

Evalueren van leerresultaten door leerkrachten in de onderbouw van het voortgezet onderwijs*

E. J. J. KREMERS

*Centraal Instituut voor Toetsontwikkeling,
Arnhem*

Samenvatting

In ons land is nog nauwelijks onderzoek verricht naar het evalueren door leerkrachten. In dit artikel doen we verslag van een verkennend onderzoek naar de werkwijze die leerkrachten in de onderbouw van het voortgezet onderwijs volgen bij het evalueren van leerresultaten. Aan het onderzoek werd meegewerkt door ruim 900 leerkrachten, afkomstig uit alle schooltypen voor voortgezet onderwijs en verdeeld over diverse vakken. De evaluatiepraktijk wordt op grond van de onderzoeksresultaten gekenschetst als 'traditioneel'. Inzichten die de laatste decennia min of meer gemeengoed zijn geworden in de evaluatieliteratuur hebben nog maar weinig invloed gehad op deze praktijk. Een mogelijke verklaring is dat in de literatuur onvoldoende rekening wordt gehouden met de voorwaarden waaraan voldaan moet worden, willen deze inzichten in het onderwijs toepasbaar zijn. Een verdere verklaring kan zijn dat deze inzichten slechts in beperkte mate bij leerkrachten bekend zijn.

1 Inleiding

Onder evaluatie wordt in algemene zin verstaan: het bepalen wat iets waard is. Toegespitst op de concrete onderwijspraktijk verstaan we onder evaluatie de activiteit waarbij de leerkracht de waarde bepaalt van onderwijsleerprocessen in het licht van de gestelde doelstellingen (vgl. De Corte, 1978; De Corte, Geerligts, Lagerweij, Peters en Vandenbergh, 1981). Evaluatie vormt een integraal onderdeel van onderwijzen en leren. Vrijwel alle

modellen voor didactisch handelen onderscheiden expliciet een evaluatiecomponent (vgl. De Corte et al., 1981; Janssens, 1983). Het is dan ook niet verwonderlijk, dat evaluatie ruime aandacht heeft gekregen in de onderwijskundige literatuur. Ook in ons taalgebied zijn de laatste jaren diverse publikaties verschenen waarin het evalueren door leerkrachten centraal staat (voor een overzicht zie Cito, 1982). Deze publikaties zijn overwegend prescriptief van aard: ze besteden aandacht aan de diverse aspecten van evalueren en schetsen hoe leerkrachten in voorkomende gevallen het beste te werk kunnen gaan. Publikaties daarentegen over de feitelijke evaluatiepraktijk van leerkrachten in ons land zijn bijzonder schaars. Een bekende uitzondering is de analyse van de wijze van cijfers geven en selecteren in het voortgezet onderwijs door De Groot in de jaren zestig (De Groot, 1966).

In dit artikel doen we verslag van een onderzoek naar de evaluatiepraktijk van leerkrachten in het voortgezet onderwijs. Het onderzoek heeft betrekking op de *onderbouw* van het voortgezet onderwijs. Aan het onderzoek werd meegewerkt door ruim 900 leerkrachten, afkomstig uit alle schooltypen voor voortgezet onderwijs (LBO, AVO (MAVO en/of HAVO), VWO, LBO-AVO (-VWO) en AVO-VWO) en verdeeld over diverse vakken.

Het onderzoek heeft een exploratief karakter. Voor zover bekend is het de eerste keer dat in ons land op een dergelijke schaal gegevens over het evalueren door leerkrachten verzameld zijn. Vergelijkbare gegevens ontbreken nagenoeg. Ook in het buitenland zijn relatief weinig studies verricht naar de evaluatiepraktijk van leerkrachten. Wel is veel onderzoek gedaan – en dan met name in de Verenigde Staten en Groot-Brittannië – naar het gebruik van gestandaardiseerde, niet door de leerkracht zelf ontwikkelde, toetsen (Janssens, 1983 en 1984). We zullen de gegevens van het onderzoek in dit artikel dan ook vooral, wanneer dat mogelijk en zinvol is, relateren aan datgene wat in de meer 'prescriptief georiënteerde' literatuur over evalueren gesteld wordt.

* Met dank aan drs. J. B. A. M. van Bergen, drs. F. J. G. Janssens en drs. R. F. van Werkhoven voor hun commentaar op het concept van dit artikel.

Het artikel kent de volgende indeling. Eerst gaan we in op de achtergrond, de opzet en de uitvoering van het onderzoek (paragraaf 2). In de daaropvolgende paragrafen besteden we aandacht aan de onderzoeksresultaten met betrekking tot de door ons onderzochte aspecten van evalueren. Achtereenvolgens komen aan de orde: de evaluatie tijdens en na het onderwijsleerproces (paragraaf 3), de evaluatie van het begingedrag (paragraaf 4), de correctie van resultaten (paragraaf 5), de cesuurbepaling (paragraaf 6) en de herhaling bij onvoldoende beheersing van de leerstof (paragraaf 7). Het artikel wordt besloten met een algemene discussie van de resultaten en een aantal aanbevelingen voor verder onderzoek (paragraaf 8).

2 Het onderzoek

2.1 Achtergrond

Het onderzoek naar het evalueren door leerkrachten in de onderbouw vormde een onderdeel van een meer omvangrijk onderzoek dat door het Cito in 1981 werd uitgevoerd. In dat onderzoek zijn we nagegaan op welke wijze leerkrachten gebruik maken van, door het Cito ontwikkelde, zogenaamde leerdoelgerichte toetsen. Om het gebruik van deze leerdoelgerichte toetsen beter te kunnen interpreteren werd het wenselijk geacht om ook inzicht te hebben in de algemene werkwijze bij het evalueren. Het is dit deel van het onderzoek waarover wij hier rapporteren.

Er zijn door het Cito voor een uiteenlopend aantal vakken voor de eerste twee à drie leerjaren (onderbouw) van het voortgezet onderwijs leerdoelgerichte toetsen geconstrueerd¹. De eerste series leerdoelgerichte toetsen zijn aan het begin van het schooljaar 1978-1979 gepubliceerd. Het betrof toetsen voor de vakken biologie, Engels, Frans, Nederlands en

wiskunde. Het onderzoek werd uitgevoerd bij leerkrachten van scholen die in de schooljaren 1978-1979 en 1979-1980 voor deze vakken toetsen besteld hadden. In deze jaren zijn door meer dan 1300 scholen voor voortgezet onderwijs (oftewel ruim 50% van alle scholen) voor een of meer van de genoemde vijf vakken toetsen besteld. Het aantal bestellers is een veelvoud van het aantal scholen dat materiaal besteld heeft en bedraagt voor de genoemde twee schooljaren ongeveer 4000.

Wij beschikten niet over de mogelijkheden om na te gaan of de bestellers van leerdoelgerichte toetsen representatief zijn voor leerkrachten in de onderbouw van het voortgezet onderwijs. Directe informatie over de representativiteit van de resultaten ontbreekt daarom. Twee vrij recent uitgevoerde onderzoeken geven hier echter wel enig uitsluitsel over (Van Galen en Bulte, 1982; Van de Perel en Schneider, 1983). In de discussie gaan we daar uitvoeriger op in².

