

Het toetsen van 'Algemene Kennis' op de basisschool

S. SANDBERGEN*

Samenvatting

'Algemene Kennis' (A.K.) is geen concreet vakgebied waaruit 'vanzelf' toetsdoelen resulteren. De onderwijsdoelstellingen bij A.K. (of onderwijs in de zaakvakken) zijn slechts vaag geformuleerd.

Het belang van dit gebied wordt echter door weinig lesgevers aangevochten. Daarom moeten toetsconstructeurs noodgedwongen zelf toetsdoelen formuleren als helder geformuleerde leerplannen ontbreken, hoewel dat strikt genomen hun werk niet is.

Bij de beschrijving en de bespreking van de gevolgde methode van doelstellinganalyse op dit weerbarstige gebied komt o.a. naar voren dat:

1. professioneel gemaakte 'objectieve studietoetsen' in sommige opzichten zeer subjectief worden samengesteld;
2. een systematische beoordeling door onderwijzers en ouders en deskundigen van de wenselijkheid van toetsonderdelen zeer nuttig is;
3. het beoordelen van de 'bereikbaarheid' van onderwijsdoelen voor een groot deel langs empirische weg kan gebeuren en dat de meningen van individuele beoordelaars hierbij van betekenisvolle waarde zijn;
4. bij de constructie van toetsen met behulp van overleg tussen gebruikers en samenstellers recht kan worden gedaan aan de vernieuwingen van het leerplan;

* De schrijver van dit artikel is werkzaam op het Research Instituut voor de Toegepaste Psychologie aan de Universiteit van Amsterdam. A's 'Psychometrisch adviseur' heeft hij sinds 1969 op het Centraal Instituut voor Toetsontwikkeling meegewerkt aan de samenstelling van subtoetsen 'Algemene Kennis'.

5. de potentiële waarde voor evaluatie en diagnose van studietoetsen in het basisonderwijs zeer groot is.

1. Inhoud

In dit verslag worden enkele ervaringen beschreven** die werden opgedaan bij het samenstellen van de subtoetsen 'Algemene Kennis' van de schooltoetsen basisonderwijs aan het C.I.T.O. te Arnhem.

'Algemene Kennis' is een weerbarstig terrein; veel minder dan bij 'Taal' of 'Rekenen' kan men hier spreken van één concreet vakgebied. A.K. op de basisschool is namelijk slechts bij uitsluiting te definiëren; ongeveer als 'Alles dat de wetenschappen en de cultuur opleveren, zonder taal en rekenen'.

De onderwijsdoelstellingen bij A.K.; (onderwijs in de zaakvakken of 'wereldverkenning') zijn bovendien nogal vaag geformuleerd.

Toch vindt bijna iedereen die met de basisschool te maken heeft, dit gebied belangrijk. Hieruit volgt dat een schooltoetsprogramma de toetsing van A.K. niet mag verwaarlozen.

** Bij het schrijven van dit artikel werd ik bijzonder gestimuleerd door het C.I.T.O. De intensieve discussies binnen de commissie Algemene Kennis hebben veel bijgedragen tot de ideeën die hier worden besproken. De onderwijzers-leden en de 'Wetenschappelijke' leden van het C.I.T.O. en het Kohnstamm Instituut worden hierbij zeer bedankt. Een eerste versie van dit verslag werd grondig bekritiseerd door J. I. van Deursen van het C.I.T.O. en dr. E. Warries van het R.I.T.P. De leesbaarheid werd daardoor sterk verhoogd!

Bij het ontbreken van helder geformuleerde, moderne leerplannen moeten toetsconstructeurs, – noodgedwongen – zelf toetsdoelen formuleren hoewel dat strikt genomen hun werk niet is.

De manier waarop deze formulering tot stand kwam en in items werd uitgewerkt, wordt beschreven en van commentaar voorzien. Daarna worden een aantal empirische gegevens besproken in verband met de 'bereikbaarheid' van onderwijsdoelen en in verband met de evaluatieve waarde van een jaarlijks schooltoetsprogramma.

Het verslag wordt besloten met een aantal konklusies.

2. *Waarom zou Algemene Kennis moeten worden getoetst?*

2.1 *Rationale voor toetsing*

Sinds 1966 (Samenwerkende Instuuten 1967) is in het programma van de schooltoetsen basisonderwijs het onderdeel 'Algemene Kennis' opgenomen, hoewel de prestaties op dit onderdeel tot nu toe (1971) niet werd 'meegeteld' in de percentielscore. De belangrijkste reden voor het opnemen van A.K. is dat de 'schooltoets' niet alleen als selectie-instrument moet worden gezien, maar bovendien de pretentie heeft om een afspiegeling te vormen van enkele einddoelen van het basisonderwijs en op die manier óók als evaluatiemiddel bruikbaar te zijn. (Warries 1971). Als de inhoud van de schooltoetsen basisonderwijs beperkt zou blijven tot de vakken 'Rekenen' en 'Taal' – populaire voorspellers overigens – dan zou men beschikken over een hulpmiddel om leerlingen te verwijzen naar de één of andere vorm van vervolgonderwijs. Maar in dat geval zouden door deze beperkte opzet géén uitspraken mogelijk zijn over andere belangrijke doelstellingen van het basisonderwijs. Het is dus belangrijk om de toetsen zo breed mogelijk op te zetten.

En 'Algemene Vorming' geldt nu eenmaal als één van de fundamentele doelstellingen van het basisonderwijs. Onderwijs in de 'zaakvakken',

opgevat als cognitief deel van deze brede algemene vorming, wordt immers gegeven om bij de kinderen een aantal basisvaardigheden te introduceren, die 'belangrijk zijn voor het vervullen van taken en het oplossen van problemen van volwassenen in een ingewikkelde samenleving'. (Sandbergen 1970).

Voor het beheersen van dit soort vaardigheden, die men noodzakelijk, nuttig of wenselijk kan vinden, is het nodig dat leerlingen feitelijke kennis en begrip van eigenschappen van de volwassen wereld hebben.

Als men dus de pretentie van de schooltoets-afspiegeling van einddoelen van het basisonderwijs – wil waarmaken moet ook dit einddoel van het basisonderwijs worden opgenomen in de toets.

2.2 *Een enquête onder de gebruikers van de toets*

Dat 'Algemene Kennis' inderdaad belangrijk wordt gevonden blijkt 'officieel' uit het leerplan van veel basisscholen; er wordt in ieder geval les in gegeven. Dat het onderdeel A.K. ook als belangrijk wordt ervaren blijkt o.a. uit de resultaten van een C.I.T.O.-enquête 'inhoud schooltoets 1970' (Peters-Sips 1970).

Als onderdeel van deze enquête werd o.m. de volgende vraag gesteld:

'De diverse subtoetsen waaruit de schooltoets bestaat, representeren onderwijsvakken of programma-onderdelen. Graag willen wij uw mening horen over deze vakken of programma-onderdelen op zich, afgezien van de wijze waarop wij ze in de toets verwerkt hebben.'

D.w.z. wij willen weten hoe belangrijk u het vindt dat dit onderdeel voorkomt op het programma van de basisschool. Deze mate van belangrijkheid die u aan de diverse onderdelen toekent, kunt u tot uitdrukking brengen door een kruisje te zetten in de kolom (voor ieder onderdeel apart) die loopt van zeer belangrijk tot zeer onbelangrijk.

Dus bijvoorbeeld zo:

Algemene Kennis (d.w.z. die kennis die traditioneel gezien onderwezen wordt in de zaakvakken en ook omschreven kan worden als kennis nodig voor de wereldoriëntatie).'

| | |
|--|-------------------|
| | zeer belangrijk |
| | belangrijk |
| | onbelangrijk |
| | zeer onbelangrijk |

In Tabel 1 wordt de verdeling weergegeven van de antwoorden op deze vraag van 309 leerkrachten van zesde klassen.

Tabel 1. Oordeel van 309 leerkrachten van zesde klassen over het belang van 'Algemene Kennis' (naar Peters-Sips 1970).

| oordeel | aantal | percentage |
|-------------------|--------|------------|
| zeer belangrijk | 147 | 47% |
| belangrijk | 150 | 48% |
| onbelangrijk | 11 | 5% |
| zeer onbelangrijk | 1 | — |

De uitkomsten van deze vraag werden verder gebruikt om alle onderdelen van de schooltoets in volgorde van belangrijkheid te plaatsen.

Deze bleek als volgt:

1. Begrijpend lezen van korte teksten.
2. Gemengde taalopgaven.
3. Begrijpend lezen van een lange tekst.
4. Algemene kennis.
5. Rekenen.
6. Spelling.
7. Woordbenoeming/Zinsontleding.

(In de oorspronkelijke enquête werd nog verschil gemaakt tussen 'deelnemers' aan het schooltoetsprogramma en niet-deelnemers. In de bovenstaande lijst worden alleen de resultaten van de deelnemers weergegeven. De rangorde was anders voor de (80) niet-deelnemers waar Algemene Kennis op de tweede plaats kwam).

Conclusies:

1. 'Algemene Kennis' wordt door 95% van de respondenten op een enquête een zeer belangrijke of belangrijke doelstelling voor de basisschool gevonden.

2. Ondanks het feit dat het onderdeel 'Algemene Kennis' niet wordt gebruikt in de advisering aan het voortgezet onderwijs, wordt de subtoets A.K., vergeleken met andere onderdelen, relatief belangrijk gevonden door de gebruikers van de toets.

3. *Wat gaat vooraf aan de samenstelling van een toets?*

Een omschrijving van 'Algemene Vorming' als: 'het verwerven van basisvaardigheden die belangrijk zijn voor het vervullen van opgaven en het oplossen van problemen in de maatschappij', is net zo pretentius als vaag.

