

Bijlage 2: Literatuur

Uit: T. Geerlings "lesgeven en zelfstandig leren"

7.7 Beheersingsvormen van cognitieve leerdoelen

Bij het nastreven van leerdoelen doen leerlingen een beroep op verschillende typen denkprocessen om leerstof onder de knie te krijgen, te beheersen. Zo gaat het bij het ene leerdoel om het zo letterlijk mogelijk weergeven van opgenomen informatie, bij andere om het toepassen van een regel, het zelf bedenken van een voorbeeld, of het geven van een oordeel.

De beheersing van leerdoelen valt in twee categorieën uiteen: reproductie en productie. De reproductieve beheersingsvorm houdt in het ophalen en weer-

geven van kennis uit het langetermijngeheugen (feiten, begrippen, verbanden, theorieën, methoden, technieken etcetera). Bij de productieve beheersingsvorm wordt niet alleen iets opgevist uit het geheugen, maar gaat het er veeleer om zelf, al nadenkend, tot de oplossing van een probleem te komen. (Denken is zoeken naar oplossingen van problemen.)

Aan de hand van onderstaande leerdoelen is geprobeerd het onderscheid tussen verschillende beheersingsvormen en de daaraan gerelateerde denkprocessen te verduidelijken.

A	B
1. De organen van het spijsverteringsstelsel van de koe kunnen opnoemen.	3. Kunnen berekenen hoeveel dollars men ontvangt voor f 300,- bij een koers van f 1,88.
2. In het Erans tot honderd kunnen tellen.	4. Een gegeven aantal levensmiddelen kunnen onderbrengen in de fabrieken van de maatschappij.
	5. Een reclamefolder voor een bepaald product kunnen ontwerpen.
	6. Gedocumenteerd schrijven (3 pagina's) over het onderwerp 'Het Nederlandse drugsbeleid'.
	7. Kunnen beoordelen of een partij zich in haar partijprogramma schuldig maakt aan racisme.
	8. De houdbaarheid van een stelling (bijvoorbeeld: Criminaliteit wordt veroorzaakt door grote inkomensverschillen) kunnen verdedigen.

De leerdoelen in kolom A betreffen het weergeven of reproduceren van kennis uit het langetermijngeheugen; bij de leerdoelen uit kolom B daarentegen gaat het om het zelf bedenken of produceren van een antwoord, oplossing of mening.

Bij de leerdoelen 3 en 4 uit kolom B richt het denkproces zich op één oplossing, die goed, gedeeltelijk fout of geheel fout kan zijn. Deze beheersingsvorm heet convergent denken (convergeren betekent: in één punt samenkomen) of logisch denken. Bij convergent denken ligt de juiste, algemeen aanvaarde oplossing van het gestelde probleem bij voorbaat vast en vaak ook de weg die naar de oplossing voert.

Bij de leerdoelen 5 en 6 zijn verschillende goede oplossingen mogelijk. Deze beheersingsvorm heet divergent denken (divergeren betekent: uiteenwijken) of creatief denken. Bij divergent denken ligt de oplossing van het probleem niet bij voorbaat vast, evenmin als de weg erheen. De diverse oplossingen kunnen van elkaar verschillen in originaliteit of bruikbaarheid.

Tenslotte de leerdoelen 7 en 8. Deze zijn gericht op het vormen van een mening of een waarde-oordeel over bepaalde informatie. Deze beheersingsvorm heet evaluatief denken (evalueren betekent: op waarde schatten) of

vorm heet evaluatief denken (evalueren betekent: op waarde schatten) of kritisch denken. Evaluatief denken is denken over informatie aan de hand van criteria die al dan niet gegeven kunnen zijn. Deze criteria zijn van tweeërlei aard:

- a. interne criteria, zoals de logische opbouw en interne samenhang, de volledigheid, nauwkeurigheid, objectiviteit of geloofwaardigheid van de informatie;
- b. externe criteria, zoals het nut of de relevantie van de informatie in het licht van een bepaald doel.

De waarde-oordelen verschillen naar inhoud en kwaliteit van de argumentatie.

Voorkennis en beheersingsvorm

Bij het aangeven van de reproductieve of productieve beheersingsvorm moet er rekening worden gehouden met de aanwezige voorkennis van de leerlingen. Dezelfde waarneembare prestatie kan geleverd worden op basis van verschillende cognitieve operaties. Er bestaat zowel reproductief als productief probleemoplossen.

Bij reproductief probleemoplossen zit het oplossingsplan kant en klaar in het langetermijngeheugen en behoeft het alleen te worden geactiveerd. Zo zal een opdracht over gevolgen van de vervuiling van de Waddenzee door een leerling die, omdat hij een enthousiast lid is van de Waddenvereniging, over een grote voorkennis beschikt, waarschijnlijk reproductief beantwoord kunnen worden. Deze leerling diept (zoals een expert dat ook doet) de oplossing kant en klaar uit zijn geheugen op, terwijl zijn medeleerlingen nadere informatie moeten verzamelen en hierover moeten nadenken. Bij productief denken (probleemoplossen) wordt tijdens het opstellen van de juiste oplossing relevante voorkennis gemobiliseerd, gecombineerd en gereorganiseerd.

7.7 Geef de beheersingsvorm aan

Opdracht

- Geef van onderstaande leerdoelen de beheersingsvorm aan: r(eproductief), c(ontvergent), d(ivergent), e(valuatief).

De leerlingen kunnen

- 1 het begrip schutkleur in eigen woorden omschrijven en met een drietal voorbeelden illustreren;
- 2 in een krantenartikel over de conjunctuur in West-Europa die zinnen, zinsdelen en begrippen onderstrepen die de conclusie, dat de conjunctuur een golfbeweging vertoont, rechtvaardigen;
- 3 met behulp van zelf ontwikkelde criteria het uitbreidingsplan van Schiphol beoordelen;
- 4 zelf een aantal redactiesommen samenstellen waarin de rekenkundige hoofdbewerkingen optellen, aftrekken en vermenigvuldigen voorkomen;
- 5 verklaren waarom twee rechterlijke instanties over eenzelfde zaak verschillend kunnen oordelen;

- | | |
|---|--------------------------|
| 6 met informatie over warmte-isolatie een gebouw ontwerpen dat een maximum aan warmte-isolatie kent; | <input type="checkbox"/> |
| 7 met een tweetal zelfbedachte voorbeelden het verschil aangeven tussen filtreren en distilleren; | <input type="checkbox"/> |
| 8 een gefundeerde mening geven over het realistisch taalgebruik in romans van J. Wolkers; | <input type="checkbox"/> |
| 9 naar aanleiding van een artikel in de krant een ingezonden stuk schrijven, dat voldoet aan de eisen van logische opbouw en eenduidige formulering; | <input type="checkbox"/> |
| 10 een oplossing bedenken voor het verminderen van koolmonoxide in de uitlaatgassen van motoren; | <input type="checkbox"/> |
| 11 wat betreft de zinsbouw een aantal verschillen opnoemen tussen de Engelse taal en het Nederlands; | <input type="checkbox"/> |
| 12 aan de hand van een aantal oorlogshandelingen in de jaren zestig, waarbij de Amerikanen betrokken waren, uitmaken of de Amerikaanse regering zich schuldig heeft gemaakt aan schending van de Rechten van de Mens. | <input type="checkbox"/> |

7.7.1 Beheersingsvorm en moeilijkheidsgraad van leerstof

De subjectief ervaren moeilijkheidsgraad van leerstof hangt enerzijds samen met capaciteiten, aanwezige relevante voorkennis en interesse van de leerlingen, anderzijds met de gestelde beheersingseisen en de beschikbare, respectievelijk daadwerkelijk geïnvesteerde studietijd. Ook vormkenmerken van de studietekst, de kwaliteit van de begeleiding en tussentijdse feedback die de leerling krijgt zijn hierop van invloed.

De objectieve moeilijkheidsgraad van leerstof die een leerling onder de knie moet proberen te krijgen, wordt vooral bepaald door de mate van complexiteit (het aantal elementen dat wordt opgevoerd en aantal en aard van de onderlinge relaties hiertussen) en de abstractiegraad (de ervaren afstand tot de waarneembare werkelijkheid) van deze leerstof.

In het omgaan met de onderscheiden beheersingsvormen moet een docent zich hoeden voor de misvatting, dat beheersingsvormen zich per definitie onderscheiden naar moeilijkheidsgraad. Hoewel reproductieve leerdoelen vaak gemakkelijker zijn dan productieve, hoeft dit niet altijd het geval te zijn. Een voorbeeld. Het zo volledig mogelijk kunnen reproduceren van Darwins evolutietheorie zal naar alle waarschijnlijkheid moeilijker zijn dan het kunnen geven van een evolutionaire verklaring voor het verschil in uiterlijk tussen een ijsbeer en een bruine beer.

Leerdoelen waarbij een beroep moet worden gedaan op divergent of evaluatief denken hoeven niet moeilijker te zijn dan leerdoelen waarin convergente denkprocessen de hoofdrol spelen. Het zelf bedenken van een som (waarbij de leerling dus divergent moet denken) kan gemakkelijker zijn dan het oplossen van een som die de leraar aanreikt (en waarvoor convergent moet worden gedacht). Elke beheersingsvorm biedt gelegenheid voor gemakkelijke en moeilijker te bereiken leerdoelen. Het kunnen uitvoeren van een renteberekening waarbij een kapitaal van f 6000,- gedurende een vol jaar tegen 6% wordt uitgezet, is een minder moeilijk leerdoel dan het maken van een berekening, waarbij een kapitaal van f 6225,50 van 1 januari tot 3 juni uitstaat tegen een rentepercentage van $5\frac{1}{4}$.

De objectieve moeilijkheidsgraad van leerdoelen wordt dus niet zozeer bepaald door de beheersingsvorm, maar veeleer door complexiteit en abstractiegraad van de leerinhoud.

7.8 Kritisch denken

Sommige lezers hebben het misschien wel eens meegemaakt. Aan de borreltafel in een buurtcafé wordt gediscussieerd over nadelen van het roken. Ineens zegt een van de stamgasten: 'Ik geloof niks van al die bangmakerij. Het loopt zo'n vaart niet. Mijn opa is 93 jaar geworden en die heeft zijn hele leven gerookt als een schoorsteen.' Het gesprek stopt omdat niemand van de aanwezigen hier iets tegen weet in te brengen. Degene die het 'opa-argument' naar voren heeft gebracht geeft een rondje en men stapt over op een ander onderwerp. Eén opa blijkt in staat om al die voortijdig aan de gevolgen van roken gestorven mensen te doen vergeten. In zo'n geval is er sprake van onkritisch denken.

Gebrek aan kritisch denken is alomtegenwoordig. Hoeveel mensen kijken niet selectief naar ondersteunend bewijsmateriaal en zijn blind voor tegenindicaties? Hoeveel mensen denken niet zwart-wit? Hoeveel mensen kunnen aanvaarden dat waarheid relatief is? Bij hoeveel mensen is het besef doorgedrongen dat autoriteiten 'valse profeten' kunnen zijn? Hoeveel mensen geloven niet in astrologie? Hoeveel goklustigen denken zich niet rijk door zich een veel grotere kans op de hoofdprijs in een loterij voor de geest te halen dan de werkelijke kans?

Twee belangrijke componenten van kritisch denken zijn:

- a kunnen selecteren van relevante informatie;
- b het waarheidsgehalte van informatie toetsen.

ad a

In de moderne samenleving staan leerlingen bloot aan een niet aflatend 'bombardement' van informatie die via televisie, radio, computer, film, telefoon, kranten, tijdschriften en boeken op hun afkomt. Informatie is gemakkelijker dan ooit te verkrijgen, maar het wordt steeds moeilijker de wassende informatiestroom te ordenen en naar waarde te schatten. Kritisch kunnen omgaan met informatie is een vaardigheid die steeds belangrijker zal worden.

ad b

Kritisch kunnen denken omvat één scala aan onderling samenhangende deeltaalvaardigheden, zoals:

- de geloofwaardigheid en juistheid van een informatiebron kunnen beoordelen;
- onderscheid kunnen maken tussen feiten en meningen;
- signaleren van vaag, meerduidig taalgebruik;
- inconsistenties en contradicties in redeneringen kunnen onderkennen;

- vooronderstellingen in een redenering kunnen blootleggen;
- onderscheid kunnen maken tussen argumenten die wel en argumenten die niet ter zake doen;
- rekening houden met tijd- en plaatsgebondenheid van informatie;
- onderscheid kunnen maken tussen feiten en gevolgtrekkingen daaruit;
- onderscheid kunnen maken tussen correlationele en causale verbanden;
- onderscheid kunnen maken tussen gerechtvaardigde en foute conclusies;
- onderscheid kunnen maken tussen verantwoorde en ongerechtvaardigde generalisaties.

Als de leraar aan een leerling vraagt een oordeel te geven over een boek dat moest worden gelezen en de leerling weet niet meer op te hoesten dan: 'Dat boek vond ik goed waardeloos', dan heeft die leerling misschien kritisch gedacht, echter wel op een onaanvaardbaar laag niveau. Leerlingen moeten voortdurend worden 'gedwongen' hun waarde-oordelen met steekhoudende argumenten te onderbouwen.

Als kritisch denken bezit heeft genomen van de persoon van de leerling, heeft het zich ontwikkeld tot een kritische attitude, die behalve in zijn denken tot uitdrukking komt in zijn gevoelens en in zijn gedrag. Een leerling met een kritische attitude heeft het zoeken naar waarheid hoog in zijn vaandel geplaatst. Hij zal er steeds naar streven eerst goed geïnformeerd te zijn en daarna pas te oordelen.

Een voorbeeld. Een leerling heeft zich verdiept in kritische literatuur over paranormale verschijnselen. Hij is dientengevolge sceptisch gaan denken over astrologie, reïncarnatie en waarzeggerij. Gevoelens van irritatie kunnen opkomen als hij beoefenaars van paranormale activiteiten betraapt op vage gemeenplaatsen, halve waarheden, misleiding of grove leugens. Een medeleerling die een voorspelling van een waarzegger op de kermis ('Je zult over een paar jaar heel rijk zijn') serieus neemt, laat hij weten dat het allemaal nep is. Op school zou meer kunnen worden gedaan om het kritisch denken van leerlingen te bevorderen. Leerstof wordt gepresenteerd als een verzameling waarheden die je moet kennen, hetgeen bij leerlingen de opvatting doet postvatten dat, als je de toets haalt, er verder niet over nagedacht hoeft te worden. Waarom zou je reflecteren over iets wat onomstotelijk vaststaat? Soms komen leerlingen met een vraag als 'waarom moeten we dit eigenlijk weten?' aanzetten. Als de leraar zo'n vraag beantwoordt met een dooddoener als 'omdat het belangrijk is voor het examen' smoort hij in feite kritisch denken van leerlingen in de kiem. Bovendien kan nadenken over het nut van leerstof de motivatie van leerlingen opkrikken. Leraren zouden dergelijke vragen juist of prijs moeten stellen, in plaats van, zoals volgens leerlingen vaak gebeurt afkappen.

7.8 Objectiviteit van leerboekteksten

De belangrijkste functie van een informatieve tekst is het overbrengen van boodschappen over feitelijkheden en probleemsamenhangen in een zo objectief mogelijke vorm en op een zo economisch mogelijke manier. Sommige leerboekteksten blijken er niet aan te ontkomen, dat er persoonlijke indrukken, meningen, waarde-oordelen ja soms uitgesproken vooroordelen in zijn opgenomen.