2.2 Opzet en uitvoering

Ten behoeve van het onderzoek werd het bestand van bestellers van leerdoelgerichte toetsen gestratificeerd naar vak en schooltype. Daarbij werden 25 subpopulaties onderscheiden. Deze werden gevormd door de combinaties van de eerder genoemde vijf vakken en vijf schooltypen. Uit elk van deze subpopulaties is een aselechte steekproef getrokken. Voor meer gegevens over de steekproef en de respons verwijzen wij naar het onderzoeksverslag (Kremers, 1982). In Tabel 1 is de samenstelling van de onderzoeksgroep weergegeven. De verschillen in aantallen tussen de afzonderlijke vakken en de afzonderlijke schooltypen hangen samen met de verhouding tussen deze vakken en schooltypen in het bestellersbestand. De verschillen tussen de schoolty-

Tabel 1 De samenstelling van de onderzoeksgroep (absolute aantallen)

Schooltype \ Vak	Bio- logie	En- gels	Frans	Neder- lands	Wis- kunde	Totaal
LBO	53	79	3	46	74	255
AVO (MAVO en/of HAVO)	110	73	76	60	109	428
VWO	3	1	5	3	5	17
LBO-AVO (-VWO)	8	11	9	12	18	58
AVO-VWO	79	37	39	17	44	216
Totaal	253	201	132	138	250	974

pen in het bestellersbestand hangen op hun beurt weer samen met het totaal aantal scholen per schooltype in ons land.

Het onderzoek is uitgevoerd met behulp van een schriftelijke vragenlijst. Deze vragenlijst bestond behalve uit vragen over de algemene werkwijze bij het evalueren, ook uit vragen over het gebruik van de leerdoelgerichte toetsen. Bij de vragen over de algemene werkwijze moesten we ons, met het oog op de omvang van de vragenlijst, een beperking opleggen.

We hebben ons alleen gericht op het evalueren van de uitkomsten, de resultaten van onderwijsleerprocessen. Deze vorm van evaluatie wordt in de literatuur aangeduid als productevaluatie (vgl. De Bruyne, 1983; De Corte, 1978; De Corte et al., 1981). Aan de hand van de leerresultaten van leerlingen wordt nagegaan in welke mate de gestelde doelstellingen zijn bereikt. Het is gebruikelijk om productevaluatie te onderscheiden van procesevaluatie, waarbij men het onderwijsleerproces zelf aan een directe evaluatie onderwerpt. Activiteiten in dat verband kunnen zijn het analyseren van de doelstellingen, de werkvormen, hulpmiddelen en dergelijke. Proces- en productevaluatie worden in de regel als aanvullende activiteiten opgevat. In het algemeen zal de noodzaak van procesevaluatie toenemen, wanneer uit de productevaluatie blijkt, dat de doelstellingen niet bereikt zijn. Aan procesevaluatie is geen aandacht besteed. Een verdere beperking was dat er alleen vragen opgenomen zijn over het evalueren met behulp van *schriftelijke* evaluatie-instrumenten. Mondelinge overhoringen zijn bijvoorbeeld buiten beschouwing gelaten³.

Over het algemeen traden er bij de resultaten geen grote verschillen op tussen leerkrachten uit de afzonderlijke vakken en afzonderlijke schooltypen. De verschillen tussen de vakken waren vrij gering en weinig consistent. De

verschillen tussen de schooltypen waren groter en vertoonden bovendien meer overeenstemming. In dit artikel ligt daarom de nadruk op de presentatie van de resultaten voor alle leerkrachten gezamenlijk. In voorkomende gevallen gaan we beknopt in op de verschillen tussen schooltypen. Voor een gedetailleerde beschrijving van de onderzoeksresultaten verwijzen we naar het al genoemde onderzoeksverslag (Kremers, 1982)⁴.

3 Evaluatie tijdens en na het onderwijsleerproces

Het is gebruikelijk om bij de evaluatie van de resultaten van onderwijsleerprocessen een onderscheid te maken tussen het verzamelen van informatie over die resultaten en het nemen van beslissingen op grond van een waardering, een beoordeling van de verzamelde informatie (vgl. De Bruyne, 1983; De Corte, 1978; De Corte et al., 1981; Gronlund, 1976). In paragraaf 3.1 gaan we in op de manier waarop leerkrachten informatie over de resultaten verzamelen. In paragraaf 3.2 gaan we in op de beslissingen die leerkrachten op basis van deze informatie nemen. (Aan de beoordeling van de verzamelde informatie besteden we in paragraaf 6 aandacht.)

3.1 Evaluatie-instrumenten

Hoe en hoe vaak verzamelen leerkrachten in de eerste twee à drie leerjaren van het voortgezet onderwijs in het algemeen informatie over de vorderingen van de leerlingen? In het onderzoek zijn door ons een drietal gangbare evaluatie-instrumenten in het voortgezet onderwijs onderscheiden. We vermelden ze hieronder met de omschrijving zoals die in de vragenlijst werd gehanteerd.

1. *Afsluitende trimesterproefwerken*: deze in-

Tabel 2 *Gehanteerde evaluatie-instrumenten en hun afnamefrequentie (percentages en gemiddelden)*

Instrument	Frequentie					\bar{x}
	nooit*	soms	tamelijk vaak	meestal	altijd	
Trimesterproefwerk	47	27	5	7	14	2.1
Proefwerk	3	11	14	30	42	4.0
Schriftelijke overhoring	5	28	23	24	20	3.3

* antwoordcategorieën en hun omschrijving:

1 = nooit; 2 = soms (d.w.z. in ongeveer een *kwart* van het totaal aantal keren dat men het betreffende instrument zou kunnen afnemen); 3 = tamelijk vaak (d.w.z. in ongeveer de *helft* ... etc.); 4 = meestal (d.w.z. in ongeveer *driekwart* ... etc.); 5 = altijd

strumenten bestrijken de leerstof van ongeveer een trimester.

2. *Proefwerken*: deze instrumenten bestrijken de leerstof van ongeveer een maand, bijvoorbeeld een hoofdstuk uit de leergang.
3. *Schriftelijke overhoringen*: deze instrumenten bestrijken de leerstof van ongeveer een of enkele weken.

Tabel 2 geeft aan hoe vaak leerkrachten bij hun evaluaties gebruik maken van deze instrumenten. Uit de tabel blijkt dat leerkrachten relatief het meest regelmatig informatie verzamelen over de vorderingen met behulp van proefwerken, gevolgd door respectievelijk schriftelijke overhoringen en trimesterproefwerken. Leerkrachten nemen gemiddeld genomen een proefwerk af in ongeveer driekwart van de keren dat zich daar de gelegenheid toe voordoet ($\bar{x} = 4.0$). Schriftelijke overhoringen nemen ze gemiddeld af in ruim de helft van het aantal keren ($\bar{x} = 3.3$) en trimesterproefwerken in een kwart van het aantal keren ($\bar{x} = 2.1$). Over het algemeen treden er tussen individuele leerkrachten geen al te grote verschillen op in de mate waarin ze van de onderscheiden instrumenten gebruik maken. De spreiding van de antwoorden over de antwoordcategorieën is niet erg groot. Alleen bij de schriftelijke overhoring zijn de antwoorden vrijwel gelijk verdeeld over de antwoordcategorieën.

3.2 *Functie(s) van evaluatie*

Voor welke doeleinden nemen leerkrachten de bovengenoemde instrumenten af? Welke beslissingen nemen zij op basis van de verzamelde informatie? De verschillende gevolgen die in zo'n beslissing besloten kunnen liggen komen tot uitdrukking in wat doorgaans wordt aangeduid als de *functie* van de evaluatie. Evaluaties kunnen diverse functies hebben zoals selectie en plaatsing, certificering, sturing van het onderwijsleerproces (formatieve evaluatie), resultaatbeoordeling (summatieve evaluatie) en diagnose (vgl. De Bruyne, 1983; Gronlund, 1976).