Pretentius: omdat het gebied dat A.K. bestrijkt, zo groot is dat slechts een beperkt aantal gebieden kunnen worden 'aangeraakt'.

Vaag: omdat deze omschrijving nog geen antwoord geeft op twee algemene vragen die traditioneel zijn bij de toetsconstructie:

1. Welke kennis, vaardigheden en inzichten dient de toets te meten (wat moet een leerling met die leerstof kunnen doen).
2. Welk gebied dient de toets te bestrijken (welke leerstof).

Zulke vragen moeten beantwoord zijn vóórdat men aan de samenstelling van een toets kan beginnen. (Bij het lesgeven gelden deze vragen óók, maar bij het maken van een toets wordt men nog concreter voor de keuze gesteld).

Nu bestaat in Nederland bij geen enkel 'vak' duidelijk overeenstemming over deze vragen. Er wordt op veel verschillende plaatsen en op veel verschillende manieren geëxperimenteerd met leervormen en leerstofinhouden, maar één algemeen aanvaarde en vooral concrete stofomschrijving die in principe geldt voor iedere leerling, ontbreekt.*

* Enkele jaren geleden werd 'Een proeve van een leerplan' gepubliceerd door het Nutseminarium voor pedagogiek aan de Universiteit van Amsterdam. Naast het respect dat een dergelijk uitgebreid werk afdwingt, kan men enkele bezwaren hebben tegen dit leerplan:

1. Uitgegaan werd van een traditionele onderwijs-

De discussie over de cognitieve doelstellingen van 'Algemene Vorming' is levendig maar dikwijls wordt in zulke algemene termen gesproken dat het niet duidelijk wordt welke specifieke vaardigheden en kennis de leerlingen op de basisschool moeten bereiken, voordat zij aan een onderwijskundige norm voldoen. Nu moet opnieuw met nadruk worden gezegd dat het leveren van een dergelijke concrete doelstellingsanalyse niet in de eerste plaats het werk is van toetsconstructeurs, maar van degenen die lesgeven (Mellenbergh, Sandbergen, Vastenhouw en De Zeeuw 1968).

Toch begonnen de toetsmakers aan dit werk omdat zij, bij gebrek aan een 'gebruiksklare lijst van toetsbare onderwijsdoelen' niet zonder meer met hun werk konden beginnen.

Dit is een direkt gevolg van twee belangrijke (algemene) uitgangspunten bij het toetsen:

I. *Bij het toetsen van onderwijsresultaten wordt er van uitgegaan dat 'leren' leidt tot verandering in het gedrag van de leerling.*

Als een onderwijsdoel geheel of gedeeltelijk bereikt is, komt dit tot uiting in een veranderd gedrag. Leerlingen die 'iets' geleerd hebben (bv. de tafels van 1-10 opzeggen of zinvolle conclusies trekken uit de krant) zullen zich dus in bepaalde situaties anders gedragen dan leerlingen die datzelfde 'iets' niet geleerd hebben. Leerresultaten zijn niet *direct* meetbaar (men kan wel constateren of iemand mentaal actief of inactief is, maar niet zien wat die activiteit inhoudt). Het is dus nodig om resultaten van leren te meten in termen van gedrag.

II. *Als men onderwijsresultaten wil toetsen, moet men de leer- en leeromstandigheden beschrijven.*

leersituatie, gestructureerd in termen van 'schoolvakken'.

2. Uit het werk wordt niet duidelijk welke uitgangspunten ten grondslag lagen aan dit leerplan.

3. Uitspraken ontbreken waaruit men kan afleiden welke leerlingen een leerstofonderdeel beheersen en welke *niet*. De leerdoelen zijn niet operationaal gedefinieerd in termen van leerdoelen.

ten deze zijn gesteld in termen van 'het oplossingsgedrag' dat nodig is voor de uitvoering van een opdracht of voor het juiste antwoord op een vraag.

Bij het maken van toetsen nu wordt verondersteld dat in een bepaald vak of onderdeel daarvan 'adekwaat oplossingsgedrag' bestaat en tijdens de vormgeving van de toets wordt voortdurend geprobeerd om gevallen te vinden waarbij de leerling die een bepaalde vaardigheid *beheerst*, het juiste antwoord wel kan geven en iemand die die vaardigheid *niet beheerst*, het juiste antwoord niet kan vinden. ('Wat moet je kunnen bij dit soort opgaven'). Concluderend:

Zo lang men geen antwoord kan geven op de twee bovengenoemde vragen, - welk gebied en welke vaardigheden - is het dus onmogelijk om een toets te maken die qua inhoud en moeilijkheid afgestemd is op het gegeven onderwijs.

4. *Toetsdoelen worden niet in het wild gevonden.*

Het probleem waarmee we in de vorige paragraaf bleven zitten, geldt voor veel vakken - wie precies weet wat bij voorbeeld productief-schriftelijke taalbeheersing is, mag het zeggen.

Bij A.K. is het probleem nog urgenter door de paradoxale situatie dat lesgevers enerzijds overtuigd zijn van het belang van bepaalde vaardigheden in het kader van een 'brede vorming' maar anderzijds minder dan bij andere vakken geneigd zijn om die vaardigheden zo concreet te beschrijven van dat zij 'vertaalbaar' zijn in toetsvragen.

Deze 'zwevende didactische toestand' is misschien voor een deel te verklaren door dezelfde drie moeilijkheden waarop men speciaal stuit als men 'A.K.' wil gaan *toetsen*:

a. Het gebied dat A.K. omvat is zo groot en zo heterogeen dat met een jaarlijkse toets slechts een zeer klein deel daarvan kan worden gedekt. Nu is de definitie van dit gebied in theorie mogelijk door een opsomming van onderwerpen, begrippen en kenniseenheden. Het is tenslotte ook mogelijk om een encyclopedie te schrijven. Maar een dergelijke benadering is

niet praktisch. Een samenstellende commissie moet dus een keuze maken, waarbij zij zich uiteraard laat beïnvloeden door persoonlijke voorkeuren van de leden, de beperkte ontwikkeling van de leden enz.

Zelfs als er een lijst van gebieden of onderwerpen zou bestaan waar iedereen vrede mee kan hebben, is zo'n opsomming van de te toetsen leerstof nog maar het begin van een oplossing.

- b. Een goede afbakening van leerstof geeft namelijk geen uitsluitsel over 'belangrijke vaardigheden binnen deze gebieden'. Als men geen keuze heeft gemaakt voor sommige vaardigheden (het oplossingsgedrag van sub II in 3) kan men met de vorm en de eigenschappen van de opgenomen vragen ook nog alle kanten uit.

Een voorbeeld:

In de toets A.K. II (1970) kwam de vraag voor:

'Hoe ontstaat wind?'

A door verschil in bewolking

(12% v/d antwoorden)

B door verschil in luchtdruk

(73% v/d antwoorden)

C door verschil in neerslag

(8% v/d antwoorden)

D door verschil in vochtigheid

(8% v/d antwoorden)

De conclusie kan zijn dat deze vraag naar kennis van de oorzaak van wind vrij goed beantwoord werd.

Deze vraag werd zo gesteld dat men uitsluitend met *verbale kennis* een juist antwoord kan geven.

Een andere vraag uit dezelfde toets, nu naar het verband tussen 'weer' en 'luchtdruk', was anders opgezet.* Er werden vier getekende barometers aangeboden. De vraag luidde:

* Kennis van barometers en van de relatie tussen 'weer en luchtdruk' is misschien wel voldoende. Waar het hier om gaat is dat er hier iets met die kennis gedaan moest worden.

Welke van deze barometers geeft de hoogste luchtdruk aan?

A (Barometer met de wijzer naar 'regen-veranderlijk') 8%

B (Barometer met de wijzer naar 'regen') 19%

C (Barometer met de wijzer naar 'mooi') 54%

D (Barometer met de wijzer naar 'veranderlijk') 21%

Bij deze tweede vraag was geprobeerd om een eenvoudige toepassingsvraag te maken waarbij de kinderen het verband tussen 'luchtdruk' en 'mooi weer' moesten gebruiken. Alléén kennis van het verbalisme 'hoge luchtdruk betekent mooi weer' was hier niet voldoende. Opvallend is hierbij overigens dat bij de eerste vraag veel meer kinderen het juiste antwoord geven dan bij de tweede vraag. Dat kan er op wijzen dat een aantal kinderen een juist antwoord kennen zonder dat zij begrijpen waarover zij praten. Eveneens opvallend is dat meer onderwijzers het eerste item 'gewenst' vonden (85%) dan het tweede item (74%).

Men kan dus binnen hetzelfde onderwerp, verschillende vaardigheden toetsen. De keus die men doet heeft invloed op de resultaten van de toetsing en op het soort vragen die men op grond van de resultaten kan beantwoorden.

- c. Een derde moeilijkheid die opgelost moet worden ligt in het feit dat de vragen 'vaardigheden' proberen te meten die speciaal belangrijk lijken voor A.K., maar dat deze vaardigheden vaak samengesteld zijn uit andere vaardigheden die men stuk voor stuk zou kunnen toeschrijven aan andere 'vakken'. Dit kan aanleiding geven tot moeilijkheden bij de gebiedsafbakening t.o.v. andere onderdelen van het toetsprogramma.

Voorbeeld:

'Om kwart over zeven 's morgens is een bank beroofd en men wil de daders zo snel mogelijk vinden. De politie weet hoe zij eruit zien, wat kan de politie het beste doen om ze gauw te vinden?'