Opdracht

De objectiviteit kan op allerlei wijzen geweld aangedaan worden. Een aantal mogelijkheden zijn:

- Ideologisch taalgebruik (bijvoorbeeld het hanteren van een begrippenkader uit een bepaalde maatschappelijke laag).
 - Benadering vanuit een eenzijdige visie (bijvoorbeeld de koude oorlog enkel zien door de bril van het Westen).
 - Feiten verwarren met meningen.
 - Het bewust of onbewust tot uitdrukking brengen van waardeoordelen of vooroordelen.
 - Het weglaten van informatie, zodat er nietszeggende of vertekende informatie wordt gegeven.
 - Het verstrekken van onjuiste informatie (bedoeld of onbedoeld).
 - Het gebruik van emotioneel gekleurde woorden.
 - Het bewijzen uit oneerlijk geselecteerde gevallen.
 - Het veronderstellen van wat bewezen moet worden.
 - Het meten met twee maten.
 -
- Kies één of meer geschikte teksten uit een gangbare methode en toets deze op objectiviteit.
Maak een verslag van dit onderzoekje en bespreek dit in de groep.

7.9 Formulering van cognitieve leerdoelen

Hoe concreter en eenduidiger leerdoelen zijn geformuleerd, hoe beter ze als richtlijn kunnen fungeren voor het lesgeven of zelfstandig leren.

Ieder leerdoel moet twee bestanddelen bevatten, namelijk de inhoud (de leerstof) en wat de leerling hiermee moet doen (het gedrag).

Tijdens de voorbereiding van lessen op de lerarenopleiding is het van belang dat de beoogde beheersingsvorm wordt overdacht. Op het lesformulier kan een student naast een concreet leerdoel de beoogde beheersingsvorm met de eerste letter van de vier cognitieve beheersingsvormen aanduiden: r, c, d of e.

Een docent kan niet waarnemen wat een leerling denkt, alleen maar hoe hij of zij zich gedraagt. Voor het omschrijven van waarneembaar, door de leerlingen te demonstreren gedrag passen termen als weten, begrijpen, inzicht hebben in, etcetera minder goed. Dergelijke begrippen verwijzen niet naar observeerbaar gedrag en zijn voor diverse interpretaties vatbaar.

Woorden die verwijzen naar waarneembaar gedrag zijn bijvoorbeeld: opnoemen, beantwoorden, opzoeken, ordenen, invullen, kiezen, onderstrepen, uitbeelden, opschrijven, vertalen, verdelen, tekenen, berekenen, construeren,

meten, oplossen, voorbeelden geven van. Deze termen zijn tevens meer eenduidig.

Welke fouten worden vaak gemaakt bij het formuleren van (cognitieve) leerdoelen?

- Vage, meerduidige omschrijvingen.
Voorbeeld: De leerlingen hebben inzicht in het wezen van het kapitalisme.
Een betere formulering zou zijn: De leerlingen kunnen de hoofdkenmerken van het kapitalisme omschrijven.
- Omschrijving enkel in termen van te behandelen leerstof.
Voorbeeld: Vitaminen.
Het leerlinggedrag ontbreekt hier.
- Omschrijving enkel in termen van beheersingsvorm.
Voorbeeld: De leerlingen behoren kritisch te leren denken.
Hierbij kan men zich afvragen over welke leerstof de leerling evaluatief moet kunnen denken en hoe hij dit moet demonstreren.
- Omschrijving in termen van leerlingactiviteit.
Voorbeeld: De leerlingen bezoeken een museum.
Zo'n omschrijving hoort in de kolom didactische aanpak thuis. Leerstof en leerlinggedrag ontbreken.

Het formuleren van leerdoelen wordt in de praktijk van het onderwijs weinig toegepast. De meeste leraren vinden het volgens de regels formuleren van leerdoelen een omslachtige werkwijze en veronderstellen dat het weinig of geen rendement zal opleveren. Toch heeft het formuleren en van te voren aan de leerlingen bekend maken van de leerdoelen de volgende pluspunten:

- naarmate de leerlingen duidelijker weten wat van hen verlangd wordt, zullen ze gericht leren;
- naarmate de leerkracht de leerdoelen nauwkeuriger voor ogen heeft, zal hij meer weloverwogen te werk gaan bij het inrichten van onderwijsleersituaties om deze doelen te bereiken;
- naarmate de leerdoelen duidelijker omschreven zijn kan het leerresultaat beter worden geëvalueerd.

7.10 Leerdoelen en zelfstandig leren

Zelfstandig en samenwerkend leren moet vanzelfsprekend richting krijgen door leerdoelen, die meestal in overleg met de docent worden vastgesteld. Toetsvragen zijn, als het goed is, eigenlijk nauwkeurig geformuleerde in vraagvorm (opdrachtvorm) gegoten leerdoelen. Het rendement van lesgeven en zelfstandig leren wordt sterk bevorderd als aan het begin aan de leerlingen bekend is gemaakt hoe ze getoetst zullen worden, welke minimumprestatie daarbij aanvaardbaar is, welke informatie en hulpmiddelen bij de evaluatie (bijvoorbeeld woordenboeken, naslagwerken, blinde kaart, tabellen) zijn toe-

gestaan en tenslotte hoeveel tijd ervoor wordt uitgetrokken.

Leerdoelen waarbij leerlingen inspraak hebben gehad blijken effectiever te zijn dan eenzijdig door een docent opgelegde leerdoelen.

Leerdoelen die door leerlingen waargenomen worden als heel gemakkelijk of uiterst moeilijk, wekken minder inzet op dan leerdoelen die als moeilijk, maar haalbaar worden gezien.

Bij zelfstandig leren moeten gemaakte vorderingen in relatie tot de gestelde leerdoelen regelmatig zichtbaar worden gemaakt. Als leerlingen ervaren dat ze vooruitgang maken, verhoogt dit hun motivatie. Succeservaringen versterken tevens de zelfeffectiviteit.

Stageopdrachten

34 De basisvorming in het voortgezet onderwijs

De met ingang van 1 augustus 1993 ingevoerde basisvorming kent de volgende uitgangspunten:

- algemene verhoging van het onderwijspeil;
- vermindering van ongelijkheid van onderwijskansen;
- vergroting van de maatschappelijke relevantie en het toepassingsgehalte van de onderwijsinhouden als uitvloeisel van het TVS-concept: toepassing, vaardigheden en samenhang.

Opdracht

- Voer gesprekken met de directie en aantal docenten van je stageschool in hoeverre ze bovenstaande uitgangspunten onderschrijven en geloof hechten aan de realisering ervan door middel van inhoudelijke (kerndoelen) en didactische vormgeving van de basisvorming op de onderscheiden schooltypen voor voortgezet onderwijs.
- Probeer vast te stellen hoe op je stageschool de uren tabel voor de verschillende vakken van de basisvorming zich verhoudt tot de adviesuren tabel en bespreek vervolgens met je stagebegeleider en/of de directie de door jouw stageschool aangebrachte eigen accenten in de uren tabel (bijvoorbeeld meer uren toegekend aan de beeldende vorming dan in de adviestabel; zie pag. 225).
- Is 'jouw' vaksectie of vakdocent van de stageschool al dan niet van mening, dat zijn/haar vak met het oog op de vormende waarde ervan in het licht van de basisvorming (voldoende) recht is gedaan met de aan het vak toebedachte uren in de adviestabel en/of toebedeelde uren in de basisvorming op je stageschool?

Maak een verslag van je bevindingen en bespreek dit met je stagebegeleider en/of vakdidacticus van je opleiding.

35 Kerndoelen voor de basisvorming

De voor de vijftien vakken van de basisvorming opgestelde kerndoelen hebben het karakter van streefdoelen. Ze zijn daarom globaal geformuleerd en bevatten geen niveau-aanduiding. Ter illustratie geven we een van de kerndoelen van het vak techniek weer: 'de leerlingen kunnen vooraf gestelde criteria voor bruikbaarheid, mate-

10.1 Entree

De wijze van evalueren heeft grote invloed op het studiegedrag van leerlingen. Leerstof die in het leerprogramma thuishoort, maar waarvan de leerlingen verwachten of weten dat deze niet zal worden getoetst, krijgt doorgaans weinig aandacht. Stel bijvoorbeeld dat leerlingen twee hoofdstukken uit een boek moeten bestuderen en van leerlingen die een jaar verder zijn horen dat bijna alle vragen in de toets uit het eerste hoofdstuk komen. Hoeveel leerlingen zullen het tweede hoofdstuk dan nog (grondig) bestuderen?

Sommige leraren hebben de merkwaardige gewoonte vragen over de 'kern' van de stof – vragen waarvan ze denken dat veel leerlingen ze goed zullen beantwoorden – maar niet in de toets op te nemen, omdat deze dan te gemakkelijk zou worden. Inderdaad zijn vragen vaak moeilijker te beantwoorden, naarmate ze meer over bijzaken en details gaan. Gevolg van deze wijze van handelen is echter wel, dat leerlingen hun aandacht zullen gaan richten op bijzaken in plaats van op hoofdzaken.

Niet alleen de inhoud, ook de beheersingsvormen waarop in de toetsing een beroep wordt gedaan hebben invloed op hoe en wat leerlingen leren. Als in toetsen het accent ligt op reproductie van kennis, zullen leerlingen vooral reproducerend gaan leren.

Eén van de centrale ideeën in de huidige onderwijsvernieuwing is, dat het kunnen toepassen van kennis belangrijker is dan het kunnen 'oplepelen' van feiten en begrippen. Omdat leerlingen hun wijze van leren afstemmen op de verwachte vorm en inhoud van de toetsing, is het dus zaak dat toetsen toepassing van kennis toetsen.

Wat leerlingen in het onderwijs opsteken is niet beperkt tot kennis en vaardigheden. Onderwijs beïnvloedt ook waarden en gevoelens.

Evaluatiemiddelen voor de niet-cognitieve doeldomeinen komen in dit hoofdstuk slechts summier ter sprake. In overeenstemming met de dagelijkse gang van zaken in scholen zal de nadruk liggen op de behandeling van evaluatiemiddelen voor cognitieve prestaties. Diverse soorten open en gesloten vraagvormen komen uitgebreid aan de orde, evenals maatregelen ter

bevordering van de objectiviteit van beoordelingen. Het gerechtvaardigde streven naar zo objectief mogelijke beoordelingen mag er niet toe leiden, dat evaluatiemiddelen die minder objectief meten, meteen uit het scala van toetsmiddelen worden weggestreept. Een voorbeeld. Het beoordelen van scripties blijkt een tamelijk subjectieve aangelegenheid te zijn. Als een scriptie door twee leraren onafhankelijk van elkaar wordt beoordeeld, is het heel wel mogelijk dat de ene leraar de loftrumpet steekt, terwijl dezelfde scriptie bij de andere docent de wenkbrauwen doet fronsen. De scriptie is echter een van de belangrijkste middelen om de schriftelijke uitdrukkingsvaardigheid te oefenen, kennis uit verschillende vakken in verband met elkaar te brengen en toe te passen en het zelfstandig leren gestalte te geven (zelf een onderwerp kiezen, zelfstandig informatie opsporen en afwegen in het licht van de probleemstelling).

Als uitsluitend zou worden geëvalueerd wat objectief meetbaar is, zou dit een ongewenste verarming van het onderwijs teweegbrengen.

Toetsresultaten kunnen tevens licht werpen op de effectiviteit van de leeromgeving. Het idee dat lage leerprestaties ook wel eens (mede) veroorzaakt zouden kunnen zijn door mankementen in het aangeboden onderwijs, wordt in veel gevallen door docenten verdrongen: slechte leerprestaties zouden met name te wijten zijn aan een te lage inzet of het beperkte intellect van de leerlingen. Ze hebben hierin natuurlijk niet altijd ongelijk. Een dergelijke opvatting leidt er nogal eens toe dat aan onderwijsevaluatie (programma-evaluatie) door leerlingen, met de bedoeling sterke en zwakke punten van het onderwijsprogramma te inventariseren, weinig belang wordt gehecht.

In scholen waar onderwijsevaluatie wel plaatsvindt, blijkt het in toenemende mate naar waarde geschat te worden.

10.2 Meten, normeren, waarderen en beslissen

Aan evalueren of beoordelen van leerprestaties zijn een viertal aspecten te onderscheiden, namelijk meten, normeren, waarderen en beslissen.

Het meten betreft het zo precies mogelijk in kaart brengen van toetsprestaties, bijvoorbeeld het tellen van het aantal goed beantwoorde vragen en het aantal gemaakte fouten. De twee belangrijkste kwaliteitseisen voor dit meten zijn betrouwbaarheid (precisie) en validiteit (geldigheid).

Het belangrijkste punt bij het normeren betreft de bepaling van de grens tussen voldoende en onvoldoende. Deze grens kan achteraf vastgesteld worden op basis van de feitelijke prestaties van de leerlingen (relatief normeren), of vooraf op grond van de door de docent beoogde doelen (absoluut normeren).

Het resultaat van de meting wordt omgezet in een waardering, doorgaans een cijfer.

Aan waarderingen worden in de meeste gevallen beslissingen verbonden. Deze beslissingen kunnen de voortgang van leerlingen betreffen (zittenblij-

ven; slagen of zakken) of leiden tot het vaststellen van kennislacunes en remediëring hiervan. Beslissingsregels geven aan welk gewicht moet worden toegekend aan verschillende toetsen en welke compensatiemogelijkheden er zijn. Bij grensgevallen, waarin sprake is van geringe verschillen in meetresultaat en waardering, kunnen de genomen beslissingen evenwel sterk verschillen, bijvoorbeeld bij een 5* of een 6.

Het spreekt vanzelf dat goede beslissingen gebaseerd moeten zijn op een verantwoorde normering en waardering en dat deze op hun beurt moeten berusten op nauwkeurige metingen.

10.3 Gebruikswijzen van beoordelingsresultaten

Evaluatieresultaten vormen de basis voor het nemen van uiteenlopende maatregelen, bijvoorbeeld een moeilijk onderdeel van de stof nog eens uitleggen, een didactische aanpak voor een hele klas wijzigen, of beslissen dat een leerling moet doubleren. Op school zijn de twee volgende gebruikswijzen gangbaar:

- tussentijdse feedback op leerprestaties;
- selectief gebruik.

Tussentijdse feedback op studieprestaties

Beoordeling met het doel leerlingen informatie te verschaffen om het eigen leerproces 'bij te sturen', heet in de literatuur *formatief*. *Formatieve evaluatie*, waarbij aan de hand van een toetsuitslag wordt nagegaan waar de lacunes in het kennen en kunnen zitten, vindt plaats als het onderwijs nog niet is afgesloten. Leerlingen kunnen vervolgens doelgericht gaan werken aan het verbeteren van hun zwakke punten. Strikt genomen mogen resultaten op een *formatieve toets* niet meetellen voor het rapportcijfer.

De effectieve docent telt niet alleen het aantal fouten, maar kijkt ook naar de soort fouten en trekt na hoe en waarom de fouten zijn gemaakt. Is de fout een *denkfout*, een *vergissing*, of is de fout gemaakt door een tekort aan informatie? Deze soorten fouten behoeven in de diverse vakken of vakkengroepen nadere specificatie, bijvoorbeeld bij talen: gebrek aan woordenkennis, spel-fouten, grammaticale fouten, slordigheidsfouten etcetera.

Uit de gemaakte fouten kan een docent vaak destilleren aan welke kennis of vaardigheid het schort of waar het denkproces bij de leerling(en) is gestagneerd of ontspoord. Soms kan een docent hier pas achter komen tijdens een nabespreking van de geleverde prestaties met de leerlingen (klassikaal, in groepjes of individueel).