In het onderzoek is vooral aandacht geschonken aan formatieve en summatieve (produkt)evaluatie. *Formatieve evaluatie* is de activiteit waarbij tijdens het onderwijsleerproces de resultaten van de leerlingen worden bepaald. *Formatieve evaluatie* is gericht op het zo nodig bijsturen, verbeteren van het onderwijsleerproces dat nog aan de gang is. *Summatieve*

evaluatie is de activiteit waarbij na afloop van het onderwijsleerproces bepaald wordt in hoeverre de leerlingen zich het onderwezene eigen gemaakt hebben. Essentieel is dat, in tegenstelling tot formatieve evaluatie, een kwaliteitsbeoordeling van de leerling plaatsvindt. Vaak gebeurt dat in de vorm van een cijfer (Bloom, Hastings en Madaus, 1971; De Bruyne, 1983; De Corte, 1978).

Rekening houdend met het bovenstaande zijn in het onderzoek de volgende functies onderscheiden:

1. het geven van een cijfer aan leerlingen;
2. het opsporen van tekorten in de beheersing van de leerstof bij (individuele) leerlingen, met het doel deze tekorten weg te werken;
3. andere doeleinden.

Leerkrachten werd gevraagd wat de functie van de evaluaties is waarbij ze de informatie verzamelen via de eerder onderscheiden trimesterproefwerken, proefwerken en schriftelijke overhoringen. Men kon daarbij meer functies per evaluatie aangeven. Tabel 3 geeft een overzicht van de antwoorden. Trimesterproefwerken worden het meest afgenomen om alleen een cijfer te geven, gevolgd door proefwerken en schriftelijke overhoringen. Het omgekeerde doet zich voor ten aanzien van de functie 'tekorten opsporen'. Combinatie van de functies 'cijfer geven' en 'tekorten opsporen' komt in ongeveer gelijke mate voor bij proefwerken en schriftelijke overhoringen; bij trimesterproefwerken komt deze combinatie aanzienlijk minder voor. De evaluaties hebben nauwelijks andere doeleinden (al of niet gecombineerd met cijfer geven en/of tekorten opsporen). Alleen schriftelijke overhoringen worden nog wel eens voor andere doeleinden afgenomen.

In het algemeen treedt het geschetste patroon op ongeacht het schooltype van de leerkracht. Leerkrachten uit het LBO en uit het LBO-AVO (-VWO) benadrukken echter wel minder het geven van cijfers en meer het opsporen van tekorten dan leerkrachten uit de overige schooltypen. Dit verschil treedt op ongeacht het instrument waarmee de evaluatie uitgevoerd wordt.

Overzien we bovenstaande gegevens dan valt op dat de functie 'cijfer geven' de meeste nadruk krijgt. Het louter opsporen van tekorten komt alleen op uitgebreidere schaal voor bij schriftelijke overhoringen (38%). Anders geformuleerd, de onderscheiden instrumenten

Tabel 3 *Functies van evaluaties (percentages).*

Instrument \ Functie	cijfer geven	tekorten opsporen	cijfer geven en tekorten opsporen	andere doeleinden*
Trimesterproefwerk	75	3	17	5
Proefwerk	48	9	39	4
Schriftelijke overhoring	13	38	34	15

* inclusief combinaties met 'cijfer geven' en/of 'tekorten opsporen'

worden vooral voor summatieve en in mindere mate voor formatieve evaluatie gebruikt. Verder valt op dat nogal wat leerkrachten hetzelfde instrument voor verschillende functies gebruiken. Proefwerken en schriftelijke overhoringen worden door meer dan $\frac{1}{3}$ deel van de leerkrachten gebruikt om zowel een cijfer te geven als om tekorten in de beheersing van de leerstof op te sporen. Bij een dergelijke werkwijze worden in de evaluatieliteratuur een tweetal bezwaren geplaatst. Een eerste bezwaar is dat de constructie van instrumenten voor deze twee functies idealiter verschillend verloopt, bijvoorbeeld wat betreft de moeilijkheidsgraad van de opgaven en de dekking van de leerstof. Een instrument dat optimaal geschikt is voor beide functies is daarom moeilijk realiseerbaar (Bloom et al., 1971; Gronlund, 1976). Een tweede bezwaar is dat een kwaliteitsbeoordeling van de door de leerling geleverde prestatie eerder bij summatieve dan bij formatieve evaluatie opportuun is (Bloom et al., 1971; De Bruyne, 1983).

Wanneer we de gegevens uit deze en de vorige paragraaf samenvoegen, ontstaat een beeld van de evaluatiepraktijk dat in sterke mate gedomineerd wordt door het geven van cijfers. In paragraaf 3.1 constateerden we dat leerkrachten de resultaten relatief het meest met behulp van proefwerken evalueren. Uit Tabel 3 valt op te maken dat bijna 90% van de leerkrachten deze instrumenten afneemt om een cijfer te geven, en dat bijna de helft ze uitsluitend afneemt om een cijfer te geven. We constateerden verder dat leerkrachten na proefwerken bij hun evaluaties relatief het meest gebruik maken van schriftelijke overhoringen. Zoals gezegd wordt dit instrument het meest gebruikt voor het uitsluitend opsporen van tekorten. Ook echter de schriftelijke overhoring wordt door bijna de helft van de leerkrachten afgenomen voor het geven van cijfers (vgl. Tabel 3).

Concluderend kan gesteld worden dat de door leerkrachten uitgevoerde evaluaties overwegend summatief van aard zijn. De beperkte plaats die formatieve evaluatie in de evaluatiepraktijk inneemt is in tegenspraak met het belang dat daaraan in de literatuur wordt toegekend. In de algemene discussie gaan we nader in op deze discrepantie tussen 'theorie' en 'praktijk'.

4 *Evaluatie van het begingedrag*

In de literatuur wordt veel waarde gehecht aan de evaluatie van het begingedrag van leerlingen. Inzicht in het begingedrag kan er toe bijdragen dat het onderwijsleerproces optimaal wordt afgestemd op de reeds aanwezige kennis en vaardigheden van leerlingen (vgl. De Corte et al., 1981; Gronlund, 1976). In het onderzoek zijn we nagegaan of leerkrachten wel eens het begingedrag van leerlingen evalueren voorafgaand aan de behandeling van een nieuw leerstofonderdeel. Er werden daarbij twee varianten onderscheiden:

1. *voortoetsing*: de evaluatie van de kennis en vaardigheden van de leerlingen, voorafgaand aan de behandeling van een nieuw leerstofonderdeel, vindt plaats om te bepalen hoeveel aandacht aan de behandeling van dit nieuwe leerstofonderdeel dient te worden geschonken;
2. *retentietoetsing*: de evaluatie van de kennis en vaardigheden van de leerlingen, voorafgaand aan de behandeling van een nieuw leerstofonderdeel, vindt plaats om te bepalen of leerstof waaraan lange tijd geen aandacht meer is besteed, maar die wel bekend wordt verondersteld bij de behandeling van dit nieuwe leerstofonderdeel, inderdaad nog beheerst wordt.

Evaluatie van het begingedrag blijkt niet erg populair bij leerkrachten. Voortoetsing komt zo goed als niet voor, retentietoetsing vindt

plaats in minder dan een kwart van het aantal keren dat zich daar de gelegenheid toe voordoet (\bar{x} bedraagt respectievelijk 1.3 en 1.7; dezelfde antwoordcategorieën zijn gehanteerd zoals omschreven bij Tabel 2). Leerkrachten uit het LBO en het LBO-AVO (-VWO) voeren beide varianten iets vaker uit dan leerkrachten uit de overige schooltypen. Gezien de bescheiden omvang van de evaluatie van het begingedrag zullen we er alleen nog in de discussie op terugkomen.

