

# **BEOORDELEN VAN ONDERZOEKS- VAARDIGHEDEN VAN LEERLINGEN**

**richtlijnen, alternatieven en  
achtergronden**

kernredactie:

Dr. K.M. Stokking  
Drs. M.F. van der Schaaf

MesoConsult B.V.  
Tilburg

juni 1999

Deze brochure is geschreven door dr. Karel Stokking en drs. Marieke van der Schaaf van ISOR, Instituut voor Onderwijsresearch van de Universiteit Utrecht. Het project waarin dit product is ontwikkeld, is gefinancierd uit het budget dat het ministerie van OC&W jaarlijks beschikbaar stelt aan de LPC voor Kortlopend Onderwijsonderzoek op verzoek van het onderwijsveld (projectnummer 97.1.2.III). De tekst is aanvankelijk verschenen bij het ISOR (als rapport 99.02). De aanvragers van het project, conrectoren van scholen die actief bezig zijn met de invoering van de tweede fase, bepleitten een ruimere verspreiding. Eén van de suggesties die ze gaven was opname van de tekst in de Studiehuisreeks van MesoConsult. LPC, ISOR en MesoConsult bleken elkaar hierover snel te kunnen vinden. We hopen en verwachten dat de tekst op scholen nuttige functies kan vervullen, en staan open voor opmerkingen en suggesties.



Capaciteitsgroep Onderwijskunde, Universiteit Utrecht  
Heidelberglaan 2, 3584 CS Utrecht

**Instituut voor  
Onderwijsresearch** © 1999 MesoConsult B.V. Tilburg

Uit deze uitgave mag niets worden veeveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

ISSN-nummer 1384-2641

**Abonneren op de Studiehuisreeks of bestellen van losse exemplaren:**

MesoConsult  
Gounodlaan 15  
5049 AE Tilburg

Telefoon 013 - 456 03 11

Fax 013 - 456 32 76

## WOORD VOORAF

Het KPC coördineert, namens de LPC (de Landelijke Pedagogische Centra), sinds 1997 een nieuwe regeling voor 'Kortlopend onderzoek'. In het kader daarvan kunnen onderwijsinstellingen jaarlijks aanvragen voor onderzoek indienen. Deze vragen moeten in hooguit één jaar worden beantwoord, en wel in de vorm van een praktisch bruikbaar product. Tegelijk moet het onderzoek rekening houden met de nieuwste wetenschappelijke inzichten en standaarden.

Voor de uitvoering van de onderzoeken die onder deze regeling vallen, zijn allianties van onderzoeksinstituten gevormd. Deze allianties dienen offertes in, waarna volgens een bepaalde procedure elk onderzoek wordt toegewezen aan één alliantie. Begin 1998 kreeg het ISOR, Instituut voor Onderwijsresearch, opdracht voor een onderzoek naar het beoordelen van onderzoeksvaardigheden van leerlingen. Deze brochure is het eindproduct van dit onderzoek.

In het onderwijs zijn op dit moment veel ontwikkelingen gaande. Een deel daarvan betreft het nastreven van nieuwe doelstellingen (onder andere meer gericht op vaardigheden), een andere didactiek (naar meer zelfstandig leren) en nieuwe manieren van beoordelen (onder andere aan de hand van praktische opdrachten).

Over het beoordelen van producten van onderzoek door leerlingen verschijnen regelmatig publicaties en producten vanuit lerarenopleidingen, de SLO, het Cito, de LPC en andere. Daarin wordt geprobeerd een brug te slaan tussen wat het landelijk beleid wil, wat de wetenschap beweert en wat scholen en docenten accepteren en aankunnen. Waarom dan ook nog deze brochure? Deze publicatie beoogt een overzicht te geven van de huidige stand van zaken op dit gebied, gebruikmakend van zowel de wetenschappelijke literatuur als reeds verschenen producten en toegepaste publicaties. Ook wordt ingegaan op aspecten en achtergronden die tot nu toe nog weinig of niet zijn belicht.

De tekst is geschreven op de situatie in de tweede fase vo. Een flink deel van de inhoud is waarschijnlijk ook relevant in het mbo.

Deze brochure is bedoeld voor docenten, sectieleiders en schoolleiders. Enerzijds worden individuele docenten aangesproken, vanuit de opvatting dat het tot de professionele taak en verantwoordelijkheid van elke docent behoort zorgvuldig te beoordelen. Anderzijds komen allerlei zaken aan de orde die direct te maken hebben met het beleid dat een vaksectie, respectievelijk een school als geheel, wil voeren. De tekst bevat een aantal richtlijnen en suggesties, maar is vooral geschreven met als uitgangspunt dat scholen, secties en docenten zelf een aantal keuzes moeten maken op basis van doelen en argumenten.

Graag bedanken we op deze plaats allen met wie we in het kader van dit project van gedachten hebben kunnen wisselen.

*Op een aantal plaatsen in de tekst zijn (in cursief) resultaten opgenomen van een onderzoek, uitgevoerd in het najaar van 1998, onder een aantal docenten natuurkunde, biologie, geschiedenis en economie, naar de manier waarop zij leerlingen onderzoek laten doen en dat begeleiden en beoordelen. Uit de uitkomsten blijkt dat secties en docenten heel verschillende keuzes maken (maar de verschillen tussen de vakken zijn meestal niet groot). Ook de docenten die aan dit onderzoek meewerkten bedanken wij graag.*

Karel Stokking en Marieke van der Schaaf



INHOUD		pagina
<i>1</i>	<i>Inleiding</i>	9
1.1	Onderzoek door leerlingen temidden van andere vernieuwingen	9
1.2	Bedoeling en opbouw van deze brochure	10
Deel I Voor docenten die direct aan de slag willen		
<i>2</i>	<i>Samenvatting van richtlijnen en suggesties</i>	12
2.1	Inleiding	12
2.2	Vaststellen wat beoordeeld gaat worden	13
2.3	Maken van een aantal keuzes	14
2.4	Een uitgewerkt model als voorbeeld	15
2.5	Plaats, functie en gebruik van beoordelingen	17
2.6	Beoordelingen en cijfers	18
2.7	Eisen waaraan beoordelingen moeten voldoen	19
2.8	Professionele ontwikkeling en samenwerking met collega's	20
Deel II Leerdoelen, begeleiding, beoordeling		
<i>3</i>	<i>Doelstellingen, eindtermen, onderzoeksvaardigheden (bij 2.2)</i>	22
3.1	Onderzoeken van leerlingen: mogelijke doelstellingen	22
3.2	Onderzoeksvaardigheden in de eindtermen; soorten onderzoek	25
3.3	Onderzoek doen: stappen en vaardigheden	29
<i>4</i>	<i>Opdrachten, begeleiding, programmering (bij 2.5)</i>	37
4.1	Doelen, opdrachten, begeleiding, beoordeling	37
4.2	Leerlijnen, didactische aanpakken, begeleiding en feedback	38
4.3	Afstemming tussen vakken, opdrachten en profielwerkstuk	46
<i>5</i>	<i>Beoordelen: functie, aanpak, context (bij 2.6)</i>	50
5.1	Meten, evalueren, toetsen, beoordelen	50
5.2	Beoordelingscriteria, scoringsmodel, normering, cijfers	52
5.3	Registreren, communiceren en verantwoorden	54

### Deel III Keuzemogelijkheden en overwegingen

6	<i>Manieren en aspecten van beoordelen (bij 2.3 en 2.4)</i>	55
6.1	Inleiding	55
6.2	Proces, product en/of presentatie	56
6.3	Logboek en/of geïntegreerd eindverslag	58
6.4	Tussentijdse beoordelingen en/of eindbeoordeling	59
6.5	Beoordelen op vakinhoud en/of onderzoek	60
6.6	Beoordelen van kennis en/of vaardigheden	61
6.7	Globale beoordeling en/of gespecificeerd scoringsmodel	62
6.8	Categorisch en/of vergelijkend beoordelen	63
6.9	Beoordeling en normering op één of meer niveaus	65
6.10	Beoordelen bij tussentijdse bijsturing door de docent	66
6.11	Individuele beoordeling van groepsprestaties	68
6.12	Beoordeling in relatie tot de keuzevrijheid van de leerlingen	70
6.13	Controleerbaarheid op juistheid van gegevens en uitkomsten	71
6.14	Vermoedens van en controle op fraude	72
6.15	Straf aftrek en/of bonuspunten	73
6.16	Leerlingen als medebeoordelaars	75
6.17	Zelfbeoordeling door leerlingen	76
6.18	Onvoldoendes en herkansingen	77

### Deel IV Beoordelen als professionele taak van docenten

7	<i>Leren beoordelen als ontwikkelproces (bij 2.7 en 2.8)</i>	79
7.1	Beoordelen als onderdeel van de taak van een docent	79
7.2	Gedragcodes voor toetsen, beoordelen en beslissen	80
7.3	Planmatig ontwikkelen van professioneel beoordelen	86
7.4	Samenwerking met collega's en schoolbeleid	89

### Deel V Wetenschappelijke achtergronden

8	<i>Kennis en vaardigheden van leerlingen (bij 3)</i>	93
9	<i>Onderzoek als kennisdomein op zich (bij 3)</i>	97
10	<i>Beoordelen als onderdeel van het onderwijsleerproces (bij 4 en 5)</i>	98

<i>11 Eisen waaraan goede beoordelingen moeten voldoen (bij 2.7 en 5)</i>	100
Epiloog	104
Referenties	105





## 1 INLEIDING

### 1.1 Onderzoek door leerlingen temidden van andere vernieuwingen

Docenten in het voortgezet onderwijs worden eind negentiger jaren geconfronteerd met een aantal vernieuwingen tegelijk. We noemen hier:

- in de kerndoelen en eindtermen zijn expliciet algemene (vakoverstijgende) vaardigheden opgenomen; in de delen A van de examenprogramma's havo/vwo van veel vakken zijn vooral allerlei informatie- en onderzoeksvaardigheden opgenomen;
- in het nieuwe schoolexamen nemen praktische opdrachten en een afsluitend profielwerkstuk een belangrijke plaats in, waartoe leerlingen een onderzoek uitvoeren of een technisch of creatief ontwerp maken; de rol van de traditionele toetsen wordt beperkt;
- van docenten wordt verwacht dat zij leerlingen leren meer zelfstandig te werken en te leren en in groepjes samen te werken.

Deze ontwikkelingen komen samen in het zelfstandig, individueel of in kleine groepen opzetten en uitvoeren van onderzoeken door leerlingen. Voor de meeste docenten brengen deze ontwikkelingen nieuwe taken en andere werkwijzen met zich mee. Naast kennis worden ook vaardigheden belangrijk, niet alleen vakspecifieke vaardigheden, maar ook vaardigheden van meer of minder vakoverstijgende aard. Voor veel docenten is het nieuw om zo expliciet aandacht te geven aan de ontwikkeling van vaardigheden bij leerlingen.


De nieuwe aanpak heeft nogal wat consequenties. Leerlingen moeten meer aan het werk worden gezet via opdrachten. Deze moeten voor een groot deel eerst nog worden ontwikkeld. De opdrachten zullen vaak een enigszins open karakter hebben: leerlingen maken gebruik van verschillende bronnen en hulpmiddelen, volgen verschillende routes en komen met verschillende resultaten. In veel gevallen zal niet slechts één antwoord goed zijn. Van docenten wordt een meer begeleidende rol verwacht en ook deze taak vergt dat zij zich veelal nieuwe kennis en vaardigheden eigen maken. En tenslotte: docenten zullen het onderzoekswerk van hun leerlingen moeten beoordelen. Deze brochure beoogt te laten zien wat daarbij de mogelijkheden zijn en waarmee men rekening moet houden om verantwoord te beoordelen.

Helemaal nieuw is het allemaal natuurlijk niet. De bètavakken hadden al hun practica, de talen hun spreekbeurten, de gammavakken hun werkstukken. Met meer open onderzoek door leerlingen hebben echter weinig docenten ervaring. In feite is 'onderzoek' als domein van kennis en vaardigheden in de mate waarin het nu in de eindtermen is opgenomen voor bijna alle vakken een nieuw onderdeel van het programma. Docenten zullen verschillen in hun affiniteit met en kennis over onderzoek (in relatie tot hun vak).

### 1.2 Bedoeling en opbouw van deze brochure

De bedoeling van deze publicatie is docenten, secties en scholen te ondersteunen bij het beoordelen van onderzoekswerk van leerlingen. Daartoe worden richtlijnen en suggesties, discussiemateriaal en achtergronden aangedragen.

*Deel I*, hoofdstuk 2, bevat een samenvatting van richtlijnen en suggesties, waarmee docenten meteen aan de slag kunnen. De auteurs hebben daartoe een aantal keuzes gemaakt, gebaseerd op aspecten, overwegingen en achtergronden



die zijn opgenomen in de delen II t/m V. Veel docenten zullen, vroeger of later, zelf hun keuzes willen maken. Zij kunnen dan bij de volgende delen te rade gaan (zie ook de inhoudsopgave).

*Deel II* bevat de belangrijkste directe achtergrond: wat wordt er beoordeeld, welke plaats heeft de beoordeling in het totale onderwijsproces, en hoe gaat een beoordeling praktisch in zijn werk?

*Deel III* beschrijft een aantal aspecten, waarbij docenten steeds verschillende keuzes kunnen maken; in de tekst staan overwegingen die daarbij een rol kunnen spelen.

*Deel IV* beschrijft hoe docenten, individueel of in onderlinge samenwerking, hun beoordelingswerk systematisch kunnen ontwikkelen.

*Deel V* geeft verdiepende informatie over wat (onderzoeks)vaardigheden zijn, en over professioneel beoordelen in het onderwijs.

Docenten kunnen individueel met deze brochure aan de slag. Vaak echter zal het voordelen bieden om over de manier van beoordelen in het team afspraken te maken en elkaar daarbij te ondersteunen. Deze brochure biedt daartoe de nodige discussiestof. Door de opbouw in delen en hoofdstukken kan men zelf kiezen welke onderdelen men aan de orde wil stellen. Paragraaf 2.8 en hoofdstuk 7 geven suggesties voor de route die men daarbij zou kunnen volgen.

## *Deel I Voor docenten die direct aan de slag willen*

### **2 SAMENVATTING VAN RICHTLIJNEN EN SUGGESTIES**

#### **2.1 Inleiding**

Dit hoofdstuk is toegevoegd nadat de rest van deze brochure al klaar was. Met enige aarzeling. Immers: rond beoordelen kunnen allerlei keuzes worden gemaakt; die moeten passen bij het beleid, de doelen en de mogelijkheden van de eigen school en sectie en de individuele docent.

Om de lezer toch enigszins op weg te helpen hebben de auteurs zelf een aantal keuzes gemaakt, die in dit hoofdstuk op een rij worden gezet. Het zijn min of meer algemene keuzes, niet gebonden aan een bepaald vak of soort onderzoek. Het gaat om een minimumset, waarmee docenten die nog weinig ervaring hebben met het beoordelen van eigen onderzoekswerk van leerlingen naar ons idee goed aan de slag kunnen gaan. De argumenten voor de door ons gemaakte keuzes staan in de rest van deze publicatie - naast andere argumenten en andere keuzes.

De in dit hoofdstuk verwerkte keuzes hebben betrekking op:

- wat beoordeeld wordt, wanneer en aan de hand waarvan (paragraaf 2.2);
- welke kennis en vaardigheden aan de hand van welke activiteiten en materialen (proces, tussenproducten, eindproduct, logboek, presentatie);
- een aantal 'technische' keuzes daarbij over de manier van beoordelen (paragraaf 2.3);
- beide voorgaande gecombineerd in een uitgewerkt model als voorbeeld (paragraaf 2.4);
- de plaats van de beoordeling en het gebruik ervan (paragraaf 2.5);
- het op grond van beoordelingen geven van cijfers (paragraaf 2.6);
- eisen waaraan beoordelingen moeten voldoen (paragraaf 2.7);
- de manier waarop een docent, respectievelijk een team, zichzelf op dit vlak verder kan professionaliseren (paragraaf 2.8).

Deze brochure probeert iets toe te voegen aan de praktijk. De lezer zal zich er even in moeten verdiepen. Een publicatie waarbij dat niet nodig zou zijn, zou geen meerwaarde hebben, en dus net zo goed niet kunnen verschijnen.

#### **2.2 Vaststellen wat beoordeeld gaat worden**

Onderzoekswerk van leerlingen kan worden beoordeeld op de vakinhoud, het onderzoek als onderzoek en het proces (samenwerking en dergelijke). Onze eerste suggestie is om zich (voorlopig) tot beide invalshoeken te beperken: de vakinhoud en het onderzoek en dus de samenwerking tussen de leerlingen vooralsnog niet mee te nemen bij de beoordeling. Verdere keuzes zijn:

- de beoordeling van het onderzoek concentreert zich op de volgende stappen of onderdelen: vraagstelling, onderzoeksopzet, werkplan, verzameling van gegevens, verwerking van gegevens, conclusies, evaluatie,

verslaglegging. Hiermee blijven 'standpuntbepaling' en 'presentatie' dus vooralsnog bij de beoordeling van onderzoeksvaardigheden buiten beschouwing;

- de beoordeling vindt plaats aan de hand van producten. Het proces speelt bij de beoordeling een rol op basis van tussentijds producten die leerlingen inleveren en een beschrijving van de belangrijkste zaken in het eindverslag. De leerlingen wordt dus niet gevraagd voor de beoordeling een logboek bij te houden en het presenteren wordt niet gerekend tot de kern van de te beoordelen onderzoeksvaardigheden;
- de beoordeling betreft uitsluitend de vakinhoud en het onderzoek. Er worden geen andere zaken 'impliciet meegenomen', zoals de samenwerking of de inzet. Veel docenten doen dit, maar daardoor wordt de beoordeling minder informatief en daarmee ook minder leerzaam voor de leerlingen, over de kennis en vaardigheden (vakinhoud en onderzoek) waarom het gaat.
- het eindoordeel wordt opgebouwd op basis van verschillende deeloordelen. Deze deeloordelen worden gecombineerd door ongewogen optelling (waardoor zwakke onderdelen worden gecompenseerd door sterke) of door per onderdeel een bepaald minimumniveau te eisen. De beoordeling vindt dus niet plaats door in één keer 'op het gevoel' een cijfer te geven. Hoewel ervaren docenten hiermee niet sterk verkeerd hoeven te zitten, is een beoordeling die wordt opgebouwd uit een aantal expliciete deeloordelen in het algemeen betrouwbaarder, objectiever en informatiever;
- aan onderzoek zitten zowel een 'doe-kant' als een 'denk-kant'. De beoordeling beperkt zich niet tot de doe-kant. Het (vakinhoudelijke) probleem en de oplossing daarvan door middel van beredeneerd onderzoek staan centraal.

*[Achtergronden: zie hoofdstuk 3, paragrafen 6.2-6.6, en hoofdstukken 8 en 9]*

### **2.3 Maken van een aantal keuzes**

De docent zorgt, als het goed is, voor consistentie tussen doel, opdracht, begeleiding en beoordeling. De opdracht maakt expliciet welke keuzes leerlingen zelf kunnen maken. Over de doelen, begeleiding en beoordeling worden zo mogelijk afspraken gemaakt tussen collega's. Vervolgens:

- de docent bespreekt vooraf met de leerlingen wat op welk moment op welke manier wordt beoordeeld, op welke criteria en met welke bedoeling;
- de beoordeling vindt plaats via deeloordelen (dus niet in één keer). De betrouwbaarheid en objectiviteit worden ondersteund door vergelijkingen, bijvoorbeeld door per criterium alle producten te beoordelen (in plaats van andersom) en/of door de producten eerst globaal te ordenen van zwak naar goed en voor de becijfering daarin ankerpunten te kiezen ('dit is de beste om die redenen en is een negen waard; deze is een stuk minder' et cetera);
- de beoordelingscriteria en de toepassing daarvan worden zoveel als mogelijk gericht op het bepalen van het inhoudelijke niveau ofwel de vakinhoudelijke kennis en onderzoeksvaardigheden. Een beoordeling moet informatie geven over wat leerlingen weten en kunnen. Niveauverschillen (bijvoorbeeld havo/vwo) spelen mee via de 'strengheid' bij elke deelbeoordeling;
- de gegeven begeleiding (hulp, bijsturing) wordt globaal verdisconteerd in de vorm van aftrek voor duidelijk tekortschietende zelfstandigheid;
- de groepjes waarin leerlingen onderzoek doen zijn heterogeen samengesteld; individualiseren van het cijfer gebeurt aan de hand van een of meer andere gegevens (bijvoorbeeld individuele bevraging). In de begeleiding besteedt de docent aandacht aan de samenwerking, maar deze wordt niet beoordeeld;

- desgewenst wordt op te laat inleveren een sanctie gesteld via cijferaftrek (mits daarnaast een inhoudelijke, gedifferentieerde beoordeling plaatsvindt);
- desgewenst wordt bij de becijfering een bonus gegeven als een onderzoek opvallend moeilijk was (complex, risicovol, bewerkelijk en dergelijke) en als de inhoudelijke uitwerking van opvallend hoog niveau en diepgang is;
- de docent ligt niet wakker van fraude: aan het onderzoek worden specifieke eisen gesteld, er is voldoende tussentijds contact, ook tussenproducten worden beoordeeld, de leerlingen moeten de uitvoering beschrijven, er wordt individueel nabesproken en/of tijdens de presentatie worden vragen gesteld.

[Achtergronden: zie hoofdstukken 4 en 5, en paragrafen 6.6-6.18]

## 2.4 Een uitgewerkt model als voorbeeld

Het voorgaande kan worden omgezet in figuur 1. Daarin staan tien onderdelen waarop een onderzoek kan worden beoordeeld. Per onderdeel is een korte toelichting opgenomen. Desgewenst kan men een of meer onderdelen splitsen in subonderdelen, die apart herkenbaar worden beoordeeld. Met tien onderdelen kan een beoordeling echter al behoorlijk gedifferentieerd plaatsvinden en bevindingen opleveren die voor leerlingen en docenten informatief zijn.

De tien onderdelen betreffen de vakinhoud en de stappen in een onderzoek. Samenwerking, standpuntbepaling en presentatie zijn hierin niet opgenomen. Het onderdeel 'vakinhoud' (nummer 10) lijkt wellicht schraaf af te tekenen bij alle stappen die op het onderzoek als onderzoek betrekking hebben. In feite echter wordt ook bij de onderdelen 1, 2, 4, 6, 7 en 8 mede op de vakinhoud (juistheid, consistentie, diepgang en dergelijke) beoordeeld. Onderdeel 10 voegt daaraan een extra criterium toe, waarin een afrondend oordeel kan worden gegeven.

	O = onvoldoende V = voldoende G = goed	beoordeling			opmerkingen
		O	V	G	
<b>1 probleemstelling</b> onderwerpkeuze, inbedding in context, vraagstelling, deelvragen, hypothesen: duidelijkheid, motivering, onderzoekbaarheid (afgebakend, eenduidig)					
<b>2 onderzoeksopzet</b> duidelijkheid, adequaatheid (gezien de probleemstelling), uitvoerbaarheid					
<b>3 werkplan</b> haalbaarheid (ook per fase), eventueel taakverdeling, voorzien in monitoring en momenten voor bijsturing					
<b>4 instrumentarium, selectie bronnen</b> technische kwaliteit, passing bij de vraagstelling					
<b>5 verzameling gegevens</b> zorgvuldigheid (veiligheid, privacy en dergelijke), controle op kwaliteit en bruikbaarheid voor het onderzoek, systematiek, volledigheid, doelgerichtheid					
<b>6 verwerking gegevens</b> betreft: bewerken, ordenen, weergeven, analyseren, interpreteren en dergelijke; criteria zoals bij onderdeel 5					
<b>7 conclusies, beantwoording vraag</b>					

duidelijkheid, correcte conclusies uit gegevens, beantwoording vraagstelling			
<b>8 evaluatie van het onderzoek</b> consistentie tussen stappen (1, 2, 4, 5, 6), sterke en zwakke punten aangegeven, verbetermogelijkheden, geldigheidsbereik conclusies, eventueel vervolgvragen			
<b>9 verlagegging</b> duidelijkheid opbouw, volledigheid, communicatieve kwaliteit (vormgeving, leesbaarheid), taalgebruik			
<b>10 vakinhoudelijk</b> inbedding probleemstelling, grondigheid van de uitwerking, niveau en diepgang			
<i>totaal aantal punten</i>			
tijdigheid inleveren			
zelfstandigheid			
moeilijkheid onderzoek			
niveau, diepgang			

Figuur 1 Beoordelingslijst voor onderzoeken van leerlingen (ISOR, 1999)  
(per beoordelingsmoment de betreffende onderdelen beoordelen)

Bij voorkeur is sprake van enkele tussentijdse beoordelingsmomenten, bijvoorbeeld na afronding van de onderdelen 1 t/m 3, 4 t/m 7, respectievelijk 8 t/m 10. In dat geval kan bij de beoordeling van het eindproduct (verslag) worden volstaan met beoordeling van de onderdelen 8 t/m 10; de andere onderdelen zijn dan immers al beoordeeld. Wie besluit om alleen het eindproduct te beoordelen, dus alle onderdelen in één keer, kan hetzelfde schema gebruiken en op basis van het verslag alle onderdelen 1 t/m 10 beoordelen.

De beoordeling geschiedt per onderdeel op een driepuntsschaal: onvoldoende, voldoende, goed. Hierin wordt mede verwerkt het niveau waarop men beoordeelt (bijvoorbeeld: havo, vwo); daarbij heeft men uiteraard normen nodig, maar die zullen eerst geleidelijk kunnen worden ontwikkeld op basis van ervaring. In het schema wordt ervan uitgegaan dat men bij de becijfering vier extra aspecten meeneemt met betrekking tot het onderzoek als geheel: het tijdig inleveren, het min of meer zelfstandig werken, de moeilijkheid van het onderzoek en het niveau of de diepgang. Beide eerste punten leiden bij afwezigheid tot aftrek, beide laatste punten leiden bij aanwezigheid tot een bonus (zie verder paragraaf 2.6).

## 2.5 Plaats, functie en gebruik van beoordelingen

Beoordelen (zie hiervoor) gaat vooraf aan en is dus iets anders dan cijfers geven (zie hierna). Beoordelingen geven primair informatie over wat een leerling weet en kan en zijn 'criteriumgeoriënteerd'. In de praktijk van het cijfers geven spelen diverse andere factoren mee, die begrijpelijk en legitiem zijn gezien de functies die cijfers vervullen, maar vaak afbreuk doen aan de kenmerken en gebruiksmogelijkheden van een goede beoordeling.

Beoordelingen en beoordelingscriteria kunnen in het onderwijs verschillende functies vervullen:

- het vooraf bekendmaken en bespreken van belangrijke beoordelingscriteria kan fungeren als een effectieve vorm van instrueren en motiveren: leerlingen maken kennis met een complexe activiteit, die buiten de school door allerlei mensen wordt uitgevoerd en aan bepaalde eisen moet voldoen en die leerlingen zich ook eigen kunnen

maken;

- inzichtelijke en objectieve beoordelingen van onderdelen en aspecten van het onderzoekswerk bieden informatie over wat de leerlingen kennen en kunnen; goede feedback bevordert het beoogde leerproces;
- een duidelijke lijst met criteria, die vanaf het begin een plaats heeft in de instructie en begeleiding en voor de leerlingen gaat leven, maakt beoordeling door leerlingen van elkaars en eigen werk mogelijk (dit kan het leerproces bevorderen en de docent enigszins ontlasten);
- expliciet gemaakte criteria en procedures voor de beoordeling van prestaties van leerlingen zijn belangrijk in de communicatie met en verantwoording aan anderen (leerlingen, ouders, collega's, schoolleiding, inspectie).

Resultaten van beoordelingen kunnen worden gecommuniceerd als 'profiel', een overzicht van de sterke en zwakke punten van de geleverde prestatie (zie bijvoorbeeld paragraaf 2.4). Op basis van beoordelingsresultaten kunnen (totaal)cijfers worden gegeven (zie paragraaf 2.6).