Formatieve zelftoetsen (met terugkoppelingsinformatie) zijn een belangrijk hulpmiddel bij zelfstandig leren. De resultaten bespreekt de leerling in principe 'met zichzelf' en hij bepaalt zelf op welke wijze hij eventuele lacunes zal opvullen.

Selectief gebruik

Selectieve beoordeling, waarbij beslissingen over voldoende of onvoldoende, zittenblijven of overgaan, zakken of slagen in het geding zijn, heet ook wel summatieve evaluatie.

De betrouwbaarheid en validiteit van summatieve toetsen moeten voldoen aan hogere eisen, naarmate de consequenties die aan de resultaten worden verbonden ingrijpender zijn. De maatschappelijke en persoonlijke consequenties van bijvoorbeeld eindexamenbeslissingen zijn zo groot, dat naar de hoogste graad van objectiviteit moet worden gestreefd.

Onderwijsevaluatie (programma-evaluatie)

Toetsresultaten kunnen vanzelfsprekend ook aanwijzingen geven voor sterke en minder sterke punten van het voorafgaande onderwijs.

10.4 Middelen voor evaluatie van leerprestaties

De bekendste evaluatiemiddelen voor de evaluatie van cognitieve leerresultaten zijn:

- mondelinge toets;
- schriftelijke toets (repetitie, proefwerk, examen), waarbij alle mogelijke vraagvormen kunnen worden gebruikt.

Op school of tijdens het huiswerk moeten leerlingen opdrachten uitvoeren. Voor wiskunde zijn het sommen, bij Nederlands (en ook veel andere vakken) spreekbeurten, bij biologie, natuur- en scheikunde practicumopdrachten, bij verzorging kookopdrachten (bijvoorbeeld een taart). In de opdrachtensfeer ligt een rijke voorraad evaluatiemiddelen gereed, die met name geschikt zijn om, naast het toepassen van leerstof, het zelfstandig en samenwerkend leren te bevorderen.

Schriftelijke opdrachten:

- vragen opstellen of stellingen formuleren over een tekst, een probleem of een boek;
- brieven opstellen (bijvoorbeeld een ingezonden brief voor de krant);
- pamflet schrijven;
- vragenlijst opstellen.

Presentaties:

- spreekbeurt (eventueel met gebruikmaking van de overheadprojector);
- demonstratie;
- forumdiscussie;
- muurkrant, affiche, collage;
- audio-visuele presentaties (video-opnames, geluidsopnames, fotoseries).

Spelvormen:

- rollenspel, simulatie;
- puzzels, kwartetten;
- kort geding;
- quiz, cabaret.

Actie-opdrachten:

- interviewen;
- tentoonstelling opzetten;
- excursie organiseren;
- organiseren van een actie.

Rapportage van verrichte activiteiten:

- practicumverslag;
- logboekaantekeningen;
- tijdbestedingsspecificatie;
- verslag van bekeken audio-visuele programma's;
- verslag van leerervaringen.

Tijdens de uitvoering van opdrachten of bij het kennisnemen van de resultaten krijgt de docent informatie over algemene vaardigheden (schriftelijke en mondelinge uitdrukkingsvaardigheid, onderzoeksvaardigheden), het sociale gedrag, de interesses van leerlingen en de mate van zelfstandigheid die leerlingen hebben bereikt. De meeste van de genoemde opdrachsvormen zijn minder geschikt voor het nauwkeurig toetsen van cognitieve doelen.

10.4 Kies en evaluatiemiddel

Opdracht

- Kies en beschrijf voor onderstaande leerdoelen het meest geschikte evaluatiemiddel. Werk het gekozen evaluatiemiddel uit en beschrijf de toetsituatie (bijvoorbeeld het te gebruiken materiaal).
Bespreek jouw keuzes met je medestudenten.

Leerdoel	Evaluatiemiddel
1 De leerlingen kunnen de afstand tussen twee plaatsen berekenen, indien de schaal van de kaart gegeven is	
2 De leerlingen kunnen aan de hand van een voedingsmiddelentabel de calorische waarde van verschillende broodsoorten aangeven	

<p>3 De leerlingen kunnen verklaren waarom poreuze stoffen slechte warmtegeleiders zijn</p>	
<p>4 De leerlingen kunnen vooroordelen, die er in Nederland ten aanzien van etnische minderheden leven, analyseren</p>	
<p>5 De leerlingen zijn in staat een eenvoudige zakelijke brief op te stellen</p>	
<p>6 De leerlingen kunnen de werking van een fietspomp uitleggen</p>	
<p>7 De leerlingen kunnen de onbepaalde telwoorden: much, many, few en little op de juiste wijze gebruiken</p>	
<p>8 De leerlingen kunnen de poëzie van de Vijftigers karakteriseren</p>	

10.5 Kwaliteitseisen voor evaluatiemiddelen

De primaire kwaliteitseisen voor evaluatiemiddelen zijn validiteit en betrouwbaarheid.

10.5.1 Validiteit

Met de validiteit van een evaluatiemiddel wordt bedoeld, dat het middel in werkelijkheid moet meten wat het zegt te meten. Wanneer bijvoorbeeld in een rekentoets de opgaven zo zijn geformuleerd, dat ze hoge eisen stellen aan tekstbegrip, meet de toets niet alleen de rekenvaardigheid, maar ook de taalvaardigheid (tekstbegrip). Deze toets meet dus behalve datgene waarvoor hij bedoeld is ook nog andere kennis of vaardigheid. Een dergelijke toets is minder valide (= geldig).

Een valide evaluatiemiddel representeert op juiste en evenwichtige wijze de leerdoelen. Validiteit heeft zowel betrekking op het gedragsaspect (de beheersingsvormen) als op de leerstof (het inhoudsaspect) van de te evalueren leerdoelen.

Vormvaliditeit betreft de vraag: komen de beheersingsvormen van de vragen en/of opdrachten in het evaluatiemiddel overeen met de beoogde, respectievelijk uitgevoerde beheersingsvormen? Indien bijvoorbeeld een toets preten-

deert convergent denken te meten, dan moet er ook inderdaad op deze beheersingsvorm een appèl gedaan worden en niet alleen op reproductie van kennis.

De inhoudvaliditeit betreft de vraag of de vragen en/of opdrachten in het evaluatiemiddel een evenwichtige afspiegeling vormen van (een representatieve steekproef zijn uit) de bestudeerde leerstof. Het evaluatiemiddel moet in bevredigende mate de stof dekken. Een mondeling literatuurtentamen over vijftien boeken, waarbij slechts vragen over één boek worden gesteld, voldoet uiteraard volstrekt niet aan deze eis.

Met een blauwdruk van een toets (ook wel toetsmatrijs geheten) is na te gaan in hoeverre het evaluatiemiddel een verzameling cognitieve doelstellingen dekt naar inhoud en beheersingsvorm:

Beheersings- vorm	r	c	d	e
Hoofdzaken van de leerstof				
Totaal				

10.5.1a Vragen/opdrachten in relatie tot leerdoelen

Opdracht

- Formuleer bij de volgende vragen/opdrachten het bijpassende leerdoel. Denk eraan dat de beheersingsvorm van het leerdoel overeenkomt met die van de vraag/opdracht.

Voorbeeld:

Vraag: Hoeveel rente ontvangt iemand die gedurende drie jaar een kapitaal van f 3575,- uitzet tegen een rentepercentage van 5%? (Beheersingsvorm: convergent denken)

Leerdoel: De leerlingen kunnen met behulp van de jaarformule te ontvangen r berekenen (convergent denken).

	Beheersings- vorm
1 Hoe denk je over het gebruik van legbatterijen? Leerdoel:	<input type="checkbox"/>
2 Als natriumoxide reageert met water ontstaat er een hydroxide. Geef de ionenvergelijking die dan optreedt. Leerdoel:	<input type="checkbox"/>
3 Construeer een enkelvoudige zin, waarin het zinsdeel – mijn moeder – meewerkend voorwerp is. Leerdoel:	<input type="checkbox"/>
4 Plaats één van de bepaalde lidwoorden: der, die, das voor de volgende zelfstandige naamwoorden Hund Haus Schornstein Garten Tür Zimmer Fenster Boden Dach Zaun Korridor Keller Leerdoel:	<input type="checkbox"/>
5 Hoe zou je het gebruik van zonne-energie kunnen bevorderen? Leerdoel:	<input type="checkbox"/>
6 Leg het verschil uit tussen een sprookje en een fabel. Leerdoel:	<input type="checkbox"/>
7 Een auto kan met een volle tank, inhoud 40 liter benzine, 480 kilometer rijden. Een liter benzine kost f 1,80. Wat zijn de benzinekosten bij een rit over een afstand van 75 kilometer? Leerdoel:	<input type="checkbox"/>
8 Beschrijf een dag uit het leven van een drugsverslaafde. Leerdoel:	<input type="checkbox"/>
9 Zoek van ieder land de bijbehorende hoofdstad 1 Noorwegen a Stockholm 2 Polen b Berlijn 3 Italië c Warschau 4 Frankrijk d Oslo 5 Oostenrijk e Londen f Rome g Bern h Parijs i Sofia j Wenen Leerdoel:	<input type="checkbox"/>
10 Welk type bloedcellen komen voor in lymfe? a alleen bloedplaatjes b alleen witte bloedcellen c alleen rode bloedcellen d	<input type="checkbox"/>

Antwoordenlijst bij het beheersingsvormspel

Vragenset A:	Vragenset B:	Vragenset C:	Vragenset D:	Vragenset E:
A ₁ - c	B ₁ - r	C ₁ - r	D ₁ - e	E ₁ - d
A ₂ - c	B ₂ - r	C ₂ - c	D ₂ - c	E ₂ - r
A ₃ - r	B ₃ - d	C ₃ - d	D ₃ - d	E ₃ - c
A ₄ - r	B ₄ - r	C ₄ - d	D ₄ - r	E ₄ - c
A ₅ - e	B ₅ - r	C ₅ - e	D ₅ - r	E ₅ - e
A ₆ - d	B ₆ - e	C ₆ - c	D ₆ - r	E ₆ - d
A ₇ - e	B ₇ - c	C ₇ - d	D ₇ - c	E ₇ - c
A ₈ - e	B ₈ - e	C ₈ - e	D ₈ - d	E ₈ - e
A ₉ - r	B ₉ - c	C ₉ - e	D ₉ - r	E ₉ - r
A ₁₀ - r	B ₁₀ - d	C ₁₀ - d	D ₁₀ - d	E ₁₀ - e
A ₁₁ - c	B ₁₁ - d	C ₁₁ - r	D ₁₁ - c	E ₁₁ - d
A ₁₂ - d	B ₁₂ - c	C ₁₂ - c	D ₁₂ - c	E ₁₂ - r
A ₁₃ - c	B ₁₃ - d	C ₁₃ - c	D ₁₃ - d	E ₁₃ - r
A ₁₄ - d	B ₁₄ - c	C ₁₄ - e	D ₁₄ - e	E ₁₄ - e
A ₁₅ - e	B ₁₅ - e	C ₁₅ - r	D ₁₅ - e	E ₁₅ - d
A ₁₆ - d	B ₁₆ - e	C ₁₆ - r	D ₁₆ - e	E ₁₆ - c

10.5.2 Betrouwbaarheid

De leraar wiskunde die een proefwerk opstelt, streeft ernaar dat de uitslag zo precies mogelijk zal aangeven hoe het op dat moment met de (op het toetsonderwerp betrekking hebbende) wiskundekennis van zijn leerlingen is gesteld. Met andere woorden: hij probeert een betrouwbaar evaluatiemiddel te maken.

Met de betrouwbaarheid van een evaluatiemiddel wordt bedoeld de mate van nauwkeurigheid, waarmee het leerresultaten meet. Hoge betrouwbaarheid impliceert weinig meetfouten. Wanneer een toetsing herhaald wordt, op een ander tijdstip of met een andere beoordelaar, moet bij benadering hetzelfde resultaat worden geboekt. (Hierom heet betrouwbaarheid ook wel reproduceerbaarheid.) De resultaten van een meting mogen niet (teveel) beïnvloed worden door storende invloeden, 'ruis'. Een leerling mag niet ten onrecht een voldoende of een onvoldoende krijgen.

Bronnen van meetfouten vinden hun oorsprong in de vormgeving van het evaluatiemiddel, de afnamesituatie en de nakijkprocedure.

a *Kwaliteit van het evaluatiemiddel*

De twee belangrijkste betrouwbaarheidsverlagende factoren in toetsen zijn de meerduidige vragen/opdrachten en te weinig toetsvragen, waardoor subjectieve interpretaties en pech of geluk een te grote rol gaan spelen. De betrouwbaarheid van een toets is hoger naarmate deze meer vragen bevat en de formulering van de vragen eenduidiger is. De vragen mogen niet te gemakkelijk, maar ook niet te moeilijk zijn. De betrouwbaarheid van de toetsing neemt toe, naarmate vaker een toetsing plaatsvindt en de resultaten hiervan worden gemiddeld.

b Afnamesituatie

Als leerlingen te weinig tijd krijgen om de gestelde vragen/opdrachten op hun gemak te kunnen maken, komt dit de betrouwbaarheid van de toetsing niet ten goede. Ook omgevingsfactoren zijn van belang. Lawaai of een te hoge of te lage temperatuur verlagen de betrouwbaarheid.

Waar leerlingen kunnen spieken of andere fraude plegen verhoogt dit (misschien) wel hun cijfer, maar niet de betrouwbaarheid van de meting.

Klachten over de afnamesituatie bij examens

'Surveillanten zorgden op een school in Weert voor irritatie: ze knipten hun nagels, praatten onderling en hadden krakende schoenen. In Coevorden, Purmerend en Driebergen hadden leerlingen last van een grasmaaier. Op een school in Amersfoort hadden examenkandidaten last van jongerejaars die op het schoolplein stonden te lachen om een paard dat luidruchtig gedekt werd. Leerlingen van een school in Haarlem grepen massaal de telefoon omdat ze problemen hadden met het krakkemikkige meubilair. Een tafeltje stortte zelfs in tijdens een examen. (...) Scholieren in Gouda werden tijdens het examen getroffen door een muggenplaag. (...) Docenten gingen de insecten met spuitbussen te lijf (...). Een jongen op een Rotterdamse school zat twintig minuten opgesloten op het toilet, omdat de deur geblokkeerd was.'

Bron: *NRC Handelsblad*, 29 mei, 1996.

c Nakijkprocedure

Er is vaak onderzoek gedaan naar de correlaties tussen beoordelingen van proefwerken met open vragen of werkstukken van een groep leerlingen die door verschillende, onafhankelijk van elkaar werkende, competente beoordelaars (docenten) zijn nagekeken. Het ging hierbij met andere woorden om de intersubjectieve overeenstemming. Het blijkt dat deze overeenstemming doorgaans behoorlijk laag is. De beoordelaarsvariabiliteit is groot. In de volgende paragraaf zijn de belangrijkste vertroebelende factoren die de docent inbrengt, op een rij gezet.

10.5.2.1 Beoordelaarseffecten

De beoordelaar is ook maar een mens

Als voormalig leraar Nederlands weet ik wat corrigeren is. Stapels werk heb ik mee naar huis genomen, opstellen, samenvattingen, scripties die ik met behulp van vakkennis, gezond verstand en frisse moed te lijf ging. Het correctiemodel zat in mijn hoofd. Waar vakkennis als een ingebrand keurmerk nog wel standhield, daalde tijdens soms nachtelijke uren het peil van het gezond verstand en de frisse moed tot het nulpunt. Van corrigeren, in de zin van 'fouten verbeteren' was in het begin nog wel iets aanwezig. Ik maakte aantekeningen in de kantlijn in de trant van 'tangconstructie', 'pleonastisch gebruik van hulpwerkwoorden', 'vermijd germanismen', dat soort voor mijn doen tamelijk opbouwende kritiek. Maar gaandeweg werden de aanwijzingen schimmiger, de strepen forser aangezet, de uitrope tekens feller totdat cynisme de overhand kreeg in toevoegingen als: 'hoe is het mogelijk?!', 'nee toch', 'hier staat echt onzin!'. Het soort commentaar waar leerlingen bij het terugkrijgen van hun werk waarschijnlijk ook weinig mee opschoten. Feedback waar niemand vet van werd. Meestal moesten die leerlingen het ontgelden die ik aan het eind van de nachtelijke correctietocht tegengekomen was.