A Een beschrijving van de daders met foto in de krant zetten (32%)

B Hun beschrijving in de uitzending van de radio nieuwsdienst van 8.00 uur 's morgens bekend maken (50%)

C Hun beschrijving in het televisie Journaal van 8.00 uur 's avonds bekend maken (11%)

D Hun beschrijving in een extra televisie uitzending van 10.00 uur 's morgens bekend maken (8%)

Deze vraag probeert te toetsen of kinderen iets van media als krant, radio of televisie weten en iets van die kennis kunnen toepassen in een concrete situatie. Maar bij zo'n vraag speelt 'goed lezen wat er staat' duidelijk mee. Het is dus mogelijk dat kinderen die wel het goede antwoord zouden kunnen bedenken, toch deze vraag missen, omdat zij geen rekening houden met gegevens die zij moeten lezen. Bij andere vragen komen weer impliciet of onbedoeld vaardigheden aan de orde als 'met getallen omgaan enz. enz.

De moeilijkheid om specifiek op één vaardigheid gerichte vragen te stellen houdt verband met een principe dat in 5.1. onder d verder wordt uitgewerkt.

De drie 'speciale moeilijkheden' bij A.K. nl. a) de heterogeniteit en de omvang van het gebied, b) de relevantie van het oplossingsgedrag en c) de specificiteit van het oplossingsgedrag, leiden tot de conclusie dat men op de één of andere manier tot een keuze moet komen bij het omzetten van onderwijsdoelen in een toets. In de volgende paragraaf wordt de gevolgde methode bij deze keuze toegelicht.

5. Hoe kan men 'toetsdoelen' vaststellen

Bij het tot stand komen van de subtoetsen A.K. in 1970 en in 1971, waren enkele theoretische principes belangrijk.

Deze uitgangspunten werden geformuleerd vóórdat werd begonnen aan een eerste beschrijving van het te toetsen gebied en vóórdat aan de eigenlijke constructie werd begonnen.

Zij worden hier opzettelijk uitgebreid weergegeven omdat zij de verdere gang van zaken sterk hebben beïnvloed en (dat is misschien boeiender)

omdat de uitgangspunten aanleiding kunnen zijn voor verschil van mening.

5.1. *Uitgangspunten bij de samenstelling van de toets.*

a. *De toets is uitsluitend bedoeld voor het meten van cognitieve onderwijsdoelen.*

Toelichting:

De toets vraagt naar feiten, begrippen en toepassingen en niet naar meningen, attitudes of overtuigingen van de leerlingen. Deze gerichtheid geldt natuurlijk voor elke studietoets, maar wordt hier extra benadrukt omdat bij sommige onderdelen de verleiding groot is voor de vragensteller om zijn visie op de mensen of op de samenleving als objectief juist voor te stellen.

In zo'n geval meten de vragen in hoeverre leerlingen dezelfde mening hebben als hij (of net doen alsof - 'social desirability' of 'test-wisness'). Volgens onze opvatting is het bij voorbeeld wel geoorloofd om te vragen naar kenmerken van een Westerse democratie, maar valt het buiten het kader van de toets om aan leerlingen te vragen of zij die - of een andere - staatsvorm de meest wenselijke vinden.

b. *De toets representeert communale doelstellingen.*

Toelichting:

Alle leerlingen die het basisonderwijs verlaten, zouden een aantal 'cognitieve vaardigheden' moeten beheersen waarmee zij:

1e. zich kunnen oriënteren in een moderne samenleving.

2e. in staat zijn om zich zelf op eigen kracht verder te ontwikkelen of dat te doen in het voortgezet onderwijs.

'Communale doelstelling' of 'minimum eis' in deze betekenis is een bijzonder eerbiedwaardig maar onhanteerbaar begrip. Het woord 'eis' houdt bij voorbeeld in dat 'het zonder die vaardigheid niet gaat'. Maar nuchter bezien

zijn er maar enkele vaardigheden die beslist nodig zijn opdat iemand zich kan oriënteren en zich verder kan scholen (bij voorbeeld lezen en schrijven). Bovendien is het voortgezet onderwijs wat betreft peil en lesrooster zo heterogeen ingericht dat het aanwijzen van bepaalde toetsdoelen vrijwel onmogelijk lijkt. Het lijkt daarom zinvoller om de einddoelen van het basisonderwijs als uitgangspunt te nemen.

Een consequentie hiervan is dat de toets zich concentreert op doelen die op zichzelf nuttig gevonden worden – maar zie d) – en vooral op doelen waarvan in redelijkheid verwacht kan worden dat zij voor alle leerlingen in principe bereikbaar zijn.

In de praktijk komt het dan minstens er op aan dat men zich bij voorgestelde toetsdoelen afvraagt:

is dit onderwerp (of begrip of vaardigheid) belangrijk voor de basisschool en is het onderwerp van de vraag gangbaar op school.*

- c. *Onderwerpen voor de toets worden niet gekozen uit de traditionele indeling in 'zaakvakken'.*

Toelichting:

Men kan belangrijke vaardigheden óók leren als het lesrooster niet in 'vakken' is gesplitst. Het is vanzelfsprekend dat de verworven kennis en vaardigheid 'later' nu eenmaal toegepast moeten worden in situaties die maar weinig lijken op de 'schoolse' (m.a.w. de wijze van kennisverwerving beïnvloedt de wijze van kennistoepassing). Voor de praktijk van het onderwijs zelf zijn er belangrijke argumenten

* De empirische moeilijkheid van items alleen geeft geen uitsluitsel bij de beslissing of een item een communale doel of minimumeis representeert. Er bestaat zo'n grote verscheidenheid van onderwerpen waaraan men aandacht zou kunnen besteden, dat hantering van het criterium 'meer dan 80 of 90% van de leerlingen moet het antwoord kunnen geven' zou leiden tot een te drastische beperking van de onderwerpen en/of het stellen van vragen op een triviaal niveau.

zoals:

1. Een strak vasthouden aan een aantal 'schoolvakken' heeft als mogelijk effect dat onderwerpen die niet reeds gemeengoed zijn moeizaam ingang vinden in het basisonderwijs.

Zo kan het voorkomen dat onderwerpen als 'maatschappelijk inzicht' ('civic education') of wereldoriëntatie vaak alleen maar geforceerd kunnen worden ingepast in de gang van zaken in de klas.

Vele leerkrachten zijn nu eenmaal niet geneigd om zich buiten 'het boekje' te begeven, zodat schoolvakken in de traditionele zin verstarrend kunnen werken.

2. Een ander argument is dat men bij het werken met een specifiek vak de relatie tussen doel en middel uit het oog zou kunnen verliezen.

Een bepaald vak, of het nu grammatica-onderwijs is, kennis van computers, moderne wiskunde, of schaken *kan* – zo maar als voorbeeld – worden ingevoerd om de leerlingen 'te leren denken'. Het gaat er m.a.w. bij sommige vakken misschien niet primair om welke stof de onderwijzer aanbiedt maar welke algemene doelstelling of opvoedingsdoel hij nastreeft. Bij het denken in termen van specifieke vakken wordt deze relatie vaak omgekeerd zodat de stof centraal komt te staan en de oorspronkelijke doelstelling uit het centrum van de aandacht verdwijnt.

Zo kan het gebeuren dat leerlingen en leerkrachten zich inspannen om bepaalde toevallige leerstofinhouden te gaan beheersen, terwijl het nut van juist *die* leerstof wellicht problematisch is.*

* Dat wil niet zeggen dat men nu maar geen grammatica-onderwijs, computerkunde, moderne wiskunde of schaken moet leren. Het gaat erom dat de min of meer toevallige keuze van een bepaalde stof voor een bepaald doel tot onbedoelde resultaten kan leiden: wel kennis van verzamelingsleer maar niet beter denken in andere dan wiskundige situaties etc.

- d. De vragen van de toets worden zoveel mogelijk gericht op het aanbieden van reële problemen binnen een levensechte context.

Toelichting:

Een studietoets heeft een beperkt doel (Sandbergen 1968): het zo goed mogelijk *meten* van onderwijsresultaten. Leerstof op de basisschool wordt zoveel mogelijk aangeboden naar aanleiding van 'levensechte situaties', waarin 'echte problemen' aan de orde komen. Nu bestaat er geen logisch verband tussen de vorm van een didactische situatie en de vormgeving van een toetsingssituatie.

Toch menen de samenstellers dat ook het toetsen van fundamentele kennis of vaardigheden zoveel mogelijk moet gebeuren in een levensechte situatie, waarbij zoveel mogelijk 'kennis' wordt gevraagd die echt gebruikt wordt. Bij het maken van vragen wordt dus geprobeerd om vooral vaardigheden te meten die kunnen dienen om reële problemen op te lossen.

Voorbeeld:

Een populaire vraag van vroeger: 'als je met een schip de kortste weg van Maastricht naar Groningen neemt, door welke wateren kom je dan?' is voor de meeste Nederlanders van nu niet reëel meer ondanks het feit dat degene die deze waterwegen vlot kan noemen, heel wat van de topografie van Nederland afweet. Een situatie waar tegenwoordig meer mensen mee te maken krijgen (die dus levensechter is) is een volgende:

Het kan gebeuren dat een automobilist zonder wegenkaart onderweg is van Hilversum naar St. Michielsgestel. Als hij via Utrecht rijdt moet hij *weten* dat St. Michielsgestel in de buurt van Den Bosch ligt om bij Utrecht op de goede manier voor te sorteren.

Als hij dus naar die plaats moet en het A.N.W.B. bord geeft de naam van zijn bestemming niet aan, moet hij kunnen kiezen uit bij voorbeeld

A richting Arnhem.

B richting Breda.

C richting Den Bosch.

D richting Nijmegen.

5.2. Een ordening in gebieden.

De algemene theoretische principes a – d uit de vorige paragraaf gingen over de inhoud van toetsvragen.* Dit soort principes is van grote praktische waarde bij het beoordelen van vragen die door itemschrijvers worden voorgesteld. Ook wordt de hoeveelheid onderwerpen door het principe van 'minimumeisen' enigszins beperkt.

