deze transformatie voor alle door scholen geproduceerde cijfers nodig zijn. Wanneer gemiddelden en standaardafwijkingen van schoolcijfers per leraar verschillen — en men vindt dit overal, zie o.a. De Groot, lit. 6 en 7 — dan zijn deze cijfers onvergelijkbaar. Men zou ze pas naast elkaar mogen zetten na ze statistisch getransformeerd te hebben. Maar waar gebeurt dit? Bovendien zal het niet steeds duidelijk zijn hoe die vergelijking tot stand moet komen.

Aan de gewenst geachte transformatie zitten verschillende problemen vast. Bijv: Moet de kans op toevalstreffers in rekening worden gebracht? Sommige auteurs doen dit door bijv. bij toetsen, bestaande uit items met 4 antwoordmogelijkheden, 25 % van het totaal aantal mogelijke goede antwoorden af te trekken. Anderen bestrijden de noodzaak van deze correctie.

Een tweede vraag is, hoe worden (al dan niet op toeval gecorrigeerde) totaal scores omgezet in punten of „schoolcijfers”?

Sommigen zetten de scores rechtlijnig om in een puntenwaardering. Zijn er bijv. 100 vragen, dan levert 0-9 goed een 1 op, 50-59 goed een 6, enz.

Deze rechtlijnige omzetting van scores in punten wordt door sommige auteurs onjuist geacht. Anderen propageren kromlijnige transformaties.

Seashore geeft in lit. 12 (zie ook lit. 3, pag. 276 e.v.) een groot aantal methoden voor de transformatie van scores, zoals deze bij tests zijn behaald.

Deze in eerste instantie verkregen scores zijn de zogenaamde ruwe scores. Hij noemt als eenvoudigste methode de vaststelling van percentielen en de omzetting van de ruwe score van ieder individu in een equivalente percentielscore. Men telt het aantal personen dat beneden elke ruwe score blijft, telt daar de helft bij van hen die precies de betreffende score hebben en deelt deze som door het totaal aantal personen in de groep.

Men kan dan zeggen: het resultaat van die en die leerling ligt op het n-de percentiel van deze groep.

Zelf werken wij met z-standaard scores. De keuze werd op deze methode bepaald omdat gemiddelde en standaardafwijking toch moeten

worden berekend en de z-score dan niet veel extra rekenwerk meer vraagt. Bovendien biedt deze methode de mogelijkheid een exacte vergelijking te maken met andere gegevens, wanneer daarvan eveneens z-scores worden bepaald.

Als basis voor de berekening van z-scores dient de standaarddeviatie van de ruwe scores,

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$

waarin

$\bar{x}$  = het gemiddelde van de ruwe scores

$x$  = de ruwe scores

$n$  = het aantal scores.

(Voor het geval  $n$  niet zeer groot is, wordt door sommige auteurs aanbevolen voor de noemer  $n - 1$  te nemen).

Verder volgen wij de vereenvoudigde methode voor de berekening van z-scores van Helen A. Heath (lit. 10), met als resultaat constanten voor gemiddelde en standaardafwijking, bijv. resp.  $M = 50$  en  $s_1 = 10$ .

$$z_1 = \frac{s_1 (x_1 - \bar{x})}{s} + M$$

$$z_2 = z_1 + \frac{s_1}{9} (x_2 - x_1)$$

$s_1$  = de aangenomen constante standaardafwijking

$M$  = het aangenomen constante gemiddelde

$s$  = standaardafwijking van de ruwe scores.

Heath berekent  $z_1$  voor de eerste ruwe score uit de reeks scores die men heeft verkregen en zet alle volgende ruwe scores om in z-scores met behulp van  $z_2$ .

Wij hebben deze methode nog enigszins verder gesimplificeerd. Uitgaande van  $M = 6$  en  $s = 1$  (empirisch bepaald) wordt  $z$  voor de hoogste en de laagste ruwe score berekend. Deze worden op millimeterpapier uitgezet. Voor de overige ruwe scores kan  $z$  worden afgelezen.

Deze transformatie wordt intussen niet als een einddoel gezien. Gestreefd wordt naar de bepaling van geslaagd of gezakt bij een examen in een vroeg stadium van de opleiding. Dit soort van beslissingen zou bijv. genomen kunnen worden bij het eerste kerstrapport. Correlatieberekeningen toonden nml aan, dat dit kerstrapport een bruikbaar krite-

rium is voor het voorspellen van slagen en falen bij de opleidingen voor het leerlingstelsel. Wanneer door verder experimenteren de p en r waarden van de items zijn verbeterd, stellen wij ons voor om voor het nemen van beslissingen over te gaan op de caesurbepaling volgens Prof. Dr. A. D. de Groot (lit. 6b). Dit kan echter pas gebeuren wanneer een examen kan worden opgesteld, dat alle cursusonderdelen op een verantwoorde wijze omvat.