Bron: Luijten, T. (1996). Het gebruik van correctiemodellen. In P. Leenheer, R.J. Simons, & J. Zuylen (Eds.). *Studiehuis in de steigers*, p. 45-58. Tilburg: Meso Consult, p. 46.

Diverse opvattingen, gewoontes en werkwijzen van beoordelaars verlagen de objectiviteit van de beoordelingen.

De belangrijkste beoordelaarseffecten zijn:

- a norminstabiliteit en persoonlijke beoordelingstendentie;
- b uitstralingseffect (of halo-effect);
- c opvattingseffect;
- d volgorde-effect;
- e contaminatie-effect.

Norminstabiliteit en persoonlijke beoordelingstendentie

Bij beoordelingen van leraren is meermalen geen sprake van constantie van beoordelingsnormen. In de praktijk past de docent de aan te leggen normen bewust of minder bewust aan aan het gemiddelde prestatieniveau van de groep leerlingen. De standaard van de beoordelaar kan tijdens het beoordelingsproces verschuiven.

Ook verschillen in persoonlijke beoordelingstendentie zorgen voor verschil in normstelling. Sommige leraren beoordelen gewoontegetrouw streng, andere mild. Sommige leraren hebben de neiging alle cijfers min of meer rond het gemiddelde te houden, andere durven meer extreem hoge en lage cijfers te geven.

Het maakt behoorlijk verschil als bijvoorbeeld een werkstuk beoordeeld wordt door een docent die meer gespreid cijfert, of door een die meer modaal cijfert. Leerlingen die een uitstekend werkstuk hebben geschreven zijn het

meest gebaat bij het eerste type beoordelaar; leerlingen die een slecht werkstuk hebben ingeleverd hebben geluk als ze de docent treffen die geen zeer lage cijfers uitdeelt.

Ook vermoeidheid kan een rol bij de beoordeling spelen. De eerste proefwerken van een grote stapel worden vaak nauwkeuriger of strenger nagekeken dan de overige.

Uitstralingseffect (of halo-effect)

Halo betekent letterlijk stralenkrans. Van een halo-effect is sprake als er een 'uitstraling' (storende invloed) uitgaat van bijkomstige kenmerken op de beoordelingsuitkomst, bijvoorbeeld:

- de leesbaarheid van het handschrift van de leerling;
 - de netheid van het ingeleverde werk;
 - spellingfouten of grammaticale fouten in bijvoorbeeld een proefwerk biologie;
 - de reputatie van de leerling;
 - sympathie- of antipathiegevoelens die de leraar jegens de leerling koestert.
- Dergelijke niet ter zake doende kenmerken zouden geen invloed op de waardering mogen hebben: de evaluatie moet zoveel mogelijk gebaseerd zijn op de relevante kenmerken van de geleverde prestatie.

Opvattingseffect

Het opvattingseffect betreft de storende invloed die uitgaat van verschillen in kennis en inzichten (opvattingen) die bij leraren kunnen bestaan bij het beoordelen, met name de verschillen in het belang dat een leraar toekent aan verschillende aspecten van de geleverde prestatie. Bekend is bijvoorbeeld dat bij het corrigeren van scripties verschil in mening over datgene, waardoor een goede scriptie zich kenmerkt, grote invloed heeft op de waardering. Enige van zulke verschillend gewaardeerde aspecten zijn: logische opbouw, taalgebruik, leesbaarheid, spelling, argumentatie etcetera.

Volgorde-effect

Het volgorde-effect ontstaat door de (onbedoelde) nawerking van eerdere beoordelingen op erop volgende. Na een aantal bijzonder slechte repetities is een redelijke prestatie een verademing voor de leraar. De voorliggende prestatie contrasteert met de eerder nagekeken repetities. In plaats van een 'redelijke' 6 is hij nu misschien geneigd om maar een 7 of nog hoger te geven. Het omgekeerde kan zich voordoen als eerst een aantal goede producten achter elkaar nagekeken is, waarna er vervolgens een middelmatig of slecht product komt, dat schril afsteekt tegen de voorafgaande en dan naar verhouding te laag kan worden gewaardeerd.

Contaminatie-effect

Bij het evalueren kunnen oneigenlijke oogmerken van een docent de objectiviteit van de beoordeling 'besmetten'. (Contaminatie betekent letterlijk be-

smetting.) Enige voorbeelden. Er zijn leraren die een goed figuur willen slaan, door zoveel mogelijk leerlingen met mooie cijfers te laten overgaan. Er zijn leraren die de status van hun vak willen verhogen en daarom laag gaan cijferen. Er zijn leraren die niet in conflict willen komen met een collega wiens cijfers gemiddeld een paar punten lager liggen; derhalve gaan ze ook maar strenger cijferen. Het komt voor dat leraren, om een leerling niet te ontmoedigen, te hoge cijfers geven en er zijn leraren die lage cijfers geven om leerlingen tot betere prestaties aan te sporen of te koeioneren.

10.5.2.1a Gedachten die de betrouwbaarheid kunnen schaden

Tijdens het beoordelen van leerlingprestaties kunnen de leraar allerlei gedachten 'door het hoofd spelen', die aan de betrouwbaarheid van de beoordeling afbreuk zouden kunnen doen.

Opdracht

- Geef van onderstaande gedachten aan op welke foutenbron(nen) ze betrekking hebben.
 - 1 'Ho, wacht even. Dat heb ik vorig jaar geloof ik steeds goed gerekend.'
 - 2 'Daar heb je Piet de lulak weer.'
 - 3 'Hè, eindelijk eens iemand die mijn uitleg prima begrepen heeft.'
 - 4 'Ik snap niet hoe mijn collega vorig jaar in deze klas zulke hoge cijfers voor Frans kon hebben.'
 - 5 'Nou, nou, zulke kanjers van taalfouten hoef je ook in een natuurkundeproefwerk niet te accepteren.'
 - 6 'Toch een leuke meid, die Jolanda.'
 - 7 'Mijn vak moet ook niet al te moeilijk overkomen, anders kiest straks niemand het in het examenpakket.'
 - 8 'Goh, 't is bijna half twaalf. Kom, snel even de laatste tien proefwerken er doorjagen.'
 - 9 'Veel zittenblijvers zou wel eens een aantal leraarlessen kunnen kosten...'
 - 10 'Deze klas is zo te zien aanmerkelijk beter dan die van het vorige jaar. Ik kan daarom ook wel wat meer van ze eisen.'
 - 11 'Ik kan toch niet iedereen een onvoldoende geven.'
 - 12 'Van zo'n handschrift word ik niet goed.'

10.5.2.1b Spieken

Opdracht

- Geef een aantal voorbeelden van spiekmethoden.
- Welke spiekmethode die je wel eens hebt meegemaakt of waar je wel eens van gehoord hebt, vind je de meest originele of meest vernuftige?
- Veel leerlingen en sommige leraren zien spieken als een soort spel of sport. Als je verliest, heb je pech gehad. Geef jouw mening als aanstaande docent over spieken.
- Bedenk een aantal manieren hoe je in het onderwijs spieken zou kunnen voorkomen.
- Wat dient er volgens jou te gebeuren, wanneer een leerling tijdens een schoolonderzoek op overduidelijke wijze op spieken wordt betrapt?

10.5.2.2 *Wenken voor betrouwbaar evalueren*

Onvoldoende kwaliteit van het evaluatiemiddel, storende invloeden in de afnamesituatie en subjectiviteit in de beoordeling zorgen voor meetfouten. Verschillende maatregelen kunnen de betrouwbaarheid van de metingen van leerprestaties verhogen.

Kwaliteit van het evaluatiemiddel

- Formuleer de vragen/opdrachten ondubbelzinnig (zodat over het verwachte antwoord geen misverstand kan rijzen).
- Zorg ervoor dat de toets uit een voldoende aantal vragen/opdrachten bestaat om de leerstof te bestrijken. De betrouwbaarheid neemt toe met de toetslengte. Immers: naarmate de steekproef van vragen representatiever is voor wat geleerd is, is de kans dat deze steekproef voor de leerling gelukkig of ongelukkig uitvalt kleiner. (Een toets mag ook weer niet veel te lang zijn, omdat dan de vermoeidheid toeslaat.)
- Het antwoord op een vraag moet niet bij een andere vraag te vinden zijn.
- Rangschik de vragen zoveel mogelijk in volgorde van oplopende moeilijkheid.
- Plaats de vragen per onderwerp bijeen.

Afnamesituatie

- Zorg ervoor dat de leerlingen tenminste enige ervaring hebben met het type vragen/opdrachten en weten welke de beste antwoordstrategie is.
- Stel de leerlingen vooraf op de hoogte van de wijze van correctie en de eventuele weging van de vragen/ opdrachten bij de beoordeling.
- Maak bekend welke hulpmiddelen (bijvoorbeeld een atlas) zijn toegestaan.
- Geef de leerlingen voldoende tijd om de opgaven te maken en stel ze vooraf op de hoogte van de beschikbare tijd. (Haast bij de beantwoording verlaagt de betrouwbaarheid.)
- Waarschuw wanneer de tijd bijna om is (bijvoorbeeld zowel een kwartier als vijf minuten voor het eind).
- Zorg ervoor dat de leerlingen niet worden gestoord.
- Let op dat afkijken of spieken niet plaatsvindt.

Beoordeling (nakijkprocedure)

- Omschrijf scoringsvoorschriften waarin de aspecten van de geleverde prestaties zijn gewogen met aantallen punten (inperking van de vrijheid van de beoordelaar verhoogt de objectiviteit):
- Kijk de geleverde leerlingprestaties per vraag (of onderdeel of aspect) na.
- Schakel zo mogelijk verschillende beoordelaars voor hetzelfde werk in en middel de uitkomsten.
- De beoordelaars moeten onafhankelijk van elkaar werken. De eerste beoordelaar moet – om de tweede beoordelaar niet te beïnvloeden – geen kanttekeningen in het werk van de leerling schrijven.

- Geef leerlingen gelegenheid hun oordeel te geven over de kwaliteit van de vragen/opdrachten. Stel op grond van het geleverde commentaar vast welke vragen of vraagonderdelen voor verwijdering in aanmerking komen (bijvoorbeeld dubbelzinnige vragen) en baseer de beoordeling op de overgebleven vragen/opdrachten.

Opmerking:

Betrouwbaarheid is een noodzakelijke voorwaarde voor de validiteit van een evaluatiemiddel. Een onbetrouwbare toetsing kan niet valide zijn, omdat deze naast de bedoelde kennis en vaardigheden storende invloeden ('ruis') meet. Echter: hoge betrouwbaarheid is geen garantie voor validiteit. Een voorbeeld. Een toets voor geschiedenis over een bepaald tijdvak waarin de vragen over jaartallen gaan, kan deze kennis van jaartallen best precies meten (betrouwbaar zijn). Als het doel van de geschiedenislessen echter niet het reproduceren van jaartallen was, maar inzicht in de betekenis van het tijdvak voor het heden, is de toets zeker niet valide. Een toets kan dus betrouwbaar zijn zonder valide te zijn.

Literatuur

Bruyne, H.C.D. de (1983). *Evalueren in de klas*. Amsterdam: Van Goor.

10.6 Overwegingen bij de keuze van een evaluatiemiddel

De belangrijkste overwegingen die bij de keuze van een evaluatiemiddel een rol spelen, zijn samengevat in de volgende vragen:

- moeten 'subjectieve' evaluatiemiddelen uit de school geweerd worden?
- wie treedt als beoordelaar op?
- moeten zich één of meer beoordelaars over de prestaties buigen?
- moet de beoordeling op meer globale dan wel meer analytische wijze plaatsvinden?
- komen naast individuele ook groepsprestaties voor beoordeling in aanmerking?

Antwoorden op de gestelde vragen komen hierna achtereenvolgens aan de orde.

Moeten 'subjectieve' evaluatiemiddelen uit de school worden geweerd?

Om fair te zijn tegenover leerlingen is het nodig dat de beoordeling van leerprestaties zo objectief mogelijk gebeurt. Deze terechte opvatting heeft aan de basis gestaan van de enorme vlucht die 'multiple-choice' examens in ons onderwijs genomen hebben. Ze elimineren de persoonlijke invloed van de beoordelaar bij het nakijken.

Naast objectiviteit en betrouwbaarheid is ook de relevantie van de meting van belang. Helaas staat de mate van levensechtheid van een evaluatiemiddel soms op gespannen voet met de betrouwbaarheid. Evaluatiemiddelen die niet

zo betrouwbaar zijn (bijvoorbeeld: rollenspel, verslag van leerervaringen, audiovisuele presentatie), hoeven om deze reden niet uit de school geweerd worden. Zodra een dergelijke 'subjectieve' meting samen met andere metingen verrekend wordt in een eindoordeel, verbleekt de subjectiviteit die verbonden is aan zo'n enkelvoudig oordeel.

Wie treedt op als beoordelaar?

Als beoordelaar in de school kunnen optreden de docent, de medeleerlingen en de leerling zelf. Deze personen kunnen afzonderlijk of in combinaties evalueren.

In de meeste gevallen treedt de docent op als beoordelaar. Bij selectief gebruikte beoordelingsresultaten is dit noodzakelijk. Bij de formatieve beoordeling ligt het anders. Het is gebleken dat leerlingen zich een redelijk tot goed oordeel kunnen vormen van het niveau van hun medeleerlingen. Ze weten soms zelfs beter dan de leraar wat een medeleerling waard is (vooral waar het de geleverde inzet betreft).

Spreekbeurten, demonstraties, rollenspelen, simulaties, producten van projectonderwijs en dergelijke lenen zich goed voor beoordeling door medeleerlingen. Omdat de beoordeling door meer dan één leerling kan plaatsvinden, kan het gemiddelde eindoordeel behoorlijk betrouwbaar zijn. Dit eindoordeel kan verrekend worden (al dan niet gewogen) met de waardering van de docent. In de praktijk blijkt, dat er meestal sprake is van een hoge mate van overeenstemming tussen het docentoordeel en de formatieve leerlingoordeelen.

Zelfevaluatie heeft in de school nog weinig wortel geschoten. Toch is een belangrijk doel van zelfstandig leren: zijn eigen prestaties realistisch kunnen inschatten en aan de juiste oorzaken toeschrijven, weten wat je wel of (nog) niet weet en kunt. Het belang van attributie en metacognitie voor zelfstandig leren kan nauwelijks overschat worden.

Het zou helemaal niet gek zijn de leerlingen zo nu en dan te vragen zichzelf, voordat het gemaakte werk door de leraar is nagekeken en teruggegeven, een cijfer toe te kennen, een soort prognosecijfer. Waar sprake is van een groot verschil tussen het prognosecijfer en het 'echte' cijfer zou een gesprekje hierover tussen leraar en leerling kunnen plaatsvinden. Dit zou de zelfkennis van de leerlingen kunnen vergroten.

Zelfevaluatie kan in complexe beoordelingssituaties een bijdrage leveren aan de oordeelsvorming, bijvoorbeeld bij de beoordeling van een stage.