Maar hiermee is nog niet voldaan aan de vraag naar een leerstofomschrijving die de onderwerpen aangeeft waaruit vragen kunnen worden gesteld en die als leidraad kan dienen voor de hoeveelheid items die per onderscheiden onderwerp worden gemaakt (de weging van onderwerpen).

De volgende lijst van 'onderwerpen' die voor de toets van 1970 werd gebruikt, geeft een zekere ordening van gebieden.**

Bij de opstelling van deze lijst werd uitgegaan van het leven van mensen in een ontwikkelde (westerse) samenleving. Een dergelijke maatschappij heeft allerlei aspecten. Een aantal aspecten zoals 'fysisch', 'natuurwetenschappelijk', 'sociaal' enz. wordt opgevat als een verzameling van gebieden.

Lijst van onderwerpen:

fysisch:

energie
geologische verschijnselen
globe, topografie
meteorologische verschijnselen, klimaat
natuurlijke hulpbronnen
ruimte
waterhuishouding

* Te onderscheiden van formele eisen als 'specificiteit', 'efficiëntie', 'objectiviteit' (zie bijv. De Groot & Van Naerssen 1969).

** In afwijking van bijv. Themans (1968) werd hier dus niet nadrukkelijk gestreefd naar leerstof die a) overal behandeld werd en b) belangrijk wordt geacht voor het V.O.

natuurwetenschappelijk:

hygiëne, volksgezondheid
mensen, planten, dieren, natuurlijk evenwicht,
bestrijdingsmiddelen
natuurbescherming
techniek, natuurkundige modellen, toepassingen,
computers

sociaal:

arbeidsverdeling beroepen
bestuur, stadsgemeente, provincie
bevolkingsgroepen, beroepen
communicatiemedia
politieke partijen
sociale verzorging
staatsvormen
verkeer
wereldbevolking
wetgeving
migratie

economisch:

arbeidsverdeling
begroting
beroepen
geldcirculatie
middelen van bestaan
ontwikkelingshulp
wereldhandel
gastarbeiders

cultureel:

godsdienst
inwinnen en evalueren van informatie
kunst
recreatie
sport
wonen

historisch:

conflicten tussen volken
consequenties van vernieuwingen
contemporaine geschiedenis
ontdekkingsreizen
ontwikkeling van de Nederlandse staatsvorm
ontwikkeling van het verkeer

Bij deze lijst kan worden opgemerkt dat de keuze van juist deze categorieën en het aantal categorieën in zekere zin arbitrair is.

De keuze van deze termen werd mede bepaald door de overweging om géén opsomming

van schoolvakken te geven en bovendien om aan te geven dat het hier géén 'leerplan' betrof. (Dat het laatste niet overbodig was bleek uit enkele reacties van onderwijzers die de verantwoording van de inhoud van de schooltoets meer als 'voorschrift' zagen dan als een verantwoording waarop kritiek of aanvulling gevraagd werd).*

Wat betreft het aantal categorieën; door verhoging van de graad van abstractie kan men waarschijnlijk met minder categorieën toe, door de vaagheid iets te verminderen, zal het aantal categorieën toenemen.

Tenslotte: bij het zoeken naar categorieën en de vulling daarvan is niet gestreefd naar volledigheid. De bedoeling was voornamelijk om een werkljst te maken die in een later stadium van de constructie zo nodig aangevuld kon worden.

5.3. *De noodzaak tot concreetheid; bezwaren van de werkwijze.*

Een lijst van onderwerpen zoals uit 5.2., die in de verantwoording voor de toets in 1970 werd opgenomen, kan bij het samenstellen van een toets slechts als eerste uitgangspunt dienen. Dit bleek al spoedig tijdens de constructiefase van de toets. Het team stuitte op de concrete moeilijkheid dat er a) hooggestemde 'Algemene principes' over de inhoud van items bestonden maar b) slechts een zeer algemene gebiedsafbakening in termen van onderwerpen bestond.

Maar: Tijdens het construeren van items is het een essentiële bezigheid dat deze twee groepen overwegingen worden gecombineerd** en dat de uitkomst van dit proces uitsluitel geeft op de punten:

* Bij Algemene Kennis gaf $\pm 35\%$ van de respondenten van de C.I.T.O.-enquête suggesties voor opname van niet getoetste leerstof.

** Op het gebied van de 'Social sciences' bestaan zeker leerplannen (de 'Proeve' van het Nutsseminarium bijv.). Het bezwaar blijft dat deze leerplannen vaak zijn geformuleerd in termen van 'enige kennis van . . . , inzicht in . . . enz., en niet nadrukkelijk zijn gekoppeld aan categorieën van vaardigheden zoals de 'taxonomie' van Bloom om er maar één te noemen.

1. Welke vaardigheden worden getoetst
2. Op welke kennisgebieden worden deze vaardigheden getoetst.

Deze verdere analyse in termen van concrete toetsdoelen is een arbeidsintensieve zaak. Dat dit proces ook zeer subjectief is wordt duidelijk tijdens de bespreking van de constructiefase.

Dit bleek bij voorbeeld gedurende de constructie van de toets 1970. In dat jaar werd een tweetal subtoetsen afgeleverd die qua vakgebied verschillend waren gericht: één deelttoets waarin voornamelijk items van 'sociale aard' waren opgenomen en één deelttoets waarin veel natuurwetenschappelijke items waren opgenomen.

In de 'sociale toets' werden de vragen voor het grootste deel gesteld n.a.v. een thema nl.: 'het leven in een stad'; in de tweede toets werden de vragen genomen uit veel verschillende natuurkundige en biologische gebieden.

In de toets over het centrale-sociale thema kwamen 'als vanzelf' allerlei punten aan de orde die in items werden omgezet. De 'thematische aanpak' leidde weliswaar tot een overvloed van onderwerpen, maar de gerichtheid op bepaalde afgesproken vaardigheden was niet bijzonder consequent.

Bij de natuurwetenschappelijke toets gold dit minder, hier werden voornamelijk eenvoudige toepassingen van principes aan de orde gesteld.

Deze manier van uitwerking in twee deelttoetsen met verschillende nadruk had één belangrijk nadeel op het punt van duidelijkheid. Het bleek namelijk dat de inhoud van de toetsen en de gemeten vaardigheden moeilijk te omschrijven waren.

In de verantwoording voor de gebruikers werden de toetsitems daarom voornamelijk gekarakteriseerd met 'trefwoorden' uit de lijst.

Als voorbeeld de opgave van toets A.K. I 1970:

| <i>onderwerp:</i> | <i>aantal vragen:</i> |
|--------------------------|-----------------------|
| topografie | 1 |
| klimaat, volksgezondheid | 2 |
| arbeidsverdeling | 2 |
| beroepen | 3 |
| middelen van bestaan | 2 |
| communicatie media | 5 |

| | |
|----------------------------|---|
| bestuur | 2 |
| wetgeving | 4 |
| verkeer | 5 |
| recreatie | 2 |
| contemporaine geschiedenis | 1 |
| geluidssnelheid | 1 |

Deze manier van werken en rapporteren had dus duidelijk nadelen. De inhoudsopgave achteraf van de toets bleek te algemeen voor 'consumenten', voor constructeurs bevatte de omschrijving te weinig houvast als men zich wilde houden aan de richtlijnen, die eerder waren afgesproken.

6. *Uitwerking in termen van vaardigheden binnen een vakgebied*

In de toetsconstructie van 1971 werd i.v.m. de tekortkomingen van de eerder beschreven werkwijze een meer uitgewerkte procedure gevolgd.

Nadat een - voorlopige - keuze was gemaakt uit de oorspronkelijke lijst van onderwerpen uit 1970, werden de gekozen 'onderwerpen' gerangschikt in een aantal rubrieken. De namen van de rubrieken (zoals 'aardrijkskundige en toeristische informatie') waren soms gelijk aan die van de lijst uit 1970. De voorlopige aanduiding die werd gegeven voor elke nieuwe rubriek gaf al enigszins weer in welke richting verder werd gedacht bij de verdere analyse in vaardigheden.

6.1 *Vijf vragen bij het identificeren van vaardigheden.*

Voordat vragen werden bedacht binnen de rubrieken, werd eerst over elk van de onderwerpen overlegd. De discussie werd over het algemeen gevoerd naar aanleiding van de volgende vragen: Vraag 1. Wat houdt dit onderwerp in: of wat is de inhoud van de rubriek.

Vraag 2. Welke onderdelen, begrippen e.d. moeten mensen dikwijls hanteren en in welke situaties?

Vraag 3. Wat voor soort eisen ten aanzien van die hantering kun je redelijkerwijze stellen aan kinderen van 11-12 jaar, na 6 jaar basisschool? m.a.w. welk gedrag kun je verlangen?

Vraag 4. Welke vaardigheid beheerst een leerling als hij aan deze 'eisen' voldoet? m.a.w. welk gedrag kan hij vertonen in dat geval?

Vraag 5. Hoe kan deze vaardigheid getoetst worden?

Tijdens de discussies volgens dit schema werden impliciet of expliciet een aantal vooronderstellingen gebruikt. Dientengevolge is het resultaat van een dergelijke analyse slechts tot op zekere hoogte 'objectief'.

Toelichting bij de vragen 1-5

In Vraag 1 kan vrijwel geen uitputtende inventarisatie gemaakt worden door een commissie; men dient zich te bepalen tot begrippen die men zelf kent. Dit is bij het basisonderwijs minder erg dan het klinkt; het heeft bij voorbeeld weinig zin om vakspecialisten te raadplegen zonder dat een geheel nieuw leerprogramma kan worden ontworpen. Hun bijdragen tot de inhoud van een begrip/onderwerp is vooral nuttig bij het vernieuwen van leergangen. Bij het concretiseren van leerdoelen op dit niveau is vooral de inbreng van didactici belangrijk.