5 Correctie van resultaten

Tot voor kort was de correctie van de resultaten vrijwel uitsluitend een zaak voor de leerkracht. Met de 'opkomst' van formatieve evaluatie – althans in de literatuur – wordt ook aan de leerling een rol toebedacht bij de correctie van de resultaten (Cito, 1978; Nuy, 1973; Warries, 1971). Het inschakelen van de leerlingen bij de correctie bij deze vorm van evaluatie kan om tweeërlei redenen wenselijk zijn. In de eerste plaats zal een leerkracht, wanneer hij de vorderingen van de leerlingen nauwgezet wil volgen, regelmatig moeten nagaan hoe het met deze vorderingen gesteld is. Een frequente afname van evaluatie-instrumenten is dan noodzakelijk. De correctie van de resultaten bij een dergelijke frequente afname kan echter een aanzienlijke belasting voor de leerkracht betekenen. Dit kan voorkomen worden door de leerlingen de correctie uit te laten voeren. De correctie door de leerlingen heeft verder als belangrijk voordeel dat zij een inzicht krijgen in hun eigen fouten en tekortkomingen. Aangezien formatieve evaluaties (idealiter) niet bedoeld zijn voor het geven van cijfers zullen leerlingen in het algemeen niet geneigd zijn bij de correctie te frauderen. Om diezelfde reden wordt echter in de literatuur het zelf laten corrigeren van de resultaten door de leerlingen bij summatieve evaluaties afgeraden.

In het onderzoek zijn drie werkwijzen bij de correctie van de resultaten onderscheiden:

1. de leerkracht corrigeert de resultaten;
2. de leerling corrigeert de resultaten⁵;
3. het varieert wie corrigeert.

Leerkrachten werd gevraagd aan te geven welke van deze werkwijzen zij volgen bij de afname van trimesterproefwerken, proefwerken en schriftelijke overhoringen. Trimesterproefwerken en proefwerken worden vrijwel uit-

sluitend door de leerkracht nagekeken (respectievelijk 98% en 97%). Voor schriftelijke overhoringen bedraagt dit percentage 74%; 22% van de leerkrachten laat de correctie van schriftelijke overhoringen zo nu en dan eens door de leerlingen uitvoeren, terwijl 4% dit altijd aan de leerlingen zelf overlaat. De verschillen tussen trimesterproefwerken en proefwerken enerzijds en schriftelijke overhoringen anderzijds treden op bij alle schooltypen. Leerkrachten uit het LBO en het LBO-AVO (-VWO) schakelen echter bij schriftelijke overhoringen leerlingen wat frequenter in bij de correctie dan leerkrachten uit de overige schooltypen. Bij beide schooltypen bedraagt het percentage leerkrachten dat resultaten uitsluitend zelf corrigeert 66%. Voor het AVO bedraagt dit percentage 76%, voor het AVO-VWO 81% en voor het VWO 94%. Overigens is ook bij het LBO en het LBO-AVO(-VWO) het percentage leerkrachten dat de schriftelijke overhoringen altijd door de leerlingen laat corrigeren gering (respectievelijk 7% en 4%).

Geconstateerd kan worden dat het alleen door de leerling laten corrigeren van de resultaten (nog) niet erg gebruikelijk is. Gezien de functie van de evaluatie waarvoor de onderscheiden instrumenten afgenomen worden kon dit tot op zekere hoogte verwacht worden. Trimesterproefwerken en proefwerken worden vooral afgenomen voor het geven van cijfers en een correctie uitsluitend door de leerkracht is dan voor de hand liggend (vgl. paragraaf 3.2). Bijna 40% van de leerkrachten neemt schriftelijke overhoringen alleen af voor het opsporen van tekorten. Op basis daarvan zou men verwachten dat bij dit instrument meer leerkrachten de correctie aan de leerlingen overlaten dan thans het geval is. Ten dele zou dit kunnen samenhangen met de door leerkrachten gehanteerde vraagvorm van de opgaven. Het laten corrigeren van gesloten (meerkeuze) opgaven door leerlingen is in de regel wat makkelijker uitvoerbaar dan de correctie van open opgaven. Een verdere verklaring kan zijn dat leerkrachten toch op de eerste plaats zelf een inzicht willen hebben in de prestaties van de leerlingen en de tekortkomingen daarin. Ook kan de behoefte aan een uniforme werkwijze een rol spelen. Trimesterproefwerken en proefwerken kijkt men meestal zelf na, dus volgt men deze werkwijze ook bij schriftelijke overhoringen.

6 Cesuurbepaling

Wanneer de informatie over de resultaten van de leerlingen verzameld en gecorrigeerd is, is de volgende stap in het evaluatieproces het beoordelen, het interpreteren van deze informatie. Centraal staat daarbij de bepaling van de cesuur, de grens tussen een voldoende en onvoldoende prestatie. In de literatuur worden in dit verband twee, nogal uiteenlopende, benaderingen onderscheiden (zie bijvoorbeeld Dousma en Horsten, 1980; Wijnen, 1977). Bedoeld zijn de absolute en de relatieve cesuurbepaling.

Bij *absolute cesuurbepaling* bepaalt de leerkracht *vooraf* vanuit de leerstof de cesuur (bijvoorbeeld: een leerling levert een voldoende prestatie wanneer hij tenminste zeven van de tien opgaven goed beantwoordt). De cesuur is dan onafhankelijk van de resultaten van de groep leerlingen. Bij *relatieve cesuurbepaling* wordt *achteraf* vanuit de resultaten van de groep leerlingen de cesuur vastgesteld (bijvoorbeeld: de cesuur wordt zodanig gekozen dat 60% van de leerlingen een voldoende en 40% een onvoldoende krijgt). Volgens de Groot (1966) worden de resultaten van leerlingen in het voortgezet onderwijs overwegend relatief beoordeeld en wel op een zodanige manier dat het percentage onvoldoendes in de tijd stabiel is. Deze gang van zaken wordt ook wel aangeduid als de 'wet van Posthumus'. Volgens deze wet blijkt de 'gemiddelde' leerkracht steeds ongeveer 25% onvoldoendes te geven.

Aan beide methodes van cesuurbepaling kleven nadelen. Zo wordt bij absolute cesuurbepaling geen rekening gehouden met beoordelingsfouten door de leerkracht van de moeilijkheidsgraad van de opgaven. Bij relatieve cesuurbepaling 'bepaalt' de groep leerlingen hoeveel kennis van de leerstof gemiddeld vereist is. Men heeft daarom benaderingen voorgesteld die elementen van beide methodes bevatten. Deze komen er op neer dat men

vooraf vaststelt hoeveel opgaven er tenminste goed moeten worden gemaakt en naderhand dit aantal corrigeert op basis van de geleverde prestatie door de leerlingen. Een feitelijke beschrijving van een dergelijke compromis-methode is echter nog niet beschikbaar (Wijnen, 1977).

In navolging van het bovenstaande werden in het onderzoek de volgende vier werkwijzen bij de cesuurbepaling onderscheiden:

1. cesuur vooraf bepalen;
2. cesuur achteraf bepalen op basis van de klasseresultaten;
3. cesuur via combinatie vooraf/achteraf bepalen;
4. cesuur op een andere manier bepalen.