*[Achtergronden: zie hoofdstukken 4 en 5, en paragrafen 6.16-6.17]*

## 2.6 Beoordelingen en cijfers

Op basis van een ingevulde beoordelingslijst (figuur 1) kan men desgewenst een cijfer geven. Het aantal van tien onderdelen maakt het mogelijk om, als men verdere weging achterwege laat, de beoordeling direct in een schoolcijfer om te zetten via de puntentelling onvoldoende = 0, voldoende = 0,5 en goed = 1. Door optelling ontstaat een cijfer tussen 0 en 10.

We nemen aan dat omzetting van de beoordeling in een cijfer alleen plaatsvindt om te voorzien in eventuele behoeften van leerlingen aan een cijfermatige waardering (zij zijn dit immers gewend) en/of om mee te tellen bij een rapport of in het kader van het schoolexamen. In alle gevallen gelden de stelregels, dat men geen meer gedetailleerde cijfers geeft dan men kan waarmaken (dus niet tot in decimalen) en dat het cijfer zoveel mogelijk blijft weergeven wat een leerling op het betreffende domein weet en kan (in plaats van ook andere zaken te herbergen).

De voorgestelde omzetting van de beoordelingen 'onvoldoende', 'voldoende' en 'goed' houdt, uitgaande van de gebruikelijke 5,5 (respectievelijk 6) als grens tussen onvoldoende en voldoende, tegelijk een bepaalde normering in. Uiteraard kan men desgewenst bepaalde onderdelen zwaarder wegen dan andere, maar dan is een aanvullende omrekening nodig. Overigens zit in de beoordeling per onderdeel al een belangrijk normerend element, waarbij men bijvoorbeeld ook kan variëren tussen havo en vwo. De suggestie om de behaalde deelscores eenvoudigweg op te tellen betekent dat zwakke onderdelen door sterke onderdelen kunnen worden gecompenseerd. Een alternatief is dat men per onderdeel of per onderzoeksfase een bepaalde minimumscore verlangt. De consequentie is dan dat leerlingen die het minimum niet halen een onvoldoende krijgen. In de praktijk zal men veelal leerlingen zodanig begeleiden dat ze zinvol met hun onderzoek verder kunnen, respectievelijk de mogelijkheid bieden om verbeteringen aan te brengen (zie ook paragraaf 6.18). De toegevoegde aspecten 'tijdigheid' en 'zelfstandigheid' scoren onvoldoende of voldoende. In het eerste geval kan men (maximaal) 1 punt aftrekken. De toegevoegde aspecten 'moeilijkheid onderzoek' en 'niveau en diepgang' scoren voldoende of goed. In het tweede geval kan men (maximaal) 1 punt als bonus geven. (In zeer uitzonderlijke gevallen behalen leerlingen daardoor een '11').

*[Achtergronden: zie hoofdstuk 5, paragraaf 6.18, en hoofdstuk 10]*

## 2.7 Eisen waaraan beoordelingen moeten voldoen

Beoordelen heeft te maken met meten (waarvoor eenduidig moet zijn wat wordt gemeten), toetsen (van wat iemand weet en kan), waarderen (met het oog op een of meer functies: geven van feedback, nemen van beslissingen en dergelijke) en evalueren (van gegeven onderwijs, opdrachten en begeleiding). Beoordelingen moeten zoveel mogelijk aan de volgende vier eisen voldoen.

### ***Betrouwbaar en valide meten***

*Betrouwbaarheid.* Als een docent enkele weken later producten nog een keer beoordeelt of als een collega de producten ook beoordeelt, mogen de uitkomsten niet teveel verschillen. De betrouwbaarheid kan worden bevorderd door het opstellen van duidelijke en eenduidige aanwijzingen die de kans op misverstanden, fouten en vergissingen zo klein mogelijk maken.

*Validiteit.* Met de beoordeling moet worden gemeten wat men wil meten en niet (ook) iets anders. De beoordeling moet aansluiten bij de behandelde stof, gegeven opdracht en begeleiding en zicht geven op de ontwikkeling van de relevante (kennis en) vaardigheden, respectievelijk op de kwaliteit van de prestatie.

### ***Adequaat beoordelen***

*Praktische bruikbaarheid.* Docenten moeten de beoordeling kunnen uitvoeren (qua bekwaamheid en beschikbare tijd) en de resulterende beoordelingen moeten de gewenste functies kunnen vervullen (zicht bieden op hoever de leerlingen zijn en wat er nog aan schort, geschikt zijn voor het meetellen bij een rapport of examen en voor communicatie met leerlingen en ouders). Kortom: de beoordeling moet uitvoerbaar zijn en betekenis hebben.

*Aanvaardbaarheid.* De procedure en de criteria moeten duidelijk en inzichtelijk zijn en kunnen worden uitgelegd aan anderen (leerlingen, ouders, collega's, schoolleiding, inspectie). Ze moeten objectief zijn en dus te verdedigen zijn. Leerlingen moeten weten waar ze aan toe zijn en moeten gelijkwaardig worden behandeld, respectievelijk gelijkwaardige kansen krijgen. In de praktijk speelt een grote rol in hoeverre docenten kunnen vermijden om 'oneigenlijke' zaken mee te laten spelen (zoals hun 'voorkennis' of 'totaalindruk' van een leerling) en erin slagen om de bedoelingen, opzet en (mogelijke) implicaties effectief te communiceren.

*[Achtergronden: zie paragrafen 5.1, 7.1, 7.2, en hoofdstuk 11]*

## 2.8 Professionele ontwikkeling en samenwerking met collega's

Docenten werken niet in een vacuüm. Ze zijn lid van een sectie en van een schoolteam. Op elk niveau kunnen keuzes worden gemaakt en deze keuzes kunnen meer of minder goed op elkaar zijn afgestemd. Individuele docenten kunnen voor zichzelf systematisch werken aan de ontwikkeling van hun beoordelingspraktijk. Zij kunnen daarvoor een eigen leerlijn uitzetten. Een stap verder en docenten maken in sectieverband afspraken. Die zullen dan om te beginnen betrekking hebben op de doelen van onderzoek door leerlingen, de geleidelijke opbouw en de didactische aanpak. Vervolgens kan men een beoordelingsmodel ontwikkelen, respectievelijk daarover consensus bereiken (criteria, scoring, weging, normering). Het verdient aanbeveling een voorlopig vastgestelde beoordelingsaanpak systematisch uit te proberen. Daarvoor moet het een en ander worden bewaard (gegeven opdrachten, daarmee beoogde doelen, aantekeningen van het verloop en de begeleiding, de gehanteerde criteria en dergelijke). Op basis daarvan kan men realistische normen gaan ontwikkelen (over wat van leerlingen valt te verwachten). Anderzijds kan men zicht krijgen op



de impact van bepaalde typen opdrachten en begeleidingsaanpakken, de eventuele verschillen tussen soorten onderzoek en dergelijke.

Ter controle op en bevordering van de betrouwbaarheid van de beoordelingen is het raadzaam een aantal producten door twee docenten te laten beoordelen: de begeleidende docent en een niet bij het onderzoek betrokken collega. Op sectieoverstijgend niveau is het van belang te komen tot afspraken over de belasting per tijdperiode van leerlingen met onderzoeksopdrachten bij de verschillende vakken. Verder is het zaak te vermijden dat docenten van verschillende vakken tegen elkaar inwerken, doordat de een de leerlingen andere zaken wil aanleren dan de ander. Op scholen waar men meer wil, respectievelijk waar meer mogelijk is, worden tussen secties afspraken gemaakt over de vaardigheden die men bij de leerlingen wil ontwikkelen en de bijdragen daaraan vanuit de verschillende vakken. Uiteindelijk hebben de doelen en didactiek rond onderzoek door leerlingen te maken met de identiteit en het klimaat van de school en dus met schoolbeleid.

*[Achtergronden: zie paragraaf 4.3 en hoofdstuk 7]*

### **3 DOELSTELLINGEN, EINDTERMEN, ONDERZOEKSVAAARDIGHEDEN (bij 2.2)**

#### **3.1 Onderzoeken van leerlingen: mogelijke doelstellingen**

Met het laten opzetten en uitvoeren van een onderzoek kan een docent verschillende doelen hebben. Hieronder staan zeven mogelijke doelstellingen. Docenten doen er goed aan bewust voor één of enkele doelen te kiezen. Dat geeft richting aan beslissingen over opdrachten, begeleiding en beoordeling.

*1 Leerlingen onderzoek laten doen als werkvorm om hun inhoudelijke kennis over het schoolvak verder te ontwikkelen.*

Dit lijkt voor de meeste vakken een haalbaar doel, mits de leerlingen vrij gericht aan het werk worden gezet. Leerlingen kunnen via onderzoek zelf bepaalde verschijnselen, processen, verschillen of verbanden ontdekken, respectievelijk daarover meer gedetailleerde kennis verzamelen en verwerken. Het gaat bij dit doel niet primair om leren onderzoeken, maar om onderzoekend leren. (Een belangrijke vraag hierbij is in hoeverre leerlingen zullen ontdekken, respectievelijk de kennis zullen ontwikkelen waar het de docent om te doen is).

*2 Leerlingen onderzoek laten doen om ze te motiveren zich actief met het vak bezig te houden en/of ze te laten oefenen met zelfstandig werken en leren.*

Ook hierbij gaat het om onderzoekend leren, maar dan gericht op andere doelen dan vakkennis. Dit doel is meer gericht op 'leren leren' en vooral de houdingskant daarbij. Leerlingen kunnen ervaren dat ze op eigen kracht kennis kunnen verkrijgen door vragen te stellen, informatie te zoeken, na te denken enzovoorts. Ook kunnen ze daardoor (meer) zelfvertrouwen ontwikkelen (wat voor bepaalde leerlingen heel belangrijk kan zijn).

*3 Leerlingen kennis laten opdoen van waar het bij onderzoek in grote lijnen om gaat: de taal, de manier van redeneren, het pendelen tussen theorie (begrippen), vraagstelling, onderzoeksopzet, gegevens en conclusies.*

Dit doel lijkt redelijk haalbaar. Voorop staat het verkrijgen van een overzicht en van het inzicht dat in onderzoek allerlei beslissingen worden genomen die onderling samenhangen. Voorzover leerlingen oefenen met het opzetten van een onderzoek, het uitvoeren en bewaken ervan en het trekken van conclusies kunnen zich op deze punten ook vaardigheden ontwikkelen.

*4 Leerlingen ervaring laten opdoen met deelvaardigheden.*

Deze doelstelling is, meer dan de vorige, expliciet op de ontwikkeling van vaardigheden gericht. Voor het opzetten en uitvoeren van een onderzoek zijn verschillende vaardigheden nodig, zoals probleemstelling formuleren, plannen, verzamelen en verwerken van informatie, rapporteren. Deze vaardigheden zullen niet alle evenveel aandacht kunnen krijgen. Docenten kunnen enkele vaardigheden kiezen die, naar hun idee, voor onderzoek (in hun vakgebied) het belangrijkste zijn. Ze kunnen bij het maken van zo'n keuze ook letten op de mate waarin de vaardigheden ook bij andere vakken aan bod komen. Daarmee kan een docent dan nog kiezen voor die vaardigheden die ook bij andere vakken aan de orde komen (want dan is de kans groter dat leerlingen een zekere vaardigheid ontwikkelen) of kiezen voor vaardigheden die bij andere vakken weinig aan bod komen (want dan leren de

leerlingen er in ieder geval bij het eigen vak iets van). Overleg tussen collega's van verschillende vakken is hierbij uiteraard zinvol.

- 5 *Leerlingen in kennis brengen met de instrumenten van een onderzoeker.*  
Dit doel betreft het doen van onderzoek als ambacht. De gereedchapskist van een onderzoeker bevat een aantal methoden, technieken en normen. Specifieke kennis hierover lijkt slechts beperkt te passen in het curriculum van het voortgezet onderwijs. Dat zou anders komen te liggen als 'onderzoek' als apart kennisdomein een plaats zou krijgen. Voor het vaststellen van de waarde van gegevens, het trekken van conclusies en het evalueren van een onderzoek is enige kennis hierover wel wenselijk.
- 6 *Leerlingen laten kennismaken met het 'ethos' van onderzoek doen.*  
Hierbij kan men denken aan zaken als vasthoudendheid, nieuwsgierigheid, systematiek, nauwkeurigheid, kritische houding, objectiviteit, eerlijkheid en controleerbaarheid. Om dit ethos ook echt te ervaren en te ontwikkelen is echter waarschijnlijk een meer intensief bezig zijn met onderzoek nodig dan in het vo realiseerbaar is.
- 7 *Leerlingen laten inzien dat kennis door mensen is en wordt ontwikkeld.*  
Dit doel heeft raakpunten met de doelen 1, 2, 3 en 6. Leerlingen kunnen leren dat kennis te maken heeft met 'de werkelijkheid', maar dat het ook gaat om denkconstructies die mensen zich over die werkelijkheid vormen. Ze kunnen leren dat het zoeken naar en ontwikkelen van kennis een open proces is (in plaats van dat kennis afgerond en statisch is). Ze kunnen leren dat kennis maatschappelijke functies vervult en ook verloren kan gaan.

Doel 2 kan desgewenst worden gekoppeld aan andere algemene vaardigheden zoals samenwerken. De doelen 3, 4 en 5 zijn wellicht het meest gericht op de voorbereiding, voor wat betreft het leren doen van onderzoek, op een studie in het hoger, respectievelijk wetenschappelijk) onderwijs. De doelen 6 en 7 kunnen misschien worden gerekend tot het gebied van de 'algemene vorming'. De onderzoeksvaardigheden zoals opgenomen in kerndoelen en eindtermen (zie paragraaf 3.2) hebben vooral betrekking op de doelen 3 en 4. Het accent dat je als docent legt zal mede afhangen van:

- de opvatting die je hebt over je eigen vak (hoe belangrijk is onderzoek daarin, in hoeverre is de vakkennis in voortdurende ontwikkeling, neemt in het vak wetenschappelijk onderzoek een belangrijke plaats in, hoeveel affiniteit heb je zelf met onderzoek);
- de verwachtingen die je hebt over wat leerlingen aankunnen en wat ze kunnen leren (valt te verwachten dat ze via eigen onderzoek inhoudelijke kennis ontwikkelen (doel 1), is het haalbaar om ze inzicht in onderzoek te laten krijgen (doel 3) of is een beperking tot deelvaardigheden (doel 4) het meest realistisch of (doel 2) tot onderzoek doen als werkvorm).

*Van een vijftigtal docenten tweede fase havo-vwo in de vakken natuurkunde, biologie, geschiedenis en economie, verspreid over enkele tientallen scholen en najaar 1998 bevraged, streven de volgende percentages in redelijke of sterke mate de genoemde doelen na:*

- 90 *onderzoek als werkvorm om op een motiverende manier kennis over het vak te laten opdoen*
- 85 *onderzoek als oefening in zelfstandig werken en leren*
- 55 *kennis laten opdoen van waar het bij een onderzoek in grote lijnen om gaat (natuurkunde, biologie vaker dan geschiedenis, economie)*
- 90 *ervaring laten opdoen met een aantal deelvaardigheden*
- 55 *kennis laten opdoen over de 'gereedchapskist' van de onderzoeker (methoden en technieken)*
- 55 *kennis laten maken met de houding en werkwijze van een onderzoeker (geschiedenis minder vaak dan de drie andere vakken)*

**3.2 Onderzoeksvaardigheden in de eindtermen; soorten onderzoek*****Onderzoeksvaardigheden in de eindtermen***

In de kerndoelen voor de basisvorming en de eindtermen havo-vwo is, bij wijze van voorbeeld, voor een aantal vakken nagegaan (in de delen A, algemene vaardigheden) welke daarvan betrekking hebben op onderzoek. Bij de meeste vakken zijn zulke eindtermen ondergebracht in de rubrieken 'informatievaardigheden' en 'onderzoeksvaardigheden'. Bij sommige vakken worden (ook) andere rubrieken gehanteerd. Dat wil overigens niet zeggen dat bij die vakken meer of heel andere typen eindtermen staan dan bij de vakken waar wel de rubrieken 'informatievaardigheden' en 'onderzoeksvaardigheden' zijn gebruikt. Dat is alleen bij de technisch-instrumentele vaardigheden bij natuurkunde en biologie het geval. Voor het overige speelt voornamelijk een rol dat de benaming van en indeling in rubrieken tussen vakken verschilt.

Deze inventarisatie is gebaseerd op de kerndoelen en eindtermen havo-vwo zoals gepubliceerd door de SLO (1996) voor de vakken natuurkunde, biologie, geschiedenis, aardrijkskunde, economie, Nederlands en wiskunde.

De indeling in de examenprogramma's verschilt van vak tot vak. Met enige goede wil kunnen echter alle in dit verband relevante eindtermen worden geordend volgens eenzelfde globale procesgang van een onderzoek. Op basis daarvan kan het overzicht worden gemaakt (figuur 2) waarin een indeling in tien stappen is aangehouden. Een dergelijk aantal lijkt een bruikbare middenweg tussen de uitersten die men in publicaties tegenkomt: een meer globale indeling van onderzoek in drie of vier fasen en een gedetailleerde lijst van twintig of meer deelactiviteiten. De nummers in het overzicht verwijzen naar de kerndoelen en eindtermen zoals opgenomen in de SLO-publicaties (1996).

	bavo	natuurkunde, biologie	aardrijkskunde	geschiedenis	economie	Nederlands	wiskunde
1 hanteren van het vakspecifieke begrippenapparaat, herkennen en formuleren van een probleem in termen van het betreffende vak		35, 36	1, 2, 7a	1, 3	6, 12-15		10, 11, 13, 14, 15
2 formuleren van een onderzoeksvraag en eventueel deelvragen, hypothesen en verwachtingen		37, 38	1, 7a, 7b	1, 6a, 6b	6		
3 maken en bewaken van een onderzoeksplan: a onderzoekspzot b planning, taakverdeling	5.1 (3.5)	39, 40	7c	6d, 6e			
4 verzamelen en selecteren van informatie/gegevens	2.3 (3.5)	15-17, 23-27, 41	3, 7e	2, 6c	7		2-4, 16
5 vaststellen van de waarde en bruikbaarheid van de gegevens (op diverse criteria)	3.1	22, 19	7f	6g	8a		1, 9, 12
6 verwerken van de gegevens (bewerken, analyseren, interpreteren, ordenen, bronnen verantwoorden)	3.1, 3.3 (3.5)	18-21	4, 7g	3, 6f, 6h	8b, 9	15, 19-21	5-8, 12, 17, 18
7 trekken van conclusies	(3.5)	42	7h	6i	8c		19
8 (a) evalueren en (b) beoordelen van het onderzoek		43	7j	6k		22, 23	20

9	bepalen en beargumenteren van een (eigen) standpunt	3,7	7	4, 7h	3	8d, 8e, 8f		21
10	(a) rapporteren (beschrijven) en (b) presenteren (overdragen)	2,2, 4,7	4-6, 8	5, 7i	4, 6j	10, 8f	14, 24, 25	21

Figuur 2 Voor onderzoek relevante kerndoelen bavo en eindtermen havo/vwo  
Informatievaardigheden zijn vooral opgenomen bij de stappen 4, 5 en 6. Onderzoekvaardigheden in engere zin zijn primair te vinden bij 2, 3, 7 en 8.

### **Soorten onderzoek**

Tot op zekere hoogte kunnen onderzoeksstappen en onderzoekvaardigheden worden onderscheiden zonder nadrukkelijk verschil te maken tussen de schoolvakken. Docenten kunnen echter geneigd zijn te stellen dat onderzoek doen bij hun vak iets heel anders is dan onderzoek doen bij een ander vak.

*In ons onderzoek najaar 1998 vond 50% van de docenten dat onderzoek op hun vakgebied niet of slechts in geringe mate afwijkt van onderzoek bij andere vakken; ruim 30% zag in redelijke mate verschil en bijna 20% wist het niet. Er was geen sprake van duidelijke verschillen tussen de docenten van de in het onderzoek betrokken vakken in de mate waarin men onderzoek op het eigen vakgebied afwijkend achtte.*

Het benadrukken van de verschillen tussen vakken kan de samenwerking tussen vaksecties bemoeilijken. Een mogelijk vruchtbare insteek kan zijn om soorten onderzoek te onderscheiden. Voorzover bij bepaalde vakken vooral bepaalde soorten onderzoek aan de orde zijn, kunnen tussen vakken verschillen optreden, maar die zijn dan een afgeleide van het soort onderzoek, terwijl toch sprake kan zijn van overeenkomst over waar het in de kern om gaat, namelijk het leren doen van onderzoek en het daarbij maken van adequate keuzes.

*In ons onderzoek bleek de vraag naar de mate waarin de docent ervaring had met verschillende soorten onderzoek door leerlingen (uitkomsten zie verderop) een van de weinige vragen te zijn waarbij duidelijke verschillen tussen de vakken naar voren kwamen. Dit ondersteunt de gedachte dat er in het algemeen veel overeenkomsten zijn tussen de vakken (wellicht meer dan docenten soms zelf denken).*

Het onderscheiden van soorten onderzoek kan aan de hand van verschillende indelingsprincipes. De volgende invalshoeken lijken zinvol:

- aan de hand van de grondvorm (bijvoorbeeld experiment, case study, survey);
- aan de hand van de vraagstelling (bijvoorbeeld bepalen van een constante, toetsen van een hypothese, zoeken naar relaties, beschrijven, verklaren);
- aan de hand van de situatie waarin het onderzoek wordt uitgevoerd (bijvoorbeeld een practicumlokaal of laboratorium, het 'vrije' veld, een archief);
- aan de hand van de gegevensverzamelingsmethode (bijvoorbeeld gebruiken van een meetinstrument, observeren, interviewen).

De eerste twee invalshoeken hangen samen met het uitgangspunt dat onderzoek doen vooral te maken heeft met een bepaalde manier van denken en redeneren (vergelijk het derde doel). De laatste twee invalshoeken hangen meer samen met het doen van onderzoek in de zin van het uitvoeren van een aantal (concrete) activiteiten (vergelijk het vierde doel). Deze twee invalshoeken liggen waarschijnlijk dicht bij verschillen tussen soorten onderzoek zoals die een rol spelen bij

de praktische organisatie en ondersteuning door de docent en de beleving door de leerlingen.

De soorten onderzoek zoals die in de praktijk worden onderscheiden zijn gebaseerd op een mengeling van de genoemde invalshoeken. Het zijn als het ware veelvoorkomende en herkenbare typen onderzoek, waarin een aantal van de bovengenoemde kenmerken geclusterd voorkomen. Aldus bekende typen onderzoek zijn onder meer bronnenonderzoek; verrichten van metingen met meetapparatuur (in de school of in het veld); experiment; gedragsstudie (observaties); gevalstudie; survey (grootschalige bevraging).

*In ons onderzoek najaar 1998 bleek dat bij alle vier in het onderzoek betrokken vakken regelmatig door leerlingen een werkstuk wordt gemaakt, het vaakst bij geschiedenis (100% van de docenten heeft in redelijke of sterke mate ervaring daarmee) en biologie (85%), minder bij natuurkunde en economie (rond 50%). Practicum en laboratoriumexperiment worden bij natuurkunde en biologie veel gedaan (practicum bij beide vakken 100%, laboratoriumexperiment bij natuurkunde 55%, bij biologie 80%). Bronnenonderzoek en literatuurstudie worden bij geschiedenis (60-70%), biologie (40-50%) en economie (30-40%) regelmatig gedaan en bij geschiedenis ook wel documentenanalyse (40%). Bij biologie worden nog veldonderzoeken (al dan niet als experiment) en gedragsstudies (observaties) gedaan (steeds rond 50% van de docenten in redelijke of sterke mate ervaring daarmee) en soms een gevalstudie of een survey. Per vak bekeken domineert bij natuurkunde het practicum en bij geschiedenis het soort onderzoek dat op teksten gericht is, terwijl biologie de grootste variatie aan soorten onderzoek laat zien en bij economie nog het minst van een onderzoekstraditie sprake is. Het ging in dit onderzoek slechts om vijftig docenten (10-14 per vak), maar het geeft toch een beeld.*

### **3.3 Onderzoek doen: stappen en vaardigheden**

De eindtermen havo/vwo met betrekking tot onderzoek door leerlingen in een aantal verschillende vakken kunnen worden ondergebracht in tien rubrieken met onderzoeksactiviteiten (zie paragraaf 3.2). Deze rubrieken kunnen worden gezien als stappen die in een onderzoek moeten worden gezet. Deze stappen worden hieronder kort nader besproken, waarbij aandacht wordt gegeven aan verschillen tussen vakken, betekenis en haalbaarheid.

Daarna wordt ingegaan op drie belangrijke vragen:

- hoe globaal of gedetailleerd kunnen en moeten in de onderwijspraktijk de fasen en activiteiten in een onderzoek worden onderscheiden;
- in hoeverre representeren de onderscheiden onderzoeksstappen ook de aan te leren *vaardigheden*;
- welke doelen zouden in de onderwijspraktijk prioriteit verdienen?

#### ***De onderzoeksstappen nader bekeken***

##### ***1 Hanteren van het vakspecifieke begrippenapparaat, herkennen en formuleren van een probleem in termen van het betreffende vak***

Bij natuurkunde en biologie wordt dit amper geëxpliciteerd. Het zit in feite opgesloten in de eindtermen 35 en 36: het herkennen en specificeren van een natuurwetenschappelijk probleem en het leggen van verbanden tussen probleemstellingen, hypothesen, gegevens en aanwezige voorkennis. Bij aardrijkskunde, geschiedenis, economie en wiskunde worden de belangrijke vakeigen begrippen in een aantal eindtermformuleringen beschreven. In de praktijk zullen leerlingen niet altijd vooraf voldoende weten welke onderdelen van de vakinhoudelijke kennis bij een gegeven onderwerp, respectievelijk probleem relevant is of kan zijn. Het bepalen daarvan lijkt echter juist de kern van wat in de betreffende eindtermen in feite wordt gevraagd.

2 *Formuleren onderzoeksvraag, eventueel deelvragen, hypothesen, verwachtingen*

Bij natuurkunde en biologie wordt verwacht dat de leerling hypothesen en verwachtingen opstelt, bij aardrijkskunde en geschiedenis worden deelvragen verwacht. Bij het formuleren van een onderzoeksvraag moet het uiteraard gaan om een onderzoekbare vraag: voldoende afgebakend en eenduidig, voor onderzoek toegankelijk en gezien de gegeven tijd en condities onderzoekbaar. Deze eisen worden in de eindtermen niet geëxpliciteerd. De moeilijkheid is dat je vrij veel kennis moet hebben van zowel de vakinhoud als van onderzoek om een en ander goed te kunnen inschatten. In de praktijk zullen leerlingen veelal moeite hebben om de begrippen helder te krijgen en om zich een beeld te vormen van het proces en het mogelijke resultaat van het onderzoek.

3 *Maken en bewaken onderzoeksplan: (a) onderzoeksopzet, (b) werkplanning*

Bij natuurkunde en biologie wordt hierbij gevraagd naar aandacht voor prioriteiten, mogelijkheden en randvoorwaarden. Bij geschiedenis is een aparte eindterm geformuleerd met betrekking tot het volgen van de planning. Opvallend is dat in de eindtermen niets wordt gezegd over de keuze van een adequate opzet voor het onderzoek en de meer methodisch-technische aspecten (steekproeftrekking, meetmethoden, meetplan, opzet gegevensanalyse). Keuzes op dit vlak raken de kern van de bestaande kennis over onderzoek. In een onderzoeksplan zullen keuzes op dit vlak samenhangen met de specifieke vraagstelling die moet worden beantwoord. De natuurwetenschappelijke vakken gebruiken daarbij een iets andere terminologie (constante bepalen, hypothese toetsen, relaties zoeken, observaties in theoretische termen beschrijven, een apparaat bouwen) dan de gammavakken (beschrijven, verklaren, interpreteren, waarderen, een oplossing ontwikkelen voor een probleem). Leerlingen moeten leren hun keuzes te beargumenteren, deze te controleren op consistentie en uitvoerbaarheid en rekening te houden met overwegingen van zorgvuldigheid, veiligheid en privacy. In de praktijk van onderzoek door leerlingen zal bij deze stap ook aandacht moeten worden gegeven aan een goede taakverdeling (bij onderzoek in een groepje), het anticiperen op benodigde hulpmiddelen, het zonedig maken van afspraken met derden et cetera. Leerlingen moeten ook leren hun eigen werk te plannen, de planning te bewaken, zonedig de werkzaamheden bij te sturen, doelmatig te werk te gaan en om te gaan met deadlines. Het zal, zeker in het begin, moeilijk zijn om de benodigde tijd (binnen zowel als buiten de lesuren) goed in te schatten.