In Vraag 2 wordt wéér een beroep gedaan op de 'algemene ontwikkeling' van de samenstellers. Ook hier is deze beperking niet zo'n ernstig probleem. Wel kan de eenzijdigheid van een commissie - zeker op den duur - een bezwaar zijn.

In Vraag 3 is een belangrijk begrip: 'communale doelstelling' of 'minimumeis'. Binnen de commissie werd 'minimumeis' ongeveer als volgt opgevat:

Een eis die men in principe aan iedere normale leerling kan stellen.

De antwoorden op de Vragen 4 en 5 liggen gedeeltelijk op het gebied van de techniek van itemconstructie. Hier zijn wel enkele hanteerbare criteria aanwezig waardoor iets meer objectiviteit kan worden betracht dan bij de discussies rond de Vragen 1, 2 en 3.

Desondanks blijft hier ook ruimte over voor persoonlijke voorkeur.

Als resultaat van het overleg in deze fase werden dus in de eerder gekozen rubrieken van on-

derwerpen of groepen van onderwerpen een aantal vaardigheden geformuleerd.

Deze vaardigheden werden als uitgangspunt gebruikt voor het schrijven van de items zelf.

De inhoudsbeschrijving in de verantwoording is nu opgesteld in twee kolommen: kolom 1, geeft aanduiding van het onderwerp. kolom 2, geeft een formulering van het gedrag dat leerlingen ten toon moeten kunnen spreiden bij het beantwoorden van een vraag. Bovendien werden voorbeelden van vragen afgebeeld die het mogelijk maakten om een concrete voorstelling te maken van deze formulering en om kritiek te leveren op de inhoud van de toets.

7. Feedback van gegevens, meningen van gebruikers

De voorgaande paragrafen schetsen een manier van toetsconstructie die in belangrijke mate subjectief is.

In vele stadia van het beschreven constructieproces zijn namelijk de gemaakte keuzen in belangrijke mate afhankelijk van de toevallige samenstelling van het constructie-team.

Dit werd van meet af aan door de samenstellers als een bezwaar gevoeld zodat naar wegen werd gezocht:

1. *de mening* van de samenstellers te toetsen aan de meningen in 'het onderwijsveld'.
2. *de meningen* te peilen van de 'consumenten' van het basisonderwijs. Deze opiniepeilingen kunnen niet los gezien worden van de resultaten van de kinderen zelf op de vragen. Empirische item-gegevens en meningen over de toets kunnen elkaar aanvullen. Onvermijdelijk hierbij is dat de twee bronnen van informatie elkaar ook wel eens tegenspreken.

7.1 Een enquête van het R.I.T.P. onder onderwijzers

Door Warries van het R.I.T.P. (project 0138) werd een tweetal enquêtes gehouden onder onderwijzers 'klachten en meningen over de schooltoets'. De enquêtevragen waren in de vorm van uitspraken die de beantwoorder kon bevestigen

of ontkennen. De uitspraken waren ontleend aan een aantal gesprekken met onderwijzers naar aanleiding van de schooltoets 1969. Jammergenoeg hadden slechts twee van de 96 uitspraken betrekking op Algemene Kennis. De enquête was namelijk in hoofdzaak gericht op andere aspecten van de toets. De twee uitspraken over A.K. worden hieronder weergegeven:

'Nieuwere opvattingen over het onderwijs in de zaakvakken worden door de toets tegengehouden'.

50% oneens

50% helemaal mee oneens.

'Het maken van de Algemene Kennis-toets is zinloos'.

75% oneens

25% helemaal mee oneens.

Het bevestigen of ontkennen van dergelijke 'krasse uitspraken', levert maar weinig gegevens op om een toets gedetailleerd te beoordelen. Een (gering) aantal uitspraken die 'spontaan' werden geleverd leveren in dit opzicht weliswaar meer informatie op, maar deze gegevens zijn moeilijk te interpreteren.

Er werd bij voorbeeld kritiek gegeven op sommige vragen van de voorgaande toets en op de presentatievorm, bovendien wordt gesignaleerd dat bij sommige onderwerpen bepaalde delen van het land boven andere delen bevoordeeld zouden worden etc. Enkele critici *missen* bepaalde onderdelen, anderen vinden vragen te moeilijk en te hoog gegrepen en ook wordt de wijze van samenstelling als 'toevallig' en niet representatief beoordeeld. Een dergelijke kritiek van de gebruikers van de toets is bijzonder nuttig en inspirerend voor de samenstellers. De moeilijkheid blijft echter bestaan dat de kritiek tot nu toe te incidenteel blijft en weinig systematisch is. Dit kan worden toegelicht aan een andere enquête die door het C.I.T.O. werd gehouden.

7.2. Een C.I.T.O.-enquête (Peters - Sips 1970)

Aan een aantal aan de schooltoets deelnemende (110) en niet deelnemende (99) scholen werd een

enquête-formulier gezonden met het onderzoek om enquête-formulieren ingevuld terug te sturen. Helaas was de 'respons' op deze enquête vrij gering, 50% van de deelnemende en slechts 21% van de niet deelnemende scholen, zonden het vragenformulier ingevuld in.*

Bij alle items werden de volgende twee vragen gesteld:

1. 'Vindt u het gewenst dat de leerstof die de leerling moet beheersen om dit item juist te beantwoorden in de basisschool onderwezen wordt?'
2. 'Denkt u dat het mogelijk is de leerstof die de leerling moet beheersen om dit item juist te beantwoorden in de basisschool met succes te onderwijzen?'

Eén van de uitkomsten was dat niet-deelnemers in het algemeen gunstiger (in de termen van wenselijkheid) oordeelden over de toetsen dan deelnemers. Een mogelijke verklaring hiervoor is dat de oordelen van leerkrachten beïnvloed worden door het feit dat de items wel of niet opgenomen zijn in een beoordelingsprocedure, waardoor deelnemers zich mogelijk 'meebeoordeeld voelen'. Dit is een minder onbevangen en vrijblijvende beoordelingssituatie dan voor niet-deelnemers.

Een bruikbaar gegeven was nog dat een aantal respondenten aangaf dat zij bepaalde onderwerpen misten in de Algemene Kennis-toets:

| | |
|------------------------------|----|
| 'Biologie' | 21 |
| 'Geschiedenis' | 20 |
| 'Aardrijkskundige begrippen' | 13 |
| 'Actuele problemen' | 14 |

7.3. Beoordelen is moeilijk

Bij het peilen van meningen van leerkrachten doen zich problemen voor die de interpretatie van de uitkomsten bemoeilijken. De mening over de wenselijkheid van bepaalde toetsonderdelen is een zeer belangrijk gegeven.

Toch moet bij de interpretatie van meningen

* Voor de schrijver reden om voor te stellen dat in de toekomst als routine alle items worden beoordeeld door de onderwijzers terwijl de leerlingen de vragen beantwoorden bij het afnemen van de toets.

rekening worden gehouden met enkele verschijnselen die in de beoordeling meespelen.

Deze bewering is te verduidelijken met een citaat uit het verslag over de reeds genoemde C.I.T.O.-enquête.

'Enkele respondenten *misten* onderwerpen die anderen *te veel* vertegenwoordigd achtten, ook werden enkele onderdelen gemist die duidelijk wel in de toets Algemene Kennis vertegenwoordigd waren.

De terminologie waarmee men de onderdelen aanduidde is niet eenduidig, velen hanteren de aanduidingen uit de 'oude' vakkensplitsing. Anderen gebruiken ad hoc bedachte aanduidingen om onderdelen te benoemen die niet goed te plaatsen zijn binnen de oude vakindeling.'

De persoonlijke voorkeur van sommige leerkrachten leidt er toe dat zij de samenstelling van de toets anders wensen. Dit is niet anders te verwachten, bovendien is overleg mogelijk als men de afwijkende meningen kent. De verwarring over termen is echter een ernstig bezwaar. Dat bovendien sommige beoordelaars vinden dat sommige onderdelen niet worden getoetst die duidelijk wel zijn opgenomen kan wijzen op een hanteren van zeer persoonlijke definities van bepaalde onderdelen. Afgezien van de vraag wie nu gelijk heeft – in het citaat boven werd de uitspraak gedaan dat de beoordelaar fout is, maar in zijn algemeenheid hoeft dat niet juist te zijn – is de begripsverwarring tussen toetsmakers en beoordelaars een groot bezwaar.

Een andere moeilijkheid is dat bij het beoordelen van de wenselijkheid van items, de faktor 'wenselijkheid' niet los gezien wordt van de faktor 'mogelijkheid'.

Dit blijkt uit het statistische verband dat wordt weergegeven in Tabel 2. Hierbij werd gebruik gemaakt van gegevens uit de C.I.T.O.-enquête.

Opvallend hierbij is het hoge verband tussen de factoren wenselijkheid en mogelijkheid en het relatieve zwakke verband tussen de wenselijkheid van sommige stukken stof en de resultaten van de kinderen. (deze tendentie is ook waar te nemen bij de andere subtoetsen).

Tabel 2: Correlaties tussen 'wenselijkheid' van items en 'mogelijkheid' en het verband met gemakkelijheid van de items (p-waarde)

| relatie tussen: | toets | |
|------------------------------------|--------|---------|
| | A.K. I | A.K. II |
| gemakkelijkheid en 'wenselijkheid' | 0,37 | 0,30 |
| gemakkelijkheid en 'mogelijkheid' | 0,27 | 0,41 |
| 'wenselijkheid' en 'mogelijkheid' | 0,91 | 0,90 |

Samenvattend kan men zeggen:

Wat men als wenselijk ziet, wordt ook mogelijk gevonden, maar wat mogelijk is, is in de praktijk lang niet altijd gerealiseerd. Dit laatste staat wellicht in verband met het probleem om de moeilijkheid van items te schatten. Dat dit niet meevalt, is overigens al lang bekend.