Een voorbeeld uit de lerarenopleiding. Een beoordelingsgesprek over een schoolstage zou prima plaats kunnen vinden aan de hand van een korte beschrijving door de student van zijn eigen oordeel over zijn wijze van omgang met de leerlingen, zijn omgang met aanstaande collega's, didactische vaardigheden, creativiteit, inzet, vakkennis en een samenvatting van zijn sterke en zwakke punten. Het zelfoordeel van de student kan aangevuld of gecorrigeerd worden door oordelen van medestudenten en de stage-begeleider.

Hoeveel beoordelaars?

De objectiviteit van de beoordeling van geleverde prestaties is gediend met het middelen van oordelen, die afzonderlijk door verschillende beoordelaars zijn toegekend. Hoe meer beoordelaars hun waardering geven, hoe beter de intersubjectiviteit gewaarborgd is. (Uiteraard hebben we hierbij competente beoordelaars op het oog. In diverse situaties kunnen dit ook leerlingen zijn.) In de schoolpraktijk komt het weinig voor dat proefwerken door meer dan één leraar worden nagekeken en dat is bij de gegeven hoge taakbelasting alleszins begrijpelijk. Examenwerk wordt doorgaans wel dubbel beoordeeld.

Moet de beoordeling op meer globale dan wel meer analytische wijze plaatsvinden?

Leerprestaties kunnen in principe op twee manieren geëvalueerd worden: globaal of analytisch.

Een globale beoordelaar evalueert het door een leerling gemaakte product als geheel. De afwegingen die aan het eindoordeel voorafgaan spelen zich af in het hoofd van de beoordelaar.

Bij de analytische manier worden allerlei aspecten afzonderlijk beoordeeld, ofwel aan de hand van een lijstje met aandachtspunten of op grond van een antwoordmodel (met scoringsvoorschrift).

De globale aanpak laat meer ruimte voor subjectieve invloeden dan de analytische methode. Naarmate de vrije interpretatieruimte toeneemt, is het oordeel onbetrouwbaarder. Subjectiviteit wordt tegengegaan door de hand te houden aan uitgewerkte beoordelingsschema's.

Docenten stellen zeer gedetailleerde beoordelingsschema's niet op prijs: het aantal verschillende beoordelingsaspecten moet gemakkelijk te overzien zijn.

Komen naast individuele ook groepsprestaties voor beoordeling in aanmerking?

Vrijwel altijd worden leerlingen beoordeeld op individueel geleverde prestaties. Dit neemt niet weg dat men zich ook situaties kan voorstellen waarin leerlingen zo nauw samengewerkt hebben aan een product (bijvoorbeeld een muurkrant, een audiovisuele presentatie) dat de individuele bijdragen moeilijk kunnen worden getraceerd. Veelal krijgen bij de beoordeling van het groepsproduct alle deelnemers dan dezelfde waardering. In gevallen waarbij sommige leerlingen zich het vuur uit de sloffen hebben gelopen, terwijl andere een minimale bijdrage hebben geleverd, kan de gelijke toedeling tot onrechtvaardigheden, frustraties en leedvermaak leiden. Een manier om hieraan het hoofd te bieden is de volgende. De docent kent de groepsprestatie een cijfer toe. Dit getal wordt met het aantal groepsleden vermenigvuldigd. Vervolgens geven alle leerlingen elkaar een cijfer, ze verdelen als het ware de 'pot'. Het gemiddelde van de elkaar toegekende cijfers moet uitkomen op het cijfer, dat de docent voor de groepsprestatie heeft gegeven. Als de leerlingen het niet met elkaar eens kunnen worden, of als er conflicten ontstaan, bemiddelt, sust en beslist de docent.

10.7 Mondelinge toetsing

In een mondelinge toets kan de docent de toetsing aanpassen aan de individuele leerling, door op bepaalde punten dieper door te vragen. De vragen zijn sneller te beantwoorden dan schriftelijke open vragen. In korte tijd kunnen dus veel vragen worden gesteld. Maar schriftelijke toetsingen met gesloten vragen zijn doorgaans objectiever, betrouwbaarder en efficiënter (bij grotere aantallen leerlingen) dan mondelinge. Een vergelijking tussen schriftelijke en mondelinge toetsing valt op de meeste punten in het nadeel van de laatstgenoemde uit.

In de interactie tussen leerling en docent spelen affectieve aspecten een rol: sympathieën, antipathieën, zenuwen. Als een docent zich ergert aan de paar-denstaart van een mannelijke leerling, komt de objectiviteit in gevaar. Het blijkt moeilijk te zijn zich objectief op te stellen tegenover kandidaten die royaal voorzien zijn van uiterlijk schoon. Zenuwachtig gedrag van leerlingen kan bij de docent de indruk wekken, dat ze er niet veel aan hebben gedaan. Verder speelt mee dat de ondervrager vaak al na korte tijd een oordeel heeft gevormd over de competentie van de leerling en hier niet meer vanaf is te brengen.

De toets kan voor de ene leerling heel anders uitvallen dan voor de andere. De docent zal, doordat leerlingen verschillend op de vragen reageren, gauw verschillende vragen stellen. Leerlingen kunnen meesturen (bijvoorbeeld door stokpaardjes van de leraar aan te snijden) en de docent kan voorbereide vragen wel eens vergeten. Leerlingen die de mondelinge toets hebben afgelegd, vertellen aan medeleerlingen 'hoe het ging' en welke vragen werden gesteld. De docent weet dit en stelt daarom per toets verschillende vragen. Een en ander kan tot gevolg hebben dat de spreiding van de vragen over de stof en de moeilijkheidsgraad van de vragen tussen de verschillende mondelinge toetsingen behoorlijk gaan uiteenlopen. (Uit onderzoek blijkt dat in de loop van een lange examendag de toegekende cijfers hoger worden.)

Mondelinge toetsen kosten veel docententijd. Soms worden leerlingen in groepjes aan een mondelinge toets onderworpen. Een pluspunt hiervan is de tijdwinst, maar de kwaliteit van mondelinge toetsen komt met zo'n werkwijze nog sterker in het gedrang.

Wanneer rekening wordt gehouden met de volgende adviezen, zal dit de betrouwbaarheid en validiteit van een mondelinge toets verhogen.

- stel de leerling op z'n gemak en begin met een paar gemakkelijke vragen;
- stel de vragen aan de hand van een lijst met (kern)begrippen;
- zorg ervoor, dat de leerlingen tijdens de toetsafnames niet in contact met elkaar kunnen komen. Wanneer dit niet mogelijk is moeten de vragen per leerling worden gevarieerd, hetgeen de onderlinge vergelijkbaarheid van de toetsingen echter niet ten goede komt;
- spreid de vragen evenwichtig over de stof. Veel korte vragen is beter dan weinig lange vragen;

- noteer meteen hoe de leerling de vragen beantwoordde (bijvoorbeeld met +, – en \pm achter de trefwoorden);
- vraag bij vage antwoorden door;
- zorg ervoor, dat je door extra toelichting niet het antwoord ‘weggeeft’;
- wijs er bij omslachtige antwoorden op dat bondigheid op prijs wordt gesteld. Maak met bluffen korte metten;
- laat je niet verleiden stokpaardjes te berijden;
- houd de tijd in de gaten.

10.8 Open vragen

Kenmerkend voor open vragen is, dat de leerling zelf het antwoord moet formuleren. Open vragen worden door leraren en leerlingen vaak als eerlijker en zinvoller ervaren dan gesloten vragen. Het is immers niet (of moeilijk) mogelijk het antwoord te raden en in de formulering van het antwoord kun je iets van jezelf kwijt. Open vragen kunnen verschillende beheersingsnormen en beheersingsniveaus aanboren; er kan gevraagd worden naar feitenkennis, naar het toepassen van kennis in praktijksituaties en naar complexe redeneringen.

De volgende soorten open vragen komen voor:

- Invulvragen, bijvoorbeeld:
De hoofdstad van Portugal is.....
De..... is de hoogste berg ter wereld.
- Korte open vragen (open vragen met korte antwoorden), bijvoorbeeld:
Aan welke rivier ligt Zwolle?
Wat is het verschil tussen een producent en een consument?
- Lange open vragen (open vragen met lange antwoorden), bijvoorbeeld:
Noem vier oorzaken waarom de agrarische beroepsbevolking in ons land in deze eeuw zo sterk is teruggelopen.

Korte open vragen beginnen vaak met:

Wie...
Wat...
Waar...
Welke...
Wanneer...

Lange open vragen beginnen doorgaans als volgt:

Vergelijk...
Geef drie oorzaken aan van...
Geef drie voorbeelden van...
Geef een samenvatting (in ongeveer tien regels) van...
Geef een overzicht van voor- en nadelen van...

Lange open vragen heten ook wel essay- of opstelvragen.

10.8.1 Richtlijnen voor het opstellen van open vragen

- Formuleer de vragen ondubbelzinnig. Alle leerlingen moeten de vraag precies zo interpreteren als de steller van de vraag heeft bedoeld. Laat eventueel een collega de vragen controleren op hun eenduidigheid.
- Omschrijf de taak van de leerling zo volledig mogelijk. De leerling moet weten hoe uitgebreid of diepgaand het verlangde antwoord moet zijn. Geef bij lange open vragen aan hoe lang het antwoord ongeveer moet zijn en hoeveel argumenten, voorbeelden, oorzaken, voor- en nadelen, aspecten etcetera de leerling moet formuleren. Vragen van het type 'Vertel wat je weet over...' zijn slechte toetsvragen. De opsteller van zo'n vraag moet zich niet verbazen als het antwoord bestaat uit een overloze ratjetoe van zaken die direct, zijdelings of in het geheel niet met het onderwerp te maken hebben.
- Controleer of de gestelde vragen (zoveel mogelijk) onafhankelijk van elkaar zijn.
- Splits een complexe vraag op in deelvragen.
- Check of de gestelde vragen een beroep doen op de beoogde beheersingsvormen (vormvaliditeit). Invulvragen en korte open vragen doen meestal een beroep op parate kennis, lange open vragen zijn in principe geschikt voor alle beheersingsvormen in het cognitieve gebied.
- Ga na of de gestelde vragen representatief zijn voor de omvang en de moeilijkheidsgraad van de bestudeerde leerstof (inhoudsvaliditeit).

10.8.1 De rol van taalvaardigheid bij examenopgaven met open vragen

Een probleem waar we bij elk examen met open vragen tegen aanlopen, is de vraag of je met deze vraagvorm in staat bent voor honderd procent de aanwezige kennis bij de leerlingen te achterhalen. In hoeverre speelt de taalvaardigheid van de leerlingen bij het maken van de toets een rol?

Dat taalvaardigheid meespeelt is onontkoombaar, maar het is ontoelaatbaar dat een toets geschiedenis ontaardt in een toets tekstverklaring, waarbij pure geschiedkundige kennis en inzicht een ondergeschikte rol spelen.

Opdracht

- Beoordeel de open vragen van een schoolonderzoek of een examentoets op de rol die de taalvaardigheid van de leerlingen speelt bij het maken van de vragen.
- Selecteer die vragen, waarvan je vindt dat de taalvaardigheid een ongewenst grote invloed heeft. Geef een duidelijke argumentatie.
- Formuleer voor de betreffende vragen één of meer voorstellen ter verbetering.

10.8.2 Nakijken van open vragen

Correctievoorschriften expliciteren de beoordelingstaak en beperken daarmee de beoordelaarsvrijheid van de docent.

Een goed correctievoorschrift bevat de volgende elementen:

- antwoordsleutel: een weergave van goede, gedeeltelijk goede en indien nodig pertinent foute antwoorden;

- scoringsvoorschrift (puntentelling): per vraag of vraagonderdeel is aangegeven hoeveel punten dit opbrengt per goed antwoord en hoeveel punten voor foute of gedeeltelijk foute antwoorden moeten worden afgetrokken;
- waardering van voor de vraagstelling minder relevante fouten (bijvoorbeeld grammaticale- en spelfouten in een proefwerk natuurkunde).

Met behulp van de omzettingstabel (omrekenformule) kan de docent de score weergeven in een cijfer.

Een mavo-examen voor Nederlands (tekstbegrip) bevatte de tekst: 'De visrijke Noordzee en de overbevissing'. De kandidaten moesten hierover twaalf vragen beantwoorden. Een vraag hieruit:

De schrijver geeft twee voorbeelden van relaties tussen vissoorten, namelijk tussen kabeljauw en haring en tussen makreel en gelegenheidsoorten. Beschrijf een mogelijke relatie tussen kabeljauw en haring als volgt:

"Als kabeljauw... dan..."

Je mag geen hele zinnen uit de tekst gebruiken.

Bij deze vraag hoorde het volgende antwoordmodel:

Uit een goed antwoord moet blijken dat er een oorzakelijk verband bestaat tussen het gedijen van de kabeljauw en de achteruitgang van de haringstand.

Bijvoorbeeld:

	Scorepunten
Als de kabeljauw met rust gelaten wordt dan (is het heel goed mogelijk dat) (of zal) er minder haring overblijft/overblijven (regels 87-90)	5 punten
Antwoorden die neerkomen op: als het de kabeljauw slecht gaat, gaat het de haring goed	geen punten
Als is opgenomen: 'dat het resultaat van een dergelijke maatregel economisch negatief is'	geen punten
Is slechts een van de beide elementen goed dan:	geen punten
Voor irrelevante of onjuiste toevoegingen 2 punten aftrekken.	

Antwoordsleutel en scoringsvoorschrift schakelen de subjectiviteit van de beoordelaar voor een flink deel uit. Soms wordt een correctievoorschrift gecompleteerd door een beoordelaarsinstructie, waarin de volgende aanbevelingen kunnen staan:

- beoordeel de antwoordpapieren per vraag en niet per leerling: de concentratie op het antwoord neemt hierdoor toe en normverschuiving treedt minder gauw op;
- probeer bij het nakijken niet tot je door te laten dringen van welke leerling het werk is;
- geef voor overtollige informatie (correcte, maar voor de vraagstelling irrelevante gegevens) geen bonuspunten;
- laat bij toetsen die beslissen over zakken of slagen ook een collega (onafhankelijk) de vragen corrigeren en aanvaard bij verschil van waardering het gemiddelde.

10.8.3 Plus- en minpunten van open vragen

Bij het beantwoorden van open vragen moeten leerlingen hun kennis over het gevraagde zelf verwoorden in een coherent 'verhaal'. Bij gesloten vragen volstaat dat ze het juiste antwoord herkennen. De verwachting van een toets met open vragen lokt bij leerlingen dan ook eerder construerend leren uit, dan een toets met gesloten vragen.

Een volgend sterk punt van de essay-vraag is, dat deze gelegenheid biedt de opbouw van de gevolgde redenering te beoordelen. De leraar kan vaak zien op welk punt de leerling eventueel 'in de fout' ging.

Korte en vooral lange open vragen meten tevens een stuk taalvaardigheid. De kans is aanwezig dat leerlingen die zich slecht uitdrukken, met een lager cijfer uit de bus komen dan ze op grond van hun werkelijke kennis van de leerstof verdienen. Het komt voor dat bij een foutief antwoord de leerling het juiste antwoord wel wist, maar het niet goed heeft kunnen verwoorden.

Een belangrijk nadeel van open vragen is, dat het nakijken speelruimte biedt voor subjectieve inzichten van de beoordelaar. De antwoorden op open vragen kennen gradaties: zij kunnen goed, gedeeltelijk goed of fout gerekend worden. Maar hoe waardeer je spel- en grammaticale fouten in andere vakken dan de talen? Hoe waardeer je juiste, maar irrelevante informatie? Wat doe je met slecht leesbare passages? Uit onderzoek komt naar voren dat een lang 'verhaal' meestal leidt tot een hoger cijfer dan een kort en bondig antwoord (met daarin ongeveer dezelfde kenniselementen). Allerlei beoordelaarseffecten kunnen hun storende invloed uitoefenen. De overeenstemming tussen leraren die dezelfde opstelvragen nakijken zonder antwoordsleutel en scoringsvoorschrift (puntentelling) blijkt vaak tegen te vallen.