Leerkrachten werd gevraagd op welk van deze manieren ze de cesuur bepalen wanneer ze de vorderingen evalueren via trimesterproefwerken, proefwerken en schriftelijke overhoringen. De antwoorden zijn weergegeven in Tabel 4. Over het algemeen zijn de verschillen tussen de instrumenten niet zo groot. Het aantal leerkrachten dat de cesuur vooraf of via een combinatie vooraf/achteraf bepaalt, schommelt zo rond de 40%. Bij schriftelijke overhoringen wordt het meest de cesuur vooraf bepaald (48%). Het aantal leerkrachten dat de cesuur uitsluitend achteraf bepaalt is vrijwel constant en bedraagt ongeveer 20%. Bepaling van de cesuur op een andere manier (4) vindt bij geen enkele instrument plaats.

Het geschetste patroon keert maar ten dele terug wanneer de resultaten voor de afzonderlijke schooltypen geanalyseerd worden. Leerkrachten uit het LBO en het LBO-AVO (-VWO) bepalen de cesuur het meest vooraf bij trimesterproefwerken en proefwerken en het minst bij schriftelijke overhoringen. Voor de andere schooltypen geldt juist het omgekeerde.

De verschillen tussen de schooltypen maken het wat moeilijk om conclusies aan de resultaten te verbinden. Gesteld kan in ieder geval worden dat het gegeven, dat slechts 20% van

Tabel 4 *Wijze van cesuurbepaling bij verschillende evaluatie-instrumenten (percentages)*

Cesuur	vooraf	achteraf	combinatie	anders
Instrument				
Trimesterproefwerk	42	18	40	0
Proefwerk	39	19	42	0
Schriftelijke overhoring	48	21	32	0

de leerkrachten de cesuur uitsluitend achteraf bepaalt, wat in tegenspraak is met de bevinding van De Groot (1966) dat leerkrachten overwegend relatief beoordelen. De Groots bevinding dateert echter al weer van ruim 15 jaar geleden en het is mogelijk dat het denken over de cesuurbepaling, niet in het laatst door De Groots eigen publikatie, een verandering heeft ondergaan⁶. Overigens blijft het de vraag, en dat is een bekend punt van kritiek van tegenstanders van de absolute methode, hoe absoluut 'absoluut' is. Anders geformuleerd, in hoeverre gaat een leerkracht, wanneer hij vooraf een beheersingsstandaard vanuit de leerstof formuleert, ook niet (impliciet) uit van het niveau van de groep leerlingen waarvan hij de voorde-ningen beoordeelt?

Een tweede opmerking betreft de wijze van cesuurbepaling bij schriftelijke overhoring. Gronlund (1976) stelt dat in principe bij formatieve én summatieve evaluatie zowel een absolute als een relatieve wijze van cesuurbepaling gevolgd zou kunnen worden. Gezien echter de verschillende functies van deze evaluaties is naar zijn mening bij formatieve evaluatie een absolute cesuurbepaling het meest voor de hand liggend. Eerder constateerden we dat schriftelijke overhoringen het meest van alle instrumenten afgenomen worden voor formatieve evaluatie. Desondanks is het aantal leerkrachten, dat bij schriftelijke overhoringen de cesuur alleen vooraf bepaalt, niet veel groter dan bij de andere instrumenten.

7 Herhaling bij onvoldoende beheersing van leerstof

Op welke manier werken leerkrachten tekorten weg wanneer ze constateren dat leerlingen de leerstof onvoldoende beheersen? Vindt dan een herhaling van de leerstof plaats en zo ja, op welke manier? Strikt genomen maakt de werkwijze bij onvoldoende beheersing van de

leerstof geen deel uit van het evaluatieproces. Aangezien echter het bepalen van tekorten bij formatieve evaluatie expliciet tot doel heeft deze tekorten weg te werken, is in het onderzoek aan dit 'weg werken' aandacht geschonken.

In paragraaf 3.2 is geconstateerd dat leerkrachten zowel trimesterproefwerken, proefwerken als schriftelijke overhoringen ten dele afnemen om tekorten op te sporen. Bij schriftelijke overhoringen benadrukken ze deze functie sterker dan bij de overige twee instrumenten. Bij de vragen over de werkwijze bij het weg werken van tekorten is in het midden gelaten met behulp van welk instrument deze tekorten vastgesteld zijn. In paragraaf 7.1 schetsen we welke werkwijze leerkrachten volgen. In paragraaf 7.2 gaan we na of leerkrachten ook evalueren of hun herhaling effectief is geweest⁷.

7.1 Werkwijze

In het onderzoek is een drietal situaties onderscheiden, al naar gelang het aantal leerlingen dat de leerstof onvoldoende beheerst.

Situatie 1:

een kleine groep leerlingen ($\frac{1}{3}$ of minder) beheerst de leerstof onvoldoende.

Situatie 2:

ongeveer de helft van de leerlingen beheerst de leerstof onvoldoende.

Situatie 3:

een grote groep leerlingen ($\frac{2}{3}$ of meer) beheerst de leerstof onvoldoende.

Leerkrachten moesten voor elk van deze situaties aangeven op welke manier zij in het algemeen de herhaling laten plaatsvinden. Daarbij konden zij een van de volgende manieren aangeven:

1. geen herhaling;
2. alleen klassikale herhaling;
3. alleen niet-klassikale herhaling;
4. klassikale herhaling in combinatie met een vorm van niet-klassikale herhaling.

Tabel 5 *Gepraktiseerde vormen van herhaling bij onvoldoende beheersing van leerstof (percentages)**

Herhaling	Aantal onvoldoendes	$\frac{1}{3}$ of minder	de helft	$\frac{2}{3}$ of meer
geen herhaling		10	2	1
alleen klassikaal		11	57	81
alleen niet-klassikaal		68	20	3
klassikaal en niet-klassikaal		11	21	15

* Als percentagebasis fungeerde het totaal aantal leerkrachten per kolom.

Bij niet-klassikale herhaling (3) werden nog de volgende mogelijkheden onderscheiden: via een vorm van interne differentiatie (bijvoorbeeld via het basisstof-verrijkingstofmodel); via de studielessen en/of via extra huiswerk; via hulp na schooltijd; op een andere manier.

In Tabel 5 is weergegeven op welke manier leerkrachten in het algemeen de herhaling laten plaatsvinden. Zoals verwacht mocht worden, neemt het aantal leerkrachten dat de herhaling alleen klassikaal laat plaatsvinden sterk toe naarmate het aantal onvoldoende beheersers groter wordt. Het omgekeerde geldt voor het aantal leerkrachten dat de herhaling alleen via niet-klassikale vormen laat verlopen. Geen herhaling komt alleen op beperkte schaal voor in situatie één (aantal niet-beheersers: $\frac{1}{3}$ of minder). Opvallend is dat ook reeds in situatie twee (aantal niet-beheersers: de helft) ruim de helft van de leerkrachten de herhaling alleen klassikaal laat verlopen; nemen we daar nog de leerkrachten bij die de herhaling deels klassikaal / deels niet-klassikaal laten verlopen, dan voert bijna 90% van de leerkrachten een vorm van klassikale herhaling uit.

In situatie één zijn de meest genoemde methodes van niet-klassikale herhaling 'hulp na schooltijd' (16%), 'interne differentiatie' (12%), 'studielessen en/of extra huiswerk' in combinatie met 'hulp na schooltijd' (10%). In situatie twee zijn de meest genoemde methodes 'interne differentiatie' (10%) en 'interne differentiatie' in combinatie met 'studielessen en/of extra huiswerk' (3%).