4 *Verzamelen en selecteren van informatie/gegevens*

Deze stap is in de eindtermen ruim vertegenwoordigd. Bij natuurkunde en biologie wordt onderscheid gemaakt tussen schriftelijke, mondelinge en audiovisuele bronnen, het kiezen en bevragen van informanten (opvallend bij deze vakken) en het halen van informatie uit grafieken, tabellen et cetera (vaak over data afkomstig uit experimenten). Bij aardrijkskunde, geschiedenis en economie wordt ingegaan op het vaststellen van de informatiebehoefte en het selecteren van relevante informatiebronnen. Bij deze stap komen de verschillen tussen bèta- en gamma-disciplines dus duidelijk naar voren. Bij de uitvoering van de gegevensverzameling worden bepaalde 'technieken' gebruikt en zijn specifieke vaardigheden van belang, zoals het halen van informatie uit bestanden met ict (bij alle vakken), het hanteren van materiaal, een apparaat of een stuk gereedschap (bij natuurkunde, biologie en eventueel aardrijkskunde), afnemen van een interview (bij aardrijkskunde, geschiedenis, economie). Tijdens de gegevensverzameling moeten de leerlingen bovendien letten op zorgvuldigheid, mogelijke foutenbronnen en dergelijke.

5 *Bepalen van de waarde en bruikbaarheid van de gegevens (diverse criteria)*

Deze stap had ook kunnen worden samengevoegd met de voorgaande. Het gaat hier om betrouwbaarheid en relevantie (bruikbaarheid) van de gegevens, bij de gammavakken aangevuld met het criterium representativiteit.

Deze activiteit is als aparte stap onderscheiden, omdat het beoordelen van de informatie op zulke criteria moet worden gebaseerd op een combinatie van keuzes en overwegingen bij de stappen 1, 2 en 4. Anders gezegd: ook de vakinhoudelijke kennis en de vraagstelling spelen hierbij een belangrijke rol.

6 *Verwerken van de gegevens*

Deze rubriek is in de eindtermen weer ruim vertegenwoordigd met de nodige verschillen tussen de vakken. Bij natuurkunde en biologie en ook economie ligt een accent op het ordenen en bewerken van gegevens en het weergeven van resultaten in grafieken en dergelijke. Bij aardrijkskunde en geschiedenis is meer sprake van analyseren en interpreteren. Bij natuurkunde en biologie valt op dat expliciet eindtermen zijn geformuleerd over het onderscheiden van hoofd- en bijzaken en het verantwoorden van feiten met bronnen. Een en ander lijkt juist ook bij de andere vakken van groot belang, maar is daar wellicht meer vanzelfsprekend. Bij deze onderzoeksstap kunnen ook eindtermen van het vak Nederlands worden geplaatst, waar het gaat om het analyseren van een betoog. Hier liggen mogelijkheden voor samenwerking. Rond het analyseren van gegevens zijn er ook samenwerkingsmogelijkheden met het vakgebied wiskunde.

7 *Trekken van conclusies*

Deze stap wordt bij alle vakken in een eindterm genoemd, maar slechts kort. Bij de gamma-vakken wordt alleen nog gevraagd dat de conclusie wordt beredeneerd, respectievelijk beargumenteerd. Met name bij de gamma-vakken kan men zich afvragen of het trekken van conclusies uit (soms veel, deels polyinterpretable en/of tegenstrijdige gegevens) niet gecompliceerd kan zijn en daarom meer aandacht verdient.

8 *(a) Evalueren, respectievelijk (b) beoordelen van het onderzoek*

Ook deze stap komt summier in één eindterm per vak aan de orde, waarbij alleen bij geschiedenis een kleine toevoeging is gemaakt ('bij gegeven criteria'). Het vak Nederlands is bij deze stap relevant via eindtermen met betrekking tot het beoordelen van een betoog. Bij verschillende vakken wordt in de eindtermen aandacht gevraagd voor de nauwkeurigheid van de resultaten en de reikwijdte van de conclusies en voor het aangeven van mogelijke verbeteringen van het onderzoek en/of zinvolle vragen voor verder onderzoek. Deze punten lijken algemeen van belang. Bij deze stap zouden alle voorgaande stappen moeten worden overzien en zou bewust lering moeten worden getrokken aangaande het doen van onderzoek. Dit lijkt daarom in de onderwijspraktijk een belangrijke stap. Leerlingen zullen moeten leren reflecteren op de kwaliteit van het uitgevoerde onderzoek, op de kwaliteit van hun eigen werk (zoals dat wordt beoordeeld door de leraar) en op hun eigen leerervaringen.

9 *Bepalen en beargumenteren van een (eigen) standpunt*

Voor deze stap zijn alleen eindtermen geformuleerd bij de gamma-vakken. Betekent dit dat waarden en normen en het komen tot een eigen stellingname slechts bij deze vakken van belang zijn? Uiteraard is er onderscheid tussen waarden en normen die onderdeel uitmaken van hetgeen is onderzocht en waarden en normen rond het onderzoek zelf (wat is het doel ervan, welke belangen worden gediend, welke consequenties kan de kennis hebben et cetera). De eerstgenoemde waarden zijn alleen bij de gamma-vakken aan de orde, de laatstgenoemde ook bij de bèta-vakken. Bij economie is er een aparte eindterm met betrekking tot het luisteren naar standpunten van anderen. Deze eindterm lijkt bij alle vakken relevant, zowel bij het bepalen van een eigen standpunt als bij het presenteren daarvan.

10 *(a) Rapporteren (beschrijven) en (b) presenteren (overdragen)*

Deze stap is in de eindtermen bij de verschillende vakken ondergebracht in verschillende rubrieken: onderzoeksvaardigheden, informatievaardigheden en taalvaardigheden. In de eindtermen Nederlands krijgt het opzetten van een correct, gestructureerd en overtuigend betoog expliciete aandacht. Bij de gamma-vakken gaat het



naast het schrijven van teksten met name om het presenteren: mondeling, audiovisueel en/of ict-ondersteund. Het is overigens opvallend dat de eindtermen niets vermelden over het verwerken van informatie uit een presentatie door de overige leerlingen.

***Hoe gedetailleerd onderscheiden we fasen en activiteiten in onderzoek?***

Tijdens de aanloop tot de invoering van de vernieuwde tweede fase zijn de afgelopen jaren allerlei indelingen gemaakt van fasen en activiteiten in onderzoek door leerlingen. In de meest globale opzet worden drie fasen onderscheiden: voorbereiding (formuleren van de probleemstelling en maken van het werkplan), uitvoering (verzamen en verwerken van gegevens) en afronding (trekken van conclusies, evalueren et cetera). In een vak als natuurkunde wordt ook gewerkt met een model in vier fasen, waarin de middelste hiervoor genoemde fase is gesplitst in 'uitvoering' (verrichten van metingen) en 'verwerking' (meetgegevens ordenen, resultaten in een grafiek weergeven en dergelijke). In hetzelfde vak worden echter ook wel acht fasen onderscheiden (met binnen de 'verwerking' apart het 'verwerken', 'interpreteren van resultaten' en 'conclusies trekken' (vergelijk APS, 1998).

In publicaties van het Cito zijn de afgelopen jaren diverse indelingen gebruikt, variërend van een model met drie fasen tot opsommingen van een twintigtal (deel-)activiteiten. (Vergelijk onder andere Groen, Van de Kerkhof, De Roo & Smeets, 1996). De in deze brochure voorgestelde tien stappen houden het midden tussen een heel globale fasenindeling en een uitgebreide lijst van (deel)activiteiten.

*In ons onderzoek najaar 1998 is gevraagd in welke mate men bij het aanleren van onderzoek de nadruk legt op de verschillende onderdelen. De meeste nadruk wordt gelegd op het trekken van conclusies (door 95% van de docenten in redelijke of sterke mate), gevolgd door het specificeren van een probleem en van een onderzoeksopzet (beide rond 85%). Rapporteren (70%), evalueren (60%), standpuntbepaling (50%) en presenteren (25%) komen achteraan. Betekent dit dat docenten zich (nog) primair richten op het 'juiste' antwoord?*

***Zijn de onderzoeksstappen ook de aan te leren (kennis en) vaardigheden?***

Indien men 'vaardigheden' opvat als de cognitieve bekwaamheden die voor de uit te voeren onderzoekstaken nodig zijn, dan kunnen de onderzoeksstappen niet worden gelijkgeschakeld aan de vaardigheden. Verschillende onderzoeksstappen kunnen immers (deels) dezelfde cognitieve activiteiten veronderstellen en voor eenzelfde stap kunnen meerdere cognitieve activiteiten nodig zijn. Lock (1989, 1990) en Brown e.a. (1996) deden onderzoek naar een aantal voor onderzoek belangrijke deelvaardigheden bij natuurwetenschappelijke vakken: onderzoeksplan maken, hanteren van apparaten en instrumenten, observeren en aflezen van meetresultaten, interpreteren van de uitkomsten en rapporteren. Uit de beoordelingen van de prestaties van de leerlingen bleek dat de vaardigheden 'onderzoeksplan maken' en 'interpreteren van de uitkomsten' met elkaar samenhangen; dit lijkt verklaarbaar uit het feit dat dit de stappen zijn die nauw te maken hebben met de vakinhoudelijke kennis van het probleem in kwestie. Observeren en rapporteren bleken meer op zich staande vaardigheden te zijn. Hieruit kan men concluderen dat de stappen in een onderzoek inderdaad niet precies overeenkomen met de aan te leren 'achterliggende' vaardigheden.

In een praktische onderwijssituatie is het uiteenrafelen van achterliggende cognitieve vaardigheden niet doenlijk. Men kan 'vaardigheden' echter ook opvatten als betekenisvolle taken in de context van een groter geheel. In deze zienswijze kan men de onderzoeksstappen beschouwen als de vaardigheden die moeten worden geleerd om een onderzoek te kunnen opzetten en uitvoeren. Deze visie komt overeen met de eindtermen, geformuleerd in de vorm van doelgerichte (deel-)activiteiten. Bovendien zullen de instructie, begeleiding en feedback die de docent de leerlingen geeft, het expliciteren van keuzes en argumenten en de reflectie op het verrichte onderzoek door de leerlingen, in de praktijk plaatsvinden in termen van de concrete stappen die in onderzoek dienen te worden gezet.

Een en ander neemt niet weg dat leerlingen zich de gewenste vaardigheden geleidelijk eigen moeten maken. Bij het beoordelen van het werk van de leerlingen naar aanleiding van de eerste praktische opdrachten die ze maken zal een docent dan ook andere criteria hanteren dan bij het profielwerkstuk. Pas bij dit laatste gaat het om het eindniveau dat leerlingen moeten behalen.

***Welke doelen verdienen in de onderwijspraktijk prioriteit?***

In paragraaf 3.1 is een overzicht gegeven van zeven mogelijke doelen met onderzoek door leerlingen. Wie uitgaat van de onderzoeksvaardigheden zoals geformuleerd in de examenprogramma's zal het vierde doel (de deelvaardigheden) centraal stellen, al dan niet in de context van het derde doel (overzicht en inzicht). De ontwikkeling van vaardigheden (vierde doel) is in het vo, waar het gaat om onderzoek, slechts beperkt mogelijk. Kennis over onderzoek (derde doel) is ook belangrijk en het lijkt dan ook aan te bevelen minimaal (ook) dit derde doel na te streven. Daarbij gaat het om 'onderzoek doen' als een manier van denken en redeneren, waarin heen en weer wordt gependeld tussen theorie, vraagstelling, onderzoeksopzet, gegevens en conclusies. Wie zich uitsluitend op de deelvaardigheden richt aan de hand van achtereenvolgens uit te voeren stappen kan bovendien de verkeerde indruk wekken dat onderzoek doen een strikt lineair proces is. Aandacht voor het derde doel kan dit helpen voorkomen. Wie het derde doel erbij betreft, is ook niet ver verwijderd van het zesde doel (kennismaking met het 'ethos' van onderzoek).

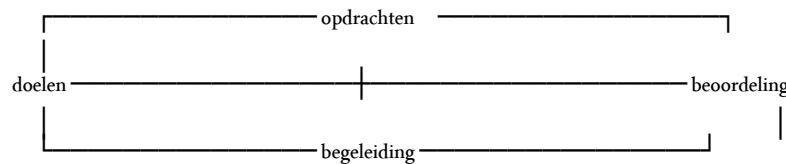
Docenten die het streven naar meer actief en zelfstandig werkende en lerende leerlingen centraal stellen, zullen het min of meer zelfstandig (individueel of in een groepje) opzetten en uitvoeren van een (min of meer open) onderzoek vooral zien als een oefening in zelfstandig werken, plannen, samenwerken en dergelijke en het tweede doel voorop stellen. De beoordeling zal dan ook primair deze algemene vaardigheden betreffen en de specifieke onderzoeksvaardigheden komen dan op het tweede plan.

Ten derde zullen er docenten zijn die het doen van onderzoek door leerlingen in de eerste plaats zien in functie van de ontwikkeling van hun kennis van het vak (eerste doel). Het zevende doel (inzicht in de ontwikkeling en status van kennis) zou dan niet ver verwijderd hoeven te zijn. Ook bij de docenten die deze keus maken, zullen de onderzoeksvaardigheden wellicht op het tweede plan komen.

## 4 OPDRACHTEN, BEGELEIDING, PROGRAMMERING (bij 2.5)

### 4.1 Doelen, opdrachten, begeleiding, beoordeling

Zowel voor de docent als voor de leerlingen is het belangrijk duidelijk te hebben welke doelen met bepaalde opdrachten worden beoogd. De opdrachten, de begeleiding en de beoordeling moeten daarop en op elkaar zijn afgestemd. Het schema in figuur 3 brengt dit in beeld.



Figuur 3 Afstemming van doelen, opdrachten, begeleiding en beoordeling

We geven enkele voorbeelden van de manier waarop deze vier componenten kunnen samenhangen. Wie bijvoorbeeld leerlingen een onderzoeksopdracht geeft om daarmee onderzoeksvaardigheden te ontwikkelen, maar het resultaat uitsluitend of vooral afmeet aan de toename van kennis over de vakinhoud of omgekeerd is niet consistent bezig. Hetzelfde geldt voor wie tijdens de begeleiding de samenwerking tussen leerlingen stimuleert, maar na afloop de leerlingen beoordeelt op de mate waarin ze individueel zelfstandig tot resultaat zijn gekomen. Wie de opdracht en de begeleiding erop afstemt dat leerlingen van stap tot stap goed weten wat ze moeten doen creëert wellicht in hogere mate condities voor de ontwikkeling van deelvaardigheden dan voor het verkrijgen van overzicht en inzicht ten aanzien van een onderzoek als geheel. Wie in doel, opdracht en begeleiding leerlingen de ruimte wil geven eigen keuzes te maken en eigen accenten te leggen, zodat het van groepje tot groepje kan verschillen welke vaardigheden vooral worden ontwikkeld, zal niet meer normgeoriënteerd beoordelen (vergelijkend tussen leerlingen), maar criteriumgeoriënteerd (vergelijkend met maatstaven voor een goede taakuitvoering). Wie er in de begeleiding voornamelijk voor zorgt dat leerlingen niet 'vast' komen te zitten, zal bij de beoordeling de gegeven begeleiding op een andere manier verdisconteren dan wie de begeleiding erop richt het onderzoek te verbeteren.

In een opdracht kunnen meer of minder zaken worden aangegeven, waaronder de keuzes die leerlingen (nog) kunnen/moeten maken, de te gebruiken bronnen en andere hulpmiddelen, de structurering in de tijd, de manier van samenwerken, de te geven begeleiding en de eventuele tussentijdse beoordelingen.

Een apart punt van aandacht betreft de binnen de verschillende schoolvakken gebruikte methode (schoolboeken). De docent zal willen aansluiten bij de methode, respectievelijk ervoor willen waken dat de aanpak en aanwijzingen voor de leerlingen met betrekking tot het doen van onderzoek, zoals verwerkt in de opdrachten en de begeleiding, niet haaks staan op wat de leerlingen uit het boek leren en gewend zijn.

*In ons onderzoek najaar 1998 gaven de docenten te kennen vrij veel eisen te stellen aan de opdrachten voor de leerlingen. Bijna alle docenten (ruim 90%) waren in redelijke of sterke mate van mening dat een opdracht van voldoende vakinhoudelijk niveau moet zijn, moet aansluiten bij het niveau van de leerlingen, voldoende uitdagend moet zijn en binnen een bepaalde tijd realiseerbaar. Daarnaast vond zo'n 85% van de docenten het belangrijk dat de opdracht*

toetsbare resultaten oplevert, rond 70% dat de opdracht aansluit bij de exameneisen, eveneens 70% dat de opdracht met het op school beschikbare materiaal kan worden uitgevoerd, en rond 65% dat de opdracht samenwerking tussen leerlingen mogelijk maakt. Beide laatstgenoemde eisen werden door docenten natuurkunde en biologie vaker aangekruist dan door docenten geschiedenis en economie.

#### **4.2 Leerlijnen, didactische aanpakken, begeleiding en feedback**

##### ***Inleiding***

Een docent kan op verschillende manieren leerlingen onderzoek laten respectievelijk leren doen. De keuzes zullen samenhangen met het beleid van de school en de sectie, de doelen van de docent, het soort onderzoek en de stappen, respectievelijk vaardigheden die daarin aan de orde zijn en de mate waarin het accent ligt op het aanleren van kennis en/of vaardigheden. De te maken keuzes hebben om te beginnen vooral betrekking op drie aspecten: een docent kan meer of minder accent leggen op de vakinhoudelijke kennis, op het opzetten en uitvoeren van onderzoek en op de samenwerking tussen de leerlingen. Verder kunnen verschillende didactische keuzes worden gemaakt.

##### ***Onderzoek als middel voor de ontwikkeling van vakinhoudelijke kennis***

Wie hiervoor kiest, wil misschien de onderwerpen als docent zoveel mogelijk zelf bepalen, passend bij de leerlijn in de methode, de exameneisen en dergelijke. Een andere mogelijkheid is dat de docent wel globaal de onderwerpen aanreikt, maar de leerlingen ruimte wil bieden om zelf specifieke vragen te kiezen, bijvoorbeeld vanuit de gedachte dat dit de motivatie van de leerlingen en daarmee de leeropbrengst kan vergroten. Een mogelijk probleem bij het door de leerlingen zelf laten kiezen van de onderwerpen is dat de docent van sommige onderwerpen minder zal afweten en het onderzoek dan minder goed vakinhoudelijk kan begeleiden, respectievelijk de gegevens en conclusies van de leerlingen minder goed kan controleren. Docenten die primair voor de vakinhoud kiezen, zijn misschien geneigd minder aandacht te geven aan de kennis en vaardigheden voor het doen van onderzoek. Het is echter de vraag of een adequate opzet en uitvoering van het onderzoek niet juist nodig is voor de gewenste ontwikkeling van vakinhoudelijke kennis.

##### ***Aanleren van kennis en vaardigheden voor het doen van onderzoek***

Ook de docent die leerlingen onderzoek wil leren doen, kan kiezen tussen meer zelf structureren en meer aan de leerlingen overlaten. Aan de ene kant kan een docent de kennis en vaardigheden die bij het opzetten en uitvoeren van een onderzoek van belang zijn doelgericht aan de orde stellen, bijvoorbeeld door een standaard stappenschema aan te reiken en te bespreken. De bijbehorende opdrachten en begeleiding zullen er dan op zijn gericht dat leerlingen zo weinig mogelijk op een verkeerde manier bezig zijn. Bij deze aanpak past ook goed het bespreken van een of meer voorbeeldonderzoeken (een potentieel leerrijke activiteit, die veel te weinig wordt gebruikt). Aan de andere kant kan een docent ervoor kiezen onderzoek te benaderen als een complexe taak, ervan uitgaande dat leerlingen zoveel mogelijk leren door er op een realistische manier mee aan de slag te gaan. Leerlingen zullen daarbij veel eigen keuzes maken in de vraagstelling en opzet, maar bijvoorbeeld ook de plaats waar het onderzoek wordt uitgevoerd en de planning. Bij deze aanpak laat de docent meer over aan de leerlingen en heeft hij de organisatie minder in de hand. Dit kan leiden tot een groter beroep op de begeleiding. Bij deze insteek zou goed kunnen passen dat de docent ook algemene vaardigheden zoals zelfstandig leren en samenwerken wil bevorderen.

*In ons onderzoek najaar 1998 bleek dat tweederde van de docenten aangeeft te werken met redelijk voorgestructureerde opdrachten. Rond een kwart laat de leerlingen min of meer zelf het onderwerp, de te gebruiken middelen en/of de plaats waar het onderzoek wordt uitgevoerd kiezen. De hoeveelheid aan het onderzoek te besteden tijd wordt door slechts 10%*

*min of meer vrijgelaten.*

### **Onderzoek in samenwerking tussen leerlingen**

In principe kunnen onderzoeken door leerlingen individueel worden opgezet en uitgevoerd en het is denkbaar dat een docent daarvoor kiest, al is het maar bij één opdracht. Motieven daarvoor kunnen zijn dat de docent zicht wil hebben op de kennis en vaardigheden van de individuele leerlingen of de leerlingen het verschil wil laten ervaren tussen individueel en samen onderzoek doen. Vaak zullen leerlingen onderzoek doen in kleine groepjes.

*In ons onderzoek najaar 1998 bleek 40% van de docenten de leerlingen bij het onderzoek altijd te laten samenwerken en nog eens 15% regelmatig; minder dan 10% deed dit nooit (er waren geen duidelijke verschillen tussen vakken). Bijna alle docenten laten leerlingen bij onderzoek samenwerken opdat leerlingen elkaar kunnen aanvullen en van elkaar kunnen leren. Stimuleren van de ontwikkeling van sociale vaardigheden wordt door ruim 60% genoemd (het meest door docenten economie, het minst door docenten natuurkunde), evenals het motief dat het laten werken van leerlingen in groepen werk bespaart (bij het begeleiden en het nakijken). Ruim 20% van de docenten noemt als motief dat op school te weinig materiaal aanwezig is om leerlingen individueel te laten werken (vooral docenten natuurkunde en ook wel biologie).*

De docent kan de samenwerking tussen de leerlingen meer of minder:

- overlaten aan de leerlingen (bijvoorbeeld om ze de ruimte te geven);
- stimuleren (bijvoorbeeld om de motivatie te bevorderen);
- structureren (bijvoorbeeld terwille van de doelmatigheid);
- doelbewust inzetten (bijvoorbeeld ter bevordering van discussie en reflectie).

De leerresultaten bij samenwerkende leerlingen en de kwaliteit van hun product zullen mede afhangen van de samenstelling van het groepje. Dit kan ertoe leiden, vanuit het oogpunt van een faire beoordeling, dat leerlingen in feite geen gelijke kansen krijgen. Een docent kan om dit te vermijden de samenstelling van de groepjes controleren of vooraf sturen (bijvoorbeeld door loting).

### **Systematisch ontwikkelen van kennis en vaardigheden bij onderzoek**

Gelet op de kerndoelen en eindtermen (zie paragraaf 3.2) zal tenminste enige expliciete aandacht moeten gaan naar het opzetten en uitvoeren van onderzoek (de tweede van de drie hiervoor beschreven opties). Hoe pak je dat als docent nu aan? Literatuur en praktijk leveren een viertal basismogelijkheden:

- a eerst introduceren en laten oefenen van afzonderlijke stappen; een motief hiervoor kan zijn dat het doelmatig en effectief kan zijn om eerst bepaalde vaardigheden te trainen, alvorens ze in te (laten) zetten in een groter geheel;
- b in een aantal na elkaar uit te voeren onderzoekjes de voorstructurering systematisch laten afnemen; motieven hiervoor kunnen zijn dat leerlingen vanaf het begin ervaring moeten opdoen met het geheel van een onderzoek en dat door bepaalde stappen eerst aan te reiken kan worden voorkomen dat leerlingen meteen vastlopen ('afbreukrisico');
- c leerlingen vanaf het begin 'in het diepe gooien', maar het proces als docent systematisch stapsgewijs begeleiden; motieven hiervoor kunnen zijn dat als leerlingen zelf de vraagstelling en de opzet ontwikkelen, het onderzoek meer iets van henzelf is, hetgeen bevorderlijk kan zijn voor de motivatie en de leeropbrengst; het afbreukrisico kan worden beperkt via de begeleiding, waarin bovendien op maat ondersteuning en feedback kan worden geboden;
- d leerlingen 'in het diepe gooien' (zoals bij c) en het van hen laten afhangen op welke momenten en in welke mate zij een beroep doen op begeleiding ('begeleiding op afroep'); motieven hiervoor kunnen zijn dat de behoefte aan

begeleiding sterk kan verschillen en dat leerlingen (ook) moeten leren te signaleren wanneer zij hulp nodig hebben en leren daarnaar te vragen.

De aanpakken b en c lijken goede mogelijkheden te bieden voor het laten ontwikkelen van kennis en vaardigheden ten aanzien van onderzoek. De kunst is om als docent de opdrachten niet teveel 'dicht te timmeren', noch teveel open te laten en in achtereenvolgende opdrachten een systematische opbouw aan te brengen van voorgestructureerd naar open en van eenvoudig naar complex. Een mogelijke aanpak is in een eerste opdracht de vraagstelling en de te volgen onderzoeks aanpak en eventueel zelfs 'de verzamelde gegevens' aan te bieden, waarna de leerlingen het onderzoek 'afmaken' en in volgende opdrachten de leerlingen steeds een stap eerder te laten beginnen om het zelf te doen (vergelijk het 'POCO-model' dat bij aardrijkskunde is ontwikkeld). Ook aan het eind van het onderzoek kan de docent in eerste instantie er meer bovenop zitten, bijvoorbeeld door samen met de leerlingen te reflecteren op het onderzoek en dit later meer van de leerlingen zelf te verwachten.

Bij het wel of niet voorstructureren kan een docent desgewenst heel precieze keuzes maken. Je kunt bijvoorbeeld onderscheid maken tussen wel of niet opgeven van variabelen, van meetprocedures, van manieren van analyseren van de gegevens, respectievelijk van punten waarop de kwaliteit van het onderzoek moet worden gecontroleerd (vergelijk bijvoorbeeld Germann, Haskins & Auls, 1996). Ook het meer aan de leerlingen overlaten van bijvoorbeeld de vraagstelling of de te volgen aanpak kan op diverse manieren vorm krijgen: door de leerlingen een lijst met opties aan te bieden om uit te kiezen, door een aantal voorbeelden te geven, door criteria te noemen waaraan een goede keuze moet voldoen of door leerlingen over door hen ingebrachte keuzes te adviseren.

Het behoeft geen betoog dat onderzoek niet alleen moeilijker wordt voor leerlingen naarmate ze meer en eerder zelfstandig beslissingen moeten nemen en activiteiten moeten uitvoeren, maar ook door de moeilijkheidsgraad van het vakinhoudelijke probleem dat ze onderzoeken. Het ligt voor de hand leerlingen eerst zelfstandig onderzoek te leren doen bij eenvoudige onderwerpen en daarna pas bij meer complexe onderwerpen (inhoudelijk moeilijker en/of met meer risico's tijdens de uitvoering), waarbij wellicht in het begin weer (tijdelijk) meer structurering en ondersteuning door de docent wenselijk is. Een systematische opbouw kan er ook toe bijdragen dat duidelijker wordt welke kennis en vaardigheden voor onderzoek nodig zijn en welke vorderingen feitelijk worden gemaakt.

