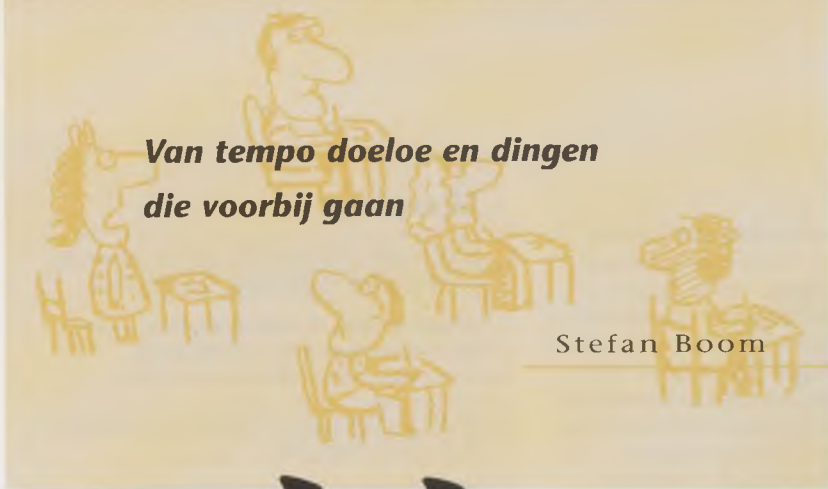




***Van tempo doeloe en dingen
die voorbij gaan***

Stefan Boom

Het havo-examen 2001

Het examen van het vorig jaar, 2000, werd zeer slecht ontvangen. De kop van het Kleio-artikel (januari 2001) waarin dit examen werd geëvalueerd, was veelzeggend:

Het havo-examen: 'oneerlijk' en 'gemeen'? Hoe is het dit jaar, in 2001, gegaan?

Hebben de examenmakers wat geleerd van de ervaringen van 2000?

Dat is de vraag die in dit artikel besproken wordt. Daarbij wordt gebruik gemaakt van de reacties op de VGN-site, de toets- en itemanalyse van het Cito en de conclusies uit een bijeenkomst van Cevo, Cito, VGN-bestuur, Commissie Havo/Vwo en de voorzitters van de regionale besprekingen.

12

Naar aanleiding van de kritiek op het examen van 2000 heeft de Cevo direct een aantal maatregelen genomen. De lengte van het examen is teruggebracht, daarmee samenhangend is de scoreschaal van 90 verkort naar 83 punten, de omvang van de bronteksten en het taalgebruik in vragen en bronnen is kritisch bekeken. Dubbele instructies in vragen zijn beperkt tot noem ... en leg uit... (of varianten daarop).

Hoe vertaalt zich dat in de analyse aan de hand van de versnelde correctie? Hebben deze maatregelen effect gehad? Om dat na te gaan leggen we de examens van 2000 en 2001 naast elkaar.¹

Vergeleken met vorig jaar zijn de eindresultaten van het examen duidelijk verbeterd. De frequentieverdeling in de tabel laat dat in één oogopslag zien. De grootste groep vragen is verschoven van de categorie p' tussen 31 en 40, naar p' tussen 41 en 50. Als we kijken naar de streefwaarden van p' ($40 < p' < 70$), dan lagen in 2000 vijftien vragen buiten deze bandbreedte (twee te makkelijk, dertien te moeilijk). Nu is dat voor negen vragen het geval (twee te makkelijk, zeven te moeilijk).

Het Duitsland-deel is in beide jaren moeilijker geweest dan het Nederland/Indonesië-deel. De p'-waarde laat zien dat het verschil tussen de twee onderwerpen minder groot is geworden. De stelopdracht (in het Duitsland onderwerp) heeft daartoe bijgedragen.

Klachten over tijdgebrek en de bewerkelijkheid van het examen hebben we dit jaar aanzienlijk minder gehoord.

Kunnen in deze gegevens de 'probleemvragen' worden teruggevonden? Dat lukt maar ten dele. Vraag vier werd moeilijk gevonden door het taalgebruik van de bron (de rede van Kennedy), maar de p'-waarde is 43, en dat valt (net) binnen de bandbreedte. De vragen met de laagste p'-waarde in het Duitsland onderwerp zijn zeven ($p'=27$), acht ($p'=18$) en negen ($p'=21$). Vraag zeven is een redeneervraag waarbij de tabel met vluchtelingen uit de DDR van een verklaring moet worden voorzien (die niet rechtstreeks uit de bron viel af te lezen, maar waarvoor wel op de aard van de getallen moest worden ingegaan), een reproductieve vraag over Schmidt (niet geleerd?) en een redeneervraag over Mitterrand (hier maar wel met dubbel r). De laatste vraag was een zogenaamde 'overlapvraag' met vwo (vraag tien), waar het ook een wat te moeilijke vraag was met een p'-waarde van 38. In het Indonesië-deel waren vraag 22 en 23 moeilijke vragen met p'-waarden

Stefan Boom is werkzaam bij de afdeling Examens van de Citogroep.