Het patroon, zoals geschetst in Tabel 5, treedt op bij alle schooltypen. Analyse van de resultaten per situatie levert echter een aantal verschillen op tussen de schooltypen. De verschillen zijn vooral opvallend in situatie één. In deze situatie wordt door leerkrachten uit het LBO en het LBO-AVO (-VWO) interne differentiatie het meest genoemd als methode van (niet-klassikale) herhaling (respectievelijk 16% en 29%). Bij het AVO neemt interne differentiatie een lage plaats in (9%), terwijl deze werkwijze in het AVO-VWO en het VWO nauwelijks genoemd wordt. Leerkrachten uit het AVO, VWO en het AVO-VWO laten de herhaling het meest plaatsvinden via hulp na schooltijd. Leerkrachten uit het VWO doen dit het meest (29%; AVO: 18%; AVO-VWO: 16%).

Ook in situatie twee zijn de verschillen nog aanzienlijk. Leerkrachten uit het LBO en het

LBO-AVO (-VWO) kiezen hier het minst voor een klassikale benadering (49% en 45%), leerkrachten uit het VWO het meest (82%). Leerkrachten uit het AVO en het AVO-VWO nemen een tussenpositie in (ongeveer 60%). Opvallend is weer het aantal leerkrachten uit het LBO-AVO (-VWO) dat de herhaling alleen laat plaatsvinden via interne differentiatie (20%). De plaats van interne differentiatie is bij het LBO nu wat minder belangrijk geworden (11%), al is deze plaats nog steeds belangrijker dan bij de andere schooltypen.

Als we bovenstaande gegevens overzien, moet geconstateerd worden dat de herhaling bij onvoldoende beheersing van de leerstof minder vaak plaatsvindt via interne differentiatie dan op grond van de aandacht voor deze werkwijze in de laatste jaren verwacht zou mogen worden (zie bijvoorbeeld: Bade en Bult, 1977 en 1981; Geerligts en Van der Veen, 1980; Nijhof en Van Hout, 1979; Nuy, 1973).

De voorkeur van leerkrachten – met uitzondering van het LBO en vooral het LBO-AVO (-VWO) – gaat vooralsnog uit naar meer vertrouwde benaderingen als hulp na schooltijd, extra huiswerk en last but not least de klassikale herhaling. Ongetwijfeld zal dit samenhangen met de organisatorische en inhoudelijke problemen die interne differentiatie dikwijls met zich meebrengt. In de algemene discussie komen we er nog op terug.

7.2 *Nameting na herhaling*

In het onderzoek zijn we tenslotte nagegaan of leerkrachten ook evalueren of een herhaling effect heeft gehad. Gemiddeld voeren leerkrachten in ongeveer de helft van het aantal gevallen een dergelijke evaluatie uit ($\bar{x} = 3.1$; dezelfde antwoordcategorieën zijn gehanteerd zoals omschreven bij Tabel 2). De onderlinge verschillen tussen de leerkrachten zijn echter aanzienlijk. Een deel van de leerkrachten voert zo'n 'nameting' meestal of altijd uit, terwijl een ander deel dat slechts af en toe doet. Er is sprake van een tweetoppige frequentieverdeling. Leerkrachten uit het LBO en het LBO-AVO (-VWO) voeren wat vaker een nameting uit dan leerkrachten uit de overige schooltypen.

Ook voor de nameting na herhaling is gevraagd op welke manier de correctie wordt uitgevoerd en de cesuur wordt bepaald (vgl. paragraaf 5 en 6). Leerkrachten volgen bij de correctie van de resultaten op nametingen on-

geveer dezelfde werkwijze als bij schriftelijke overhoringen: 74% kijkt nametingen uitsluitend zelf na, 18% laat het nakijken zo nu en dan eens over aan de leerlingen, terwijl 8% dit altijd aan de leerlingen overlaat. Wat betreft de cesuurbepaling, leerkrachten bepalen bij nametingen na herhaling veel vaker de cesuur vooraf dan bij trimesterproefwerken, proefwerken en schriftelijke overhoringen: 57% bepaalt de cesuur vooraf, 16% achteraf en 25% bepaalt de cesuur via een combinatie vooraf/achteraf. Op zich is dat niet verwonderlijk. De nameting na herhaling vormt de tweede evaluatie van dezelfde leerstof. In die situatie zal de leerkracht vaak over voldoende informatie beschikken om de cesuur vooraf te bepalen.

8 Discussie en aanbevelingen

8.1 Algemene discussie

Uit het onderzoek komt een beeld naar voren van de evaluatiepraktijk in de onderbouw van het voortgezet onderwijs dat men 'traditioneel' zou kunnen noemen. Traditioneel in die zin dat inzichten, die de laatste twee decennia min of meer gemeengoed zijn geworden in de evaluatieliteratuur, nog maar weinig invloed lijken te hebben op deze praktijk. We wijzen op de beperkte plaats van formatieve evaluatie en evaluatie van het begingedrag. We noemen verder het gering aantal leerkrachten dat, ook bij formatieve evaluatie, de correctie van de resultaten door de leerlingen laat uitvoeren. We vermelden daarnaast de constatering dat herhaling van de leerstof veelal klassikaal plaatsvindt; interne differentiatie is nog weinig ingeburgerd. Alleen de werkwijze die leerkrachten volgen bij de cesuurbepaling past niet zonder meer in dit beeld.

Opvallend is dat dit beeld over het algemeen optreedt ongeacht het vak of het schooltype van de leerkracht. Anders geformuleerd, de evaluatiepraktijk in de diverse schooltypen en bij de diverse vakken vertoont een grote mate van overeenstemming. Wel gaan leerkrachten uit het LBO en het LBO-AVO(-VWO) in de regel wat minder traditioneel te werk dan leerkrachten uit het AVO, het AVO-VWO en het VWO.

We zullen het geschetste beeld nu vanuit een drietal gezichtspunten bediscussiëren, te weten de geldigheid van het beeld, de representati-

teit voor leerkrachten in de onderbouw van het voortgezet onderwijs en de oorzaken voor de discrepantie tussen 'theorie' en 'praktijk'.

Het onderzoek had betrekking op het evalueren met behulp van *schriftelijke* instrumenten. De ruimte ontbrak om ook vragen te stellen over het evalueren met behulp van mondelinge evaluatietechnieken. Bekend in het voortgezet onderwijs zijn de mondelinge overhoring aan het begin van de les en de mondelinge beurt tijdens de les. Het is mogelijk dat leerkrachten frequenter de resultaten formatief evalueren of dat ze vaker een evaluatie van het begingedrag of van het effect van een herhaling uitvoeren dan in het onderzoek geconstateerd is, doordat ze deze evaluaties ook met behulp van mondelinge technieken uitvoeren. Het geschetste beeld zou dan misschien maar ten dele opgaan. De Bruyne (1983) merkt in dit verband overigens op, dat schriftelijke beantwoording door de leerlingen in het algemeen de voorkeur verdient boven mondelinge beantwoording. Mondeling uitgevoerde evaluaties dragen in de regel een subjectiever, inefficiënter en minder systematisch karakter dan schriftelijk uitgevoerde evaluaties. Vervolgonderzoek is gewenst om meer zicht te krijgen op de functie en frequentie van evaluaties die met behulp van mondelinge technieken uitgevoerd worden.