*In ons onderzoek najaar 1998 bleek dat 70% van de docenten leerlingen eerst deeltaalvaardigheden aanleert; dit verschilt echter van vak tot vak: bij natuurkunde en biologie doet meer dan 85% dit, bij geschiedenis 65% en bij economie 35%. Ongeveer de helft begint met 'kookboekonderzoeken' (gesloten), waarna de opdracht steeds opener wordt; ook dit gebeurt meer bij natuurkunde en biologie en veel minder bij de beide andere vakken. Eveneens de helft laat leerlingen eerst onderzoek doen bij makkelijke en daarna bij moeilijker onderwerpen (hierbij is niet veel verschil tussen de vakken). Ongeveer 20% 'gooit leerlingen in het diepe' en geeft begeleiding op maat.*

White en Frederikson (1998) beschrijven een aanpak (bij het vak natuurkunde), waarin de klas als geheel fungeert als groep onderzoekers en het leren over de vakinhoud en het leren over onderzoek, respectievelijk het leren onderzoeken hand in hand gaan. De klas werkt onder begeleiding van de docent op basis van een expliciete cyclus voor het doen van onderzoek en met behulp van een lijst met criteria waaraan een goed onderzoek moet voldoen. De docent brengt voorbeelden van (goede) onderzoeken in en stelt regelmatig gerichte vragen en veel tijd wordt besteed aan discussiëren en reflecteren over het onderzoek en het onderwerp (probleem) in kwestie. Leerlingen beoordelen elkaars en hun eigen werk aan de hand van de lijst met criteria. Volgens White en Frederikson is het voor de leerlingen motiverend dat ze duidelijk weten wat er wordt verwacht en waarop hun werk wordt beoordeeld. Leerlingen blijken elkaars en hun eigen werk redelijk goed te kunnen beoordelen. Doordat de beoordeling gefundeerd moet zijn en moet worden

beargumenteed, draagt het beoordelen op zich bij aan het leerproces. Leerlingen blijken de onderzoeks aanpak vrij vlot onder de knie te krijgen en ook vakinhoudelijke kennis te ontwikkelen. Het vak gaat meer voor ze leven en krijgt meer samenhang. Tenslotte blijkt dat leraren en leerlingen door de grote mate waarin het doen en beoordelen van onderzoek expliciet is gemaakt ook beter met elkaar over het onderzoek kunnen communiceren.

Een mogelijke variant op de klassikale aanpak die White en Frederikson beschrijven, zou kunnen zijn dat een onderzoek wordt verdeeld over meerdere groepjes leerlingen. Groepjes kunnen bijvoorbeeld werken aan verschillende deelvragen (parallel) of een volgend groepje kan verder gaan waar een vorig groepje is gebleven (serieel).

### ***Begeleiding en feedback***

Docenten geven leerlingen niet alleen opdrachten waarmee ze aan de slag gaan, maar ook begeleiding tijdens de rit. Deze begeleiding kan meer of minder intensief en gestructureerd zijn, mede afhankelijk van de gekozen basisaanpak. Tegen de achtergrond van de gekozen opbouw en aard van de opdrachten is in de begeleiding nog veel afstemming mogelijk van de bedoelingen van de docent op de keuzes, mogelijkheden en ervaringen van de leerlingen. Een docent kan tijdens de fase van de formulering van de probleemstelling en de ontwikkeling van een onderzoeksopzet erop toezien en zondig erop aansturen, dat het onderzoek voor de betreffende leerlingen van een adequaat moeilijkheidsniveau is, zodat ze er zoveel mogelijk van zullen leren. De ontwikkeling van cognitieve vaardigheden voor de uitvoering van een complexe taak zoals onderzoek wordt verder bevorderd als de docent:

- expliciteert welke (denk)stappen in een onderzoek worden gezet;
- leerlingen ruimte en gelegenheid biedt dingen te proberen en te oefenen;
- zicht houdt op de activiteiten en voortgang van de leerlingen en op maat (en in geleidelijk afnemende mate) ondersteuning en feedback geeft;
- bevordert dat leerlingen reflecteren op hun werk (proces en resultaat).

De docent kan variëren in de mate en aard van de ondersteuning (bijvoorbeeld door bepaalde suggesties te doen of hulpmiddelen ter beschikking te stellen). Ook de feedback die de docent aan de leerlingen geeft op hun keuzes, voortgang en (tussentijdse) resultaten is van groot belang. De docent kan hierbij variëren:

- meer gericht op de vakinhoud of meer op het onderzoek als onderzoek;
- meer globaal of meer gedetailleerd;
- meer procesgericht of meer productgericht;
- convergerend (toewerkend naar beslissingen, conclusies) of divergerend (opties genererend);
- meer gericht op de concrete activiteiten, dus de doe-kant van onderzoek of meer gericht op de denkkant: begrippen, argumenten, conclusies, relaties tussen het onderzoek en de verkregen kennis, reflectie op gemaakte keuzes en de consequenties daarvan en 'hoe het ook anders had gekund'.

Leerlingen zullen veelal moeten afleren dat vragen om begeleiding hun cijfer drukt. Een optie is om bij het begin van het onderzoek enkele begeleidingsmomenten in te plannen, waar ze al dan niet gebruik van kunnen maken.

*In ons onderzoek najaar 1998 bleek dat docenten op verschillende wijzen proberen in te spelen op verschillen tussen leerlingen. Ongeveer 80% van de docenten (bij biologie het meest, bij economie het minst) doet dit door leerlingen keuzevrijheid te geven ten aanzien van het onderwerp van hun onderzoek en rond 70% door te variëren in de begeleiding; een kleine 10% varieert (ook) in de beoordeling. Een meerderheid van de docenten voert tussentijdse besprekingen*

(individueel of met groepjes leerlingen). Individuele besprekingen komen het vaakst voor bij de gammavakken (rond 80%), minder bij biologie (50%) en het minst bij natuurkunde (35%). Klassikale besprekingen en schriftelijke feedback komen beide voor bij een kwart van de docenten. Als feedback geven docenten vaak (90%) een aanwijzing, suggestie of tip of verwijzen ze naar bepaalde bronnen. Daarnaast worden ook tegenvragen gesteld en voorbeelden gegeven (tussen 60% en 70%) en een aanpak of heuristiek aangeboden (30%).

Docenten zullen, naarmate zij meer ervaring krijgen in het begeleiden van (open) onderzoek, zicht krijgen op de problemen die vaak bij leerlingen voorkomen. Zij zullen proberen voor veel voorkomende problemen een structurele oplossing te ontwikkelen. Mogelijke problemen zijn:

- tekort aan vakinhoudelijke kennis;
- vraagstelling te breed, niet concreet genoeg, niet onderzoekbaar;
- verkeerde inschatting benodigde tijd, geen realistische planning;
- onvoldoende methodische know-how en besef van normen bij onderzoek;
- te weinig zorgvuldigheid (veiligheid, privacy et cetera);
- onvoldoende nauwkeurigheid van werken;
- te weinig diepgang van analyse en interpretatie;
- geen overzicht over het onderzoek als geheel;
- problemen met de taakverdeling en samenwerking.

*In ons onderzoek noemde 85% van de docenten als problemen voor leerlingen het specificeren van een probleem en het formuleren van een onderzoeksvraag, 70% het maken van een onderzoeksopzet en werkplan, het trekken van conclusies, het evalueren van het onderzoek en het beargumenteren van een eigen standpunt en 50% het rapporteren en het mondeling presenteren.*

#### **4.3 Afstemming tussen vakken, opdrachten en profielwerkstuk**

De nieuwe opzet schoolexamen is een examendossier, waarin een belangrijke plaats is ingeruimd voor beoordelingen van door leerlingen uitgevoerde opdrachten en gemaakte producten. Er wordt onderscheid gemaakt tussen praktische opdrachten en een afsluitend profielwerkstuk, waarin tenminste twee vakken moeten worden geïntegreerd. Docenten moeten beoordelingen geven van het proces, het product en de presentatie. De beoordelingscriteria moeten van te voren bij de leerlingen bekend zijn. De opdrachten worden zoveel mogelijk zelfstandig (individueel of in groepjes) gemaakt. Zonodig stuurt de docent tussentijds bij. De praktische opdrachten worden gewaardeerd met een cijfer. Gelet op de beschikbare uren aantallen zullen per vak twee tot vijf praktische opdrachten 'voor een cijfer voor het schoolexamen' kunnen worden gemaakt (afhankelijk van de zwaarte van het vak en de door de docent gekozen omvang van de opdrachten). Het profielwerkstuk wordt beoordeeld met 'onvoldoende', 'voldoende' of 'goed' (een tenminste voldoende beoordeling is een voorwaarde voor deelname aan het centraal eindexamen). Praktische opdrachten en profielwerkstukken zullen vaak de vorm hebben van het opzetten, uitvoeren en rapporteren van een onderzoek.

Tijdens de tweede fase is het in principe mogelijk om in de ontwikkeling van onderzoeksvaardigheden bij leerlingen een geleidelijke opbouw aan te brengen. Een idee achter de combinatie van praktische opdrachten en profielwerkstuk is ook dat het werken aan praktische opdrachten de leerlingen voorbereidt op het maken van een profielwerkstuk. Een aantal



van twee tot vijf praktische opdrachten is uiteraard zeer gering om een complex vaardigheidsdomein als onderzoek enigszins onder de knie te krijgen. Docenten kunnen leerlingen echter vaker opdrachten geven, ook 'oefenopdrachten' die niet meetellen voor het schoolexamen. Ook zijn combinaties mogelijk met het handelingsdeel van het schoolexamen en de vrije ruimte. Docenten die gericht willen werken aan het ontwikkelen van onderzoeksvaardigheden bij leerlingen zullen ook geneigd zijn zulke combinaties te maken. Docenten die met onderzoek door leerlingen primair andere doelen nastreven, zullen wellicht volstaan met een beperkt aantal praktische opdrachten en op het vlak van de ontwikkeling van onderzoeksvaardigheden minder verwachten. Het is de vraag in hoeverre onderzoeksvaardigheden vakgebonden of vakspecifiek zijn en in hoeverre vakoverstijgend of algemeen. Waarschijnlijk ligt de waarheid niet precies in het midden, maar wel ergens tussen beide uitersten, afhankelijk van de vaardigheden en de vakken waarover het gaat. Docenten binnen bijvoorbeeld eenzelfde profiel kunnen hun bijdrage aan de ontwikkeling van algemene vaardigheden op elkaar afstemmen of hun taken op dit vlak over de vakken verdelen. *Afstemmen* kan bijvoorbeeld door gezamenlijk af te spreken dat alle docenten in een bepaalde periode gericht aandacht geven aan algemene vaardigheid X, in een volgende periode aan vaardigheid Y et cetera (uiteraard op basis van enige consensus over de aard van die vaardigheden). *Taken verdelen* kan onder meer door de vaardigheden zodanig toe te wijzen aan de vakken, dat elk vak de trekker is voor een bepaald algemeen vaardigheidsdomein.

Specifiek kan bij de onderzoeksvaardigheden worden nagegaan in hoeverre ze:

- specifiek zijn voor bepaalde vakken of vakkenclusters (alfa, bèta, gamma), in die zin dat ze bij andere eenvoudigweg niet van toepassing zijn (zoals het ontwerpen en bouwen van een apparaat voor een natuurkundig experiment);
- in bepaalde vakken of gebieden een groter belang hebben of (traditioneel) meer aandacht krijgen dan in andere;
- in diverse vakken in andere combinaties en/of volgorde worden toegepast;
- in bepaalde vakken of gebieden een specifieke betekenis hebben; dit zal vrij vaak het geval zijn, maar de vraag is of dat het (terwille van de leerlingen) gaan hanteren van één gemeenschappelijke term in de weg staat;
- in de diverse vakken worden verbonden met specifieke vakinhoudelijke kennis; dit zal min of meer per definitie het geval zijn, maar de vraag is of dat het gaan hanteren van één gemeenschappelijke term in de weg staat.

Enkele belangrijke verschillen tussen vakken zijn de mate waarin sprake is van goed dan wel slecht gedefinieerde problemen en de mate waarin een onderzoek slechts één bepaalde (goede) uitkomst kan hebben of meerdere uitkomsten.

Voor een betrouwbare beoordeling van onderzoeksvaardigheden zijn minimaal ongeveer vijf metingen nodig, gespreid over vakken en onderwerpen. Het is dus beter meer kleine opdrachten te geven dan minder grote en beter de beoordeling tussen vakken te coördineren dan om dit niet te doen.

#### *Intercollegiale afspraken en onderlinge toetsing (ook opgenomen 7.4)*

Docenten kunnen rond onderzoeken van leerlingen samenwerken binnen de eigen vaksectie en tussen vaksecties (binnen een vakkencluster of profiel). Samenwerking kan onder meer betreffen:

- afspreken welke doelen centraal staan;
- bepalen aan welke (onderzoeks)vaardigheden vooral wordt gewerkt;
- afspreken welke opbouw en didactische aanpak wordt gevolgd;
- verdelen van taken bij het kiezen of ontwikkelen van geschikte opdrachten;

- afspreken op welke manier het werk van de leerlingen wordt beoordeeld.

Bij de beoordeling gaat het op z'n minst om de ontwikkeling en vaststelling van de criteria, respectievelijk het scoringsmodel. Tijdens de ontwikkeling van een beoordelingsaanpak kunnen docenten de werkbaarheid en betrouwbaarheid ervan nagaan door op bescheiden schaal met twee of meer docenten dezelfde producten van leerlingen te beoordelen, de beoordelingen te vergelijken en de ervaringen en (ongetwijfeld aanwezige) verschillen te bespreken. Als eenmaal een aanpak is afgesproken, kunnen docenten elkaar blijven ondersteunen. Op gezette tijden kunnen zij steekproeven van onderzoekswerk van elkaars leerlingen mede beoordelen om na te gaan in hoeverre de beoordelingen nog vergelijkbaar zijn, de aangehouden normering stabiel is, de afgesproken criteria nog toepasbaar en bevredigend zijn en dergelijke. Niet alleen per sectie, maar ook tussen vaksecties zijn afspraken nodig. Voorkomen moet worden dat leerlingen ongecoördineerd in dezelfde periode voor verschillende vakken onderzoeken moeten uitvoeren. Daarnaast is het wenselijk dat de docenten op een vergelijkbare manier met onderzoek omgaan, zodat leerlingen niet bij het ene vak iets leren wat ze bij een het andere weer moeten afleren. En voor het profielwerkstuk zal men tussen vakken tot overeenstemming moeten komen over de wijze van beoordelen.

*In ons onderzoek najaar 1998 bleek dat binnen de meeste vaksecties afspraken zijn gemaakt over de doelen bij onderzoek door leerlingen, de keuzevrijheid die ze krijgen, de beoordelingscriteria en de normering. In mindere mate zijn (al) afspraken gemaakt over de begeleiding, het gebruik van hulpmiddelen en dergelijke. Afspraken tussen verschillende vaksecties zijn nog slechts volgens zo'n 10% van de docenten gemaakt; ze zijn het meest gemeld door docenten economie.*

## 5 BEOORDELEN: FUNCTIE, AANPAK, CONTEXT (bij 2.6)

### 5.1 Meten, evalueren, toetsen, beoordelen

Terminologie is niet het belangrijkste, maar voor effectieve communicatie wel onmisbaar. Helaas is de terminologie op dit terrein niet helemaal eenduidig. Hieronder worden enkele lijnen getrokken.

*Meten* is een algemene term die wordt gebruikt voor het op een bepaalde schaal toekennen van numerieke waarden aan kenmerken van objecten. Een goede meting moet aan bepaalde eisen voldoen, afhankelijk van het meetmodel dat men hanteert. Dit kan als volgt worden toegelicht. In het onderwijs worden leerresultaten gemeten voor het toekennen van cijfers en daarop gebaseerde beslissingen over bevordering en diplomering. Op grond van diploma's ontstaan rechten op toegang tot vervolgonderwijs. In het onderwijs gebeurt toekenning van cijfers vaak op grond van een impliciet meetmodel, waarin lage scores voor het ene onderdeel kunnen worden gecompenseerd door hoge scores voor het andere onderdeel. Dit gebeurt dan in feite vaak ongeacht de onderlinge samenhang en het relatieve belang van de betreffende kennis- en vaardigheidsonderdelen. In beroepsopleidingen worden kennis en vaardigheden gemeten op grond waarvan beroepsbeoefenaren kunnen worden gecertificeerd. Zij krijgen dan het recht de betreffende beroepsmatige handelingen te verrichten. In deze beroepsopleidingen en bij de certificering van beroepsbeoefenaren worden vaak expliciet niet-compensatorische modellen gebruikt. Een procesoperator, een accountant, een arts enzovoort, zij dienen op elk relevant deelgebied apart aan bepaalde minimumeisen te voldoen, omdat slechts op basis daarvan een competente (en verantwoorde) beroepsuitoefening mogelijk is.

*Evalueren* betekent: de waarde van iets bepalen. Evalueren gebeurt meestal met het oog op bepaalde beslissingen. Evaluaties in het onderwijs kunnen bedoeld zijn om zicht te krijgen op de mogelijkheden om het onderwijs te verbeteren, om inzicht te krijgen in de kwaliteit van een onderdeel en om verantwoording af te leggen aan derden. Voor evaluaties worden ook vaak leerlingresultaten verzameld en verwerkt. Wat geëvalueerd wordt zijn echter niet de leerlingen, maar (onderdelen of aspecten van) het onderwijs.

*Toetsen* is een term die in het algemeen wordt gebruikt voor het controleren van beweringen of verwachtingen, meer specifiek voor het controleren van leervorderingen. De term is op te vatten als werkwoord, maar ook als meervoudsvorm van een zelfstandig naamwoord. In deze tweede betekenis zijn toetsen gestandaardiseerde meetinstrumenten voor kennis en vaardigheden.

*Beoordelen* is een term die wordt gebruikt voor metingen die zijn gebaseerd op zorgvuldige beoordeling door kundige personen. Hierbij kunnen diverse methoden en hulpmiddelen worden gebruikt. Voor het beoordelen van kennis, vaardigheden en houdingen van leerlingen kan gebruik worden gemaakt van: gestandaardiseerde toetsen, zelf gemaakte proefwerken, mondelinge beurten, observaties, meer en minder gesloten opdrachten, demonstraties door leerlingen, werkstukken, gemaakte producten, beoordelingen door leerlingen van elkaars werk, beoordelingen door leerlingen van hun eigen werk. Beoordelingen kunnen op allerlei zaken betrekking hebben. In het onderwijs kan het gaan om beoordeling van bijvoorbeeld de kwaliteit van een curriculum, leermiddel of examenprogramma, het klimaat op een school, de professionaliteit van een leraar en de kennis, vaardigheden en houdingen van leerlingen. Hulpmiddelen bij het beoordelen zijn onder meer meetinstrumenten en beoordelingsmodellen. Bij het beoordelen van kennis en vaardigheden van leerlingen kunnen ook de hiervoor genoemde toetsen een rol spelen. Het beoordelen van kennis,

vaardigheden en houdingen van leerlingen kan verschillende functies hebben: selectie, plaatsing van leerlingen in bepaalde stromen, diagnose van leerproblemen, feedback aan leerlingen, evaluatie van het onderwijs, geven van cijfers, beslissingen over bevordering, diplomering.

*In ons onderzoek najaar 1998 bleek dat rond 85% van de docenten met het beoordelen beoogt om leerlingen te motiveren en te stimuleren serieus aan het werk te gaan, om leerlingen feedback te geven (over wat ze goed en fout hebben gedaan) en om een cijfermatige interpretatie te kunnen geven van de prestaties van de leerlingen. In mindere mate (60%) streven docenten met het beoordelen na dat in kaart wordt gebracht in hoeverre leerlingen zich bepaalde vaardigheden eigen hebben gemaakt en waar het nog aan schort (docenten economie geven dit vaker aan dan de docenten van de andere vakken).*

## **5.2 Beoordelingscriteria, scoringsmodel, normering, cijfers**

De criteria waarop een docent het onderzoekswerk van leerlingen beoordeelt, zullen verschillen alnaargelang de doelen die met onderzoek door leerlingen worden nagestreefd en de stappen of vaardigheden in een onderzoek waarop eventueel het accent is gelegd.

Een goede beoordeling is bovendien afgestemd op de gekozen didactische aanpak. De docent die bijvoorbeeld leerlingen stapsgewijs dingen aanleert en laat oefenen, veronderstelt een stapsgewijze opbouw van kennis en vaardigheid en zal dan ook vaak per stap beoordelen. Wie leerlingen meer 'in het diepe gooit', zal eerder geneigd zijn het onderzoekswerk van de leerlingen als één geheel te beoordelen.

De docent zal moeten beslissen welke aspecten of onderdelen van het werk zullen worden beoordeeld en aan de hand van welke criteria dat gebeurt. Deze essentiële informatie over de wijze en criteria van beoordeling behoort vooraf aan de leerlingen te worden meegedeeld.

Vervolgens moet worden besloten welke vorm de deeloordelen zullen hebben (v-tjes (voldaan); minnetjes en plusjes; onvoldoende, voldoende, goed; cijfers). Tenslotte is de vraag hoe de deeloordelen zullen worden gecombineerd (bijvoorbeeld ongewogen of met een bepaalde weging) en uitgedrukt in één bepaalde maat (bijvoorbeeld: onvoldoende, voldoende of goed; of een cijfer).

Beoordelen van werk van leerlingen valt dus niet samen met het geven van een cijfer, maar gaat eraan vooraf. Beoordelen gebeurt in termen van de kennis en vaardigheden waar het om gaat en leidt tot inzicht in sterke en zwakke punten. Het geven van één eindcijfer vergt dat de deeloordelen worden gecombineerd en dat de combinatie wordt geprojecteerd op de voor schoolcijfers gebruikelijke schaal (van 1 tot 10).

Het vooraf expliciteren van een scoringsmodel of scoringsvoorschrift lijkt wellicht overdreven, maar is nodig om inzichtelijk, betrouwbaar en objectief te kunnen beoordelen.

Het *combineren van deelscores tot een totaaloordeel* gebeurt vaak door optelling. Daardoor kunnen deelscores voor elkaar compenseren. Is deze werkwijze wel verstandig? De bij onderzoek relevante deelvaardigheden (zoals: opstellen van een werkplan, verzamelen van gegevens, trekken van conclusies, rapporteren) hangen niet alle duidelijk met elkaar samen. Een leerling kan goed zijn in het ene en zwak in het andere, maar door onderlinge compensatie hetzelfde

(voldoende) eindcijfer krijgen. Aangezien een onderzoek slechts met succes kan worden uitgevoerd en afgerond als alle relevante vaardigheden in voldoende mate worden beheerst, is het wenselijk op z'n minst ook te verlangen dat per deelvaardigheid een voldoende oordeel wordt behaald.

Docenten zullen hun beoordelingen ook moeten (leren) *normeren*: welk prestatieniveau kan, respectievelijk moet worden verwacht? Omdat in veel vakken bij veel docenten open onderzoek door leerlingen een nieuw fenomeen is, zullen docenten er eerst meer ervaring mee moeten opdoen. Vanuit de normen die gelden voor wetenschappelijk onderzoek kan wel worden aangegeven op welke criteria moet worden gelet. Dat zijn in feite de te onderscheiden vaardigheden. De vraag 'wanneer is het goed genoeg' kan echter niet in absolute zin worden beantwoord. Realistische normen daarvoor ontstaan in de praktijk bij bekwame docenten en gemotiveerde leerlingen.

Gezien de condities waaronder docenten onderzoeken van leerlingen moeten beoordelen, is het niet zo zinvol cijfers te geven in decimalen. Eindtotalen kunnen wel op zo'n 'nauwkeurig' cijfer uitkomen, maar de betreffende vaardigheden zijn gewoon niet zo precies te meten. Anderzijds is een simpele driedeling (onvoldoende, voldoende, goed) wellicht wat grof. De praktijk moet uitwijzen tot welke mate van differentiatie zinvol beoordelingen kunnen worden gegeven en gebruikt.

*In ons onderzoek najaar 1998 bleek 50% van de docenten cijfers in decimalen te geven, 33% in halve punten, 17% in hele punten en 4% in termen van onvoldoende, voldoende en goed. Er waren hierin geen duidelijke verschillen tussen de in het onderzoek betrokken vakken.*

### **5.3 Registreren, communiceren en verantwoorden**

De bij beoordelingen gevolgde aanpak en verkregen resultaten moeten op de een of andere manier worden vastgelegd (al dan niet in aansluiting op de manier waarop andere gegevens worden geregistreerd, bijvoorbeeld in een leerlingvolgsysteem). De beoordelingen moeten ook kunnen worden gecommuniceerd naar de leerlingen, ouders en collega's. Het verdient in het algemeen aanbeveling een scoringsmodel te hanteren waarin op een niet te grove, maar ook niet te gedifferentieerde manier een beoordeling plaatsvindt van een aantal onderdelen of aspecten (deelvaardigheden), zodat inzicht wordt verkregen in de sterke en zwakke punten van de leerling. Zulke bevindingen zeggen meer dan alleen een eindoordeel of totaalcijfer en vormen daardoor meer zinvolle feedback. Met een dergelijk scoringsmodel kan ook de manier waarop de beoordelingen tot stand komen goed worden verduidelijkt en verantwoord, niet alleen aan de leerlingen, ouders en collega's, maar zonnodig ook aan de schoolleiding en de inspectie.

*In ons onderzoek najaar 1998 bleek dat 80% van de docenten de beoordelingscriteria schriftelijk vastlegt; ongeveer de helft legt meer vast dan alleen de criteria, zoals de te behalen scores per criterium en het antwoordmodel. De gegeven beoordelingen worden bijna altijd vastgelegd. Ze worden evenwel (slechts) door ongeveer de helft van de docenten altijd of meestal nabesproken (de overigen doen dit regelmatig, soms, of nooit). Bij ruim 80% van de docenten tellen de beoordelingen altijd of meestal mee voor het rapport (bij ongeveer 10% nooit) en bij ruim 60% voor het schoolexamen (bij een kleine 20% nooit). Er waren hierin geen duidelijke verschillen tussen de in het onderzoek betrokken vakken.*

## 6 MANIEREN EN ASPECTEN VAN BEOORDELEN (bij 2.3 en 2.4)

### 6.1 Inleiding

Deel II ging in op onderwijskundige achtergronden. Deel III (hoofdstuk 6) beschrijft een aantal alternatieven en argumenten, direct aansluitend op de richtlijnen en suggesties in deel I (hoofdstuk 2).

In de navolgende zeventien paragrafen passeren kort evenzovele aspecten de revue waarbij docenten kunnen kiezen wat en hoe ze onderzoekswerk van leerlingen beoordelen. Enkele paragrafen (6.4 en 6.5) vloeien direct voort uit hetgeen eerder in deze brochure is behandeld, aangaande de doelen die de docent nastreeft (zie paragraaf 3.1). Enkele andere paragrafen (6.6, 6.7, 6.16) zijn ontleend aan wetenschappelijke literatuur over beoordelen, de betrouwbaarheid ervan en de integratie met het onderwijsleerproces (zie de hoofdstukken 10 en 11).

Veel van de (overige) paragrafen komen direct voort uit de taken die in de tweede fase aan docenten zijn toebedeeld en de problemen die de uitvoering daarvan in de praktijk met zich mee blijkt te brengen. De wettelijke taken van de docent gaan niet veel verder dan wat in 6.2 wordt aangestipt: de beoordeling moet zich uitstrekken over het proces, het product, respectievelijk de presentatie. Het *beoordelen van het proces* is lastig als leerlingen zelfstandig in groepjes (ook) buiten aanwezigheid van de docent aan hun onderzoek werken. Hierover gaan de paragrafen 6.3, 6.4, 6.10, 6.11, 6.14. Een andere 'hot topic' betreft het *verdisconteren van de begeleiding* die de docent tussentijds heeft gegeven. Dat komt aan de orde in 6.4, 6.10, 6.12, 6.15. Een derde notoir probleem is het geven van waarderings aan *individuele leerlingen* als het werk is gedaan in een groep. Zie 6.3-6.5, 6.11 en 6.15. De mogelijke positie van de *vakinhoudelijke kennis* bij de beoordeling van onderzoekswerk van leerlingen komt aan bod in 6.5, 6.6, 6.13, 6.15 en 6.17. Bepaalde maatregelen kunnen *de betrouwbaarheid en aanvaardbaarheid van de beoordeling* vergroten. Zie daarvoor 6.2, 6.3, 6.7, 6.8, 6.16, 6.17. Een laatste insteek: de mogelijkheden om beoordelingen een rol te laten spelen in *het leerproces* zelf. Zie daarvoor de paragrafen 6.4, 6.7, 6.12 en 6.15-6.17.