Derksen-Mögelin (1968) toonde aan dat de onderlinge overeenstemming tussen verschillende (onderwijzers-) beoordelaars bij het schatten van de moeilijkheid van items vrij gering is (wat de één moeilijk vindt, vindt de ander weer gemakkelijk).

Een ander interessant resultaat uit dit onderzoek was dat de onderwijzers de vragen in het algemeen als moeilijker beoordelen dan zij in werkelijkheid zijn. Als men dus aan leerkrachten vraagt om de moeilijkheid van vragen te schatten blijkt dikwijls dat de vragen als gemakkelijker of moeilijker worden beoordeeld dan ze in feite zijn.

In het volgende overzichtje wordt dit geïllustreerd voor de commissie Algemene Kennis. (Van Deursen 1970).

Tijdens de constructiefase van de items voor 1971 werd namelijk de commissieleden gevraagd om te schatten, hoe moeilijk de goedgekeurde concept-items zouden zijn. (zie Figuur 2a).

Voor de beoordeling werden drie gradaties gebruikt: gemakkelijke vraag, aangegeven met één kruisje, middelmatig moeilijk item (twee kruisjes) 'moeilijk item' (3 kruisjes).

Deze voorspellingen werden getoetst aan de percentages juiste antwoorden (p-waarden), die bleken uit een proef-afname van de geconstrueerde vragen. Ook hier werden drie gradaties gebruikt: méér dan 80% juist, van 60-80% juist en minder dan 60% juiste antwoorden.

Tabel 3: Verband tussen voorspelde en gevonden moeilijkheid bij A.K.

| empirische moeilijkheid | voorspelde moeilijkheid | | | |
|----------------------------|-------------------------|----|-----|----|
| | x | xx | xxx | |
| > 80 | 10 | 2 | | 12 |
| 60-80 | 2 | 5 | | 7 |
| < 60 | 2 | 3 | 4 | 9 |
| | | | | 28 |

68% juist voorspeld

25% voorspeld met 1 klasse verschil

7% voorspeld met 2 klassen verschil

Nu bleek de commissie Algemene Kennis overigens nog de beste voorspelling te doen van de moeilijkheidsgraad van de items van de proef-toets.

Echter evenals bij andere commissies (Tabel 4) blijkt dat zelfs bij het gebruiken van vrij grove klassen, vrij veel 'fouten' worden gemaakt.

Tabel 4. Verband tussen voorspelde en gevonden moeilijkheid bij stillezen, spelling en hoofdrekenen (gecombineerd).

| | x | xx | xxx | |
|-------|----|----|-----|-----|
| > 80 | 19 | 22 | 9 | 50 |
| 60-80 | 9 | 17 | 16 | 42 |
| < 60 | | 8 | 7 | 15 |
| | | | | 107 |

De gedemonstreerde overschatting of onderschatting van de leerlingen wordt bij het samenstellen van de schooltoets wel weer hersteld omdat de items van de eindversies van de toetsen óók worden gekozen op grond van de gevonden moeilijkheid bij een proefafname.

In de constructie van een toets leidt een discrepantie tussen de 'werkelijke' (empirische) moeilijkheidsgraad en de schatting door onderwijzers dus niet noodzakelijk tot ernstige gevolgen. Als men voldoende items heeft, kan men de moeilijk-

heid van de toets wel degelijk in de hand houden.

Bij de beoordeling in het kader van een enquête is echter duidelijk dat de uitkomsten vertekend kunnen worden door het feit dat individuele beoordelaars slechts een zeer beperkt aantal gevallen kunnen waarnemen (of weinig ervaring met de toetsvorm hebben). Deze beperkte waarneming kan er toe leiden dat de beperkte ervaring gelijk gesteld wordt aan de gehele werkelijkheid zonder dat rekening gehouden wordt met variabelen die men zelf in zijn praktische situatie niet hanteert.

Deze laatste opmerking kan worden toegelicht met een tweetal onderzoeken uit het R.I.T.P.

Warries (1967) wijst er op dat oordelen over de moeilijkheid van vragen samenhangen met de vormgeving: items met een goede vormgeving zijn vaker kernitems* dan items met veronderstelde fouten in de vormgeving'. (p. 16).

In een ander onderzoek (Warries 1970) - een van de veel te zeldzame gevallen waarin de ouders van de kinderen gevraagd worden om vragen op wenselijkheid te beoordelen - werd voor A.K. een zwak verband gevonden tussen de eigen opleiding en de positie van de beoordelaars enerzijds en de eisen die zij aan de kinderen stellen anderzijds.

M.a.w. 'wat iemand zelf geleerd heeft, moeten de kinderen ook leren'. Nu hebben 'ouders' in het algemeen het meest beperkte waarnemingsveld t.a.v. dit soort zaken. Zij zien meestal alleen hun eigen kinderen. Onderwijzers hebben een iets groter veld n.l. de klas en hun vorige klassen, maar hier zijn factoren zoals het milieu van de kinderen, de mogelijkheden en hulpmiddelen op school en hun persoonlijke voorkeur werkzaam. Een conclusie die voortvloeit uit deze interpretatie is dat degenen die over empirische gegevens van de gevraagde items beschikken ook degenen zijn die de meeste informatie hebben over de 'mogelijkheid' en 'de moeilijkheid' van vragen.

Hun beperking ligt vanzelfsprekend in het feit dat zij slechts beschikken over gegevens over een

* Kernitems ongeveer in de betekenis van items die een begrip representeren dat 'niet gemist kan worden'.

zeer gering aantal vragen per keer.

Zonder een systematisch landelijk evaluatie-onderzoek waarbij een onderdeel van een leerplan, of zelfs een geheel leerplan wordt doorge-licht (survey) zijn dus voorlopig slechts conclusies mogelijk over een beperkt gebied, n.l. het gebied waarover de vragen worden gesteld.

Op dit ogenblik lijkt de conclusie mogelijk dat het zeer zinvol is om de vraag te stellen 'welke onderwerpen vindt u de moeite waard om te toetsen'. Deze vraag moet men stellen aan ouders, onderwijzers en 'deskundigen'. De vraag naar de 'bereikbaarheid' van een bepaald doel moet echter voornamelijk langs empirische weg worden beantwoord. Deze conclusie sluit aan bij die van Mellenbergh (1971). Mellenbergh vat een aantal onderzoeken (ook over de beoordeling van de 'kwaliteit') over de waarde van itembeoordelingen als volgt samen: 'De resultaten van de onderzoeken zijn van dien aard dat men niet een te groot vertrouwen mag hebben in beoordelingen van items' (p. 99).

8. Terugblik op enkele vragen uit 1970

Dat het afnemen van studietoetsen op zeer grote schaal evaluatiemogelijkheden van het onderwijs biedt, wordt geïllustreerd met enkele opmerkingen over items die waren opgenomen in de schooltoets 1970.

Onder *evaluatie* wordt hier verstaan:

Een empirische vaststelling van de mate waarin de gestelde onderwijsdoelen bereikt zijn.

Dit vaststellen kan op verschillende niveaus gebeuren. Op landelijk niveau of op regionaal niveau kan een landelijke of plaatselijke instantie zich verantwoordelijk voelen voor de stand van zaken in het onderwijs. De empirische gegevens kunnen antwoord geven op vragen als: 'Welk soort vragen en vaardigheden worden in het algemeen goed of slecht beheerst'. Gegevens kunnen ook conclusies mogelijk maken in de trant van: 'Op onderdeel x zijn de prestaties goed genoeg; bij onderdeel y zou extra aandacht nodig zijn'.

Dit soort conclusies is bijzonder waardevol omdat zij houvast geven voor het nemen van gerichte maatregelen.

Jammer genoeg wordt deze gebruiksmogelijkheid in Nederland nauwelijks benut, door 'instanties'. Voor evaluatie binnen de school gaat dit ook op. Mogelijk is dit te verklaren vanuit de nadruk die gelegd wordt op de selecterende functie van schooltoetsen: voldoende of onvoldoende voor het HAVO bij voorbeeld.

Deze selectie-functie en de denkwijze die daarin is verwerkt, betekent in feite dat een toets-score uitsluitend gevolgen heeft voor de leerlingen.

Deze denkwijze komt ook tot uiting in een vaak beluisterde kritische opmerking dat evaluatie weinig zin heeft 'omdat de leerlingen toch van school gaan en het is te laat om nog iets te veranderen'.

Maar evaluatie heeft *wel* bijzonder veel zin omdat toetsgegevens gebruikt kunnen worden voor het onderwijs dat in de toekomst zal worden gegeven. In een publicatie van Van Deursen en Jansen (1971) wordt overigens een aantal concrete aanwijzingen gegeven die een dergelijke evaluatie mogelijk maken binnen een bepaalde school.*

Op het terrein van 'Algemene Kennis' is het misschien interessant sommige resultaten te bespreken van de kinderen die in 1970 aan de vragen hebben gewerkt. Voor een uitputtende analyse ontbreekt hier helaas de gelegenheid. De opmerkingen worden weergegeven in twee delen:

Betreffende

1. Algemene Kennis I 'Sociaal'
2. Algemene Kennis II 'Natuurwetenschappelijk'

De toetsen waren vrij heterogeen samengesteld zodat de opmerkingen betrekking hebben op losse vragen.*

* Dat evaluatie nog niet erg populair is blijkt o.a. uit de geringe belangstelling van scholen voor de genoemde handleiding. Er waren namelijk slechts 70 aanvragen voor deze handleiding binnengekomen op het moment dat dit verslag werd geschreven (juni 1971).

* De vragen zijn niet volledig afgedrukt. Desgewenst kan men de volledige toetsen aanvragen bij het C.I.T.O., postbus 1034, Arnhem.

In het algemeen komen slechts vragen aan de orde die door minder dan 60% van de leerlingen juist werden beantwoord.