Het zou verbetering van de methodiek en efficiency betekenen, wanneer een objectief geconstrueerd „Kerstrapport” tot het nemen van beslissingen zou kunnen voeren. Daarna zou alle aandacht voornamelijk op het geven van onderwijs kunnen worden gericht en zou selectie in het opleidingsproces geen plaats van betekenis meer mogen innemen. Voor dit verdere onderwijs zouden de LT's en KT's uitsluitend als hulpmiddel bij het leerproces worden gebruikt en KT niet meer als prestatietesting voor het nemen van beslissingen. (Zie lit. 6 en 7).

## 10. STAND VAN ZAKEN

### 10.1 Omvang

Momenteel wordt met een groep van 9 leraren aan en met toetsen gewerkt. Er zijn  $\pm 100$  lesstoftoetsen in gebruik (totaal  $\pm 1000$  items). Enkele honderden items zijn in bewerking (1 of 2  $\times$  geprobeerd). Het betreft hier op mechanisch gebied onderwerpen zoals: beitels, boren, zagen, vijlen, ruimen, lagers, koppelingen, riemen en snaren, pompen, afsluiters, appendages, hefbomen, hoogovenproces, aftekenen van werkstukken, kettingen, meten en meetgereedschap, micrometers, neutrale lijnberekening, passingstelsels, tandwielen- en kasten, tekenen.

Elektrotechnische onderwerpen: gereedschapsleer voor elektriciens, installatiematerialen, vaktheorie, kabels leggen en afwerken, beveiligingen, transformatoren, gelijkstroommachines, draaistroom.

Het aantal lesonderwerpen en items is inmiddels voldoende groot om zonder veel moeite enkele tientallen kerntoetsen samen te stellen. Per maand worden gemiddeld 2 per leerling verschillende soorten lesstoftoetsen en 1 kerntoets afgenomen. Een LT heeft 10-40 items, een KT 40-50. Aan LT werken dan ongeveer 80 cursisten.

### 10.2 Persoonlijke ervaringen

De leraren, die met objectieve toetsen als hulpmiddel werken, zijn ingenomen met dit middel. Ook in de gevallen waarbij een eerste confrontatie met de uitkomsten „schokkend” was, in de zin van veel meer fouten dan men verwacht had. In enkele gevallen moeten lessen volledig worden herhaald.



### 10.3 Toetsen — voor- en nadelen

#### Voordelen:

- de prestatie­meting is betrouwbaar;
- het afnemen kost leerlingen en leraar weinig tijd, dit betekent efficiënter benutten van de lestijden;
- scoring en waardering in cijfers kost de leraar geen tijd, want dit gebeurt op „kantoor”;
- leerling en leraar kunnen in enkele minuten een inzicht krijgen in wat wel en niet van de gepresenteerde leerstof „is overgekomen”. Dit kan het leerproces versnellen;
- prestaties zijn objectief te scoren;
- prestaties van groepen (klassen, leerling­lichtingen, enz.) worden objectief vergelijkbaar;
- met objectief scorebare toetsen met een empirisch-statistische basis kan het verband met externe criteria gemakkelijker worden onderzocht, dan mogelijk is met proefwerken en dergelijke;
- de methode dwingt meer dan de klassieke proefwerken tot grondige bezinning op de leerstof (zie par. 2).

#### Bezwaar:

De constructie van de toetsen vraagt voorbereidingstijd en vindingrijkheid van:

- de leraren;
- statistisch geschoolde kracht;
- het team dat de toetsen vorm geeft („technicus”, statisticus, neerlandicus).

De in de toetsconstructie gestoken tijd wordt zinvoller, naarmate het gebruik van de toets frequenter wordt.

Vergeleken met de klassieke wijze van proefwerken, mag de vele moeite niet als een specifiek voor toetsen geldend nadeel worden gezien. Want proefwerken zouden, teneinde evenals de toetsen betrouwbaar en objectief te zijn, eveneens empirisch-statistisch bewerkt moeten worden. Wat hier als bezwaar wordt genoemd, is niet meer dan een algemene moeilijkheid optredend bij elke verantwoorde fundering van prestatie­metingen: het vele werk.