### 6.2 Proces, product en/of presentatie

Volgens de aan het onderwijs verstrekte aanwijzingen moeten docenten bij onderzoekswerk door leerlingen drie zaken beoordelen: proces, product en presentatie.

Ten aanzien van het *proces* is menigeen geneigd zich primair af te vragen *hoe* daarop zicht kan worden verkregen. De antwoorden daarop liggen voor de hand: observeren (dat kan alleen voorzover leerlingen in aanwezigheid van de docent aan het onderzoek werken), voeren van tussentijdse gesprekken (maar dat is arbeidsintensief), leerlingen een logboek laten bijhouden (maar daaraan hebben veel leerlingen een hekel), leerlingen tussenproducten laten maken (deze oplossing is misschien het meest werkbaar en biedt bijkomende voordelen) (zie 6.4). Er wordt weinig aandacht gegeven aan *wat* moet worden beoordeeld.

- Indien het daarbij zou gaan om zaken als de zelfstandigheid of de onderlinge samenwerking kan men zich afvragen in hoeverre die bij het beoordelen van *onderzoeksvaardigheden* een rol moeten spelen.
- Voorzover het zou gaan om de (methodische) keuzes die leerlingen maken en de mate waarin ze die beargumenteerd maken, betreft het zaken die in een goed verslag zullen worden opgenomen en dus aan de hand van het product kunnen worden beoordeeld.
- Een derde mogelijkheid is dat het de bedoeling is om de 'authenticiteit' te controleren. Dan echter betreft het geen aspect dat op zich moet worden *beoordeeld*, maar meer een randvoorwaarde voor de geldigheid van de uit te brengen beoordeling.

De docent moet steeds bedenken dat het proces mede afhangt van de gegeven opdracht en begeleiding. Naarmate die meer gestructureerd zijn, zal in de meeste gevallen het proces meer gestandaardiseerd verlopen (en neemt de zin van een aparte beoordeling van het proces af).

Vaak zal het onderzoekswerk van de leerlingen uitmonden in een *product* in de zin van een verslag. Er valt veel voor te zeggen in principe altijd tenminste een verslag te laten maken. Het op papier onder woorden moeten brengen van opzet, uitvoering en resultaten van een onderzoek kan de discussie en reflectie stimuleren. Bovendien is rapporteren één van de onderzoeksvaardigheden. De docent zal de leerlingen wel vooraf duidelijk moeten maken aan welke eisen het product moet voldoen. Het ligt voor de hand dat in een verslag zowel het *onderzoek* (als onderzoek) als de relevante *vakinhoud* aan de orde komt, terwijl het verslag als tekst en *communicatief product* ook aan bepaalde eisen moet voldoen. Bovendien lijkt het verstandig ook bepaalde eisen te stellen aan de vorm en de omvang; dit scheidt niet alleen duidelijkheid voor de leerlingen, maar draagt ook bij aan een meer gelijkwaardige en daardoor meer verantwoorde beoordeling.

De beoordeling van de *presentatie* valt samen met die van het product, als het schriftelijke product de gekozen vorm van presenteren is. Mondelinge presentaties zijn een verhaal apart. Daarvoor geldt hetzelfde als voor bijvoorbeeld de samenwerking, namelijk dat men zich kan afvragen in hoeverre het daarbij gaat om *onderzoeksvaardigheden*. Eerder lijkt mondeling presenteren een vaardigheid op zich.

Voorzover de docent aparte beoordelingen geeft van de onderdelen proces, product en presentatie ontstaat de vraag hoe deze tot een totaal kunnen worden gecombineerd. Het Cito heeft voorgesteld de deelscores gewoon op te tellen, ook al verschillen de aantallen aspecten per onderdeel (Zie Van der Krogt en Sinkeldam, 1998). Het gewicht dat de onderdelen krijgen is dan echter afhankelijk van het aantal onderscheiden aspecten, terwijl de onderdelen bovendien voor elkaar compenseren. Bij het opzetten en uitvoeren van een onderzoek en het mondeling presenteren is echter sprake van duidelijk verschillende vaardigheden, waarbij de ene leerling goed kan zijn in het een en zwak in het ander en de andere leerling omgekeerd. Wat is het informatiegehalte van een totaaloordeel waarin dit zonder meer wordt gecombineerd?

*In ons onderzoek najaar 1998 bleek dat een kwart van de docenten het proces niet meeneemt in de beoordeling, de helft wel, maar impliciet en het resterende kwart zegt het proces expliciet mee te nemen.*

### 6.3 Logboek en/of geïntegreerd eindverslag

Voor het beoordelen van het proces wordt wel voorgesteld de leerling een logboek te laten bijhouden met daarin de planning, taakverdeling, afspraken, gemaakte keuzes en argumenten daarvoor en dergelijke. Men kan zich afvragen wat

in de praktijk de zin van zo'n logboek kan zijn gezien de mate waarin leerlingen het serieus zullen invullen en de relevantie van de informatie voor de beoordeling. Een logboek kan een functie vervullen voor de leerlingen om zicht te houden op de afspraken en de voortgang en het speelt misschien een rol bij de communicatie met de docent. Daarmee is het echter nog geen document dat een rol kan of moet spelen bij de beoordeling. Onze suggestie:

- een logboek zien als een hulpmiddel, waarvan men het gebruik aan leerlingen overlaat of hooguit een functie geeft in de communicatie;
- alle meer inhoudelijke zaken zoals keuzes en argumenten daarvoor laten beschrijven in het verslag; zulke informatie hoort daarin ook thuis;
- de eventuele knelpunten en daarvoor bedachte oplossingen eveneens laten beschrijven in het product en wel in de hoofdtekst voorzover het voor het onderzoek relevante zaken betreft (zoals problemen bij het verzamelen van voldoende betrouwbare en relevante gegevens) en in een toegevoegde zelfevaluatie, waar het de samenwerking tussen de leerlingen betreft;
- in een aparte toevoeging laten beschrijven welke activiteiten door welke leerlingen zijn uitgevoerd (en welke dingen echt samen zijn gedaan), aan te vullen met een korte zelfbeoordeling daarvan; dit kan de reflectie door de leerlingen stimuleren en nuttige informatie opleveren voor het eventueel verbijzonderen van de beoordeling naar individuele leerlingen.

*In ons onderzoek najaar 1998 is door een kleine 40% van de docenten aangegeven dat zij de leerlingen geen logboek laten bijhouden, ruim 30% doet het af en toe en een kwart doet het regelmatig of meestal. Degenen die het doen, hebben meestal als doel zelf zicht te houden op de voortgang en daarnaast dat leerlingen voor zichzelf dingen vastleggen. Een aantal doet het (ook) ter controle op een eerlijk en serieus verloop en een minderheid doet het om de reflectie bij de leerlingen te stimuleren. In overeenstemming hiermee worden in het logboek het vaakst de planning en taakverdeling vastgelegd en daarnaast problemen en vragen aan de docent, terwijl bij een minderheid beslissingen, argumenten en resultaten in het logboek worden opgenomen.*

#### **6.4 Tussentijdse beoordelingen en/of eindbeoordeling**

Een docent kan eventueel volstaan met een beoordeling van het eindproduct. Als de relevante gegevens met betrekking tot het proces ook in dat product zijn opgenomen, kan het eindproduct een voldoende basis zijn voor de beoordeling. Een docent kan het onderzoek ook enkele malen tussentijds beoordelen. Dat biedt enkele duidelijke voordelen:

- het afbreukrisico wordt verkleind; leerlingen krijgen tussentijds feedback, kunnen zaken zonodig bijstellen en kunnen met vertrouwen verder; hun mogelijkheden om zinvol te oefenen met volgende stappen worden vergroot;
- er kan meer gedifferentieerde feedback worden gegeven; het ene groepje heeft meer problemen met het komen tot een onderzoekbare vraagstelling, een ander groepje zit vast met de gegevensverwerking et cetera; een docent kan dit effectiever begeleiden door enkele tussentijdse momenten in te lassen;
- de docent houdt beter zicht op de samenwerking in het groepje en de bijdragen van de individuele leerlingen;
- de beoordeling door de docent wordt in het algemeen betrouwbaarder, als deze wordt gebaseerd op meerdere deelbeoordelingen.

Gelet op het onderzoeksproces ligt het voor de hand om op twee of drie momenten een tussentijdse beoordeling in te lassen na:



- keuze en formulering van onderwerp, probleem, onderzoeksvraag;
- keuze en uitwerking van de te volgen onderzoeksopzet (inclusief het instrumentarium, de keuze van de gegevensbronnen, en dergelijke);
- verwerking van gegevens en vervaardiging van een conceptverslag.

Wie wil werken met tussentijdse beoordelingen moet daarvoor uiteraard criteria opstellen en vantevoren aan de leerlingen bekend maken. Tussentijds beoordelen kost tijd. Deze tijd zal beperkter zijn naarmate men werkt met een beperkter aantal duidelijke en goed gekozen criteria en naarmate men de groepjes groter laat zijn. Er valt ook tijdwinst te boeken. Ten eerste zal bijsturing door de docent problemen in een later stadium voorkomen en de dan benodigde begeleiding beperken. Ten tweede kunnen onderdelen die tussentijds zijn beoordeeld worden uitgespaard in de eindbeoordeling.

*In ons onderzoek najaar 1998 bleek dat bijna alle docenten het werk beoordelen na het indienen van het eindverslag. Een kwart doet het (pas of ook) na de bespreking daarvan met de leerlingen. Ruim 40% beoordeelt ook het werkplan en ruim 5% geeft na elke onderzoeksstap een beoordeling.*

### 6.5 Beoordelen op vakinhoud en/of onderzoek

Een docent kan rond onderzoek door leerlingen verschillende doelen nastreven (vergelijk paragraaf 3.1). De ene docent zal veel nadruk leggen op de vakinhoudelijke kant, de andere meer op het proces van onderzoek doen als zodanig. Het ligt voor de hand om dit ook in de beoordeling tot uitdrukking te laten komen. Bij een aantal vakken zijn in de examenprogramma's expliciet eindtermen opgenomen over het adequaat hanteren van het vakeigen begrippenkader. Ook bij vakken waar dat in de eindtermen niet zo duidelijk aan de orde komt, zullen docenten mede of met name daarop willen beoordelen.

Er kunnen zich nu (tenminste) twee situaties voordoen. (a) Docenten kunnen van mening zijn dat de aard en ontwikkeling van kennis in hun vak dermate is verweven met onderzoek, dat beide niet goed kunnen worden gescheiden. (b) Docenten kunnen er ook geen enkele moeite mee hebben de vakinhoud en het onderzoek doen als twee apart te beoordelen zaken te zien of dit zelfs vanzelfsprekend vinden. Dit hangt wellicht samen met de vakken. Het is voorstelbaar dat bijvoorbeeld docenten natuurkunde meer de eerste mening zijn toegegaan en docenten economie meer de tweede mening. Ook kan binnen eenzelfde vak sprake zijn van verschillende benaderingen. Een algemene stelregel zou zijn dat wie duidelijk onderscheid maakt tussen het juist hanteren van de vakspecifieke begrippen en het correct opzetten en uitvoeren van een onderzoek, deze componenten ook afzonderlijk beoordeelt. Dit nog afgezien van de vraag in hoeverre het dezelfde leerlingen zullen zijn die op beide hoog, respectievelijk laag scoren. Juist door vakinhoud en onderzoek apart te beoordelen kan dit worden vastgesteld. Docenten doen er juist goed aan beide apart te beoordelen om te vermijden dat hun oordeel over het een hun oordeel over het andere gaat beïnvloeden. Een gescheiden beoordeling zou ook beter recht doen aan de mogelijkheid dat leerlingen beter kunnen zijn in het ene en zwakker in het andere. Een ongedifferentieerde totaalscore is dan immers niet zo informatief. Verder kan het onderscheid de samenwerking tussen vaksecties bevorderen.

*In ons onderzoek najaar 1998 gaf driekwart van de docenten aan het inhoudelijke niveau van het verslag het zwaarst mee te tellen; tweederde telt (ook) de probleemstelling/onderzoeksvragen en de conclusies het zwaarst mee (ook nog sterk inhoudelijk!) en ruim de helft (ook) de opzet van het onderzoek, de uitvoering van het onderzoek en de opbouw van het verslag.*

## 6.6 Beoordelen van kennis en/of vaardigheden

Het complexe domein 'onderzoeksvaardigheden' betreft zowel kennis als vaardigheden en bestrijkt een aantal componenten. We noemen:

- vakinhoudelijke kennis;
- kennis over en inzicht in onderzoek (onderzoek als kennisdomein);
- diverse deelvaardigheden: formuleren van een onderzoekbare vraagstelling, opstellen van een werkplan et cetera, tot en met rapporteren over het onderzoek;
- de vaardigheid in het bewaken ('monitoren') van de voortgang van een onderzoek, ervoor zorgen dat het onderzoek op koers blijft;
- de vaardigheid om achteraf te reflecteren op uitgevoerd onderzoek.

Eventueel kan men ook meer randvoorwaardelijke vaardigheden toevoegen:

- de vaardigheid in het reguleren van het eigen handelen (jezelf motiveren om door te zetten, dwingen nauwkeurig te werken, niet laten afleiden et cetera);
- de vaardigheid in het samenwerken met anderen.

Dat bij het doen van onderzoek diverse kennis- en vaardigheidsgebieden aan de orde zijn, wil nog niet zeggen dat een docent deze alle moet beoordelen. Enerzijds maakt de docent keuzes. Een docent die de kennis over onderzoek niet centraal stelt, hoeft die kennis in principe ook niet te beoordelen. Anderzijds is het nog niet zo eenvoudig om naast kennis nu ook vaardigheden te gaan meten en beoordelen. Bovendien is het totale domein van kennis en vaardigheden dat voor onderzoek van belang is eenvoudigweg te omvangrijk om van docenten te kunnen verwachten dat zij dat allemaal wel even meenemen in hun beoordelingen van onderzoekswerk van leerlingen. Het motto is dus: beperk de doelen, maar werk daar vervolgens dan ook gericht aan, zowel in de opdrachten en begeleiding als bij de beoordeling. Dat is voor de docent het meest bevredigend (wie zich beperkt, heeft meer kans iets te bereiken) en naar leerlingen toe het meest duidelijk en effectief. Vanuit het leerproces van de leerlingen gezien ligt het misschien het meest voor de hand als docenten zich tijdens de eerste jaren dat zij werken met min of meer zelfstandig uit te voeren open onderzoek door leerlingen richten op de eerste drie bovengenoemde componenten: de vakinhoudelijke kennis, de basiskennis over onderzoek en de meest relevante onderzoeksdeelvaardigheden. Dat betekent niet dat de docent tijdens de begeleiding bijvoorbeeld geen aandacht schenkt aan de samenwerking tussen de leerlingen, maar zolang dat geen expliciet doel is, blijft dat buiten de beoordeling.

## 6.7 Globale beoordeling en/of gespecificeerd scoringsmodel

Het is voor een docent het minste werk om het eindproduct door te lezen en vervolgens 'op gevoel' een 6, een 7 of wat dan ook uit te delen. Een dergelijke manier van beoordelen is echter vatbaar voor toevalligheden en subjectiviteiten (ook als de docent dat zelf niet wil). Een globaal oordeel is ook weinig informatief: de sterke en zwakke punten blijven zo buiten beeld. In het algemeen verdient het de voorkeur de te beoordelen aspecten, respectievelijk criteria, enigszins te specificeren. Dat wil niet zeggen dat een docent lange lijsten moet gaan aanleggen. Een aantal van vijf tot tien zal vaak adequaat zijn, dat wil zeggen: inhoudelijk redelijk dekkend en voldoende om te kunnen leiden tot een meer betrouwbare, objectieve en informatieve beoordeling. Alleen al het ontwikkelen van een dergelijke lijst is een belangrijke exercitie. Daardoor komt boven tafel wat men (kennelijk) belangrijk vindt. Voor leerlingen is een lijst met criteria een belangrijk

houvast. Wie zo'n lijst uitreikt en bespreekt, is in feite bezig de leerlingen effectief te instrueren. Werken met een gespecificeerd model kan er ook aan bijdragen dat de onderzoeken van verschillende (groepjes) leerlingen op een vergelijkbare manier worden beoordeeld oftewel dat men als beoordelaar consistent blijft. Docenten die werken met een gespecificeerde lijst met criteria kunnen ook gemakkelijker de leerlingen zelf een rol geven bij de beoordeling. Het ligt voor de hand de specificatie te baseren op de te onderscheiden stappen. De uitgebrachte deelbeoordelingen kan men als gedetailleerd *profiel van de sterke en zwakke punten* communiceren naar de leerlingen. Hoe kan dit leiden tot een cijfer? Een veel gebruikte manier bij toetsen van kennis is te vertrekken vanuit de perfecte prestatie (alles goed is een 10). Bij vaardigheden is echter niet zo'n duidelijk perfecte prestatie aan te wijzen en moeten bovendien vaak alle relevante onderdelen elk voor zich op een bepaald niveau worden beheerst. De gebruikelijke werkwijze *punten aftrekken* kan men handhaven voor 'vanzelfsprekende' onderdelen, maar verder beter vervangen door *punten geven*, namelijk voor wat goed ging.

*In ons onderzoek najaar 1998 bleek dat 15% van de docenten een globaal cijfer geeft voor het onderzoek als geheel, 25% een globaal cijfer voor elk onderdeel (op basis waarvan een eindoordeel wordt bepaald), 40% een cijfer voor het onderzoek als geheel op basis van scoring van een aantal aspecten en 25% een cijfer voor elk onderdeel op basis van scoring van aspecten.*

### 6.8 Categorisch en/of vergelijkend beoordelen

Het waarden van de prestaties van leerlingen op toetsen waarmee kennis wordt gemeten is relatief eenvoudig: antwoorden zijn goed of fout en dat is meestal eenduidig vast te stellen. Het beoordelen van vaardigheden, zeker als het cognitieve vaardigheden betreft en als de beoordeling gebeurt aan de hand van producten, is minder simpel. Docenten die nog niet vaak resultaten van onderzoekswerk van leerlingen hebben beoordeeld, zullen er waarschijnlijk moeite mee hebben te bepalen hoe ze een gegeven product moeten waarden. Daarvoor zijn geen duidelijke normen en deze docenten hebben ze ook nog niet in de praktijk voor zichzelf kunnen ontwikkelen.

Beoordelen op een categorische manier, dat wil zeggen zonder vergelijkingsmateriaal (normen, of andere producten), is voor mensen lastig, lastiger dan beoordelen op een vergelijkende manier. Een docent die alle producten van een klas naast elkaar legt, kan moeite hebben om te bepalen of een bepaald product nu een 7 of een 8 waard is, maar zal zich wel behoorlijk zeker voelen bij het ordenen van deze producten in betere, minder goede en zwakke. Doordat vergelijkend beoordelen voor mensen gemakkelijker is, maken zij daarbij ook minder fouten en zijn dergelijke oordelen gemiddeld genomen betrouwbaarder.

De moraal van dit deel van het verhaal is dat docenten die nog ervaring moeten opdoen met het beoordelen van onderzoekswerk tot betere beoordelingen kunnen komen als zij alle producten bijvoorbeeld eerst grof sorteren op onvoldoende en (tenminste) voldoende en vervolgens de tweede groep of beide groepen nader onderverdelen. Als de docent zich vervolgens afvraagt op welke gronden hij een product aan de ene of andere stapel toewijst, zullen de criteria die hij kennelijk belangrijk vindt boven water komen. Deze criteria kunnen worden gebruikt bij de ontwikkeling van een vaste lijst met aandachtspunten, die de volgende keer kan worden gebruikt.

Naarmate de onderdelen of aspecten waarop moet worden gelet en de criteria waaraan daarbij moet zijn voldaan concreter zijn gespecificeerd, wordt een categorische beoordeling minder problematisch. Dat komt deels doordat de deelbeoordelingen meer zijn gericht op specifieke punten en deels doordat bij het combineren van deelbeoordelingen

eventuele beoordelingsfouten naar boven en naar beneden vaak min of meer tegen elkaar wegvallen.

Het vergelijkend beoordelen van producten (alle producten naast elkaar, zie boven) wordt lastig als de beoordeling gebeurt aan de hand van een aantal specifieke aandachtspunten. Meestal wordt dan *per product de hele beoordelingslijst* afgewerkt. Dat heeft als voordeel dat men zich goed in het product kan verdiepen en eventueel bij bepaalde deelbeoordelingen rekening kan houden met wat men weet over andere aandachtspunten. Deze manier van werken kan de validiteit van de beoordeling ten goede komen, doordat men het product zo volledig en afgewogen mogelijk tot z'n recht laat komen. Een mogelijk nadeel is dat bij de beoordeling een bepaalde totaalindruk kan gaan overheersen (ondanks de gespecificeerde lijst met criteria). Een alternatief is om niet per product alle criteria langs te lopen, maar *per criterium alle producten*. Dat is lastiger, doordat men steeds opnieuw in elk product (onderzoek) moet duiken, maar biedt de mogelijkheid per criterium de producten te vergelijken, dus vergelijkend te beoordelen, waardoor de kwaliteit van de beoordeling kan toenemen. Deze alternatieve aanpak is werkbaarder naarmate de producten (verslagen) een min of meer vaste opbouw hebben. Een docent zou op z'n minst af en toe deze alternatieve aanpak (per criterium alle producten langslopen) kunnen toepassen om daarmee een check te verkrijgen op de objectiviteit (respectievelijk 'normvastheid') van zijn beoordelingen.

*In ons onderzoek najaar 1998 gaf ruim de helft van de docenten aan de normering te bepalen aan de hand van de leerstof, respectievelijk de exameneisen; eveneens ruim de helft baseert de normering (ook) op de prestaties van de rest van de klas of van parallelklassen of voorgaande leerjaren en een kwart doet het (ook) op basis van de vorderingen van de individuele leerlingen. Veel docenten gaan bij het normeren dus vergelijkend te werk.*

#### **6.9 Beoordeling en normering op één of meer niveaus**

Bij het geven van cijfers passen veel docenten zich aan bij het niveau van de leerlingen. Daarin ligt een probleem: de beoordelingen van dezelfde prestaties kunnen anders uitvallen alnaargelang de groep (klas, jaargang) waarvan deze leerlingen deel uitmaken. Een relatief goede leerling in een zwakke klas kan bij een objectief gezien gelijke prestatie een hogere waardering verkrijgen dan een relatief zwakke leerling in een goede klas et cetera. Dit risico zal kunnen afnemen als zich geleidelijk duidelijker normen gaan ontwikkelen voor de kwaliteit van onderzoekswerk van leerlingen, respectievelijk voor het niveau dat van leerlingen kan worden verwacht.

Naarmate de ervaring van docenten met het geven van geschikte opdrachten en het begeleiden toeneemt en ze meer consistent samenwerken en elkaars pogingen om leerlingen kennis en vaardigheden te laten ontwikkelen meer systematisch ondersteunen en naarmate leerlingen in de onderbouw (basisvorming) al meer zelfstandig leren werken en leren en meer onderzoeksvaardigheden (en andere algemene vaardigheden) ontwikkelen, kunnen de verwachtingen aangaande het niveau van hun onderzoek in de tweede fase toenemen en kan de lat geleidelijk hoger worden gelegd.

Het lijkt voor de hand te liggen om de lat voor havo-leerlingen minder hoog te leggen dan voor vwo-leerlingen. De formuleringen van de eindtermen ten aanzien van de algemene vaardigheden in de examenprogramma's van de verschillende vakken zijn weliswaar voor havo en vwo geheel of grotendeels gelijk, maar het lijkt terecht om te veronderstellen dat de opstellers toch impliciet wel van een zeker niveauverschil zullen zijn uitgegaan. Zoals in het algemeen ten aanzien van zelfstandig door leerlingen uitgevoerd onderzoek zich nog normen moeten ontwikkelen, geldt dit ook voor het onderscheid tussen havo en vwo. Docenten die zowel in havo als in vwo werken, zullen hierop goed zicht kunnen krijgen en scholen zullen hierover afspraken kunnen (en moeten) ontwikkelen. De vraag is dan wel waar men stopt met

differentiatie in de beoordeling. Moeten dan niet ook leerlingen in het ene profiel anders worden beoordeeld dan leerlingen in het andere profiel of leerlingen die wetenschappelijk onderwijs gaan volgen anders dan leerlingen die dit niet gaan doen?

### 6.10 Beoordelen bij tussentijdse bijsturing door de docent

Bij het begeleiden van zelfstandig onderzoek door leerlingen zullen docenten vaak tussentijds commentaar geven, (groepjes) leerlingen ondersteunen en het proces bijsturen als duidelijk wordt dat leerlingen bezig zijn om vast te lopen of het onderzoek dreigt te mislukken.

Men kan stellen dat de docent bij het (formeel) beoordelen van het werk van de leerlingen rekening moet houden met de mate waarin tussentijds hulp is gegeven, respectievelijk is bijgestuurd. Deze hulp of bijsturing zal immers verschillen tussen onderzoeken en daardoor zou de beoordeling van het werk van verschillende groepjes niet eerlijk plaatsvinden. Er zijn ook argumenten om hierin terughoudend te zijn:

- Zodra leerlingen weten dat vragen om hulp kan leiden tot aftrek van punten, zullen zij daar minder snel toe overgaan, terwijl:
  - het signaleren wanneer men bij een ander om advies of steun moet vragen en het ook durven vragen daarvan op zich kunnen worden beschouwd als wenselijke vaardigheden;
  - het tussentijds doorspreken van de voortgang het afbreukrisico kan beperken (en veel frustratie kan vermijden);
  - het tussentijds geven van begeleiding kan bijdragen aan de leermogelijkheden en het leerproces van de leerlingen (met betrekking tot zowel de vakinhoud als het doen van onderzoek).
- Het valt te verwachten dat leerlingen die onderweg veel hulp nodig hebben toch niet met een eindresultaat zullen komen dat aan hoge normen voldoet; door op de beoordeling van het eindproduct ook nog eens de gegeven hulp in mindering te brengen worden zij 'dubbel gepakt'.

Een soort tussenoplossing is 'normale' hulp niet in rekening te brengen, maar zware bijsturing wel. Veel zal afhangen van de aard van de vraag van leerlingen (is deze wel nodig of lopen ze wel erg snel naar de docent toe). Verder is erg belangrijk op welke manier de docent begeleidt. Wie vooral procesgerichte feedback geeft, heeft minder aanleiding om de gegeven begeleiding in rekening te brengen dan wie in feite een deel van het werk voor de leerlingen zelf heeft uitgevoerd.

Er is ook een argument om juist geen koudwatervrees te hebben voor het verdisconteren van de begeleiding, met name in de situatie dat de docent tussentijdse beoordelingen geeft: als leerlingen bij een bepaalde stap duidelijk hulp van de docent nodig hebben en *daarom* voor die stap een lagere tussenbeoordeling krijgen, zullen zij met de gegeven hulp hun voordeel kunnen doen en des te beter kunnen scoren bij de volgende stap, hetgeen de eerder gegeven 'aftrek' compenseert.

In feite is de situatie echter niet zo simpel, dat het alleen gaat om een soort principiële keus tussen wel of niet verdisconteren van de gegeven begeleiding. Leerlingen, respectievelijk groepjes leerlingen zullen verschillen in ambitieniveau, in het risico dat zij met een onderzoek willen lopen en in assertiviteit bij het inschakelen van anderen als hulpbronnen. En de begeleiding door de docent zal naar aanleiding van deze en andere factoren ongetwijfeld verschillen van groepje tot groepje (en daar is op zich niets mis mee). Welnu: waar de gang van zaken bij de verschillende groepjes leerlingen-onderzoekers op een aantal punten uiteen zal lopen, is het minder makkelijk verdedigbaar één punt, zoals de

gegeven hulp of sturing, te verdisconteren in de beoordeling.