Een tweede punt van discussie betreft de representativiteit van de resultaten. Directe informatie over de representativiteit ontbreekt (vgl. paragraaf 2.1). Twee vrij recent uitgevoerde onderzoeken geven hier echter wel enig uitsluitsel over. Uit een door het ITS uitgevoerd onderzoek blijkt, dat wanneer scholen voor voortgezet onderwijs vernieuwingsactiviteiten ontplooiën, ze dit het minst doen op het gebied van evaluatie. Slechts 5% van de scholen onderneemt activiteiten op dit gebied. Dit in tegenstelling tot bijvoorbeeld de leerlingbegeleiding, de vormgeving van de brugperiode, de aansluiting basisonderwijs - voortgezet onderwijs en dergelijke. Aan het onderzoek werd meegewerkt door 1600 scholen (alle schooltypen) en het mag als representatief beschouwd worden (Van Galen en Bulte, 1982). In het onderzoek zijn geen concrete vragen gesteld over de werkwijze bij het evalueren. Het tweede onderzoek betreft een onderzoek door het Licor naar de behoefte aan een geautomatiseerde opgavenbank bij scholen in het voortgezet onderwijs. In het kader van dit

onderzoek zijn aan een beperkt aantal leerkrachten uit het AVO-VWO (± 50) ook vragen gesteld over de werkwijze bij het evalueren van leerresultaten. De leerkrachten waren afkomstig uit scholen die belangstelling hadden voor geautomatiseerde opgavenbanken. Het onderzoek leverde een vergelijkbaar beeld op van de evaluatiepraktijk als het onderhavige onderzoek (Van de Perel en Schneider, 1983).

Gegeven deze resultaten zijn wij van mening dat de resultaten van ons onderzoek waarschijnlijk redelijk representatief zijn voor leerkrachten in de onderbouw van het voortgezet onderwijs. Vervolgonderzoek zal echter moeten aantonen dat deze veronderstelling juist is.

Als de signaleerde discrepantie tussen de evaluatieliteratuur en de evaluatiepraktijk correct is, waardoor wordt zij dan veroorzaakt? Deze vraag is niet eenvoudig te beantwoorden. We beperken ons tot een opsomming van een aantal mogelijke verklaringen waarom leerkrachten vooralsnog vrij 'traditioneel' te werk gaan bij het evalueren.

Een eerste verklaring kan zijn dat bij de voorgestelde werkwijzen in de evaluatieliteratuur onvoldoende rekening wordt gehouden met de voorwaarden waaraan voldaan moet worden, willen deze werkwijzen in het onderwijs toepasbaar zijn. Zo is het opsporen van tekorten in de vorderingen van de leerlingen als ook het wegwerken van deze tekorten een tijds- en arbeidsintensieve aangelegenheid. Hetzelfde kan gezegd worden van het afnemen van voorttoetsen en retentietoetsen en het uitvoeren van nametingen na herhaling. Ook met het bepalen van een cesuur vooraf is meer tijd gemoeid dan met het bepalen van een cesuur achteraf. Zelfs al zou een leerkracht meer tijd aan het evalueren willen besteden, dan heeft hij deze in het vigerende onderwijssysteem lang niet altijd ter beschikking. De leerstof die per jaar behandeld moet worden is meestal vrij precies omschreven en is van een zodanige omvang dat een leerkracht weinig speelruimte heeft. Hij moet voort en heeft geen tijd om al te lang bij tekorten stil te staan. Naast dit gebrek aan tijd moet ook de 'druk' op de leerkracht om rapportcijfers te geven genoemd worden. Voor de bepaling daarvan zal hij in de loop van het schooljaar voldoende cijfers moeten verzamelen. In dat verband is het niet zo verwonderlijk dat schriftelijke overhoringen zowel afgeno-

men worden om tekorten op te sporen als om cijfers te geven. Voor een uitgebreidere beschrijving van de hier aangestipte aspecten van de lessituatie van leerkrachten in het voortgezet onderwijs verwijzen we naar bijvoorbeeld Adviesgroep Landelijk Brugklasproject (1978), Nuy en Timmer (1979), Van de Perel en Schneider (1983).

We hebben de indruk dat de eerder genoemde verschillen tussen het LBO en LBO-AVO(-VWO) enerzijds en de overige schooltypen anderzijds ten dele met de hier gegeven verklaring samenhangen. In de eerst genoemde schooltypen wordt in de regel de (voortgang door de) leerstof wat minder centraal gesteld en het onderwijs meer afgestemd op de specifieke capaciteiten van de leerlingen dan in de overige schooltypen.

Een aanvullende verklaring kan zijn dat leerkrachten op zich hun evaluatiepraktijk wel zouden willen en kunnen veranderen, maar dat vernieuwingen op andere terreinen vooralsnog belangrijker gevonden worden. We verwijzen naar de uitkomsten van het eerdergenoemde ITS-onderzoek. Een verdere verklaring kan zijn dat leerkrachten onvoldoende kennis hebben van de evaluatieliteratuur. Uit buitenlands onderzoek blijkt dat deze kennis vaak ontbreekt (Janssens, 1984). Het gevolg hiervan is dat leerkrachten hun evaluatiepraktijk niet als problematisch ervaren, of dat, wanneer ze dat wel doen, ze niet weten hoe ze deze praktijk kunnen veranderen. Wellicht dat in dit verband ook een rol gespeeld heeft dat een instituut als het Cito zich tot nu toe overwegend heeft bezig gehouden met de constructie van toetsen en de verspreiding daarvan in het onderwijs. De laatste jaren heeft inmiddels de kennisproductie en -verspreiding bij leerkrachten met betrekking tot evalueren meer nadruk gekregen in de werkzaamheden (Cito, 1983).

8.2 Aanbevelingen voor verder onderzoek

In een vervolgonderzoek zou nagegaan moeten worden of de resultaten, zoals voorzichtig verondersteld, representatief zijn voor leerkrachten in de onderbouw van het voortgezet onderwijs. Dit onderzoek zou zich ook kunnen uitstrekken over vakken die niet in het onderzoek opgenomen zijn, bijvoorbeeld aardrijkskunde, geschiedenis, natuurkunde e.d. Overwogen zou verder kunnen worden om het onderzoek uit te breiden naar de bovenbouw van het voortgezet onderwijs. Vooralsnog denken

we niet dat dit een ander beeld zal opleveren, of het zou een nog traditioneler moeten zijn. In de bovenbouw wordt de druk van de te behandelen leerstof onder invloed van het naderende examen alleen nog maar groter. De beschikbare tijd voor minder traditionele werkwijzen neemt dan nog verder af.

In het onderzoek is aandacht besteed aan de algemene werkwijze bij evalueren. In een vervolgonderzoek zou ingegaan kunnen worden op de gedetailleerde werkwijze die leerkrachten volgen bij hun evaluaties. Te denken valt aan aspecten als de constructie en inhoud van de evaluatie-instrumenten (omvang, onderdelen, vraagvorm), de correctie van de resultaten (foutenanalyse), de cesuurbepaling (concretisering van de absolute en de relatieve methode), de (rapport) cijferbepaling en dergelijke. Een aanzet hiertoe vindt men bij Van de Perel en Schneider (1983). Verder is al betoogd dat in een vervolgonderzoek nagegaan zou moeten worden wat de functie en frequentie is van mondeling uitgevoerde evaluaties.

In de vorige paragraaf is een aantal verklaringen gegeven voor het traditionele karakter van de evaluatiepraktijk in het voortgezet onderwijs. Toetsing van deze verklaringen op hun realiteitswaarde is gewenst. Alleen dan kan men gerichte maatregelen treffen voor het verbeteren van deze praktijk.