Een laatste minpunt is dat in een essay-toets minder vragen over de stof gesteld kunnen worden dan wanneer gesloten vragen zouden zijn gebruikt, waardoor de geluksfactor een grotere rol gaat spelen (de dekking van de bestudeerde stof is kleiner).

10.9 Gesloten vragen

Bij gesloten vragen (of geprecodeerde vragen) zijn er een aantal antwoordmogelijkheden gegeven, waaruit de leerling de juiste moet kiezen. In tegenstelling tot open vragen meten zulke vragen dus niet onbedoeld tevens formulevaardigheid.

- Rangschikkingsvragen: de leerling moet een rangschikking opstellen op grond van een bepaald criterium, bijvoorbeeld chronologische volgorde.
Voorbeeld:
De volgend gebeurtenissen uit de Tweede Wereldoorlog staan in een verkeerde tijdsvolgorde. Geef de juiste volgorde met een cijfer (1-5) aan.

slag om Stalingrad	...
D-day	...
'Battle of Britain'	...
slag bij El-Alamein	...
aanval op Pearl Harbour	...

- Sorteervragen: de leerling moet juiste combinaties uit twee rijen bij elkaar plaatsen.

Voorbeeld:

Welke stad ligt in welk land?

1 Oslo	1 a.b.c.d.e.f.g.h	a Italië
2 Sint Petersburg	2 a.b.c.d.e.f.g.h	b België
3 Milaan	3 a.b.c.d.e.f.g.h	c Zwitserland
4 Barcelona	4 a.b.c.d.e.f.g.h	d Rusland
5 Genève	5 a.b.c.d.e.f.g.h	e Spanje
6 Dublin	6 a.b.c.d.e.f.g.h	f Noorwegen
7 Antwerpen	7 a.b.c.d.e.f.g.h	g Frankrijk
8 Lyon	8 a.b.c.d.e.f.g.h	h Ierland

- Meerkeuzevragen: de leerling moet uit verschillende gegeven alternatieven het enige juiste (of het beste) aankruisen.

Voorbeeld:

Een kaarsvlam blijft behalve door het kaarsvet, branden door

- a koolzuur
- b vuur
- c zuurstof
- d waterdamp

De meerkeuzevraag is opgebouwd uit een stam, een goed alternatief en (meestal drie) foute alternatieven. Deze laatste heten ook wel afleiders. Een meerkeuzevraag hoeft niet per se vier antwoordmogelijkheden te hebben: het kunnen er minder of meer zijn.

- Juist/onjuist-vragen: de leerling moet bij een gegeven uitspraak aangeven of deze juist of onjuist (goed of fout, waar of onwaar) is.

Voorbeeld:

Gegeven:

In het kortetermijngeheugen kunnen slechts een beperkt aantal informatie-eenheden opgeslagen worden

Vraag:

Het maximale aantal ligt doorgaans dichter bij de zeven dan bij de tien

juist/onjuist

De rangschikkingsvraag en de sorteervraag hebben gemeen dat ze voornamelijk een beroep doen op pure feitenkennis. De hogere beheersingsvormen vallen moeilijk binnen hun bereik. Dit nadeel is in mindere mate verbonden aan de meerkeuzevraag en de juist/onjuist vraag.

Stappen bij het maken van een toets met gesloten vragen

Een toets, die de kwalificatie 'goed' verdient, moet tenminste een juiste afspiegeling vormen van zowel de inhoud van de bestudeerde stof, als van de geoefende en beoogde beheersingsvorm. Wat betreft de beheersingsvorm is het zaak, bij gesloten vragen het onderscheid tussen kennis (reproduceren) en inzicht in het oog te houden. De leraar kan bij het maken van de toetsvragen de volgende stappen nemen:

- denk na over de hoofdzaken van de leerstof;
- maak een overzicht van de leerstof in de vorm van een lijst met trefwoorden;
- geef bij de trefwoorden aan of het gaat om feitenkennis of probleemoplossen;
- formuleer de vragen (als de toets selectief wordt gebruikt, streef dan naar tenminste 50 vragen);
- controleer de kwaliteit van de vragen aan de hand van de gegeven vuistregels en breng waar nodig verbeteringen aan;
- ga na of de verzameling vragen de hoofdzaken van de leerstof in bevredigende mate dekt;
- laat de toets door een vakcollega maken en becommentariëren.

10.9 Rangschikken en sorteren

Opdracht

- Construeer voor de tekst over bestuiving en bevruchting bij zaadplanten één rangschikkingsvraag en een tweetal sorteervragen.
 - Geef bij elke vraag de beheersingsvorm aan.
- Uiteraard kan men ook een andere tekst kiezen.

Bestuiving en bevruchting bij zaadplanten

Het overbrengen van stuifmeel op de stamper heet bestuiving. Stuifmeel kan uit zichzelf niet van een meeldraad naar een stamper komen. Bestuiving kan op verschillende manieren gebeuren: onder andere door insecten en door de wind.

Bloemen waarbij stuifmeel door insecten overgebracht wordt, heten insectenbloemen. Insectenbloemen hebben meestal vrij grote en opvallend gekleurde kroonbladeren.

Hierdoor worden insecten aangelokt. Dit gebeurt bovendien door de geur en vooral door het zoete sap, de nectar, onder in de bloem. Uit dit nectar maken de bijen honing. Om deze nectar te verzamelen moeten ze diep de bloem in. Ze strijken daarbij met hun lichaam langs de meeldraden en stamper. Het stuifmeel is ruw en kleverig. Het hecht zich gemakkelijk aan het insect, waardoor het vaak bij het kruipen in een andere bloem bestuiving kan veroorzaken. Er gaat weinig stuifmeel verloren bij insectenbestuiving. Insectenbloemen maken dan ook weinig stuifmeel.

Bij windbloemen wordt het stuifmeel door de wind van de meeldraden naar de stamper geblazen. Windbloemen zijn vaak onopvallend gekleurd en tamelijk klein.

Denk maar aan de bloempjes van de brandnetels en grassen. De kans dat een stuifmeelkorrel op de stamper van dezelfde soort plant terecht komt is maar heel klein. Daarom maken deze bloemen erg veel stuifmeel. Dit is glad en licht, zodat het makkelijk zweeft. De grote meeldraden en stampers steken vaak ver uit de bloempjes naar buiten zodat de wind er goed bij kan.

Als de stuifmeelkorrel op de stempel van dezelfde bloem komt, of op een bloem die op dezelfde plant zit, dan spreekt men van zelfbestuiving.

Als de stuifmeelkorrel terecht komt op een stempel van een bloem die op een andere plant zit, dan spreken we van kruisbestuiving. Die andere plant moet natuurlijk wel van dezelfde soort zijn.

Als de stuifmeelkorrel op de stempel is gekomen, kan de stuifmeelkorrel uitgroeien naar een zaadbeginsel. Er ontwikkelt zich dan een stuifmeelbuis, die door de stijl naar een zaadbeginsel groeit. In het zaadbeginsel bevindt zich de eicel. De kernen van stuifmeelkorrel en eicel versmelten.

De versmelting van de kernen van twee voortplantingscellen noemen we bevruchting.

Na de bevruchting vinden in de bloem allerlei veranderingen plaats. Uit de zaadbeginsels ontstaan na de bevruchting de zaden. Je kunt dat goed zien bij een appel.

Uit de zaadbeginsels ontstaan de pitten, de zaden van de appelboom. Uit het vruchtbeginsel ontstaat de vrucht van de boom. Een vrucht (appel) bevat dus een aantal zaden (pitten). Uit elk zaad kan na kieming weer een nieuwe plant tot ontwikkeling komen.

10.9.1 Zijn gesloten vragen objectief?

In de literatuur worden gesloten vragen onder andere aanbevolen, omdat ze volstrekt objectief zouden zijn. Inderdaad is bij de correctie van gesloten vragen sprake van volledige objectiviteit: het maakt niet uit wie nakijkt, de score zal – vergissingen daargelaten – dezelfde zijn. Meningsverschillen tussen beoordelaars spelen bij het nakijken geen rol, omdat de scoringsleutel ondubbelzinnig aangeeft welke antwoorden goed en welke fout gerekend worden.

Echter: tijdens het opstellen van de vragen en het vaststellen van de antwoordsleutel speelt het subjectieve oordeel van de docent wel degelijk een rol. Een docent kan vinden dat hij een vraag eenduidig heeft geformuleerd, terwijl bij de beantwoording van leerlingen kan blijken dat de vraag toch voor tweeërlei uitleg vatbaar was. De docent bepaalt welk alternatief het beste of het juiste is. Soms zijn er allerhande nuanceringen tussen goed en fout. Een andere docent – of leerlingen – kunnen soms op goede gronden een andere opvatting zijn toegedaan.

Hoe kan nu de objectiviteit bij het opstellen van de vragen en het vaststellen van de antwoordsleutel het best gewaarborgd worden? De remedie bestaat uit twee stappen:

- a goede kwaliteitscontrole vooraf, door de toets op te stellen met of te laten natrekken door een of meer collega's;
- b leerlingen gelegenheid geven commentaar te leveren op de kwaliteit van de vragen. Als hun kritiek hout snijdt, moet dit uitlopen op verandering van de antwoordsleutel of verwijdering van de discutabele vragen.

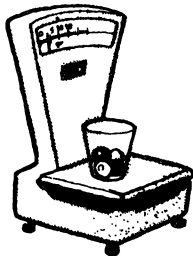
10.9.2 Meerkeuzevragen

Het gebruik van toetsen met meerkeuzevragen heeft sedert de jaren zeventig in ons land een hoge vlucht genomen. Op zijn minst twee feiten hebben hieraan bijgedragen: het verschijnen van het boek 'Vijven en zessen' door A.D. de Groot in 1966 en de oprichting van het Centraal Instituut voor Toets Ontwikkeling (CITO) te Arnhem in 1968.

De Groot keerde zich in zijn boek tegen wat hij noemt de grondfout van ons onderwijs: de selectie met rekbare maatstaven. Deze maatstaven wilde hij uniformeren met meerkeuzevragen.

Het CITO construeert en verspreidt toetsen met meerkeuzevragen. (Overigens houdt dit instituut zich ook bezig met de ontwikkeling van allerlei andere evaluatiemiddelen.)

Hier volgen twee voorbeelden van meerkeuzevragen, ontleend aan een eindtoets basisonderwijs.



Op deze weegschaal staat een bekertje met 4 knikkers die precies even zwaar zijn. Het gewicht van het bekertje met de knikkers is 4 gram. Als Piet 2 knikkers uit het bekertje haalt, geeft de wijzer een gewicht van 3 gram aan. Wat wegen 2 knikkers samen?

- A $\frac{1}{2}$ gram
- B 1 gram
- C 2 gram



Pim gaat een kelder in met een stormlamp. Plotseling valt de deur in het slot en is de ruimte luchtdicht afgesloten. Hij weet niet hoe lang het zal duren voor Koos met de sleutel komt. Hij doet de lamp uit. Waarom is dit een verstandig besluit?

- A De lucht blijft zuiverder
- B Het zal niet zo warm worden in de kelder
- C Hij bespaart daardoor olie
- D Hij bespaart daardoor zuurstof

De belangrijkste vuistregels voor het opstellen van meerkeuzevragen volgen hieronder:

- a voor de hele vraag:
 - nauwkeurig, eenduidig formuleren (slechts voor één uitleg vatbaar);
 - formuleer zo eenvoudig en bondig mogelijk;
 - stam en alternatieven moeten grammaticaal en logisch op elkaar aansluiten;
 - één vraag mag niet het oplossen van een andere vraag vergemakkelijken;
 - vermijd strikvragen;
 - vermijd zoveel mogelijk vragen met ontkenningen;
- b voor de alternatieven:
 - één alternatief moet volledig goed zijn, de andere moeten onjuist of minder juist zijn; dit moet ondubbelzinnig vaststaan;
 - de onjuiste alternatieven moeten voor de leerlingen die de stof niet (goed) beheersen, aantrekkelijk (plausibel, attractief, acceptabel) zijn (Voorbeeld: in vragen die gaan over de geschiedenis van de negentiende eeuw, hoort Titus niet in een van de afleiders te figureren; in vragen over de Oudheid daarentegen, kan deze naam wel opgevoerd worden.);
- c voor de lay-out (opmaak):
 - let op dat een vraag in z'n geheel op één bladzijde staat;
 - onderscheid de verschillende vragen van elkaar door er voldoende wit tussen te laten;
 - begin niet met de moeilijkste vragen, dit om ontmoediging tegen te gaan;
 - let er op dat de juiste alternatieven in willekeurige volgorde zijn gezet.

De toets als geheel moet in bevredigende mate de onderwijsdoelen dekken. Zorg er dus voor, dat de inhoud van de vragen ontleend is aan hoofdzaken van de leerstof en dat de beheersingsvorm en de moeilijkheidsgraad van de vragen overeenstemmen met de leerdoelen. De toets moet uit een voldoende aantal vragen bestaan en de leerstof evenwichtig bestrijken.

Welk aantal alternatieven verdient de voorkeur?

Het opstellen van meerkeuzevragen verloopt vaak moeizaam, doordat aannemelijke alternatieven – aannemelijk voor leerlingen die de stof niet goed onder de knie hebben – vaak niet voor het oprapen liggen. Voor sommige onderwerpen zijn afleiders, die voor leerlingen attractief zijn, doodeenvoudig niet voorhanden. Soms staat de vraag maar één zinnige afleider toe, wat door veel leraren te weinig wordt gevonden.

Afleiders die er met de haren bijgesleept zijn, worden echter ofwel door de leerlingen moeiteloos als fout alternatief herkend ('schertsafleiders'), ofwel zo'n afleider houdt een overschrijding in van datgene wat moest worden bestudeerd, waardoor de vraag inboet aan validiteit.

In tegenstelling tot wat vaak gedacht wordt, is het bij meerkeuzevragen niet noodzakelijk steeds naar vier alternatieven per vraag te streven. Slechte alternatieven kan de docent beter 'weggooien'. Er blijven dan vragen over met drie, of soms maar twee alternatieven. Een combinatie van tweekeuze-, driekeuze- en vierkeuze-vragen in één toets is zowel voor het beantwoorden als voor het scoren geen probleem.

De raadkans bij twee- of driekeuzevragen is wel groter dan bij vierkeuzevragen, maar dat kan de docent corrigeren door meer vragen in de toets op te nemen. Een bijkomend voordeel van een langere toets is, dat de dekking (inhoudsvaliditeit) van de leerstof hoger is.

Beantwoordingsstrategie

Voor de betrouwbaarheid van een meerkeuze-toets is het van belang dat de leerlingen enige ervaring hebben opgedaan met een goede strategie bij het beantwoorden van zulke vragen. De volgende richtlijnen voor leerlingen kunnen hierbij van nut zijn:

- laat je niet ontmoedigen, als je in het begin een of meer vragen niet weet. Ga eerst door en denk later nog eens rustig over die moeilijke vragen na;
- streep niet meer dan één alternatief aan;
- bekijk alle alternatieven; soms lijken er meer 'goed' te zijn, maar je moet de beste kiezen. Denk niet: deze is vast goed, de rest hoef ik niet te lezen!
- als je niet weet welk alternatief je moet kiezen, probeer dan eerst uit te vinden welke beslist niet goed zijn;
- laat geen vraag onbeantwoord;
- kijk je werk na; concentreer je op de vragen waarbij je hebt getwijfeld.