Men kan ook redeneren dat een docent tijdens de fase van de formulering van de probleemstelling en de ontwikkeling van een onderzoeksopzet erop toeziet en zonodig erop aanstuurt dat het onderzoek voor de betreffende leerlingen van een adequaat moeilijkheidsniveau is (niet te moeilijk, maar ook niet te gemakkelijk), waarna, als het goed is, niet veel begeleiding meer zal hoeven te worden gegeven.

Een punt op zich is nog de manier waarop de docent zicht houdt op de begeleiding die de groepjes krijgen. Een mogelijke optie is de van de docent verkregen feedback tot vast onderdeel te maken van wat de leerlingen in hun verslag moeten beschrijven, verantwoorden en evalueren. Een andere optie is dat de docent er zelf een logboek van bijhoudt.

### 6.11 Individuele beoordeling van groepsprestaties

Praktische opdrachten, respectievelijk onderzoeken doen leerlingen in veel gevallen in een klein groepje. Moet de waardering voor het werk van zo'n groepje niet worden gedifferentieerd naar de bijdragen van de individuele leerlingen, en zo ja, hoe zou dat dan kunnen? Net als ten aanzien van een aantal andere punten die in deze brochure worden besproken, zijn hierbij diverse standpunten mogelijk.

Allereerst kan men stellen: waarom zou je differentiëren? Alle leerlingen van een groepje krijgen gewoon dezelfde beoordeling. Als de leerlingen dit maar van tevoren weten, is dat acceptabel. Een bezwaar kan zijn dat de groepjes verschillen van samenstelling. Een leerling die lid is van een 'beter' groepje heeft dan voordeel. Als leerlingen weten dat het hele groepje dezelfde waardering krijgt, zullen 'betere' leerlingen wellicht slechts 'bij elkaar' in een groepje willen zitten en de groepjes nog duidelijker in niveau gaan verschillen. Dit bezwaar zou men kunnen proberen te ondervangen door bij het beoordelen (van de groepsprestatie) hiermee rekening te houden. Andere opties zijn dat de docent erop toeziet dat de groepjes voor een onderzoek heterogeen en dus gelijkwaardig worden samengesteld (bijvoorbeeld door simpelweg te loten), of dat de groepjes in de loop van de tijd steeds van samenstelling veranderen. Dan is er op dit punt geen probleem: in het eerste geval niet per onderzoek, in het tweede geval niet op de lange duur.

Een iets ander standpunt is dat het vreemd is, respectievelijk onmogelijk bij groepsprestaties individuele oordelen te geven. Immers: leerlingen zullen in een groepje hoe dan ook van elkaar leren en elkaar beïnvloeden. Hoe intensiever de samenwerking en onderlinge discussies, hoe minder de individuele bijdragen zullen kunnen worden gereconstrueerd. Webb (1995) geeft aan dat in groepen waarin men elkaar helpt, elkaar dingen uitlegt, samen dingen bediscussieert en gelijkwaardig betrokken is meer wordt geleerd en gemiddeld genomen een beter groepsproduct ontstaat. In heterogene groepen die zo functioneren leren ook de sterkere leerlingen veel, niet *hoewel* maar juist *doordat* zij als trekker, uitlegger en dergelijke fungeren. Twee duidelijke aanwijzingen uit de door Webb besproken onderzoeken zijn, dat (a) meelifgedrag slecht is voor het leren van de leerlingen in een groep en bij open opdrachten ook voor het groepsproduct en (b) het in de groep verdelen van taken (in plaats van alle taken samen aanpakken) gunstig kan zijn voor het groepsproduct, maar slecht is voor het leerproces van de deelnemers. Resumerend: heterogene groepjes waarin gelijkwaardig wordt samengewerkt hebben drie voordelen: de leerlingen leren meer (qua kennis en vaardigheden), de groepsprestatie zal hoger kunnen zijn (NB: in deze optiek wordt goede samenwerking bij een bonus daarvoor dus in feite dubbel beloond) en het geven van eenzelfde beoordeling aan alle leden van de groep is geen probleem.

Een ander standpunt is dat leerlingen die er persoonlijk weinig aan hebben gedaan (de situatie die het meest aanleiding

is voor differentiatie van de beoordeling naar individuele leerlingen) ook minder zullen hebben geleerd, hetgeen zich zal wreken bij proefwerken en cetera. Een eventueel voor zo'n leerling te hoge (groeps)beoordeling corrigeert zich in de loop van de tijd dan vanzelf.

Dan is er tenslotte het standpunt dat de beoordeling van een door een groepje uitgevoerd onderzoek wel moet worden gedifferentieerd naar de individuele leerlingen. Wegen waarlangs dit zou kunnen, zijn:

- de leerlingen doen een belangrijk deel van de onderzoeksactiviteiten op school in aanwezigheid van de observerende docent;
- de docent vraagt direct na afronding van het onderzoek individueel door over het *onderzoek*, respectievelijk het *product*;
- de docent geeft mondelinge beurten over hetzelfde *onderwerp*;
- de docent verlangt dat in het eindproduct ook aandacht wordt besteed aan (en eventueel zelfs reflectie, evaluatie en/of zelfbeoordeling plaatsvindt op) de bijdragen van de afzonderlijke leden van de groep.

Een specifieke optie is de volgende. Bepaal wat het werk waard is, vermenigvuldig dat cijfer met de omvang van het groepje en laat de leerlingen het totaal onder elkaar verdelen alnaargelang ieders inbreng. In de praktijk blijkt dit lang niet altijd goed te werken door voor de hand liggende oorzaken: de ene leerling is assertiever dan de andere en beter in staat voor zichzelf te zorgen, een leerling kan minimaal een bepaald cijfer nodig hebben om op het volgende rapport voldoende te staan en door de overigen daarbij worden geholpen en dergelijke. Het voorstel is ook niet aanvaardbaar, doordat de docent de beslissing welke cijfers de individuele leerlingen krijgen hierbij helemaal in handen van de leerlingen geeft.

## 6.12 Beoordeling in relatie tot de keuzevrijheid van de leerlingen

Bij min of meer zelfstandig uit te voeren min of meer open onderzoek hebben de leerlingen een zekere keuzevrijheid. Dit heeft gevolgen voor de beoordeling. Ten eerste kan een onderzoek een kant opgaan waarbij de docent zich minder deskundig weet (qua inhoudelijk onderwerp, qua soort onderzoek en dergelijke) en daardoor ook minder gerichte begeleiding kan bieden. Dit laatste moet de leerlingen uiteraard duidelijk zijn, voordat zij definitieve keuzes maken. Het is overigens denkbaar dat docenten van een vaksectie verschillen qua expertise en dat de begeleiding van de groepjes over de docenten wordt verdeeld mede op grond van onderwerpen die zijn gekozen en/of de soorten onderzoek. Ten tweede zullen de onderzoeken van de verschillende groepjes leerlingen onderling gaan verschillen, waardoor strikt genomen de vergelijkbaarheid van de beoordelingssituatie kan verminderen. Ten derde kan het zijn dat door bepaalde keuzes (bijvoorbeeld voor een originele, maar ook riskante onderzoeksopzet) een of meer onderdelen of aspecten van het onderzoek minder goed uit de verf komen, waardoor min of meer 'buiten de schuld van de leerlingen' een lagere beoordeling dreigt. Ook deze laatste consequentie moet de leerlingen uiteraard duidelijk zijn, alvorens zij definitieve keuzes maken. Ten vierde kan door bepaalde keuzes een onderdeel of aspect, dat de docent wel rekent tot de kennis en vaardigheden die de leerlingen bij onderzoek zouden moeten leren eenvoudigweg niet van toepassing zijn of niet aan bod komen. Een groepje heeft bijvoorbeeld ontdekt dat de vraagstelling van hun onderzoek kan worden beantwoord door gebruik te maken van al beschikbare gegevens en daarop uit te voeren analyses; zo'n ontdekking, respectievelijk keuze is op zich zonder meer positief te waarderen, maar bepaalde vaardigheden in het verzamelen van gegevens blijven hierdoor buiten schot.

In al deze gevallen zal de docent ervoor moeten zorgen dat het onderzoekswerk op een adequate en verantwoorde

manier beoordeelbaar blijft. Voorzover het geven van keuzemogelijkheden aan de leerlingen een goede beoordeling in de weg staat, zou kunnen worden besloten dergelijke mogelijkheden niet te bieden. Voorzover keuzemogelijkheden wel open staan, maar nadelen kunnen hebben zoals boven geïllustreerd, moet dit voor de leerlingen vooraf duidelijk zijn.

### **6.13 Controleerbaarheid op juistheid van gegevens en uitkomsten**

Bij open onderzoek kunnen leerlingen onderwerpen, vraagstellingen, onderzoeksopzetten en gegevensbronnen kiezen waarmee de docent niet of niet goed bekend is. Daardoor kunnen in het verslag of de presentatie gegevens, uitkomsten en conclusies verschijnen die de docent niet goed kan controleren om te zien of een fout of vergissing is begaan (wat iets anders is dan fraude). Dit probleem kan enigszins worden ondervangen door te vragen dat leerlingen hun bronnen nauwkeurig specificeren (een in het algemeen terechte eis), respectievelijk de gebruikte informatie bijvoegen. De docent kan een en ander dan zelf vergelijken. Dat kost echter extra tijd (de docent doet als het ware een stukje van het onderzoek nog eens over) en is ook niet altijd uitvoerbaar. Als de leerlingen bijvoorbeeld interviews hebben afgenomen (maar niet op band opgenomen), is alleen een protocol beschikbaar waaraan niet kan worden afgelezen hoe correct dat is. Dat laatste kan worden opgelost door een formele flattering door de betreffende respondent of informant. Het eind (van wat een docent kan willen controleren) raakt op deze manier echter wel zoek.

Aan dit probleem zitten twee te onderscheiden kanten. Ten eerste is de informatie die leerlingen presenteren beschrijvend van aard (ook meetresultaten die zijn afgelezen van een meetinstrument of uitkomsten van een analyse door een computerprogramma zijn dat). Beschrijvingen zijn in feite nooit toetsbaar (controleerbaar op juistheid). Men kan wel de onderzoeksactiviteit nogmaals uitvoeren, maar dan hebben we het niet meer over wat de leerlingen beweren de eerste keer te hebben waargenomen. Er kan alleen worden verlangd (en dat moet ook) dat de *procedures* zorgvuldig worden beschreven en verantwoord.

Ten tweede kunnen de resultaten worden beoordeeld op de mate waarin ze plausibel zijn gegeven wat we (respectievelijk de docenten) weten. Voorzover dat zo is, heeft de docent in feite geen reden om te veronderstellen dat de leerlingen fouten of vergissingen hebben begaan (tenzij de docent uit het verslag kan opmaken dat de leerlingen op een verkeerde manier te werk zijn gegaan; dan geeft juist het verkregen hebben van 'normale' resultaten te denken).

Het zal in de praktijk niet veel voorkomen dat leerlingen komen aanzetten met resultaten die afwijken van wat had mogen worden verwacht. Als dat gebeurt, kan dat worden aangegepen om samen na te gaan wat er aan de hand is (een buitenkans voor wie leerlingen in aanraking wil brengen met het ethos van het doen van onderzoek en het proces van kennisontwikkeling).

### **6.14 Vermoedens van en controle op fraude**

Leerlingen die min of meer zelfstandig een onderzoek moeten uitvoeren, daarbij een zekere keuzevrijheid hebben en als eindproduct een verslag moeten uitbrengen, zouden in de verleiding kunnen komen gebruik te maken van min of meer kant en klare teksten, van Internet of anderszins. Docenten die op dit punt niet te veel gelegenheid willen geven, zullen om te beginnen de keuzevrijheid van de leerlingen inperken. Naarmate het onderwerp, de vraagstelling en/of de aard van het uit te voeren onderzoek aan specifiekere eisen moeten voldoen, wordt de kans kleiner dat leerlingen geschikte voorbeelden zonder meer kunnen overnemen. Een tweede maatregel is te zorgen voor regelmatige tussentijdse contacten



in het kader van de begeleiding, al dan niet gekoppeld aan tussentijdse beoordeling (zie ook paragraaf 6.4). Een derde optie is te verlangen dat de leerlingen in hun verslag hun werkwijze en activiteiten beschrijven en verantwoorden (vergelijk paragraaf 6.3). Verder kent menig docent zijn leerlingen goed en kan hij op basis van het gedrag en het ingeleverde verslag een redelijke inschatting maken van de 'authenticiteit' van het geleverde werk. Een laatste en in potentie effectieve controle is mogelijk door over het onderzoek, het product en/of het onderwerp de individuele leerlingen afzonderlijk enkele vragen te stellen (vergelijk paragraaf 6.11). De eventuele fraudeur die dan nog niet door de mand valt, moet wel slim zijn (in welk geval de docent ook niet wakker hoeft te liggen over een mogelijk ten onrechte tot stand gekomen positieve beoordeling van een onderzoek). Deze laatste optie (stellen van vragen) kan ook worden gecombineerd met een presentatie van het onderzoek voor de rest van de klas. De 'lastige' vragen kunnen dan ook door andere leerlingen worden gesteld (en de presentatoren zullen zich voor het front van de klas liever niet willen blameren).

### 6.15 Straftrek en/of bonuspunten

Bij de beoordeling van prestaties zoals de opzet, uitvoering en verslaggeving van een onderzoek spelen naast de primaire beoordelingscriteria, die verband houden met de doelen die de docent vooral wil bereiken, meestal ook nog wel andere criteria een rol op grond waarvan de uiteindelijke beoordeling net even iets lager of iets hoger kan uitvallen.

*Aftrek* zal alleen plaatsvinden voor wat in principe vanzelfsprekend zou moeten zijn en waarvan het ontbreken een serieus minpunt betekent. In het kader van praktische opdrachten aan leerlingen, waarbij leerlingen een product moeten inleveren, is het gangbaar voor die inlevering een deadline te stellen. Maar dan: wat te doen als leerlingen die deadline overschrijden? Hier gaan docenten verschillend mee om, mede afhankelijk van de cultuur in de school, het aantal leerlingen dat het betreft, de mate waarin het steeds dezelfde leerlingen zijn et cetera. Een aantal docenten hanteert een strikte en zware sanctie op te laat inleveren, bijvoorbeeld in de vorm van een forse onvoldoende (ongezien de kwaliteit van het werk). Vanuit het criterium dat beoordelingen aanvaardbaar moeten zijn, kan een dergelijke sanctie alleen worden verantwoord vanuit het standpunt dat het op tijd inleveren van een verslag onderdeel uitmaakt van de kern van de vaardigheid die de leerlingen moeten ontwikkelen. Nu is het op tijd nakomen van aangegane verplichtingen in het algemeen belangrijk. Bij contracten in het bedrijfsleven staat op overschrijding van termijnen ook vaak een sanctie. Deze dient echter wel in overeenstemming te zijn met de mate en ernst van de overschrijding. Bij het niet op tijd inleveren van een onderzoeksverslag door leerlingen lijkt het hooguit te verdedigen daarvoor enige aftrek toe te passen (uiteraard als dit van tevoren bij de leerlingen al bekend was), maar niet om dan meteen maar een '1' of iets dergelijks uit te delen, laat staan zonder het werk zelf alsnog serieus te beoordelen.

Opvallend positieve prestaties kunnen worden gewaardeerd met een zekere *bonus*. Ook hierbij moeten de aanleidingen daarvoor in relatie staan tot de kern van waar het om gaat en moet de omvang ervan in verhouding blijven. Aspecten die bij opvallend presteren in aanmerking zouden kunnen komen voor een zekere bonus zijn:

- kenmerken die iets zeggen over de moeilijkheid van het onderzoek: complexiteit, genomen risico's, mate van toegankelijkheid van bronnen en dergelijke;
- het gerealiseerde niveau, de inhoudelijke diepgang, de originaliteit;
- kenmerken van het proces, zoals samenwerking, inzet en tijdsinvestering;
- de mate van zelfstandigheid, respectievelijk mate waarin geen hulp en bijsturing nodig was;

Een mogelijke vuistregel voor wat betreft het handhaven van de juiste verhouding zou kunnen zijn dat door dergelijke

bonuspunten het cijfer (op de bekende schaal van 1 tot 10) voor het onderzoek in totaal ten hoogste met 1 punt kan toenemen.

*In ons onderzoek najaar 1998 bleken docenten in hun beoordeling van onderzoeken op vrij veel aspecten te letten. De twee aspecten waar de minste docenten op letten, zijn de hoeveelheid tijd die de leerlingen erin hebben gestoken (30% let hier niet op, 10% doet het juist expliciet) en het met het onderzoek genomen risico (20% doet dit niet, 20% expliciet). Voor alle aspecten geldt dat 40-65% van de docenten er impliciet op let. Er expliciet op letten gebeurt het vaakst met de complexiteit van het onderzoek en de originaliteit (beide bij 50%). Op de tweede plaats komen de aspecten zelfstandigheid, werken volgens planning, nauwkeurigheid, werkhouding/inzet en samenwerking in de groep (alle bij rond 30%). De eerste groep (complexiteit, originaliteit) heeft primair te maken met het methodische en inhoudelijke niveau van het onderzoek, de tweede (zelfstandigheid t/m samenwerken) met algemene vaardigheden die niet specifiek zijn voor alleen onderzoek.*

### **6.16 Leerlingen als medebeoordelaars**

Als een docent werkt met een lijst specifieke beoordelingscriteria (wat in het algemeen vanuit beoordelingsoogpunt de voorkeur verdient boven een globale beoordeling) en als deze criteria reeds bij de start aan de leerlingen bekend worden gemaakt (wat vanuit het oogpunt van aanvaardbaarheid zonder meer wenselijk is), kunnen de beoordelingscriteria ook een belangrijke functie vervullen in het onderwijsleerproces. Leerlingen leren mede daardoor wat de diverse onderdelen en aspecten van een onderzoek zijn en waaraan moet worden voldaan om een onderzoek met succes op te zetten, uit te voeren en af te ronden. Met behulp van de criteria kan de reflectie van de leerlingen op het onderzoek worden gestimuleerd en gericht. Dit type gebruik van de beoordelingscriteria kan plaatsvinden na afloop van het onderzoek, als het eindproduct er ligt, maar ook rond het concept daarvan (zodat de inzichten en conclusies die de leerlingen opdoen nog in het eindproduct kunnen worden verwerkt). Het kan zelfs tussentijds op in principe alle momenten dat tijdens het onderzoek begeleiding plaatsvindt. Leerlingen kunnen dan mede aan de hand van de criteria toewerken naar een optimaal resultaat.

De beoordelingscriteria kunnen ook worden gebruikt door leerlingen aan de hand daarvan elkaars werk te laten beoordelen. Door de begeleiding zo op te zetten, dat bijvoorbeeld enkele malen tussentijds de diverse groepjes leerlingen-onderzoekers de stand van zaken in hun onderzoek kort voorleggen of presenteren aan enkele of alle medeleerlingen kunnen zij elkaars werk aan de hand van de criteria beoordelen en elkaar op grond daarvan feedback geven (onder supervisie, respectievelijk begeleiding van de docent). Docenten kunnen de tijd hiervoor vinden door deze aanpak een expliciete plaats te geven in hun programma en in plaats daarvan iets minder tijd aan instructie te besteden. Uit onderzoek blijkt dat leerlingen er veel van kunnen opsteken en dat door deze aanpak docenten ook zelf meer inzicht ontwikkelen in onderzoek en er beter met leerlingen over kunnen communiceren (vergelijk Wolf e.a., 1991; White en Frederikson, 1998). Een mogelijke aanpak is de beoordelingscriteria te verdelen over individuele leerlingen of tweetallen, zodat leerlingen zich kunnen concentreren op één aspect (de criteria kan men in de loop van de tijd uiteraard laten rouleren). Naarmate leerlingen beter leren onderzoekswerk aan de hand van de criteria te beoordelen, kan ook worden overwogen de leerlingen een rol te geven bij de door de docent vast te stellen beoordelingen. Bij de beoordeling van presentaties die zijn gericht aan een bepaalde doelgroep, zoals een mondelinge presentatie aan de medeleerlingen, ligt het zelfs zonder meer voor de hand die doelgroep te betrekken bij de beoordeling. Immers: een van de criteria zal dan zijn dat de presentatie duidelijk en informatief is en de meningen van de ontvangers zijn daarbij een belangrijke maatstaf. Ook bij de beoordeling van (schriftelijke) producten kunnen medeleerlingen een medebeoordelende rol vervullen. Ten aanzien van het eindverslag van een onderzoek kan, voorzover dat eveneens moet worden beoordeeld op

de communicatieve waarde, dezelfde redenering gelden als hiervoor gegeven voor mondelinge presentaties. De docent zou het oordeel van medeleerlingen, uiteraard mits dat op een eenduidige, inzichtelijke manier gebeurt, kunnen meewegen bij de totstandkoming van de eindbeoordeling.

Docenten die werken met tussentijdse deelbeoordelingen kunnen overwegen om, als zij die momenten benutten om leerlingen elkaars werk te laten becommentariëren, de beoordelingen van de leerlingen te gebruiken als de tussentijdse beoordelingen en de eigen beoordeling te concentreren op het eindproduct.

*In ons onderzoek najaar 1998 bleek dat driekwart van de docenten (vrijwel) nooit andere leerlingen inschakelt als medebeoordelaars; een kwart doet dit soms (op de vijfpuntsschaal van (vrijwel) nooit tot (vrijwel) altijd). Deze potentieel nuttige activiteit komt dus nog weinig voor.*

### 6.17 Zelfbeoordeling door leerlingen

Veel van wat hiervoor is gezegd over medebeoordeling door medeleerlingen geldt ook voor zelfbeoordeling ofwel beoordeling van het eigen werk. Ook dat kan de reflectie en daarmee het leren bevorderen. Daarbij moet wel worden bedacht dat het systematisch, objectief en kritisch reflecteren op het eigen werk voor veel leerlingen waarschijnlijk moeilijker zal zijn dan het beoordelen van producten van andere leerlingen. De extra complicatie is immers dat enige afstand moet worden genomen van wat men zelf heeft gedaan. Men kan leerlingen in hun eindproduct ook een zelfevaluatie of zelfbeoordeling laten opnemen. Dit kan tenminste de volgende drie bedoelingen hebben:

- leerlingen evalueren zelf hun eigen onderzoek, in de zin dat ze onderkennen wat de sterke en zwakke punten zijn, wat de waarde is van de conclusies, hoe het beter zou kunnen, respectievelijk had gekund;
- leerlingen reflecteren op hun eigen leerproces, onderkennen wat ze zelf hebben geleerd en aan welke punten ze nog moeten werken;
- leerlingen geven een beoordeling van hun eigen werk, die eventueel kan worden vergeleken met de beoordelingen van medeleerlingen en van de docent en eventueel kan worden betrokken bij de eindbeoordeling.

Hier gaat het om het laatste: leerlingen krijgen invloed op de beoordeling van hun eigen werk. Dat lijkt in strijd met alle regels en iets wat een docent vooral niet moet doen. Als de docent echter verlangt dat leerlingen hun oordelen beargumenteren, krijgen leerlingen die invloed niet voor niets. Het moeten beargumenteren van de zelfbeoordeling kan op zich weer het gewenste leerproces bevorderen en het kan leerlingen stimuleren te proberen aan deze voorwaarde te voldoen. Leerlingen, respectievelijk groepjes leerlingen die in staat zijn om hun eigen werk op gefundeerde wijze te beoordelen leveren ook een prestatie die op zich gewaardeerd mag worden, juist omdat het gaat om een prestatie die direct de kern raakt van waar het om gaat, namelijk leren goed onderzoek te doen. Dus het meetellen van zo'n zelfbeoordeling is dan aanvaardbaar. En als de argumenten steekhoudend zijn, heeft de docent ook geen gronden de betreffende oordelen *niet* over te nemen. Docenten die hun leerlingen leren hun eigen werk te beoordelen, slaan daarmee twee vliegen in één klap: zij bereiken de doelen waar het eigenlijk om ging en kunnen een deel van het beoordelingswerk aan de leerlingen zelf overlaten.

### 6.18 Onvoldoendes en herkansingen

Het is de vraag in hoeverre bij de uitvoering van een praktische opdracht, respectievelijk een onderzoek onvoldoendes

zullen voorkomen. Een docent kan op het standpunt staan dat hij elk groepje 'doorbegeleid' tot een voldoende. Bij het profielwerkstuk in de tweede fase, waarbij een tenminste voldoende resultaat een voorwaarde is voor deelname aan het centraal examen, kan een in eerste instantie onvoldoende beoordeling ook alleen op die manier worden behandeld: doorgaan tot een voldoende. Bij praktische opdrachten kan een docent onderscheid maken tussen opdrachten 'voor een cijfer dat meetelt voor het schoolexamen' en opdrachten waarbij dat niet het geval is (maar bijvoorbeeld wel voor het eerstvolgende rapport). Opdrachten die meetellen voor het examen zouden eveneens kunnen worden doorbegeleid naar een voldoende. Het afronden van een onderzoek met een onvoldoende kan slechts voorkomen als de docent de leerlingen in hoge mate zelfstandig aan het werk laat. Zodra begeleiding wordt gegeven, zeker als daarbij 'go/no-go'-beslissingen vallen, zal het weinig voorkomen dat het onderzoek tot een onvoldoende leidt. Voorzover een onderzoek toch als onvoldoende moet worden beoordeeld, kan men redeneren dat er genoeg gelegenheid is deze onvoldoende te compenseren door bij volgende onderzoeken een beter resultaat te behalen. Juist bij een onderzoek zijn er mogelijkheden passende vervolgoopdrachten te geven, op grond waarvan het totaal toch nog met een voldoende beoordeling kan worden afgerond. Zulke vervolgoopdrachten kunnen verschillende vormen aannemen, bijvoorbeeld:

- een bepaalde deelvraag nader laten onderzoeken;
- aanvullende gegevens laten verzamelen, verwerken en bij de conclusie betrekken;
- op basis van de bij de leerlingen beschikbare gegevens hen de verwerking laten overdoen;
- het eigen onderzoek zorgvuldig en beargumenteerd op papier laten beoordelen aan de hand van de lijst aandachtspunten en beoordelingscriteria;
- aan de leerlingen te vragen een nieuw onderzoeksplan op te stellen waarmee hun onderzoeksvraag wel goed zou zijn te beantwoorden.

Ook hierbij gelden weer de voorwaarden dat de docent de beoordeling daarvan systematisch aanpakt en dat de leerlingen er vantevoren van op de hoogte zijn (zodat ze hiertegen geen bezwaar kunnen maken).

## *Deel IV Beoordelen als professionele taak van docenten*

### **7 LEREN BEOORDELEN ALS ONTWIKKELPROCES (bij 2.7 en 2.8)**

#### **7.1 Beoordelen als onderdeel van de taak van een docent**

In de VS, waar de ontwikkelingen die de aanleiding vormen voor deze brochure al iets langer gaande zijn, constateerde men in de 80-er jaren dat veel leraren moeite hebben met het kiezen, ontwikkelen en gebruiken van beoordelingsinstrumenten en -procedures (zie bijvoorbeeld Stiggins en Bridgeford, 1985). De problemen waren voor een deel organisatorisch van aard (het vinden van voldoende tijd, het realiseren van samenwerking met collega's) en betroffen voor een ander deel het correct vaststellen en interpreteren van resultaten. Stiggins en Bridgeford kwamen tot de schatting dat in het voortgezet onderwijs driekwart van de leraren serieuze zorgen had op dit terrein, dat de beoordelingspraktijk van een substantieel deel van de leraren grote risico's met zich meebracht voor de kwaliteit van de beoordelingen en de deskundigheid van leraren op dit vlak diende te worden vergroot.

Bijna tien jaar later signaleerden Seyfert, Simon and Schlesinger (1994) dat leerlingen en ouders problemen kunnen hebben met 'zachte' metingen en beoordelingen van 'zachte' vaardigheden, voor een deel uit zorg over de vraag wat leerlingen nog leren en of de traditionele kennis en basisvaardigheden niet teveel worden verwaarloosd dan wel normen worden verlaagd. Ook kunnen ouders wantrouwen koesteren over de eerlijkheid en objectiviteit van beoordelingen als geen gebruik wordt gemaakt van gestandaardiseerde toetsen en leerlingen zelfs onderling mogen samenwerken.