HAVO 2001-1, vergeleken met havo 2000-1**p' waarde tussen****vraag (nieuwe stijl) 2000****vraag 2001**

10-20	20, 22	8
21-30	5, 13, 24, 25	7, 9, 13,
31-40	4, 10, 14, 15, 16, 18, 19	5, 22, 23
41-50	2, 6, 9, 17, 21	2, 3, 4, 12, 14, 15, 16, 18, 19
51-60	7, 12, 23	10, 17, 21
61-70	1, 3	1, 6
71-80	8, 11	11, 20
gemiddelde p'-waarde		
totaal:	42,1	46,9
Nederland/Indonesië:	46,3	48,8
Duitsland:	37,6	45,0

van 31 en 35. In vraag 22 vinden we de beroemde 'grievens', – met een moeilijke bron, vraag 23 is een leer-vraag uit het einde van de stofomschrijving. Vragen waarover veel is geklaagd (vraag 18 met de onduidelijke *tempo doelo* foto en 21 met de andere foto) zaten beide binnen de bandbreedte: vraag 18 had een p'-waarde van 45, vraag 21 had een p'-waarde van 58.

De N-term

Hoewel er een aanzienlijke verbetering in de resultaten is geweest in vergelijking met het vorige jaar, is toch besloten tot het hanteren van een andere N-term dan 1. Als er van een N-term =1 zou zijn uitgegaan, dan werd het gemiddelde cijfer een 5,2 – met een onvoldoende voor 58% van de leerlingen. Dat vindt de Cevo onaanvaardbaar; het streven is een onvoldoendepercentage van maximaal één derde van het aantal kandidaten. Bij een N-term van 1,8 waren er 31% onvoldoendes, dat is acceptabel.

Het latere commentaar van sommige docenten was, dat door toepassing van deze N-term wel erg veel leerlingen op hoge eindresultaten kwamen. Dat kan te maken hebben met de dubbelwerking van een aantal processen. De versnelde correctie geeft een wat te somber beeld. In de haast waarmee dit gepaard gaat, worden de correctievoorschriften waarschijnlijk door de meeste leerkrachten streng geïnterpreteerd. De regionale examenbesprekingen zijn nog niet geweest, steun van collega's bij het vinden van alternatieve juiste antwoorden hebben de meeste leerkrachten nog niet gehad. De aard van ons vak laat nu eenmaal meer interpretaties en antwoordmogelijkheden toe dan in het correctievoorschrift aangegeven kunnen worden. Dit komt tot uiting in de formulering *voorbeeld van een juist antwoord*. Voorbeeld van een juist antwoord is daarom een

goede formule voor docenten om, mits er zorgvuldig mee wordt omgegaan, meerdere antwoordmogelijkheden dan het antwoordmodel noemt, te belonen.²

Veel leraren kijken na de regionale examenbesprekingen het werk nog een keer na. Na die tweede ronde krijgen veel leerlingen er punten bij. Kortom: de basis onder de normering is (hoeveel weten we niet) eerder pessimistisch dan optimistisch gekleurd.

De commentaren van geschiedenisdocenten

Opmerkelijk in de kritiek op de VGN-site was, dat veel docenten zagen dat de examenmakers geluisterd hadden naar de opmerkingen uit 2000. Maar na die aanvankelijke tevredenheid ontstond de twijfel bij het nakijken. Hoe kwam dat? Volgens sommigen waren de vragen onduidelijk en het correctievoorschrift te vaag. De foto bij vraag 18 (*tempo doelo*) bleek onduidelijk en nog meer multi-interpretabel dan de vragenmakers hadden beoogd. Het taalgebruik bleek voor havisten moeilijk, zij kenden het woord grievens bijvoorbeeld niet. Maar vervolgens kwamen er in de commentaren veel opmerkingen voor, die meer op de veranderingen van de tweede fase ingingen dan op het examen zelf.

Misverstanden

Uit de reacties vallen een aantal misverstanden over het examen te destilleren. Om maar direct een eind te maken aan één zo'n misverstand: ook in de tweede fase dient de stof 'gewoon geleerd' te worden. De Cevo vraagt in haar opdracht om een havo-examen waarvan circa 40% van de punten behaald moet worden met reproductieve vragen. Het zou dan moeten gaan om ongeveer 32 punten.³

Er blijkt echter (naar schatting – want het is afhankelijk van wat in het gebruikte katern staat wat de leerling heeft geleerd) bij ongeveer zes à zeven vragen rechtstreeks naar het weergeven van de stofomschrijving gevraagd. In het examen is bij elke vraag voortdurend sprake van het toepassen van het geleerde op de bron die wordt aangeboden, of van het in een nieuw verband beschrijven van de geleerde stof. Dat verklaart misschien waarom een aantal leerkrachten opmerkt dat bij hen 'een mavo-instroom leerlinge', 'die het vooral van hard werken en uit het hoofd leren moet hebben' het hoogste scoort. Dat zij dat jammer vinden laat ik geheel voor hun rekening. Tegelijkertijd echter wordt door sommigen aangegeven dat hun 'twee-afzakkers', 'leerlingen die dan niet zoveel schoolwerk hadden', maar wel over meer dan voldoende inzicht voor de havo beschikken, ook hoog scoren. Dat ook deze leerkrachten dat erg vinden laat ik wederom voor hun rekening. Blijkbaar toetst dit examen zowel kennis als inzicht. Ontbreekt het ene, dan is dat te compenseren met het andere.