10.9.2 Verbeter meerkeuzevragen

Opdracht

- Ga na welke fouten er aan onderstaande items kleven.
- Verbeter de fouten.

- 1 De Zuiderzee is afgesloten om:
 - a de zoetwatervisserij daar mogelijk te maken
 - b meer recreatiegebied te hebben
 - c nieuwe landbouwgronden te kunnen ontginnen
 - d overstromingen tegen te gaan
- 2 De Noormannen waren afkomstig uit:
 - a De Middellandse zeestaten
 - b Noorwegen, Zweden en Denemarken
 - c Germanië
 - d de Keltische landen
- 3 Katoen komt van een...
 - a dier
 - b mineraal
 - c plant
 - d schaap

- 4 De meest gewaardeerde Nederlandse staatsman sinds het begin van deze eeuw is:
 - a Colijn
 - b Den Uyl
 - c Gerbrandy
 - d Thorbecke
- 5 Een rondeel is:
 - a een vierregelig gedicht
 - b een dichtvorm zonder vast rijmschema
 - c een oude dichtvorm, die uit niet meer dan twee rijmwoorden bestaat en waarbij een tweetal versregels een aantal keren herhaald wordt
 - d een vorm van directe beeldspraak
- 6 In de tweede helft van de vorige eeuw kwam als krachtbron in de eerste plaats:
 - a elektriciteit
 - b atoomkracht
 - c paarden
 - d stoomkracht
- 7 In welke regel staan de woorden in de juiste volgorde?
 - a arm – straatarm – rijk – schatrijk
 - b rijk – straatarm – arm – schatrijk
 - c straatarm – arm – rijk – schatrijk
 - d straatarm – arm – schatrijk – rijk
- 8 Het is niet juist te beweren, dat de klassieke psychoanalyse zich niet bezighoudt met:
 - a geconditioneerde reflexen
 - b mathematische gedragsmodellen
 - c testconstructie
 - d theorievorming

10.9.2.1 Analyse van meerkeuzevragen

Bij de constructie van toetsen op het CITO wordt vaak een voorlopige toetsversie op enkele scholen uitgetoetst (proefafname). Na analyse van de toets worden slechte vragen verwijderd of herzien. Een leraar in de gewone schoolpraktijk komt niet toe aan proefafnames. Wel kan hij na afname van een door hemzelf gemaakte meerkeuzevraag-toets de kwaliteit van de vragen analyseren en slechte meerkeuzevragen verbeteren voor een volgende afname of weggooien.

De moeilijkheidsgraad van een meerkeuzevraag kan achterhaald worden door het berekenen van de p-waarde. De p-waarde van een meerkeuzevraag is het getal dat aangeeft, welk percentage leerlingen de vraag goed beantwoordde. De berekening vindt als volgt plaats:

$$p = \frac{\text{aantal leerlingen dat de vraag juist beantwoordde}}{\text{totaal aantal leerlingen}}$$

Voorbeeld:

van de 30 leerlingen hebben er 18 een goede antwoord aangekruist.

$$\text{De p-waarde} = \frac{18}{30} = .60$$

P-waarden kunnen variëren van 0 tot 1. Als $p = 0$ heeft geen enkele leerling de vraag goed beantwoord, als $p = .50$ heeft de helft van de leerlingen de vraag goed gemaakt. $P = .80$ als 80 procent van de leerlingen de vraag goed heeft beantwoord, enzovoort. Naarmate de p-waarde hoger is, was de vraag gemakkelijker.

Bij selectief gebruik zullen bij meerkeuzevragen met vier alternatieven de p-waarden liggen tussen de .25 en .75. Hoe strenger de toets is, des te lager zijn de p-waarden. Als $p = 1$ hebben alle leerlingen de vraag goed beantwoord. Welke p-waarden een docent acceptabel vindt, hangt samen met de gebruikswijze van de toets.

De aantrekkelijkheid van de afleiders van een meerkeuzevraag wordt aangegeven met de a-waarde. (Een meerkeuzevraag met vier alternatieven bezit dus één p-waarde en drie a-waarden.)

De a-waarde van een afleider is het getal, dat aangeeft welk percentage van de leerlingen het foute alternatief koos.

De berekening is als volgt:

$$a = \frac{\text{aantal leerlingen dat de afleider koos}}{\text{totaal aantal leerlingen}}$$

Voorbeeld:

In een klas met 30 leerlingen waren de antwoorden op de vier alternatieven van een meerkeuzevraag als volgt verdeeld:

$$\text{alternatief a (juist)} \quad 15 \text{ leerlingen } p = \frac{15}{30} = .50$$

$$\text{alternatief b} \quad 6 \text{ leerlingen } a = \frac{6}{30} = .20$$

$$\text{alternatief c} \quad 6 \text{ leerlingen } a = \frac{6}{30} = .20$$

$$\text{alternatief d} \quad 3 \text{ leerlingen } a = \frac{3}{30} = .10$$

A-waarden liggen, evenals p-waarden, tussen 0 en 1. Is $a = 0$, dan heeft geen enkele leerling de afleider gekozen. Hoe hoger de a-waarde, des te aantrekkelijker was de afleider voor de leerlingen. Acceptabele a-waarden dienen bij benadering te liggen tussen .10 en .40. Voor de verschillende afleiders van een meerkeuzevraag mogen de a-waarden niet te ver uiteenlopen. Afleiders met extreme waarden behoren te worden verwijderd of herschreven.

Betekenis van p- en a-waarden

Wat kan een lage p-waarde nu betekenen? Enkele voor de hand liggende oorzaken zijn de volgende:

- 1 Er is een fout in de sleutel geslopen, zodat een verkeerd alternatief als het juiste gerekend is. (Dit komt vrij geregeld voor!)
- 2 Een of meer van de andere alternatieven kunnen als het juiste verdedigd worden. Deze alternatieven zullen dan door vele leerlingen gekozen zijn (hoge a-waarden!). (...)
- 3 Bij het onderwijs is het betreffende onderdeel van de stof niet (voldoende) behandeld. Dit gebeurt wel eens wanneer items van vorige jaren in de toets zijn opgenomen.
- 4 Het onderwijs van dit gedeelte van de stof is niet goed overgekomen, of het leerboek is hier onduidelijk. De docent krijgt in dit geval informatie over zijn onderwijs, en bovendien een aanwijzing bij welk deel van de stof dit verbeterd moet worden.
- 5 In het oog van de leerlingen gaat het item over een onbelangrijk deel van de stof, bijvoorbeeld voetnoten, of kleine letters. Vindt de docent – ook bij nadere beschouwing – dit onderdeel wel van belang, dan zal hij dat in het vervolg bij de behandeling ervan duidelijk moeten maken aan de leerlingen.
- 6 De leerlingen hebben afgesproken dit stuk stof niet te leren.
- 7 De leerlingen hebben de stof slecht geleerd. In dit geval zullen vrijwel alle items een lage p-waarde hebben.
- 8 Het item blijkt bij nader inzien te moeilijk voor het niveau van de leerlingen.

Wat kunnen verder de a-waarden ons leren? Het komt soms voor, dat de a-waarde, behorende bij een bepaald alternatief, aanzienlijk lager is dan de andere twee a-waarden van dat item. Dit kan betekenen, dat de leerlingen in ieder geval geleerd hebben, dat het bedoelde alternatief onmogelijk juist kan zijn (goed onderwijs). Het kan echter ook betekenen, dat het alternatief weinig plausibel is, ook voor slechte leerlingen; zo'n alternatief kan ook een indicator bevatten. In dat geval kan men het alternatief beter vervangen.

Een hoge a-waarde (ongeveer even hoog als de p-waarde) wijst er meestal op, dat het alternatief als goed antwoord te verdedigen is, of dat het item een strikvraag is. In beide gevallen mogen we het item in het algemeen niet onveranderd laten.

Bron: Groot, A.D. de, & Naerssen, R.F. van (Eds.) (1973). *Studietoetsen. Construeren, afnemen, analyseren. Deel I*. Den Haag: Mouton, p. 87-88.

10.9.2.1a Analyse van een zelf afgenomen meerkeuzevraag

Een tweetal analyses van dezelfde vraag leverden de volgende waarden op.

Vraag:

In Noorwegen wordt de belangrijkste delfstof gevormd door:

- | | | |
|---|-----------|---------------------------|
| a | bauxiet | .17 |
| b | ijzererts | .38 |
| c | olie | .44 (het juiste antwoord) |
| d | graniet | .01 |

Dit item werd herschreven met het resultaat de volgende a- en p-waarde(n)

- | | | |
|---|-----------|---------------------------|
| a | bauxiet | .09 |
| b | ijzererts | .37 |
| c | olie | .34 (het juiste antwoord) |
| d | aardgas | .20 |

Opdracht

- Geef de reden aan waarom werd besloten tot itemrevisie
- Geef je oordeel over de moeilijkheidsgraad van het originele en het herschreven item.
- Welke conclusie kun je trekken met betrekking tot het gegeven onderwijs over de in dit item opgenomen leerstof, indien je de a-waarden en de p-waarden van het originele item vergelijkt met die van het herschreven item?

10.9.2.1b Analyse van een zelf afgenomen meerkeuzevragentoets

Om een goed overzicht te krijgen van hoe zowel een hele klas als een individuele leerling een meerkeuzevragentoets heeft gemaakt, is het handig de resultaten in een toetsmatrix onder te brengen. De leraar kan nu in één oogopslag zien welke items goed en welke slecht uitpakken.

Opdracht

- Bereken voor elk item de p-waarde en de a-waarden.
- Bepaal op grond van de verkregen p- en a-waarden welke items herschreven dienen te worden.
- Beoordeel de moeilijkheidsgraad van de toets.
- Bereken de score per leerling en de gemiddelde score.
- Beoordeel de cesuur (bepaal zelf of je bepaalde items op grond van de gevonden waarden al of niet wilt handhaven voor de bepaling van de cesuur).
- Ken elke leerling een cijfer toe.
- Bereken hoeveel procent van de leerlingen een onvoldoende heeft.
- Vergelijk jouw gegeven cijfers met die welke door andere studenten zijn gegeven.

10.9.3 Juist/onjuist vragen

Van de juist/onjuist vraag (ook wel ja/nee vraag of waar/onwaar vraag genoemd) zijn twee typen bekend:

- a de enkelvoudige juist/onjuist vraag, waarbij een leerling een bewering of stelling met juist of onjuist moet beantwoorden.

Voorbeeld:

De hoofdstad van Hongarije is Boekarest juist/onjuist

- b de meervoudige juist/onjuist vraag, waarbij een gemeenschappelijke stam wordt gevolgd door twee of meer afzonderlijke beweringen, die elk met juist of onjuist beantwoord moeten worden.

Voorbeeld:

Gegeven:

Cara is in ons land de meest voorkomende longaandoening.

Vragen:

- | | |
|--|---------------|
| 1 Cara betekent: chronische a-specifieke respiratoire aandoeningen | juist/onjuist |
| 2 Cara wordt gekenmerkt door de klachten hoest en kortademigheid | juist/onjuist |
| 3 Cara komt bij kinderen en oudere mensen minder vaak voor dan bij andere leeftijdsgroepen | juist/onjuist |

(Voor de liefhebbers volgt hier de sleutel: 1. juist, 2. juist, 3. onjuist)

In de strikte zin van het woord is een juist/onjuist vraag een meerkeuzevraag. Je hebt immers de keuze uit twee, of eigenlijk drie alternatieven: juist, of onjuist, of je laat de vraag onbeantwoord.

Pluspunten van juist/onjuist vragen

In de Nederlandse scholen wordt betrekkelijk weinig gebruik gemaakt van de onderhavige vraagvorm. Juist/onjuist vragen hebben de naam dat ze slechts simpele parate kennis kunnen toetsen. Dit is een onjuiste veronderstelling. Ze kunnen wel degelijk ook voor het evalueren van productief denken ingezet worden. Een paar voorbeelden:

- De bewering, dat het communisme zich in een land het eerst zal ontwikkelen onder de arme boeren is in overeenstemming met de leer van Marx juist/onjuist
- Als de nationale productie groeit, terwijl de gemiddelde arbeidsproductiviteit constant blijft, zal de werkgelegenheid toenemen juist/onjuist
- Gegeven:
'My father has bought a new car.'
Vraag:
Gesteld in de lijdende vorm luidt bovenstaande zin:
'A new car has been bought by my father.' juist/onjuist

Een volgend bezwaar betreft de hoge raadkans: als je blind gokt, heb je toch 50 procent kans op het goede antwoord. De toets zou ontaarden in een

raadspelletje. Dit bezwaar kan eenvoudig worden ondervangen door toevals-correctie toe te passen.

De juist/onjuist vraagvorm heeft vergeleken met de meer gangbare meerkeuzevragen pluspunten, die een frequenter gebruik in het onderwijs rechtvaardigen. Juist/onjuist vragen zijn relatief gemakkelijk te maken. Over plausibele alternatieven (afleiders) hoeft niet te worden nagedacht. Onderwerpen waarbij geen acceptabele alternatieven voor het oprapen liggen, kunnen wel in de vorm van een stelling gegoten worden. Leerstof is immers doorgaans binair opgebouwd: er wordt iets gesteld of ontkend.

Het gaat er bij toetsingen om, dat de vragen ontleend zijn aan hoofdzaken van de bestudeerde leerstof. Voor diverse onderwerpen is het niet gemakkelijk of onmogelijk acceptabele afleiders te vinden. Een mogelijk gevolg hiervan is, dat enerzijds bepaalde hoofdzaken van bestudeerde leerstof bij de evaluatie niet aan bod komen, terwijl anderzijds voor details, waarvoor wel gemakkelijk aannemelijke afleiders konden worden gevonden, een te grote plaats is ingeruimd. Dit probleem speelt niet bij juist/onjuist vragen.

Vuistregels voor het opstellen van juist/onjuist vragen

De belangrijkste vuistregels voor het maken van juist/onjuist vragen zijn:

- formuleer de vraag in de vorm van een bewering, liefst voorafgegaan door een 'gegeven';
- vermijd dubbele beweringen; als in de ogen van een leerling beide beweringen niet tegelijk juist of onjuist zijn, welk antwoord moet de leerling dan kiezen?
- deskundigen moeten het eens zijn over de juistheid of onjuistheid van de bewering;
- formuleer de bewering kort, zakelijk, precies en ondubbelzinnig. Woorden met een vaag omlinjende betekenis zoals 'gewoonlijk', 'zelden', 'weinig', 'vaak', 'soms', 'nu en dan', 'goed', 'spoedig', 'lang', 'kort', 'snel', 'langzaam' en dergelijke, zijn uit den boze. Dus: frequentie moet een percentage worden, een hoeveelheid moet precies zijn gekwantificeerd en tijdsduur moet in minuten tot jaren zijn gegeven. Ook het maken van een vergelijking is vaak een goede oplossing;
- vermijd al te absolute en al te relativerende formuleringen; maak geen gebruik van woorden als 'altijd', 'nooit', 'alle' omdat op de meeste regels uitzonderingen bestaan en het antwoord derhalve 'onjuist' zal luiden; vragen waarin naar een mogelijkheid wordt gevraagd, de zogenaamde 'kan'-vragen, hebben als antwoord meestal 'juist': immers, heel veel is ooit wel eens voorgevallen of mogelijk gebleken;
- vraag niet naar waarde-oordelen;
- pas op voor 'weggevers' (vragen die zonder bestudering van de stof gemakkelijk kunnen worden beantwoord);
- vraag naar hoofdzaken, weersta de neiging details in de toets op te nemen;
- let erop, dat er ongeveer een even groot aantal juiste als onjuiste beweringen in de toets is opgenomen.