Hoewel we deze laatste geluiden moeten zien in hun Amerikaanse context, waar het onderwijs veel meer en langer dan in Nederland een cultuur van frequente gestandaardiseerde toetsing heeft gekend, zijn de genoemde vragen van leerlingen en ouders bij de beoordelingen die leraren uitbrengen over de resultaten van praktisch werk ook in ons land actueel. In de media worden discussies gevoerd over de zin en onzin van algemene vaardigheden en de eerste juridische processen over verondersteld oneerlijke beoordelingen van praktische opdrachten (met als uiteindelijk effect zelfs het zakken voor het examen) zijn reeds gevoerd. Docenten zullen dus aan het beoordelen voldoende aandacht moeten schenken.

#### **7.2 Gedragscodes voor toetsen, beoordelen en beslissen**

In de VS hebben de hiervoor beschreven ontwikkelingen geleid tot het opstellen van 'Standards for Teacher Competence in Educational Assessment of Students' door de American Federation of Teachers e.a., 1990. Een samenvatting van deze Standards is opgenomen in figuur 5. In Nederland is in 1997 een Gedragscode ontwikkeld voor het toetsen, beoordelen en beslissen in het vo, samengevat in figuur 4. De kwaliteitsstandaarden in deze Gedragscode zijn ontwikkeld met behulp van aanwijzingen en commentaren van leraren en (andere) deskundigen en heeft tot doel dat toetsen, beoordelingen en beslissingen zorgvuldig tot stand komen. Daarmee wordt bedoeld dat een beoordelingsprocedure aanvaardbaar moet zijn voor de betrokkenen en dus doelmatig en billijk. Daartoe moet worden voldaan aan eisen als objectiviteit, transparantie, validiteit en betrouwbaarheid. De kwaliteitsstandaarden richten zich op zogenoemde

niveaubepalende toetsen (meetellend bij het geven van cijfers) van kennis en vaardigheden en zijn van toepassing op alle vakken. De samenstellers realiseren zich dat de autonomie van de leraar steeds meer afneemt, maar de gedragscode richt zich desalniettemin tot de individuele leraar. "Deze blijft (...) zelf verantwoordelijk voor zijn toetsen, beoordelingen en beslissingen; hij kan op deze verantwoordelijkheid ook worden aangesproken door de schoolleiding, leerlingen en hun ouders, en (ook) behoort het tot zijn verantwoordelijkheid dat hij de taken gewoonlijk in overleg met de collega's uitvoert." (Creemers-van Wees, Knuver, Vos, en Van der Linden, 1997, p. 81).

Creemers-van Wees, Knuver, Vos, en Van der Linden (1997) ontwikkelden een gedragscode voor docenten in het voortgezet onderwijs ten aanzien van toetsen, beoordelen en beslissen. Creemers en Knuver (1998) publiceerden daarvan een samenvatting (figuur 4). Onderstaande samenvatting is op beide publicaties gebaseerd.

<b>Gedragscode voor toetsen, beoordelen en beslissen (OCTO)</b>	<b>(samenvatting)</b>
<i>Algemeen</i>	
1a	De leraar neemt de regels van de school en vaksectie in acht, voor zover deze niet in strijd zijn met de wet of deze gedragscode. Een slechte of onzinnige regel moet door een leraar worden aangekaart bij de schoolleiding.
1b	De leraar maakt tijdig aan zijn leerlingen bekend welke werkwijze hij hanteert bij het toetsen, beoordelen en beslissen. Hiermee wordt het beoordelingsproces doorzichtig gemaakt.
1c	De leraar zorgt ervoor dat bij het beoordelen van toetsresultaten gelijke gevallen gelijk worden behandeld. Persoonlijke voorkeur van leraren mag nooit doorklinken in toetsresultaten. De houding of motivatie van de leerlingen mag niet meetellen.
1d	De leraar overlegt waar nodig over toetsen, beoordelingen en beslissingen over leerlingprestaties met collega's of andere leden van de schoolorganisatie.
<i>Inhoud en vorm van toetsen</i>	
2a	De leraar is zorgvuldig in de afstemming van de gekozen of zelf ontwikkelde toets op de inhoud en het niveau van de leerstof.
2b	De leraar kiest een adequate toetsvorm, bijvoorbeeld een proefwerk of een werkstuk, afgestemd op de soort kennis en/of vaardigheden die moeten worden getoetst.
<i>Instructie en afname van toetsen</i>	
3a	De leraar zorgt voor een zorgvuldige instructie voorafgaand aan de toetsafname. De leerlingen moeten tijdig en duidelijk op de hoogte worden gebracht van de aard van de toets, de kennis en/of vaardigheden die worden getoetst en de beoordelingsprocedure.
3b	De leraar zorgt voor een zorgvuldige afname van de toets, gericht op het verkrijgen van betrouwbare en valide gegevens. Dit betekent onder meer dat leerlingen voldoende tijd moeten krijgen en dat indien leerlingen aanwijzingen krijgen deze gelijk zijn voor alle leerlingen.
3c	De leraar voorkomt oneigenlijk gebruik van toetsen, dat wil zeggen gebruik dat niet is gericht op het beoordelen van kennis en vaardigheden van leerlingen.
<i>Het scoren van de antwoorden</i>	
4a	Bij het gebruik van gestandaardiseerde toetsen zorgt de leraar voor het correct toepassen van de voorgeschreven scoringsregels en voor correcte interpretatie van de resultaten.

4b	Voor zelfgemaakte toetsen stelt de leraar zorgvuldige scoringsaanwijzingen op, zodanig, dat deze ook kunnen worden gebruikt door andere beoordelaars. In twijfelgevallen wordt de toets zonedig door twee beoordelaars beoordeeld.
<i>Het bepalen van het cijfer en de cesuur</i>	
5a	De leraar zet de verkregen score zorgvuldig om in een cijfer. De leraar is zich bewust van de consequenties daarvan voor de leerlingen. Wanneer leerlingen zelf een bijdrage leveren aan de beoordeling, ziet de leraar erop toe dat deze bijdrage zorgvuldig tot stand komt.
5b	Indien leerlingen en/of ouders daarom vragen, maakt de leraar de procedure voor het omzetten van scores in cijfers bekend.
5c	De leraar maakt een eventuele bijstelling van een vooraf vastgestelde norm aan de leerlingen bekend.
5d	De leraar voorkomt dat er leerlingen door normbijstelling worden benadeeld.
<i>De rapportage van de toetsresultaten</i>	
6a	De leraar moet binnen een redelijke termijn na de toetsafname de beoordeling vaststellen en bekend maken.
6b	De leraar zorgt ervoor dat de toetsresultaten aan de leerlingen inhoudelijk worden toegelicht, door informatie over de beoordeling van onderdelen en aanwijzingen over wat wel en niet goed ging. De leraar biedt leerlingen de mogelijkheid tot nabespreking.
6c	De leraar biedt de leerlingen de mogelijkheid hun werk en de beoordeling ervan in te zien.
6d	De leraar is bereid om de beoordeling te bespreken, als een leerling of collega of de directie dat wenst.
6e	De leraar voorkomt oneigenlijk gebruik van toetsresultaten, waaronder het op grievende wijze meedelen van resultaten aan leerlingen en het zonder noodzaak en toestemming verstrekken van resultaten aan derden.
<i>Het nemen van beslissingen op grond van toetsresultaten</i>	
7a	De leraar ziet erop toe dat beslissingen over individuele leerlingen zorgvuldig tot stand komen. In rapportenvergaderingen brengt de leraar zijn kennis en standpunt naar voren.

Figuur 4 Gedragscode voor toetsen, beoordelen en beslissen

Enkele artikelen van de Gedragscode verdienen in deze brochure speciale aandacht:

- *Leraren stemmen de toetsvorm af op de te meten soort kennis en/of vaardigheden (art. 2b).* Dit veronderstelt dat leraren expliciet maken welke kennis en/of vaardigheden zij willen toetsen, respectievelijk beoordelen en weten welke toetsvorm daarvoor geschikt is.
- *Leraren delen leerlingen tijdig en duidelijk mee: de aard van de toets, de kennis en/of vaardigheden die worden getoetst en de beoordelingsprocedure (art. 3a).* Dit vergt dat de leraar inhoud en procedures rond afname en beoordeling expliciteert en naar de leerlingen openheid van zaken geeft.
- *Leraren werken met behulp van zorgvuldige scoringsaanwijzingen, die zonedig ook door andere beoordelaars kunnen worden gebruikt; leerlingen kunnen ook een bijdrage leveren aan de beoordeling, mits de leraar erop toeziet dat dit zorgvuldig gebeurt (art. 4b, 5a).* Een en ander vraagt van de leraar dat deze ook de scoring volledig inzichtelijk maakt voor zichzelf en voor anderen.
- *Leraren zorgen voor een inhoudelijke toelichting op de resultaten aan de leerlingen, door informatie over de beoordeling van onderdelen en aanwijzingen over wat wel en niet goed ging (art. 6b).* Dit betekent dat leraren niet volstaan met het meedelen van één totaaloordeel, maar dit specificeren.
- *Leraren zorgen voor een zorgvuldige afname van de toets, gericht op het verkrijgen van betrouwbare en valide gegevens. Dit betekent onder meer dat leerlingen voldoende tijd krijgen en dat eventuele aanwijzingen voor alle*

*leerlingen gelijk zijn (art. 3b).*

Het eerste deel van dit artikel houdt in dat ongeacht het doel van de toets of de beoordeling (het gebruik of de functie) sprake moet zijn van een betrouwbare (nauwkeurige) en valide (accurate) meting. Het tweede deel verdient ons inziens discussie. Dat leerlingen voldoende tijd krijgen, lijkt in het algemeen correct, maar leerlingen moeten ook leren werk op tijd af te hebben, respectievelijk te werken in een situatie waarin voor een bepaalde activiteit slechts een bepaalde tijd beschikbaar is. En dat aanwijzingen voor alle leerlingen gelijk zouden moeten zijn, lijkt slecht te passen bij het op maat begeleiden van (groepjes) leerlingen, hetgeen voor het leerproces juist wenselijk kan zijn. En voor het creëren van gelijke kansen kan juist differentiatie nodig zijn.

- *Leraren overleggen waar nodig over toetsen, beoordelingen en beslissingen met collega's en anderen binnen de school (art. 1d).* Dit artikel spreekt voor zich.
- *Leraren houden zich aan de regels van de school en vaksectie, behalve als deze in strijd zijn met de wet of deze gedragscode; slechte of onzinnige regels worden aangekaart bij de schoolleiding (art. 1a).* Dit artikel vraagt van leraren een kritische en actieve opstelling en onderstreept hun (mede-)verantwoordelijkheid voor de totstandkoming en handhaving van zorgvuldige beoordelingen en beslissingen.

#### **Standards for Teacher Competence in Educational Assessment of Students**

*1 Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.*

*2 Teachers should be skilled in developing assessment methods appropriate for instructional decisions.*

Deze twee standaarden beschrijven dat leraren adequate keuzes moeten maken uit een breed scala van beoordelingsmethoden, zowel als zij uit reeds beschikbare methoden kiezen als in het geval zij zelf een methode ontwikkelen. Zij moeten daarbij rekening houden met de gewenste functie(s) van de beoordeling (in relatie tot het onderwijs en de leerlingen), de technische kwaliteit van de methode (in termen van betrouwbaarheid, validiteit en eerlijkheid) en de mogelijk ongewenste consequenties van verkeerde keuzes. Van zelf ontwikkelde methoden moeten zij de kwaliteit kunnen vaststellen.

*3 The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.*

Deze standaard stelt dat leraren uiteenlopende toetsen en andere beoordelingsmethoden op een juiste wijze moeten kunnen afnemen en scoren en de uitkomsten interpreteren. Ten aanzien van het afnemen is onder meer van belang dat dit zodanig gebeurt, dat leerlingen worden bemoedigd en niet ontmoedigd. Ten aanzien van het scoren moeten leraren kunnen omgaan met diverse typen scores en berekeningen daarmee. Ten aanzien van het interpreteren moeten leraren in staat zijn uit de resultaten af te leiden wat de sterke en zwakke punten van de leerlingen zijn en om eventuele onduidelijkheden op te lossen alvorens op de uitkomsten beslissingen te baseren.

*4 Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.*

Deze standaard stelt dat leraren op basis van resultaten van beoordelingen juiste beslissingen moeten kunnen nemen over de gewenste voortgang van de leerlingen en de inhoud en uitvoering van het onderwijs. Ook dienen zij op de hoogte te zijn van onderzoeken en evaluaties op landelijk, regionaal en plaatselijk niveau en adequaat gebruik daarvan bij verbetering van het onderwijs in de klas/school en in het plaatselijke en landelijke beleid.

*5 Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.*

Deze standaard geeft aan dat leraren de juiste aanpakken moeten kennen en toepassen bij het normeren en geven van cijfers, bij de verschillende typen beoordelingsmethoden en bij het baseren van samenvattende (rapport)cijfers op de resultaten van verschillende beoordelingen. Leraren moeten de cijfers die zij geven kunnen verantwoorden en verdedigen. Zij moeten onjuiste



<p>werkwijzen, zoals het gebruiken van cijfers om te straffen, herkennen/vermijden en hun eigen werkwijzen kunnen evalueren/verbeteren.</p> <p><i>6 Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.</i></p> <p>Deze standaard stelt dat leraren beoordelingsresultaten effectief moeten kunnen rapporteren aan en bespreken met leerlingen, ouders, collega's en anderen. Zij moeten de betekenis, beperkingen en implicaties van beoordelingen kunnen aangeven, uitleggen en verantwoorden (inclusief de betekenis voor de verdere ontwikkelingsmogelijkheden van de leerlingen) en verkeerde interpretaties en verkeerd gebruik voorkomen.</p> <p><i>7 Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.</i></p> <p>Deze standaard beschrijft dat leraren al hun activiteiten met betrekking tot beoordelingen moeten baseren op zorgvuldig en eerlijk professioneel handelen (waaronder het vermijden van discriminatie, in verlegenheid brengen van leerlingen en schenden van privacy). Leraren moeten, als zij onverantwoorde beoordelingspraktijken signaleren, deze trachten te beëindigen. Ook moeten leraren eventueel niet adequaat gebruik van leerlingresultaten voor de bepaling van de effectiviteit van leraren kunnen onderkennen.</p>
---

Figuur 5 Standards for Teacher Competence in Educational Assessment of Students

De zeven standaarden (figuur 5) zijn opgesteld door de American Federation of Teachers, de National Council on Measurement in Education en de National Education Association in 1990 met betrekking tot de bekwaamheid van leraren in het beoordelen van prestaties van leerlingen. Deze standaarden zijn bedoeld een rol te gaan spelen in de opleiding en nascholing van leraren en de toekomstige certificering van leraren. Ter vergelijking met de in Nederland ontwikkelde beroepscode (zie figuur 4) hebben wij deze Standards hier samengevat.

### 7.3 Planmatig ontwikkelen van professioneel beoordelen

*Gevraagd welke knelpunten men ervaart, rept 80% van de docenten in ons onderzoek najaar 1998 over tijdgebrek voor het begeleiden, voor het beoordelen en 70% over gebrek aan moderne media op school. Daarnaast heeft 50% gebrek aan geschikte onderzoeksopdrachten, signaleert rond 40% bij zichzelf tekortschietende vaardigheden en noemt 30% gebrek aan motivatie bij leerlingen. Doorgevraagd in welke mate men zichzelf voldoende voelt toegerust, blijkt dat 60% van de docenten zich niet of slechts in geringe mate voelt toegerust voor de beoordeling van onderzoekswerk van leerlingen.*

#### ***Inleiding***

Beoordelen van werk van leerlingen, anders dan in de vorm van de traditionele kennistoetsen en proefwerken, is voor veel docenten relatief nieuw. Gezien de ontwikkelingen, ook internationaal, valt te verwachten dat docenten prestaties van leerlingen in de toekomst geleidelijk steeds meer langs deze weg zullen moeten waarderen. De accenten in kerndoelen en eindtermen op vaardigheden (waarvoor andere manieren van toetsing nodig zijn) en de richtlijnen voor het nieuwe schoolexamen zijn waarschijnlijk niet van voorbijgaande aard. De professionaliteit van docenten zal zich mede in deze richting moeten ontwikkelen. Om onderzoeken van leerlingen adequaat te kunnen begeleiden en beoordelen, moeten docenten veel weten van hun eigen vak, kennis hebben over onderzoek en kennis hebben over beoordelen, terwijl zij hun doelen, opdrachten, begeleiding en beoordeling ook op elkaar moeten afstemmen binnen en tussen vaksecties. Het ontwikkelen van meer deskundigheid en bekwaamheid met betrekking tot beoordelen kan en hoeft niet van de ene dag op de andere. Docenten kunnen hier systematisch aan werken en van jaar tot jaar verbeteringen aanbrengen, en ze hoeven het bovendien niet alleen te doen: juist bij een onderwerp als dit kunnen ze

vruchtbaar met elkaar samenwerken. In deze brochure wordt ervan uitgegaan dat docenten werken met expliciete opdrachten, criteria en beoordelingen. Dat is in feite een basisvoorwaarde om op een aanvaardbare manier te kunnen beoordelen. Het is dan weinig ingrijpend als de docent een en ander systematisch bewaart. Daarnaast is het aan te bevelen op de een of andere manier vast te leggen welke begeleiding de leerlingen krijgen en wat daarbij de ervaringen zijn.

Aldus gewapend kan een docent vervolgens de aan een groep leerlingen gegeven opdrachten, begeleiding en beoordelingen met elkaar gaan vergelijken. Wie eenmaal zover is, kan een volgende stap zetten en verschillende klassen en/of leerjaren met elkaar gaan vergelijken. Het ontwikkelen van een goede opzet en hulpmiddelen en het opdoen van ervaring met het beoordelen van onderzoeken van leerlingen is een proces dat vier à vijf jaar in beslag kan nemen. In zo'n periode kunnen zich in de praktijk ook de nodige normen ontwikkelen ten aanzien van wat van leerlingen kan worden verwacht en wat van docenten. We werken het voorgaande hieronder verder uit.

#### ***Bewaren van opdrachten, criteria en beoordelingen***

Een eerste maatregel die een docent kan nemen om eraan te werken dat hij meer deskundigheid en bekwaamheid in het beoordelen van werk van leerlingen ontwikkelt, is het systematisch bewaren van de gegeven opdrachten, de beoordelingscriteria en de gegeven beoordelingen. Door daarbij ook (kopieën van) de producten van de leerlingen te bewaren - legt de docent in feite vast hoe coulant of streng hij heeft beoordeeld en welke criteria hij lichter of zwaarder heeft laten wegen. Aan de hand hiervan kan de docent desgewenst besluiten de criteria en/of de normen aan te passen. Het kan bijvoorbeeld blijken dat een bepaald criterium vaak niet van toepassing is of niet kan worden beoordeeld of dat bepaalde normen zelden door de leerlingen blijkt te worden gehaald of juist altijd (zodat het niet informatief is). De combinatie van criteria, producten en beoordelingen geeft ook houvast voor de volgende keer dat weer onderzoeken moeten worden beoordeeld. Dit kan voorkomen dat men ongemerkt en ongewenst lichter of zwaarder gaat beoordelen of andere accenten gaat leggen.

#### ***Vastleggen van begeleiding en ervaringen***

Docenten slaan uiteraard veel gebeurtenissen en ervaringen in hun geheugen op. Na enige tijd zullen ze zich echter een aantal zaken niet meer zo precies kunnen herinneren. Wie systematisch de eigen deskundigheid en bekwaamheid wil ontwikkelen, kan er baat hebben op gezette tijden dingen kort op te schrijven. Daarvoor zou eventueel een standaardlijstje met aandachtspunten kunnen worden gebruikt. Zaken die de moeite van het vastleggen waard kunnen zijn:

- vragen en problemen waar leerlingen mee komen;
- soorten aanwijzingen die men als docent geeft;
- hoe de leerlingen daarop reageren en wat ze ermee doen;
- indrukken die men als docent opdoet, bijvoorbeeld over de impact van een bepaalde opdracht, van een bepaalde samenstelling van een groepje leerlingen, van een bepaalde manier van begeleiden op het werk van de leerlingen;
- de hoeveelheid tijd die met bepaalde activiteiten is gemoeid (nettotijd respectievelijk doorlooptijd) bij de leerlingen en bij zichzelf;
- ideetjes die de docent al werkend krijgt 'voor een volgende keer';
- de mate van aansluiting van opdrachten en onderzoeken bij de methode;
- punten waarover een keer moet worden gesproken met collega's.

#### ***Vergelijken van opdrachten, begeleiding en beoordelingen***

Op het moment dat een docent beschikt over de beoordelingen van verschillende onderzoeken van dezelfde leerlingen, uitgevoerd aan de hand van verschillende opdrachten, kan hij enkele nadere analyses uitvoeren. Enkele op voorhand relevante vragen zijn:

- in hoeverre leiden verschillende opdrachten (meer of minder open, onderwerp X of onderwerp Y) tot duidelijk verschillende rangordes in de beoordelingen van de leerlingen;
- in hoeverre hangen de beoordelingen op verschillende momenten bij tussentijdse beoordelingen met elkaar samen;
- in hoeverre hangen de beoordelingen op verschillende criteria bij gespecificeerde beoordelingen met elkaar samen;
- in hoeverre hangen kenmerken van de gegeven begeleiding samen met de gegeven beoordelingen?

Wie in een cursusjaar aan dezelfde klas van bijvoorbeeld 24 leerlingen driemaal een steeds andere onderzoeksopdracht geeft, waarbij gewerkt wordt in groepjes van bijvoorbeeld gemiddeld vier leerlingen, heeft gegevens over achttien onderzoeken. Bij een dergelijk aantal kunnen analyses zoals hierboven genoemd al zinvol zijn. Wie dezelfde leerlingen ook het daaropvolgende cursusjaar heeft en dan nog eens drie onderzoeksopdrachten geeft, kan aan het dubbele aantal komen.

Per onderzoek zullen meerdere deelbeoordelingen beschikbaar zijn, bijvoorbeeld van twee tussentijdse beoordelingen (elk op bijvoorbeeld drie criteria) en een eindbeoordeling (op bijvoorbeeld zes criteria), dus in totaal bijvoorbeeld twaalf deeloordelen. Dergelijke aantallen gegevens passen nog op één A4-tje en kunnen eventueel nog met de hand worden verwerkt door bijvoorbeeld met een rekenmachientje gemiddeldes te bepalen. Een stapje professioneler zou zijn dat de gegevens even worden ingevoerd en voor de analyse een of ander computerprogramma wordt gebruikt. Het opzetten van een dergelijke manier van analyseren van gegevens die docenten zelf verzamelen vergt enige investering (vooral in tijd). Door dit bijvoorbeeld als vaksectie aan te pakken (of als vakkencluster binnen hetzelfde profiel) kan de belasting worden gespreid. Eenmaal opgezet zal deze investering zichzelf kunnen terugverdienen. Uit dit soort analyses kunnen docenten in principe veel lering trekken. Zij kunnen daardoor deskundigheid ontwikkelen, die niet via nascholing of vakliteratuur kan worden verkregen.

#### ***Vergelijken van klassen en leerjaren***

De voorgaande redenering kan worden doorgezet ten aanzien van het vergelijken van klassen en leerjaren. Het vergelijken van klassen in de zin van opeenvolgende cohorten kan inzicht geven in de mogelijke effecten van wat de leerlingen tijdens de basisvorming hebben geleerd, met name als het onderwijs in de basisvorming op de school in de loop van de jaren verandert. Het vergelijken van leerjaren bij dezelfde groep leerlingen kan inzicht opleveren in de mate waarin kennis en vaardigheden met betrekking tot onderzoek zich daadwerkelijk cumulatief ontwikkelen. Dit laatste moet immers nog blijken.

### **7.4 Samenwerking met collega's en schoolbeleid**

#### ***Intercollegiale afspraken en onderlinge toetsing (ook opgenomen in 4.3)***

Docenten kunnen rond onderzoeken van leerlingen samenwerken binnen de eigen vaksectie en tussen vaksecties (binnen een vakkencluster of profiel). Samenwerking kan onder meer betreffen:

- afspreken welke doelen centraal staan;
- bepalen aan welke (onderzoeks)vaardigheden vooral wordt gewerkt;
- afspreken welke opbouw en didactische aanpak wordt gevolgd;

- verdelen van taken bij het kiezen of ontwikkelen van geschikte opdrachten;
- afspreken op welke manier het werk wordt beoordeeld.

Bij de beoordeling gaat het op z'n minst om de ontwikkeling en vaststelling van de criteria, respectievelijk het scoringsmodel. Tijdens de ontwikkeling van een beoordelingsaanpak kunnen docenten de werkbaarheid en betrouwbaarheid ervan nagaan door op bescheiden schaal met twee of meer docenten dezelfde producten van leerlingen te beoordelen, de beoordelingen te vergelijken en de ervaringen en (ongetwijfeld aanwezige) verschillen te bespreken.

Als eenmaal een aanpak is afgesproken, kunnen docenten elkaar blijven ondersteunen. Op gezette tijden kunnen zij steekproeven van onderzoekswerk van elkaars leerlingen mede beoordelen om na te gaan in hoeverre de beoordelingen nog vergelijkbaar zijn, de aangehouden normering stabiel is, de afgesproken criteria nog toepasbaar en bevredigend zijn en dergelijke.

Niet alleen per sectie, maar ook tussen vaksecties zijn afspraken nodig. Voorkomen moet worden dat leerlingen ongecoördineerd in dezelfde periode voor verschillende vakken onderzoeken moeten uitvoeren. Daarnaast is het wenselijk dat de docenten op een vergelijkbare manier met onderzoek omgaan, zodat leerlingen niet bij het ene vak iets leren wat ze bij een het andere weer moeten afleren. En voor het profielwerkstuk zal men tussen vakken tot overeenstemming moeten komen over de wijze van beoordelen.

Uiteindelijk moeten beslissingen van docenten en secties passen in, respectievelijk gedragen worden door een beleid dat de school als geheel voert.

#### ***Beoordeling door meer beoordelaars (docenten)***

Eén van de eisen voor een verantwoorde beoordeling is dat deze betrouwbaar is, dat wil zeggen dat als dezelfde docent enkele weken later de producten nogmaals beoordeelt of als een collega de producten ook beoordeelt de uitkomsten niet teveel mogen verschillen. Docenten die een (nieuw) beoordelingsmodel willen gaan gebruiken, doen er dan ook goed aan om de betrouwbaarheid te controleren. Menig docent zal zich ook enige tijd later nog wel min of meer herinneren wat hij van een product vond, waardoor de controle door herhaalde beoordeling door dezelfde docent uiteraard minder waard wordt. Het verdient dan ook aanbeveling bij voorkeur de tweede manier te gebruiken, waarbij docenten producten van elkaars leerlingen beoordelen. De hoeveelheid werk die daaraan vastzit kan beperkt blijven door dat slechts te doen voor een steekproef, bijvoorbeeld enkele in de ogen van de eerste docent goede, enkele middelmatige en enkele zwakke producten. Het spreekt voor zich dat de tweede beoordelaar in eerste instantie alleen het product onder ogen krijgt met daarbij de gegeven opdracht en de te hanteren criteria en zondig korte informatie over het betreffende groepje leerlingen en de gegeven tussentijdse begeleiding. Pas na afloop van de tweede beoordeling worden beide beoordelingen met elkaar vergeleken. Een dergelijke dubbele beoordeling is in eerste instantie heel waardevol tijdens de ontwikkeling van een (nieuw) beoordelingsmodel. Door naar aanleiding van beide beoordelingen de afwijkingen met elkaar door te nemen kan het beoordelingsmodel worden aangescherpt en van onduidelijkheden ontdaan. Daarna hoeft niet bij elke beoordeling een dubbele beoordeling te worden toegepast. Zolang men blijft werken met hetzelfde beoordelingsmodel volstaat het om op gezette tijden elkaars beoordelingen weer even te dubbelen ter controle of men nog op dezelfde eenduidige wijze beoordeelt, of dat wellicht enige verschuiving of verwatering is opgetreden. De mate waarin de betrouwbaarheid een probleem kan vormen, zal des te kleiner zijn naarmate het beoordelingsmodel meer eenduidig en concreet is gespecificeerd.