Alleen als in andere items één of twee 'afleiders'

(foute antwoorden) opvallend veel worden gekozen, worden de desbetreffende vragen in het volgende overzicht genoemd.

| Toets en itemnr. | Veronderstelde vaardigheid die nodig is voor het juiste antwoord | presentatievorm | percentage juiste antw. |
|------------------|---|---|-------------------------|
| AK I 3 | inzicht van de noodzaak van de bereikbaarheid van een overheidsgebouw opm. 23% 'een raadhuis behoort aan een plein te staan'. | verwijzing naar een tekening, verder verbaal | 68% |
| AK I 5 | De betekenis van een veelgebruikt symbool kennen opm.: 40% 'een waag' | afbeelding van 'vrouw justitia' met weegschaal, zwaard en blinddoek | 39% |
| AK I 7 | Toepassing van een eigenschap van een massa-medium (snelheid van radio). opm.: 32% kiest 'in de krant zetten' Het begrip snel had mogelijk nog duidelijker in de vraag moeten staan | verbaal, het verspreiden van een signalement | 50% |
| AK I 9 | Bemiddelende functie van een bestuurder onderkennen. opm.: 43% kiest voor het gebruik van machtsmiddelen (politie, leger), dit item is niet voor 100% objectief | verbaal, wat is de beste manier om de vrede te herstellen | 54% |
| AK I 12 | Kennen van bronnen van inkomsten voor enkele beroepen opm.: 29% kiest: 'de bankbediende' | verbaal (wie krijgt zijn salaris van de gemeente?) | 67% |
| AK I 13 | Toepassing van het begrip 'recreatie' opm.: De term 'recreatie' is waarschijnlijk bij velen niet bekend. 19% verwarring met onderwijs | verbaal: welke maatregel bevordert de recreatie? | 61% |
| AK I 14 | Beslissingsbevoegdheid van het gemeentebestuur kennen opm.: objectiviteit is niet onaanvechtbaar | verbaal: 'wie kan besluiten?' | 48% |
| AK I 18 | In een verkeersprobleem, de beste oplossing aan kunnen geven opm.: 59% kiest voor een oplossing die op grond van de stam beslist niet mogelijk is. | verbaal met verwijzing naar een tekening | 28% |

| Toets en itemnr. | veronderstelde vaardigheid die nodig is voor het juiste antwoord | presentatievorm | percentage juiste antw. |
|------------------|--|--|-------------------------|
| AK I 20 | De relatie tussen natuurlijke omstandigheden en de vestigingsplaatsen voor industrie begrijpen met het oog op de — gezondheid. Opm.: 34% neemt geen notitie van een gegeven in de stam | verwijzing naar tekening 'is deze plaats juist?' | 54% |
| AK I 25 | De functie van een overheidsdienst kennen | verbaal, koppelen van een dienst aan een activiteit — welke koppeling is fout? | 43% |
| AK I 27 | De betekenis begrijpen van de term 'achterland' opm.: alle respondenten vonden deze vraag gewenst, maar het 'echte' begrip ontbreekt vaak bij de kinderen | verbaal: wat is . . . | 32% |
| AK I 28 | Begrip 'geluidssnelheid' kunnen toepassen opm.: 38% van de kinderen is in de war met weerkaatsing | verbaal naar verklaring vragen van een verschijnsel | 42% |
| AK I 29 | Kennis van een methode voor het zenden van T.V.-beelden naar een ander continent | verbaal Hoe doet men dit? | 48% |
| AK I 30 | Het begrip 'sociale wet' kunnen toepassen | verbaal Welke is een sociale wet? | 57% |
| AK II 1 | Toepassing van een eenvoudig hefboomprincipe | tekening Welke tang kun je het beste . . . | 54% |
| AK II 7 | Toepassing van het begrip 'zwaartekracht' opm.: 62% verwarring met 'gewichtloos' | Verbaal voorspelling van 'gewicht' | 31% |
| AK II 9 | Kennis begrip 'energiebron' | Welke is de oudste manier om energie op te wekken | 56% |
| AK II 12 | Inzicht in het gedrag van een gas in een ruimte. Opm.: 48% van de kinderen geeft een antwoord dat een precies omgekeerde toepassing is van het principe | tekening: Welke fluitketel gaat het eerst fluiten? | 49% |

| Toets en itemnr. | veronderstelde vaardigheid die nodig is voor het juiste antwoord | prestatievorm | percentage juiste antw. |
|------------------|---|--|-------------------------|
| AK II 15 | Toepassing van het begrip 'verdamping' en snelheid van verdamping. Opm.: 43% van de leerlingen denkt dat water sneller verdampt dan benzine | tekening. Welke vloeistof is het eerst verdampt? | 43% |
| AK II 19 | Eenvoudige toepassing van een elektromagneet verklaren. Opm.: 43% van de leerlingen weet niet waar een elektromagneet voor gebruikt kan worden. | verbaal | 57% |
| AK II 20 | Toepassing van het begrip 'luchtdruk' in een barometer | tekening barometer | 54% |
| AK II 21 | Kennis van de toepassing van een hygrometer | verbaal. Welk instrument heb je nodig? | 49% |
| AK II 22 | Kennis van de term 'edel metaal' | verbaal, definitie | 43% |
| AK II 26 | Inzicht hebben in het verband tussen waterverbruik en industrialisatie. Opm.: 28% meer waterverbruik door de uitbreiding van havens | verbaal verklaring hoe komt dat | 59% |
| AK II 27 | Toepassing van de term luchtverontreiniging | verbaal waar is . . . het ergst? | 45% |
| AK II 28 | Toepassing van het begrip sterilisatie | verbaal, hoe bacterievrij maken | 59% |
| AK II 29 | Eigenschappen van bacteriën in verband brengen met voedselbereiding | verbaal bacteriën juist nodig voor het voedsel | 36% |

Een algemene opmerking naar aanleiding van deze itemgegevens is dat sommige begrippen of onderwerpen kennelijk veel bekender zijn dan andere, waarom dit zo is blijkt niet altijd uit de onderwerpen zelf. Het begrip 'zuurstof' levert bij voorbeeld weinig moeilijkheden op maar de begrippen 'electromagneet' en 'verdampen' geven veel moeilijkheden.

De nadruk die op bepaalde begrippen gelegd

wordt, is vanzelfsprekend bijzonder afhankelijk van de voorkeur, tijd en mogelijkheden op een bepaalde school. In zoverre kan de simpele vaststelling van percentages juiste antwoorden in een land, een streek of een school nooit op zichzelf aanleiding geven tot de conclusies dat 'kinderen onvoldoende weten'.

Wel kunnen dit soort empirische gegevens aanleiding zijn om zich af te vragen of de uitkomsten

stroken met de bedoelingen van de onderwijzers of opstellers van leerplannen.

Een andere opmerking

Afgezien van de behandeling van het onderwerp of de leerstof in kwestie wordt de indruk gewekt dat er één vaardigheid is die belangrijk is voor het oplossen van veel vragen. Bij het toetsen van 'Algemene Kennis' wordt veelal prijs gesteld op het toetsen van andere zaken dan alleen 'weetjes' en 'feiten'. Dit brengt met zich mee dat men genooddakt is om dikwijls 'situaties te beschrijven', waarin gegevens worden verwerkt waar leerlingen rekening mee moeten houden om de vraag op te lossen. In sommige gevallen blijkt dat vragen moeilijk zijn omdat kinderen bijv. *niet lezen wat de vraag is, geen rekening houden met de gegevens of eisen* die overigens duidelijk zijn geformuleerd.

Voorlopige ervaringen met eenvoudige routebeschrijvingen, eenvoudige dienstregelingen, het lezen van kaarten enz. ondersteunen deze opmerking. (De conclusie dat A.K. dus 'eigenlijk stil-lezen is', is overigens onjuist maar op deze kwestie wordt op een andere plaats teruggekomen).

Hoofdzaak in deze paragraaf was echter om de mogelijke bruikbaarheid te illustreren van itemgegevens voor beslissingen in de onderwijspraktijk.

9. 'Eerlijkheid' en 'wenselijkheid' bij het opnemen van vragen

Bij het construeren van toetsen is een belangrijk uitgangspunt dat men probeert inderdaad te toetsen wat op school onderwezen wordt (of zou

kunnen worden). Daarnaast is het belangrijk dat men ervoor zorgt dat men bepaalde groepen kinderen of scholen niet benadeelt of bevoordeelt bij het uitkiezen van vragen.

Een kritische opmerking van een onderwijzer aan de samenstellers van de toets 'Algemene Kennis' luidde dat een toets:

'ontzettend ver doorgevoerd (is) in de interesse-sfeer van de jongens' en (in) sommige items te weinig rekening (is) gehouden met de leeftijd en het milieu van de leerlingen. Men kan op een 'volkschool' minder ver gaan dan op een 'élite-school'.

Dergelijke opmerkingen slaan direct op dit principe. Blijkt het zo te zijn dat de vragen een of meer groepen leerlingen bevoordelen of benade-len dan verdient deze kwestie nadere aandacht.

Bij 'Algemene Kennis' zijn inderdaad bepaalde items aan te wijzen die tot dergelijke opmerkin-gen aanleiding kunnen geven.

Zo werd bij een bespreking van de concept-toetsen voor 1970, door collega's uit andere con-structieteams bezwaar gemaakt tegen de 'natuur-wetenschappelijke' toets omdat deze te moeilijk zou zijn. Andere opmerkingen bij een andere subtoets signaleerden een mogelijke bevoorde-ling van jongens boven meisjes en een bevoorde-ling van stadskinderen boven dorpskinderen.

Het is gedeeltelijk eenvoudig om na te gaan in hoeverre deze opmerkingen gerechtvaardigd zijn. Voor een andere deel is dit probleem van 'eerlijk-heid' moeilijker op te lossen.