Opmerking: Steeds weer blijkt dat anderen gebreken in vragen ontdekken, die door de opsteller zelf over het hoofd zijn gezien. Vraag daarom een collega om een 'second opinion', of maak samen vragen.

Hieronder volgen enkele juist/onjuist vragen die constructiefouten bevatten.

a	Heeft Amsterdam de meeste inwoners?	juist/onjuist
b	In ons land zal nooit meer oorlog gevoerd worden	juist/onjuist
c	De Mount Everest, de hoogste berg ter wereld, ligt in het Andesgebergte	juist/onjuist
d	De daling van het geboortecijfer is veroorzaakt door de pil	juist/onjuist
e	Hyena's eten alleen kadavers	juist/onjuist
f	De apartheidspolitiek in Zuid-Afrika was verwerpelijk	juist/onjuist
g	Baby's slapen korter dan kleuters	juist/onjuist
h	In de laatste kwart eeuw vertoont het geboortecijfer een dalende trend	juist/onjuist
i	Kanker neemt de laatste jaren toe	juist/onjuist
j	De belangrijkste manier om de industrie te stimuleren is het verschaffen van investeringspremies	juist/onjuist
k	Kernenergie heeft de toekomst	juist/onjuist
l	Het is niet gebleken dat griep niet wordt veroorzaakt door een virus	juist/onjuist
m	Elke schoolfobie gaat vaak gepaard met leerstoornissen	juist/onjuist

Commentaar op deze vragen:

- De vraag is niet geformuleerd in de vorm van een stelling. (Ook is niet aangegeven met welke stad of steden het inwonertal moet worden vergeleken.)
- Voorspelling die nog moet blijken.
- Dubbele bewering, waarvan één juist (de hoogste berg ter wereld) en de andere onjuist is. De leerling wordt hierdoor in verwarring gebracht.
- De pil is niet de enige oorzaak. De aanduiding voor welk(e) land(en) dit geldt ontbreekt.
- Ze eten hoofdzakelijk aas. Het woord 'alleen' staat geen enkele uitzondering toe, zodat de vraag te gemakkelijk wordt.
- Hier wordt naar een opvatting gevraagd (waarmee hopelijk iedere lezer eens is).
- Dit is een 'weggever', iedereen weet dat dit onjuist is.
- Hier wordt niet aangegeven in welk(e) land(en) deze trend zich voordoet
- Er zijn vele soorten kanker; sommige vormen nemen toe (bijvoorbeeld longkanker), andere komen minder voor dan vroeger (bijvoorbeeld maagkanker), althans in ons land. Ook de tijdsaanduiding ('de laatste jaren') is meerduidelijk.
- Het is een van de vele mogelijke manieren (andere zijn bijvoorbeeld: verbetering van vervoersmogelijkheden, exportsubsidie, loonsverlaging).

Welke manier de belangrijkste is valt in het algemeen niet te zeggen: het verschilt per land, per industrietak en... per deskundige.

- k Het gaat hier om een bepaalde mening over een uitspraak, die van alles kan betekenen.
- l Dubbele ontkenning, die de vraag nodeloos ingewikkeld maakt.
- m 'Elke' staat geen enkele uitzondering toe, 'vaak' wel: de vraag bevat dus een tegenstrijdigheid en is daarom onbeantwoordbaar. Bovendien zijn 'elke' en 'vaak' woorden die voor juist/onjuist vragen ongeschikt zijn: 'elke' is te absoluut en 'vaak' te vaag.

Ter afsluiting van deze paragraaf een viertal juist/onjuist vragen over leerstof uit dit hoofdstuk:

- 1 Als een toets valide is, is de toets ook betrouwbaar juist/onjuist
- 2 Gegeven:
Een item uit een toets met meerkeuzevragen heeft een p-waarde=1
Vraag:
Dit item maakt onderscheid tussen slechte en goede leerlingen juist/onjuist
- 3 De betrouwbaarheid van toetsen bestaande uit open vragen is meestal lager dan die van toetsen met gesloten vragen juist/onjuist
- 4 Met het oog op de inhoudsvaliditeit van een toets voldoen open vragen beter dan gesloten vragen juist/onjuist

¹ (De sleutel van deze vier vragen is te vinden op pagina 387.)

10.9.3 Juiste 'misbaksels'

Opdracht

- Stel in groepen voor een te kiezen tekst een aantal juist/onjuist vragen op die zondigen tegen de gegeven vuistregels.
- Probeer iedere vuistregel tenminste één keer met voeten te treden.
- Bespreek de 'misbaksels' plenair: laat de gemaakte fouten door medestudenten ontdekken en verbeteren.

10.9.4 Scoringsregels en beantwoordingsstrategie

Scoringsregels bepalen hoe foute antwoorden en vragen die niet ingevuld zijn worden geteld. De twee meest gebruikte scores zijn de goed-score en de goed-min-fout-score.

De goed-score

Bij de goed-score wordt het aantal keren 'goed' (of 'juist') geteld op de volgende wijze:

goed	=	1 punt
fout	=	0 punten
geen antwoord	=	0 punten

De beantwoordingsstrategie die bij deze score de beste resultaten geeft, luidt: beantwoordt zonder uitzondering alle vragen; als je iets beslist niet weet, kun je beter raden dan het antwoord open laten. De leerling die op deze wijze bezig is, maakt maximaal gebruik van zijn kennisbezit. Ook oppervlakkige (globale) en/of gedeeltelijke (partiële) kennis wordt aangesproken. De toevals-correctie (kansscore) bij de waardering bedraagt 50 procent bij vragen met 2 alternatieven, ruim 33 procent bij vragen met 3 alternatieven enzovoort. Bij een toets van 60 vierkeuzevragen is de kansscore dus 15 goed beantwoorde vragen.

De goed-min-fout-score

De goed-min-fout-score wordt als volgt berekend:

$$\text{goed} = 1 \text{ punt}$$

$$\text{fout} = - \frac{1}{a-1} \quad (a = \text{aantal alternatieven})$$

$$\text{geen antwoord} = 0 \text{ punten}$$

Voor foute antwoorden worden punten afgetrokken, meestal één punt. Een leerling die niets weet en alleen maar gokt, komt hierdoor, gemiddeld genomen, uit op een score van nul.

Blind raden zet bij de goed-min-fout scoringsregel dus weinig zoden aan de dijk: de kans op een 'treffer' is even groot als die op een 'misser'. In het wilde weg raden komt overigens niet zo veel voor, want meestal weet de leerling wel iets over het antwoord op de vraag. Op beredeneerde wijze raden kan zijn nut hebben. Tussen weten en niet weten is vaak sprake van een vloeiende overgang. Vaak weet een leerling er wel 'iets' van. Dit 'iets' betreft partiële kennis of globale kennis. (Ook in het dagelijks leven worden veel beslissingen genomen op basis van partiële kennis.) Soms weet degene die een fout antwoord geeft er meer van dan de leerling die de vraag onbeantwoord laat. Desondanks wordt hem een punt afgetrokken; zijn tekort aan kennis wordt in zo'n geval 'overgecorrigeerd'.

Sommige leerlingen zullen een vraag reeds beantwoorden als ze slechts weinig zeker zijn van het goede antwoord (bijvoorbeeld voor ongeveer 30 procent), andere zullen de vraag pas beantwoorden als ze er voor tenminste 80 procent zeker van zijn. De eerstgenoemde leerlingen zullen over het geheel genomen hun kennisbezit beter tot uitdrukking laten komen in de toetsuitslag (omdat ze al snel hun partiële en globale kennis durven gebruiken) dan

de leerlingen die de vraag pas invullen als ze zeker of vrijwel zeker zijn van het goede antwoord. De goed-min-fout-scoringsregel bevoordeelt dus de zelfbewuste, 'goklustige' en 'testslimme' leerlingen.

Conclusie: de goed-score levert het beste beeld van de opgedane kennis (inclusief partiële en globale kennis) op. De antwoordstrategie die een leerling moet volgen is: beantwoordt alle vragen. Zelfs als een leerling helemaal niets van de vraag weet, is het beter te raden dan de vraag open te laten.

Van score naar waardering

Wanneer een docent meerkeuzevraagtoetsen absoluut normeert (vooraf van een waardering voorziet) moet hij rekening houden met de raadkans (prestatie bij blind raden). Pas boven de kansscore wordt echte kennis weergegeven. De kansscore is afhankelijk van de gehanteerde scoringsregel en het aantal alternatieven per vraag.

Bij de goed-min-fout-scoringsregel is de kansscore nihil. Bij de goed-scoringsregel is de kansscore afhankelijk van het aantal alternatieven. Voor juist/onjuist vragen is de raadkans 50 procent. Bij meerkeuzevragen met meer dan twee alternatieven ligt de kansscore lager.

Het absoluut bepalen van de cesuur bij de goed-scoringsregel moet plaatsvinden op basis van het aantal goede antwoorden minus de kansscore. Bij de relatieve normering doet de kansscore er niet toe.

10.9.5 *Vergelijking tussen open vragen en gesloten vragen*

Vergelijking tussen open vragen en gesloten vragen

	open vragen	meerkeuzevragen en onjuist/juist vragen
kenmerk	de leerling moet het antwoord zelf formuleren	de leerling moet het juiste antwoord kiezen
activiteit van de leerling	de meeste tijd wordt besteed aan denken en schrijven	de meeste tijd wordt besteed aan lezen en denken
denkproces	vaak kan het denkproces van de leerling worden getraceerd	het denkproces is onbekend
voorbereiding en correctie	voorbereiding is verhoudingsgewijs gemakkelijk, de correctie tijdrovend en moeilijk	voorbereiding is verhoudingsgewijs moeilijk en tijdrovend, het nakijken gemakkelijk (hoe groter de groep leerlingen, des te efficiënter is deze vraagvorm)*
inhoudsvaliditeit	verhoudingsgewijs moeilijk inhoudsvalide te maken	hoge inhoudsvaliditeit
vormvaliditeit	in beginsel geschikt voor elke beheersingsvorm	met name geschikt voor het beoordelen van reproduceren en convergent denken
objectiviteit	beoordelaarseffecten spelen een rol	objectieve score
raadkans	nihil (leerling kan wel bluffen door een omslachtig verhaal te schrijven)	gokkans die afneemt naarmate het aantal alternatieven toeneemt
invloed op studeergedrag	stimulans voor construerend leren	stimulans voor reproducerend leren

* Het omslagpunt waarboven het maken van een toets met gesloten vragen qua doelmatige tijdsinvestering de voorkeur verdient boven open vragen ligt bij circa 50 leerlingen. Als vragen herhaald worden gebruikt, is dit aantal snel bereikt.

**DE BEOORDELING VAN TOETSRESULTATEN:
HET BEPALEN VAN DE CESUUR VIA EEN ABSOLUTE METHODE.**

De cesuur is de score waarmee een leerling nog juist een voldoende resultaat behaalt.

Bij een absolute methode wordt deze cesuur vastgesteld op basis van de leerstof en de toetsvragen. De cesuur kan vooraf worden bepaald.

(Dit in tegenstelling tot een relatieve methode, waarbij de grens tussen voldoende en onvoldoende afhangt van de toetsresultaten, of bv. van het aantal beschikbare plaatsen in een vervolgopleiding)

MEERKEUZETOETS: DE METHODE VAN NEDELSKY.

Je gaat uit van een fictieve leerling die nog juist een voldoende verdient. Bij elke meerkeuzevraag bepaal je welke foute alternatieven deze leerling in elk geval moet kunnen afwijzen.

Tussen de overgebleven alternatieven zal de leerling raden welke het juiste antwoord is. De kans dat de vraag goed wordt beantwoord is dan

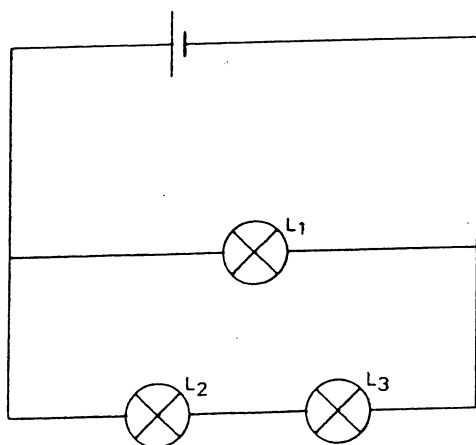
1

(aantal alternatieven waartussen de leerling raadt)

Dit is de score die de leerling op deze vraag zal behalen. Door van elke meerkeuzevraag de op deze manier verkregen scores op te tellen vind je de totale toetsscore van de juist voldoende leerling. Deze score is de cesuur.

Een voorbeeld:

Drie identieke (=dezelfde) lampen zijn geschakeld als in de onderstaande figuur.



Als lamp L₁ wordt losgedraaid

- A gaat lamp L₂ feller gloeien.
- B blijft lamp L₂ even fel gloeien.
- C gaat lamp L₂ zwakker gloeien.
- D gaan alle lampen uit.

Als je vindt dat de juist voldoende leerling in elk geval alternatief D moet afwijzen, maar niet A en C, dan zal hij raden tussen A, B en C en is de kans op een juist antwoord 1/3. De bijdrage van deze vraag aan de cesuur is dan 1/3.

OPEN VRAGEN TOETS: DE EIS-ITEM METHODE.

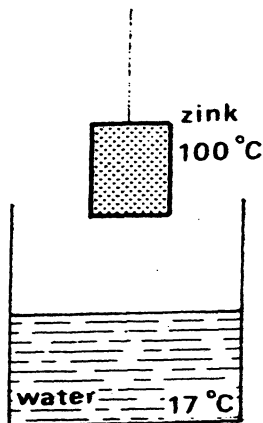
Je gaat ook hier uit van een fictieve leerling die nog juist een voldoende verdient. Je gaat de puntentoekenning na die je hebt opgesteld voor je toets. Bij elk punt beslis je of de juist voldoende leerling dit punt wel of niet moet kunnen behalen. Door alle punten die de juist voldoende leerling moet kunnen halen bij elkaar op te tellen vind je de totale toetsscore van de juist voldoende leerling. Deze score is de cesuur.

Een voorbeeld:

Soortelijke warmte

Dolf wil de soortelijke warmte van zink bepalen.
Eerst hangt hij een blokje zink van 100 g in kokend water. Het zink heeft nu een temperatuur van 100 °C.
Dan brengt hij het blokje zink in een bekersglas met 200 g water van 17 °C.
Zie figuur 14.

figuur 14



Na enige tijd is de temperatuur van het water en het zink 20 °C geworden.
De soortelijke warmte van water is 4,2 J/g·°C.

Verwaarloos bij de volgende berekening warmte-uitwisseling met het glas en de omgeving van het bekersglas.

- 29 Bereken de soortelijke warmte van zink.

Maximumscore 5

- 29 verdeling van de punten bijvoorbeeld als volgt:

- . gewerkt met $Q_{op} = Q_{af}$
- . gebruik van $Q = c \cdot m \cdot (T_e - T_b)$
- . hierin correct ingevuld voor water
- . uitwerken

1
1
1
2

Stel nu dat je meent dat de juist voldoende leerling de warmtebalans $Q_{op} = Q_{af}$ moet kunnen gebruiken; $Q = c \cdot m \cdot (T_e - T_b)$ moet kunnen gebruiken; correct moet kunnen invullen voor water; maar de uitwerking geheel mag missen.

Dan wordt zijn score 1+1+1+0 = 3 op deze vraag.
De vraag draagt dan 3 punten bij aan de cesuur.