### 10.4 Wat nog te doen staat

Naast uitbreiding van de objectieve toetsmethode tot al de lesonderwerpen die zich voor deze werkwijze lenen, staat de vaststelling van de kwalificatie geslaagd-gezakt op uitwerking te wachten. De vraag is nml of een gediplomeerd cursist ook een goed praktijkman is. Maar wil men examencijfers correleren met de praktijk, dan zal men voor dat slagen

of falen in de praktijk over een maatstok moeten beschikken. Naar zo'n praktijkkriterium moet worden gezocht, want personeelsbeoordelingen zijn voor ons doel te onbetrouwbaar (zie lit. 14).

Het vinden van een ander criterium zal niet meevallen (zie lit. 14 en lit. 3., p. 210 e.v.). Gedacht wordt aan een vergelijking met resultaten bij voortgezette opleidingen, snelheid van promotie of salaristoename, analyse van werkkaarten, zelfbeoordelingen en beoordelingen door klasgenoten of collega's.

Op den duur zal ook de samenstelling van lesstof op een meer empirische grondslag moeten plaatsvinden dan nu het geval is. Voorts zal moeten worden nagegaan welk type items (juist-onjuist, open-ended, 3, 4 of 5 afleiders) het beste past bij een bepaald soort lesstof.

Dit artikel pretendeert niet meer dan in grote lijnen aangeven wat op het gebied van lesstoftoetsen bij het opleidingscentrum Hoogovens gaande is en wat de werkwijze is. Daarbij wordt de hoop gekoesterd te kunnen komen tot een uitwisseling van gegevens en ervaringen met andere scholen en/of bedrijfsopleidingen.

#### LITERATUUR

1. ADKINS, D.: Construction and analysis of achievement tests. H. Maj. Stat. London.
2. BRADLEY, J. I. and MC. CLELLAND, J. N.: Basic Statistical Concepts: A Self-Instructional Text. Chicago: Scott, Foresman, 1963). (Geeft statistische Grondslagen voor tests).
3. DRENTH, DR. P. J. D.: De psychologische test. Van Loghum Slaterus/Arnhem 1966.
4. Educational Testing Service: Building a test. E.T.S., Princeton.
5. GROOT, PROF. DR. A. D. DE: Methodologie. Mouton & Co., Den Haag 1961.
- 6a. GROOT, PROF. DR. A. D. DE: Waaraan voldoet een „onvoldoen” prestatie niet? Paed. Studiën 41e jaargang, no. 9, pag. 394-409.
- 6b. GROOT, PROF. DR. A. D. DE: De Kern-item-methode voor de bepaling van de caesuur voldoende/onvoldoende. Paed. Studiën 41e jaargang, no. 10, pag. 425-440.
7. GROOT, PROF. DR. A. D. DE: Vijven en zessen. Wolters, Groningen.
8. GUILFORD, J. P.: Fundamental Statistics in Psychology and Education. Mc. Graw-Hill Book Comp. Inc. New York, 1956.
9. GUILFORD, J. P.: Psychometric Methods. Mc. Graw-Hill Book Comp. Inc. New York.
10. HEATH, HELEN A: Educ. and Psych. Measurement, Vol. 25, no 2., 1965, pag. 323.
11. LINDQUIST, E. F. (editor): Educational Measurement. American Council on Education. Washington 1955.

12. SEASHORE, H. G.: Methods of expressing test scores. Psychol. Corp. Test Service Bull, no. 48, 1955.
13. SPITZ, J. C.: Statistiek voor Psychologen, Pedagogen, Sociologen. N.V. Noord-Hollandsche Uitgeversmaatschappij, Amsterdam, 1965.
14. WOLFF, DR. C. J. DE: Personeelsbeoordeling. Swets & Zeitlinger, Amsterdam 1963.
15. NAERSSSEN, DR. R. F. VAN en BEAUMONT, R. C. VAN: De coëfficiënt van Long als een benadering van de biseriële  $r$  in de itemanalyse. Ned. Tijdschrift voor de Psych. Mei '66, p. 308.

### *Curriculum Vitae*

H. F. Mulder, MO-Pedagogiek A en B, Doctoraal examen Sociale Psychologie en Soc. Pedagogiek. Psycholoog bij Hoogovens, IJmuiden, o.a. adviserend t.b.v. opleidingen bij Hoogovens.

L. v. Genderen, NO-akten Nj en NI, assistent-chef technische opleidingen bij Hoogovens, IJmuiden.