*In ons onderzoek najaar 1998 gaf 10% van de docenten aan (vrijwel) altijd collega's in te schakelen als medebeoordelaar;*

*ruim 10% doet dit meestal, ruim 15% regelmatig, rond 50% soms en ongeveer 15% (vrijwel) nooit.*

***Vergelijken van docenten en vakken; docenten als onderzoekers***

In paragraaf 7.3 is aandacht besteed aan mogelijkheden om als docenten meer professionaliteit in het beoordelen te ontwikkelen door af en toe zelf gegevens, die al werkend toch al beschikbaar komen, nader te analyseren.

De meest vergaande vorm van gezamenlijke analyse doet zich voor als binnen een vaksectie meerdere docenten hierbij samenwerken (zodat ook de eventuele verschillen tussen docenten zichtbaar kunnen worden), respectievelijk als dit binnen het grotere geheel van een profiel gebeurt (zodat ook verschillen tussen vakken kunnen worden nagegaan). Uiteraard zullen bij zulke vergelijkingen niet alleen de gegeven beoordelingen worden vergeleken, maar ook de doelen, typen opdrachten, soorten onderzoek, begeleidingsaanpakken en de aanpakken en criteria bij de beoordeling.

Naarmate doelen, aanpak en criteria bij de beoordeling duidelijker tussen docenten zijn afgesproken en meer met elkaar overeenkomen, kunnen vergelijkingen die op opdrachten, soorten onderzoek en begeleidingsaanpakken betrekking hebben beter worden gemaakt.

Docenten zijn, deze lijn volgend, in feite zelf onderzoekers. Dat kan op zich al leerwinst opleveren, doordat zij net als de leerlingen te maken krijgen met het formuleren van een vraagstelling, het verwerken van gegevens en het trekken van conclusies.

Het op deze manier aanpakken van de gewenste professionalisering kan een effectieve manier zijn om meer greep te krijgen op de begeleiding en de beoordeling van onderzoeken van leerlingen en als vaksectie of vakkencluster hierop samen te werken en beleid te ontwikkelen.



## 9 ONDERZOEK ALS KENNISDOMEIN OP ZICH (bij 3)

In lijn met de huidige trend waarin het belang van vaardigheden naar voren wordt geschoven, zijn in de kerndoelen en eindtermen voor wat betreft onderzoek allerlei onderzoeks*vaardigheden* opgenomen. Vaardigheden ontwikkelen zich echter goeddeels op basis van kennis en het flexibel benutten van vaardigheden vergt eveneens kennis. In de eindtermen ontbreekt evenwel het gezichtspunt dat onderzoek ook een kennisdomein vertegenwoordigt. Leerlingen leren onderzoek op te zetten en uit te voeren kan niet zonder aandacht voor *de kenniskant van onderzoek doen*: de voornaamste begrippen en hun onderlinge samenhang, de belangrijkste methodologische principes en normen en hun belang en de know-how omtrent technieken voor de verzameling en verwerking van gegevens. Zonder volledig te zijn noemen we de volgende begrippen:

- het begrip 'kenmerk' of 'variabele';
- het meetniveau van variabelen (kwalitatieve categorieën, dichotomieën, rangordegegevens, discrete kwantitatieve variabelen, continue kwantitatieve variabelen);
- het aan elkaar gerelateerd zijn van variabelen; oorzaken en gevolgen;
- de functie van variabelen in onderzoek (verklarend of 'onafhankelijk', te verklaren of 'afhankelijk', interveniërend, constant gehouden);
- deduceren, interpreteren, verklaren, voorspellen;
- hypothese (algemeen), verwachting (voor bepaald onderzoek);
- aggregatieniveau, (des)aggregeren;
- de schaal waarop wordt gemeten, respectievelijk de schaal waarop meetresultaten worden afgebeeld;
- onderscheidend vermogen;
- variabiliteit, toeval, meetfout, nauwkeurigheid, betrouwbaarheid;
- steekproef, replicatie, herhaalbaarheid, schatting;
- accuratesse, validiteit, vertekening;
- eerlijke test, intern geldige onderzoeksopzet;
- toetsen van een hypothese;
- generaliseren, generaliseerbaarheid;
- geloofwaardigheid, objectiviteit.

Bij de beslissing welke begrippen in de tweede fase minimaal aan de orde moeten komen is afstemming met het vervolgonderwijs wenselijk.

## 10 BEOORDELEN ALS ONDERDEEL VAN HET ONDERWIJSLEERPROCES (bij 4 en 5)

Bij toetsing van onderzoeksvaardigheden aan de hand van zelfstandig onderzoek is sprake van keuzevrijheid voor en samenwerking tussen leerlingen en tussentijdse hulp van de docent en zijn onderwijs en toetsing niet langer nadrukkelijk gescheiden. Dit past in een internationale trend: 'performance assessment'. De gedachte daarbij is aan te sluiten bij de actieve, zelfstandige leerling voor wie opdrachten en beoordelingen betekenisvol moeten zijn, levensecht, uitdagend en motiverend. Bij zulke opdrachten is meestal geen sprake meer van één goed antwoord, is er geen strikte tijdslimiet en mogen allerlei informatiebronnen en hulpmiddelen worden gebruikt. Zulke opdrachten zouden met name (algemene) vaardigheden toetsen en meer valide zijn dan andere manieren van meten. Naar de condities waaronder zulke verwachtingen realistisch zijn, wordt nog volop onderzoek gedaan.

De toenemende aandacht voor dergelijke vormen van beoordelen gaat gepaard met het benadrukken van de ondersteunende functies die beoordelingen voor het onderwijsleerproces kunnen hebben (diagnose van problemen, geven van feedback) en van het meten van duidelijk omschreven kennis en vaardigheden van leerlingen op basis waarvan uitspraken kunnen worden gedaan over wat de leerlingen nu feitelijk beheersen (criteriumgeoriënteerd meten in plaats van normgeoriënteerd meten, waarin het rangordenen van leerlingen op basis van onderlinge vergelijking centraal staat).

Wat zijn de mogelijkheden en beperkingen om beoordelingen een meer geïntegreerde plaats in het onderwijsleerproces te geven? Dat hangt af van het curriculum, de aard en functie van de beoordeling en de mogelijkheden om de relevante kennis en vaardigheden te meten (vergelijk Nitko, 1989). Het meer integreren van beoordeling en onderwijsleerproces heeft in potentie effecten op de leeractiviteiten, leerprocessen en leerresultaten van de leerlingen. Het verstrekken van duidelijke beoordelingscriteria kan bijdragen aan verduidelijking van de beoogde leerdoelen en hun belang. Op basis van zulke criteria kan realistische en betekenisvolle feedback worden gegeven. Bovendien kunnen duidelijke doelen en realistische feedback leerlingen motiveren, doordat zij ervaren dat ze iets tastbaars en waardevols leren (vergelijk onder andere Crooks, 1988; Snow & Lohman, 1989; Boekaerts, 1991; Rowe & Hill, 1996; Wolfe, 1996). Of zulke effecten optreden hangt uiteraard mede af van de verwachtingen en percepties van de leerlingen. Deze worden echter in belangrijke mate mede gestuurd door vooraf duidelijke en uitdagende (maar in principe wel haalbare) standaarden te verstrekken, respectievelijk te bespreken.



## 11 EISEN WAARAAN GOEDE BEOORDELINGEN MOETEN VOLDOEN (bij 2.7 en 5)

Traditioneel worden aan *metingen* twee hoofdeisen gesteld: betrouwbaarheid en validiteit. Deze eisen zijn gericht op het verkrijgen van meetresultaten die nauwkeurig en accuraat zijn. Deze eisen gelden ook voor metingen van kenmerken van mensen, waaronder hun kennis en vaardigheden.

Aan *beoordelingen* van prestaties van leerlingen, waaraan voor hen diverse consequenties kunnen zijn verbonden, worden aanvullend ook eisen gesteld die zijn gericht op een juiste interpretatie en verantwoord gebruik van de resultaten conform bekende en legitieme bedoelingen.

Over deze materie is nationaal en internationaal veel gepubliceerd, zie bijvoorbeeld De Groot, 1970; Eggen en Sanders, 1993; Linn, 1989, 1994; Linn e.a., 1991; Messick, 1984, 1994, 1995.

Een en ander is natuurlijk gemakkelijker gezegd dan gedaan. In de praktijk blijkt het moeilijk te zijn om te komen tot beoordelingen die consistent zijn en voorspellende waarde hebben voor verwante andere of toekomstige prestaties van leerlingen. Toch kunnen we er niet omheen, dat ook beoordelingen van praktisch werk, waaronder relatief zelfstandig uit te voeren onderzoeken, zoveel mogelijk aan deze eisen moeten voldoen. Dit geldt overigens niet alleen als er cijfers voor worden gegeven, maar ook als de beoordeling beperkt blijft tot bijvoorbeeld 'onvoldoende', 'voldoende' en 'goed' zoals bij profielwerkstukken.

De hierboven kort aangeduide eisen waaraan goede beoordelingen moeten voldoen, zijn in de literatuur nader uitgesplitst. Opgeteld over verschillende auteurs kan een lijst van vijftien à twintig aspecten worden samengesteld. Deze aspecten kunnen echter zonder veel verlies worden samengevat in een viertal rubrieken. We lichten de eisen in figuur 6 nader toe en geven daarna enkele voorbeelden.

<p><i>BETROUWBAAR EN VALIDE METEN</i></p> <ul style="list-style-type: none"> <li>• <b>betrouwbaarheid:</b> als dezelfde docent enkele weken later de producten nog een keer beoordeelt of als een collega de producten ook beoordeelt, mogen de uitkomsten niet teveel verschillen;</li> <li>• <b>validiteit:</b> met de beoordeling moet min of meer worden gemeten wat men wil meten (en niet iets anders); de beoordeling moet: <ul style="list-style-type: none"> <li>- aansluiten bij de behandelde stof, gegeven opdracht en begeleiding;</li> <li>- de ontwikkeling van de relevante (kennis en) vaardigheden zichtbaar maken;</li> <li>- valide onderscheid maken tussen meer en minder bekwame leerlingen;</li> <li>- duidelijk maken waartoe de leerlingen in staat zijn;</li> </ul> </li> </ul> <p><i>ADEQUAAT BEOORDELEN</i></p> <ul style="list-style-type: none"> <li>• <b>praktische bruikbaarheid:</b> <ul style="list-style-type: none"> <li>- docenten moeten de beoordeling kunnen uitvoeren qua bekwaamheid en beschikbare tijd;</li> <li>- de resulterende beoordelingen moeten de gewenste functies kunnen vervullen, zoals: <ul style="list-style-type: none"> <li>• zicht bieden op hoever de leerlingen zijn en wat er nog aan schort;</li> <li>• geschikt zijn voor het meetellen bij de vaststelling van (rapport)cijfers;</li> <li>• geschikt zijn voor communicatie met leerlingen en ouders;</li> </ul> </li> </ul> </li> <li>• <b>aanvaardbaarheid:</b></li> </ul>
---

- de beoordelingsprocedure en de beoordelingscriteria moeten duidelijk en inzichtelijk zijn, kunnen worden uitgelegd aan anderen (leerlingen, ouders, collega's, schoolleiding, inspectie);
- de procedure en criteria moeten objectief zijn, tegenover anderen te verdedigen;
- leerlingen moeten weten waar ze aan toe zijn;
- leerlingen moeten gelijkwaardig worden behandeld, respectievelijk gelijkwaardige kansen krijgen (onder andere qua beschikbare tijd, hulpmiddelen, begeleiding).

Figuur 6 Eisen aan goede beoordelingen

De *betrouwbaarheid* spreekt waarschijnlijk min of meer voor zich. De betrouwbaarheid van beoordelingen kan worden bevorderd door het opstellen van duidelijke en eenduidige aanwijzingen, die de kans op misverstanden, fouten en vergissingen zo klein mogelijk maken.

De *validiteit* is een eis met meerdere facetten. Het 'aansluiten bij de behandelde stof, de gegeven opdracht en begeleiding' betekent bij het beoordelen van *vaardigheden* in feite dat de beoordeling is afgestemd op de *uitvoering* van de betreffende taak of taken. Als een docent bijvoorbeeld werkt met een bepaalde indeling voor het doen van onderzoek in achtereenvolgende stappen, dan is het wenselijk dat de beoordeling daarbij aansluit, bijvoorbeeld door per onderzoeksstap een deelbeoordeling te geven. Dit moet dan heel concreet in het scoringsvoorschrift tot uitdrukking komen: elke onderzoeksstap moet met tenminste één aandachtspunt zijn vertegenwoordigd. Eventueel kan het beoordelingsmodel ook nog aansluiten bij de manier waarop de betreffende vaardigheden zich *ontwikkelen*. Met het laatste kan men 'de ontwikkeling van de relevante (kennis en) vaardigheden zichtbaar maken'. De eis dat wordt gemeten wat men beoogt te meten en niet (ook) iets anders betekent concreet dat als de docent bijvoorbeeld nastreeft dat leerlingen overzicht en inzicht krijgen ten aanzien van wat onderzoek doen inhoudt (het derde doel, zoals beschreven in paragraaf 3.1):

- een beoordeling daarvan aan de hand van een praktisch uitgevoerd onderzoek niet tot heel andere resultaten moet leiden dan bijvoorbeeld een onderzoektoets;
- de beoordeling niet vooral betrekking moet hebben op de samenwerking tussen de leerlingen.

Het 'valide onderscheid maken tussen meer en minder bekwaam leerlingen' wil slechts zeggen dat er iets niet klopt als leerlingen die in feite weinig bekwaam zijn toch hoge scores behalen en omgekeerd. Als dit het geval is, wordt wellicht iets anders gemeten dan is bedoeld, speelt het toeval een te grote rol of is wellicht sprake van fraude. Het 'duidelijk maken waartoe de leerlingen in staat zijn' betekent dat de docent op basis van de beoordelingsresultaten eigenlijk enigszins moet kunnen voorspellen hoe goed de leerlingen het zullen doen bij nieuwe, andere opdrachten op het examen en/of in het vervolgonderwijs.

De *praktische bruikbaarheid* komt erop neer (naast de uitvoerbaarheid voor de docent), dat de *betekenis* van de beoordelingsresultaten duidelijk is. Dan ook kunnen de resultaten een rol spelen in het onderwijsleerproces zelf, doordat de docent zicht krijgt op de eventuele knelpunten en de leerlingen zinvolle feedback krijgen.

De *aanvaardbaarheid* tenslotte hangt op aspecten als eenduidigheid en eerlijkheid. In de praktijk speelt hierbij een grote rol in hoeverre docenten kunnen vermijden om 'oneigenlijke' zaken mee te laten spelen (zoals hun 'voorkennis' of 'totaalindruk' van een leerling) en erin slagen de bedoelingen, opzet en (mogelijke) implicaties effectief te communiceren.

Uiteraard is het voldoen aan deze eisen een kwestie van meer of minder, mede afhankelijk van de bedoelingen van de docent met de beoordeling en de condities waaronder de beoordeling moet plaatsvinden. Wel gelden alle eisen als in principe bij elke beoordeling van belang. Je bent er dus niet als de ene beoordeling aan het ene criterium voldoet en de

andere aan het andere. Die vlieger gaat niet op.

Beoordelingsmethoden voor vaardigheden worden wel 'vaardigheidstoetsen' genoemd. Deze kunnen variëren in 'levensechtheid'. Straetmans (1993) onderscheidt drie groepen: (1) 'hands-off'-toetsen (met pen en papier), (2) simulaties, en (3) 'hands-on'-toetsen (waarbij de lerende de vaardigheid in volle omvang en complexiteit in een zo reëel mogelijke context moet demonstreren). Bij hands-off-toetsen is het meestal gemakkelijker om te voldoen aan eisen van betrouwbaarheid, terwijl op basis van hands-on-toetsen beter de toekomstige prestaties kunnen worden voorspeld (zie ook Elshout-Mohr en Meijer, 1996). Simulaties nemen een tussenpositie in.

Waar op deze dimensie bevindt zich het beoordelen door leraren van onderzoekswerk? Aan de ene kant kan hands-on-toetsing worden genoemd, want de leerlingen beantwoorden niet vragen over onderzoek, maar doen daadwerkelijk onderzoek. Aan de andere kant blijft de situatie een schoolse, waarin de leraar de condities bepaalt en ondersteuning geeft. Heeft de leraar hierdoor 'het beste van twee werelden' of is sprake van een beoordeling die noch betrouwbaar noch levensecht is? Dat hangt af van wat scholen, secties en individuele docenten ervan gaan maken.

## EPILOOG

In het voortgezet onderwijs komt meer aandacht voor vaardigheden, waaronder onderzoeksvaardigheden en vaardigheden in het zelfstandig werken en leren. Docenten laten leerlingen vaker zelfstandig onderzoek doen, ook als onderdeel van het nieuwe schoolexamen in de tweede fase havo/vwo. Docenten kunnen hierbij min of meer zelf bepalen welke doelen ze nastreven, welke opdrachten en begeleiding ze daartoe geven en hoe ze het werk van de leerlingen beoordelen. Veel docenten hebben hiermee nog weinig ervaring.

Deze publicatie behandelt, tegen de achtergrond van mogelijke doelen en aanpakken, de beoordeling van onderzoekswerk van leerlingen. Er wordt een aantal richtlijnen en suggesties gegeven. Docenten, secties en scholen kunnen echter ook andere keuzes maken. De keuzes moeten worden afgestemd op de eigen doelen en situatie. Bij veel keuzes kunnen verschillende argumenten worden gehanteerd. Deze publicatie geeft hiervan een overzicht.

Het zorgvuldig beoordelen van producten van leerlingen is van groot belang. Een goede beoordeling biedt leerrijke feedback en van beoordelingen die meetellen voor een rapport of voor het examen hangt voor leerlingen veel af. Beoordelen is daarom, naast instrueren en begeleiden, een professionele taak en verantwoordelijkheid van docenten. Deze publicatie geeft aanwijzingen voor de ontwikkeling daarvan en voor het daarbij samenwerken met collega's.

Deze brochure is niet alleen bedoeld voor docenten, maar ook voor sectieleiders en schoolleiders. Keuzes van docenten moeten worden gedragen door het beleid van de secties en van de school. Voorzover dat er nog niet is, kan deze publicatie ook worden gebruikt om zulk beleid te ontwikkelen.

## REFERENTIES

- American Federation of Teachers, National Council on Measurement in Education, National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. Washington, D.C.
- APS (1994). *Natuur- en scheikunde: Practicum en vaardigheden*. Utrecht.
- APS (1998). *Leerlingen helpen bij onderzoek*. Verslag kaderconferentie. Utrecht.
- Boekaerts, M. (1991). *Subjective competence, appraisals and self-assessment*. Learning and Instruction 1, 1-17.
- Beckers, M. & Tooren, M. van (1998). *Toetsing vaardigheden vraagt veel van docenten*. PMVO-journaal 27 (16-9-98), 11-12.
- Brown, C.R., Moore, J.L., Silkstone, B.E., & Botton, C. (1996). *The construct validity and context dependency of teacher assessment of practical skills in some pre-university level science examinations*. Assessment in Education 3 (3), 377-391.
- Codename Future Project (1997, 1998). *Docentenhandleiding en dossierkrakers*. Den Haag: Stichting Codename Future.
- Collins, A., Brown, J.S. & Newman S.E. (1989). *Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics*. In: L.B. Resnick (Ed.). *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale: Erlbaum, 453-494.
- Creemers, L. & Knuver, A. (1998). *Leraren positief over invoering gedragscode voor cijfers geven*. Didaktief&School 4 (mei), 4-8.
- Creemers-van Wees, L.M.C.M., Knuver, J.W.M., Vos, H.J. & Linden, W.J. van der (1997). *Toetsen, beoordelen en beslissen in het voortgezet onderwijs*. Enschede: OCTO.
- Crooks, T.J. (1988). *The impact of classroom evaluation practices on students*. Review of Educational Research 58 (4), 438-481.
- Eggen, T. & Sanders, P. (Eds.) (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Elshout-Mohr, M. & Meijer, J. (1996). *Instrumentconstructie voor de meting van algemene vaardigheden VOCL '93*. Amsterdam: SCO-Kohnstamm Instituut.
- Elshout-Mohr, M. & Meijer, J. (1998). *Hoe beoordeel je algemene vaardigheden?* Didaktief&School 2 (maart), 40-41.
- Germann, P.J., Haskins, S. & Auls, S. (1996). *Analysis of nine high school biology laboratory manuals: promoting scientific inquiry*. Journal of Research in Science Teaching 33 (4), 475-499.
- Gott, R. & Duggan, S. (1996). *Practical work: its role in the understanding of evidence in science*. International Journal

of Science Education 18 (7), 791-806.

Groen, H., Kerkhof, A. van de, Roo, L. de & Smeets, L. (1996). *Vaardigheidstoetsing in het studiehuis. Een verslag van exploratief ontwikkelingswerk waarin verschillende varianten de revue passeren*. Arnhem: Cito.

Groot, A.D. de (1970). *Some badly needed non-statistical concepts in applied psychometrics*. Nederlands Tijdschrift voor de Psychologie 25, 360-376.

Groot, A.D. de (1986). *Wat neemt de leerling mee van onderwijs? Gedragsrepertoires, programma's, kennis-en-vaardigheden*. In: A.D. de Groot, Begrip van Evalueren. Den Haag: VUGA, 63-82.

Hambleton, R.K. & Murphy, E. (1992). *A psychometric perspective on authentic measurement*. Applied Measurement in Education 5 (1), 1-16.

Hennessy, S. (1993). *Situated cognition and cognitive apprenticeship: Implications for classroom learning*. Studies in Science Education 22, 1-42.

Keeves, J.P. (1994). *Methods of assessment in schools*. In: T. Husén, & N. Postlethwaite (Eds.). International encyclopedia of education (2nd ed.). New York: Pergamon, 362-370.

Krogt, M. van der & Sinkeldam, R. (1998). *Handleiding praktische opdrachten*. Arnhem: Cito.

Lijnse, P.L. (1994). *Probleemoplossen en algemene vaardigheden: een poging tot discussie*. Tijdschrift voor Didactiek der  $\beta$ -wetenschappen 12 (3), 246-260.

Linn, R.L. (Ed.) (1989). *Educational Measurement* (3rd ed.). New York: Macmillan.

Linn, R.L. (1994). *Performance Assessment. Policy promises and technical measurement standards*. Educational Researcher 23 (9), 4-14.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). *Complex, performance-based assessment: expectations and validation criteria*. Educational Researcher 20 (8), 15-21.

Lock, R. (1989). *Assessment of practical skills. Part 1. The relationships between component skills*. Research in Science & Technological Education 7 (2), 221-233.

Lock, R. (1990). *Assessment of practical skills. Part 2. Context dependency and construct validity*. Research in Science & Technological Education 8 (1), 35-52.

Messick, S. (1984). *The psychology of educational measurement*. Journal of Educational Measurement 21 (3), 215-237.

Messick, S. (1994). *The interplay of evidence and consequences in the validation of performance assessments*.

- Educational Researcher 23 (2), 13-23.
- Messick, S. (1995). *Validity of psychological assessment. Validation of inferences from person's responses and performances as scientific inquiry into score meaning*. American Psychologist 50 (9), 741-749.
- Moerkerke, G. & Dochy, F. (1997). *Het toetsen van vaardigheden*. In: G. ten Dam, H. van Hout, C. Terlouw, J. Willems. (Red.) Onderwijskunde Hoger Onderwijs. Handboek voor docenten. Assen: Van Gorcum, 214-236.
- Nitko, A.J. (1989). *Designing tests that are integrated with instruction*. In: R.L. Linn (Ed.). Educational Measurement (3rd ed.). New York: Macmillan, 447-474.
- Novak, J.R., Herman, J.L. & Gearhart, M. (1996). *Establishing validity for performance-based assessments: An illustration for collections of student writing*. The Journal of Educational Research 89 (4), 220-233.
- O'Sullivan, R.G. & Johnson, R.L. (1993). *Using performance assessments to measure teachers' competence in classroom assessment*. Paper AERA, Atlanta, april 12-16.
- PMVO (1998). *Nader beschouwd: Praktische tips voor het aanleren, oefenen en beoordelen van vaardigheden*. Den Haag.
- PMVO (1998). *Nader beschouwd: Examendossier: ervaringen uit het netwerk*. Den Haag.
- Rowe, K.J. & Hill, P.W. (1996). *Assessing, recording and reporting students' educational progress: the case for 'subject profiles'*. Assessment in Education 3 (3), 309-352.
- Ruiz-Primo, M.A. & Shavelson, R.J. (1996). *Rhetoric and reality in science performance assessments: an update*. Journal of Research in Science Teaching 33 (10), 1045-1063.
- Seyfart, J.T., Simon, D.J. & Schlesinger, J. (1994). *Assessing student performance: are our assumptions valid?* Paper presented at the annual meeting of the American Association of Colleges of Teacher Education. Chicago, February 16-19.
- Shavelson, R.J., Baxter, G.P. & Gao, X. (1993). *Sampling variability of performance assessments*. Journal of Educational Measurement 30 (3), 215-232.
- Sinkeldam, R. (Red.). (1998). *Handleiding profielwerkstuk*. Arnhem: Cito.
- SLO (1996). *Herziene kerndoelen basisvorming*. Enschede.
- SLO (1996). *Eindexamenprogramma's havo/vwo*. Enschede.
- Sluyter, C. (1998). *Proefwerken op de helling*. In: Van twaalf tot achttien, juni, 40-41.
- Snow, R.E. & Lohman, D.F. (1989). *Implications of cognitive psychology for educational measurement*. In R.L. Linn (Ed.). Educational measurement (3rd ed.). New York: Macmillan, 263-331.

- Stiggins, R.J. & Bridgeford, N.J. (1985). *The ecology of classroom assessment*. Journal of Educational Measurement 22 (4), 271-286.
- Stokking, K.M. (1997). *Algemene vaardigheden in het curriculum. Verkenning en programma*. Leuven/Apeldoorn: Garant.
- Stokking, K.M. (1997). *Algemene vaardigheden zijn steeds belangrijker*. Didaktief & School 1997 (27,8), pp. 36-38.
- Stokking, K.M. (1998). *Algemene vaardigheden: een complexe uitdaging*. Pedagogische Studiën 75 (2), pp. 110-120.
- Stokking, K. & Voeten, R. (1999). *Valid classroom assessment of complex skills*. In: P. Simons, J. van der Linden, T. Duffy. New Learning. Boston, Dordrecht, London: Kluwer Academic Publishers.
- Straetmans, G.J.J.M. (1993). *Het vaststellen van competentie met vaardigheidstoetsen*. Tijdschrift voor Hoger Onderwijs 11 (3), 188-203.
- Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs (1997). *Het examendossier*. Den Haag.
- Swanson, D.B., Norman, G.R. & Linn, R.L. (1995). *Performance-based assessment: Lessons from the health professions*. Educational Researcher 24 (5), 5-11, 35.
- Tilburg, P.A. van, Verloop, N. & Vermunt, J.D. (1998). *Teachers' knowledge and beliefs about the teaching of inquiry skills*. Paper NARST, San Diego, april 19-22.
- VaPro (1997). *Eindtermenboek Laboratoriumtechniek*. Leidschendam: Stichting Vakopleiding Procesindustrie.
- Vonderen, J. van (1998). *Vaardigheden zijn niet aan één vak gebonden*. Uitleg 16 (17-6-98), 11-15.
- Webb, N.M. (1995). *Group collaboration in assessment: Multiple objectives, processes, and outcomes*. Educational Evaluation and Policy Analysis 17 (2), 239-261.
- White, B.Y. & Frederiksen, J.R. (1998). *Inquiry, modeling, and metacognition: Making science accessible to all students*. Cognition and Instruction 16 (1), 3-118.
- Wolf, D., Bixby, J., Glenn, J. & Gardner, H. (1991). *To use their minds well: Investigating new forms of student assessment*. Review of Educational Research 17, 31-74.
- Wolfe, E.W. (1996). *Student Reflection in Portfolio Assessment*. Paper, Annual Meeting of the National Council on Measurement in Education, New York, April 9-11.