Wat betreft de moeilijkheid van de items zijn er gegevens bekend over 4 jaren schooltoets. (zie Tabel 5).

Tabel 5. De moeilijkheid van subtoetsen „Algemene Kennis” uitgedrukt in gemiddelde percentages juiste antwoorden vergeleken met „Totaal taal” en „Totaal rekenen” over 1967-1970. (naar Jansen 1970.

| Subtoets | 1967 | 1968 | 1969 | 1970 |
|--------------------|------|------|------|------|
| Algemene Kennis I | 75 | 70 | 67 | 67 |
| Algemene Kennis II | 67 | 75 | 68 | 63 |
| Totaal Rekenen** | 65 | 64 | 67 | 61 |
| Totaal Taal*** | 84 | 78 | 73 | 73 |

** d.w.z. de subtoetsen Hoofdrekenen I en II en rekenvraagstukken.

*** d.w.z. de subtoetsen: Spelling, Gemengde Taalopgaven I en II, Stillezen lang stuk, Stillezen korte stuk-ken en woordbenoeming/zinsontleding.

Uit Tabel 5 blijkt dat Algemene Kennis in moeilijkheid ligt tussen rekenen en taal. Bovendien blijkt, dat de 'natuurwetenschappelijke' toets Algemene Kennis II inderdaad iets moeilijker is dan de 'sociaal-economische' subtoets Algemene Kennis I. *Iets moeilijker* behoeft echter nog niet te betekenen: *te moeilijk*. De vraag naar de beste moeilijkheidsgraad is uitsluitend te beantwoorden tegen de achtergrond van het gebruik van de toets. Als men deze vragen zou gebruiken om in de huidige Nederlandse situatie 'minimum-eisen' te toetsen m.a.w. te toetsen voor 'beheersing' of 'mastery' (Warries 1970) zouden deze items zonder meer te moeilijk zijn. In een selectiesituatie zouden de vragen weer van een 'goede' moeilijkheid zijn.

In het onderhavige geval waarbij de uitslagen van de kinderen niet worden opgenomen in de totale percentielscore kan men een dergelijke uitslag gebruiken om te zien hoe de resultaten in het algemeen zijn van kinderen op dit onderdeel.

Op een tweede probleem n.l. dat van de benadeling of bevoordeling van sommige groepen kinderen is minder gemakkelijk te antwoorden.

Het is allang bekend dat jongens op de meeste vragen van een schooltoets iets betere resultaten behalen dan meisjes. Dit geldt echter voor vrijwel alle subtoetsen.

Hieruit kan men zowel de conclusie trekken dat of de gehele toets zo geconstrueerd is dat meisjes over de hele linie benadeeld worden of dat meisjes in de realiteit op deze leeftijd iets mindere prestaties leveren.

De laatste conclusie lijkt het meest aannemelijk. Jongens – als groep – zijn op deze leeftijd iets beter in schoolwerk dan de meisjes. Maar als er inderdaad sprake zou zijn van een specifieke jongenstoets, zou men maatregelen moeten nemen bij voorbeeld door inschakeling van meer vrouwelijke itemschrijvers. In het tweede geval ligt een meer fundamentele verandering, namelijk van het onderwijs zelf, voor de hand.

Dit zou daarom zo'n fundamentele verandering zijn omdat men dan naar wegen moet zoeken om de verschillen in schoolprestaties tussen de groepen te verminderen. Daarmee zouden we waarschijnlijk terechtkomen op het gebied van

verwachtingen over de sociale rollen van meisjes en verandering van belangstellingspatronen. Maar ook zonder zo'n analyse is het mogelijk om een standpunt in te nemen. Bij toetsing van beheersing van sommige leerstofonderdelen (minimumeisen of communale doelstellingen) kan men namelijk stellen dat de vraag naar sekse, sociale klasse of woonplaats niet relevant is.

Iedereen moet kunnen rekenen, lezen of iets van aardrijkskunde afweten.

Sommige onderdelen zijn dus zo belangrijk dat men deze vraag nadrukkelijk negeert.*

Iedereen moet dus bepaalde doelen bereikt hebben wil het onderwijs zijn doel bereikt hebben. In een selectie – of plaatsingsituatie kan men dan altijd afzonderlijke normen voor de verschillende groepen gebruiken als men vindt dat bepaalde factoren niet mogen meetellen. Dit is echter de taak van degenen die plaatselijk de beslissingen moeten nemen. Beleidsinstanties moeten hun beleid in zo'n geval duidelijk maken en hun wensen kenbaar maken.

De taak van toetsconstructeurs is wel om er op te letten dat er geen duidelijk 'oneerlijke' vragen worden gesteld.

Bij vragen over 'het kunnen kaartlezen' bij voorbeeld worden kinderen uit Arnhem bevoordeeld als *alle* kinderen vragen moeten beantwoorden over een plattegrond van Arnhem en omstreken, tenslotte kan men niet van alle kinderen verwachten dat zij de weg weten in die stad.

Als daarentegen kinderen uit het oosten van het land geen eenvoudige vragen over zoiets algemeen als een haven kunnen beantwoorden, mag men de fout niet zoeken in de vragen, maar is het nuttiger om dit onderwerp op school meer te benadrukken.

Het criterium voor dit soort beslissingen blijft echter vaag 'het belang van de stof', en deze be-

* Dit wil nadrukkelijk niet zeggen dat men daarmee maar moet berusten in 'de bestaande toestanden'. Het is – volgens deze schrijver althans – volkomen terecht dat men via 'verrijkingprogramma's' e.d. een ongewenste situatie probeert te verbeteren.

slissingen blijven aanvechtbaar zolang men geen duidelijk overzicht heeft van de minimumeisen waaraan een leerling van de lagere school moet voldoen.

De waarde van zulke uitkomsten blijft uiteraard beperkt doordat een 'absolute maatstaf' in vele gevallen ontbreekt. De term 'minimumeis' zal dan ook op allerlei gebieden nader geoperationaliseerd moeten worden.

In het Research Instituut voor de Toegepaste Psychologie zijn hierover onderzoeksplannen in een vergevorderde staat van voorbereiding.

Maar zolang we verstoken blijven van de resultaten van zulk onderzoek en van een instituut dat zich systematisch bezig houdt met periodiek evaluatieonderzoek (een leerplan veroudert snel), moet met onvolledige gegevens verder gewerkt worden aan de toetsontwikkeling.

Literatuur

- Derksen-Mogelin, M., (1968), *Beoordeling van L.O. Toetsitems*, Ongepubliceerd Doctoraal werkstuk, Universiteit van Amsterdam 1968 (ook besproken in Mellenbergh 1971).
- Deursen, van J. I. en G. G. H. Jansen (1971), *Gebruik van de scores van de schooltoetsen basis-onderwijs 1971*. Arnhem, C.I.T.O.-publicatie nr. 9 Centraal Instituut voor Toetsontwikkeling.
- Deursen, van J. I. (1970), Verslag proeftoets-1970. Arnhem, C.I.T.O. memo nr. 7. 1970.
- Groot, A. D. de en R. F. van Naerssen, *Studiетоetsen; construeren, afnemen, analyseren*. 's-Gravenhage, Mouton, 1969.
- Jansen, G. G. H. (1970), Psychometrische Analyse van de Schooltoets 1970. Arnhem, C.I.T.O.-memo nr. 13. Centraal Instituut voor Toetsontwikkeling.
- Mellenbergh, G. J., S. Sandbergen, J. Vastenhouw en G. de Zeeuw (1968). Vaardigheden en tekorten een strategie voor het analyseren van onderwijsdoelen. *N.T. Psychologie*, 23.10.1968.
- Mellenbergh, G. J. (1971), *Studies in Studietoetsen*, Academisch proefschrift Psychologisch laboratorium, Universiteit van Amsterdam.
- Peters-Sips, M. W. H. (1970), Verslag van de enquête 'inhoud schooltoets 1970'. Arnhem, Centraal Instituut voor Toetsontwikkeling. C.I.T.O.-memo nr. 5.
- Samenwerkende instituten Nutsseminarium voor Pedagogiek en Research Instituut voor de Toegepaste Psychologie; (1967), *Amsterdamse schooltoetsen*; verslag van het eerste onderzoek 'L.O.-schooltoets Amsterdam, 1966 Groningen, Wolters 1967. Empirische studies over onderwijs nr. 3.
- Sandbergen, S. (1968), Studietoetsen; noodzaak, nut en verdiensten. *Vernieuwing van opvoeding en onderwijs*. 27 (1968) nr. 263 pag. 14-29.
- Sandbergen, S. (1969), Verantwoording van de toets Algemene Kennis 1970 in: *Schooltoetsen basis-onderwijs, verantwoording inhoud toets 1970*. Arnhem, Centraal Instituut voor Toetsontwikkeling. C.I.T.O.-uitgave Dec. 1969.
- Sandbergen, S. (1970), *Verantwoording Algemene Kennistoets 1971*, in: Verantwoording Inhoud 1971. Arnhem, Centraal Instituut voor Toetsontwikkeling. C.I.T.O.-1970.
- Warries, E. (1970), Het relatief meten van leerprestaties in het Onderwijs, *Nederlands Tijdschrift voor de psychologie* (D.P.O.) XXV, no. 6, juni 1970.
- Warries, E. (1971), Drie redenen om te toetsen in het onderwijs. *Pedagogische Studiën*, april 1971.
- Warries, E. (1971), De toets getoetst; *Ouders van Nu*, november 1970, p. 34-41.
- Themans, B. (1968), Ervaringen opgedaan bij het werken met studietoetsen op grote schaal, Schaal I en II. *Pedagogisch Forum* 2 (1968) nr. 6, 283-289 en 3 (1969) nr. 1, 28-35.