Noten

1. Voor een beschrijving van de constructie en de gebruiksmogelijkheden van deze toetsen verwijzen we naar de bijbehorende docentenhandleiding (Cito, 1978). Informatie over de constructie vindt men ook bij Van Bergen en Peters-Sips (1980) en Kuhlemeier en Van Werkhoven (1983).
2. Het is niet aannemelijk dat de algemene werkwijze bij het evalueren, zoals wij die hier rapporteren, beïnvloed is door het bestellen van leerdoelgerichte toetsen. Van de leerkrachten die aan het onderzoek meewerkten had (op dat moment) slechts 15% regelmatig van deze toetsen gebruik gemaakt. Bij deze groep van regelmatige gebruikers was het bovendien eerder zo dat men de leerdoelgerichte toetsen inpaste in de algemene werkwijze die men reeds bij het evalueren volgde, dan dat het omgekeerde het geval was (Kremers, 1982).
3. De termen 'mondeling' en 'schriftelijk' hebben geen betrekking op de wijze van vragenstellen, maar op de wijze van beantwoording door de

- leerlingen. Men spreekt daarom ook van schriftelijke instrumenten, wanneer de vragen mondeling gesteld worden en de antwoorden schriftelijk gegeven worden (vgl. De Bruyne, 1983).
4. De resultaten zijn ook nog per vak per schooltype geanalyseerd. In het algemeen keerde het antwoordpatroon voor de schooltypen voor alle leerkrachten tezamen ook terug bij elk vak afzonderlijk. Er was slechts incidenteel sprake van interactie tussen vak en schooltype. (Kremers, 1982).
 5. Bij deze werkwijze zijn in principe nog twee varianten mogelijk: 1. de leerling kijkt zijn eigen werk na; 2. de leerling kijkt het werk van een van zijn medeleerlingen na. Voor een dergelijke uitwerking van de vraagstelling ontbrak de ruimte.
 6. Ook in de geactualiseerde tiende druk van De Groots publicatie uit 1966 wordt verondersteld dat het relatieve beoordelen aan omvang heeft ingeboet (De Groot en Wijnen, 1983).
 7. Leerkrachten die te kennen gaven alle genoemde instrumenten uitsluitend af te nemen om leerlingen een cijfer te geven, hoefden de vragen bij dit onderdeel niet te beantwoorden. Het ging hierbij om minder dan 10% van het totaal aantal leerkrachten.

Literatuur

- Adviesgroep Landelijk Brugklasproject, *Eindadvies*. Amsterdam: LPC, 1978.
- Bade, J. J. & J. Bult, *Leren met verschillen (reeks)*. Hoevelaken: CPS, 1977.
- Bade, J. J. & J. Bult, *De praktijk van de interne differentiatie*. Nijkerk: Callenbach, 1981.
- Bergen, J. B. A. M. van & M. W. H. Peters-Sips, *Inventarisatie van leerboeken in het eerste en tweede leerjaar van het voortgezet onderwijs*. Arnhem: Cito, Algemene publicatie nr. 13, 1980.
- Bloom, B. S., J. Th. Hastings & G. F. Madaus, *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill, 1971.
- Bruyne, H. C. D. de, *Evalueren in de klas*. Amsterdam: Van Goor Zonen, 1983.
- Cito, *Handleiding leerdoelgerichte toetsen*. Arnhem: Cito, 1978.
- Cito, *Handleiding leerdoelgerichte toetsen*. Arnhem: Cito, 1982 (tweede gewijzigde uitgave).
- Cito, *Plannen van het Cito voor 1984 tot en met 1988*. Arnhem: Cito, 1983.
- Corte, E. De, *Evaluatie van het didactisch handelen. Handboek voor de onderwijspraktijk*. Deventer: Van Loghum Slaterus, 1978.
- Corte, E. De, C. T. Geerligns, N. A. J. Lagerweij, J. J. Peters & R. Vandenberghe, *Beknopte didaxologie*. Groningen: Wolters-Noordhoff, 1981.
- Dousma, T. & A. Horsten, *Tentameneren*. Utrecht: Spectrum, 1980.

- Galen, A. M. van & J. A. Bulte, *Veranderen is gewoon*. Nijmegen: ITS, 1982.
- Geerligs, T. & Tj. van der Veen, *Lesgeven. Interne differentiatie in de praktijk*. Assen: Van Gorcum, 1980.
- Gronlund, N. E., *Measurement & evaluation in teaching*. New York: MacMillan, 1976.
- Groot, A. D. de, *Vijven en zessen*. Groningen: Wolters, 1966 (eerste druk).
- Groot, A. D. de & W. H. F. W. Wijnen, *Vijven en zessen*. Groningen: Wolters-Noordhoff, 1983 (tiende druk).
- Janssens, F. J. G., *De relatie tussen evalueren en onderwijs*. Arnhem: Cito, Documentatiereeks nr. 110, 1983.
- Janssens, F. J. G., *Toetsgebruik getoetst; een literatuurverkenning*. Arnhem: Cito, Documentatiereeks nr. 122, 1984.
- Kremers, E. J. J., *Gebruikers- en behoeftenonderzoek project Leerdoelgerichte Toetsen*. Arnhem: Cito, Documentatiereeks nr. 68 en 69, 1982.
- Kuhlemeier, J. B. en R. F. van Werkhoven, *Inventarisatie van leerboeken talen in de onderbouw*. Arnhem: Cito, Algemene publikatie nr. 32, 1983.
- Nijhof, W. J. & J. van Hout, *Differentiatie in het onderwijs – Bijdragen tot de Onderwijsresearchdagen 1979*. Den Haag: SVO, SVO-reeks nr. 23, 1979.
- Nuy, M. J. G., *Achtergronden (reeks): 0. Van klassikaal naar meer gedifferentieerd onderwijs. 1. Mastery learning. 2. De 'learning for mastery strategy' van B. S. Bloom als model voor individualisering van het onderwijs*. Den Bosch: KPC, 1973.
- Nuy, M. J. G. en J. Timmer, Enkele ervaringen bij de introductie van mastery learning in het voortgezet onderwijs. In: E. Warries (Red.), *Beheersingsleren – een leerstrategie*. Groningen: Wolters-Noordhoff, 1979.
- Perel, R. van de & H. Schneider, *Een geautomatiseerde opgavenbank – Een hulpmiddel in het onderwijs?* Leiden: LICOR, 1983.
- Warries, E., Drie redenen om te toetsen in het onderwijs. *Pedagogische Studiën*, 1971, 48, 152-161.
- Wijnen, W. H. F. W., Toetsen en beslissingen. *Losbladig Onderwijskundig Lexicon*. Alphen a. d. Rijn: Samsom, 1977.

Curriculum vitae

E. J. J. Kremers (geb. 1952) studeerde ontwikkelingspsychologie aan de Katholieke Universiteit van Nijmegen (doctoraalexamen 1978). Sedert 1977 is hij werkzaam bij het Cito. Momenteel is hij als onderwijskundig medewerker verbonden aan de afdeling Beginfase Voortgezet Onderwijs van dit instituut. Hij is daar onder meer betrokken bij de constructie van leerdoelgerichte toetsen. Publiceerde onder andere over affectieve doelstellingen in het onderwijs.

Adres: Cito, Afdeling Beginfase Voortgezet Onderwijs, Postbus 1034, 6801 MG Arnhem.

Manuscript aanvaard 4-9-'84

Summary

Kremers, E. J. J. 'Evaluation of learning outcomes by teachers in secondary schools'. *Pedagogische Studiën*, 1984, 61, 496-508.

In general, little research has been done on the evaluation practices of teachers. In an exploratory study a questionnaire was used to investigate how teachers evaluate achievement in the first three years of secondary school. More than 900 teachers of various subjects from every type of school participated in the study. The results of this study indicate that evaluation practice can be characterized as traditional. The literature on evaluation appears to have had little influence. Summative evaluation is emphasized in contrast to formative evaluation or evaluation of entry behavior. One possible explanation is that in the literature too little attention is paid to the practical application of the methods proposed. Another explanation is that teachers are not familiar with the literature.