Een tweede misverstand is de aandacht die aan casussen moet worden besteed. De invoering van casussen (de eerste in 1999 bij het 'Fin de Siècle') verdient geen schoonheidsprijs. Zonder dat duidelijk werd wat er precies mee werd bedoeld, werden examenmakers opgezadeld met de opdracht casussen te toetsen. Bestonden de casussen bij het examenonderwerp van 1999 nog uit een verdieping van de stof, bij het onderwerp van 2000 ging het om uitstapjes buiten de stof (vooral over de DDR) en bij Indonesië in 2001 over nogal onduidelijk geformuleerde vraagstukken.

De drie stofomschrijvingscommissies van deze onderwerpen hebben elk een andere omschrijving van het begrip casus en de bedoeling ervan geformuleerd.

leerd (zie de desbetreffende stofomschrijvingen).

Wel duidelijk was, dat een casus omvangrijk en met gebruik van bronnen diende te zijn. Toetstechnisch levert dat het volgende bezwaar op: als in één en hetzelfde onderwerp zowel een casus (van zo'n 6 tot 8 punten) als een stelopdracht (van zo'n 12 tot 15 punten) zit, dan is er geen sprake meer van voldoende spreiding van de vragen over de stof. Daardoor zou het examen niet voldoende inhoudsvalide meer zijn. Om dit gevaar te ondervangen is besloten dat in het onderwerp waarin de stelopdracht valt (meestal het onderwerp dat voor het tweede jaar getoetst wordt) geen casus wordt afgevraagd ('septembermededeling van de Cevo'). Als docenten dit hadden geweten vóór dit examen, dan had hen dat wellicht een hoop tijd gescheeld.

Een derde misverstand is de aanwijzing: *Gebruik bron ...* Dat betekent geenszins dat de kandidaat de eigen kennis van het onderwerp even totaal moet vergeeten en alle aspecten van een antwoord uit de bron moet kunnen halen. Voor alle vragen bij bronnen heb je kennis nodig. Brongerelateerde vragen stellen we niet, maar in sommige gevallen dwalen we toch wat verder van de bron af, meestal met het doel de leerling beter op weg te helpen.⁴ Bijvoorbeeld in vraag 16 (over de Lombokexpeditie) waar (in de A-vraag, bedoeld als 'opstap' naar de volgende vragen) expliciet vermeld wordt dat hiervoor de bron niet gebruikt hoeft te worden. Criterium daarbij is: kan de vraag helemaal zonder de bron beantwoord worden? Als we dit criterium toepassen bij vraag 7 (over de DDR-vluchtelingen) dan heeft de kandidaat hier de bron wel nodig, om te zien dat deze slechts kwantitatieve gegevens bevat. Dat is ongetwijfeld voor een havist een moeilijke exercitie geweest. Maar mogen dit soort vragen dan helemaal niet meer gesteld in het examen?

Een laatste misverstand zit rondom de stelopdrachten. Gaat het daarbij op de havo altijd om inleving in een historisch figuur? Of om een briefopdracht? Nee dus. Examenmakers worstelen met het zoeken naar een goede vorm voor de stelopdrachten. Aanvankelijk is ervan uitgegaan dat de verschillen tussen havo- en vwo stelopdrachten zouden moeten liggen in de mate van inleving in een historisch figuur. Zo hebben de afgelopen generaties havisten zich

moeten inleven in de doopsgezinde patriot Menno Nauta, de middeleeuwse broeder Pierre, de negentiende eeuwse veenarbeider Jarich Nooitgedacht en de orthodox-protestant Gerrit Gerritse. Dat stuitte op steeds grotere bezwaren. Wat was de functie van die inleving? Is inleving niet veel moeilijker (en daarom meer op vwo-niveau) dan een 'straight' betoog?⁵

De examenmakers proberen aan deze bezwaren tegemoet te komen door het aantal lagen wat voor die inleving in een stelopdracht moet zitten, zoveel mogelijk te beperken. De havist moet snel aan het schrijven kunnen, zonder zich eerst door een ellenlange inleiding te worstelen, waar telkens naar teruggekeken moet worden. Om dat te bereiken wordt de inleving verplaatst van een persoon naar een concrete situatie. In havo 2001-1 is de vorm van een debat in de klas gekozen. In havo 2001-2 moest de kandidaat zich inleven in een fotoredacteur die zijn keuze voor bepaalde foto's in een boek over bondskanseliers verdedigt.⁶

Oorzaken

Een flink aantal reacties op de VGN-site ging in op de structurele verandering in de leerling-populatie van ons vak. Geschiedenis als keuzevak in de oude stijl kende weliswaar geen honderd procent enthousiaste en bewuste kiezers, maar toch relatief veel werkelijk geïnteresseerde leerlingen.

Aan die goede oude tijd is door de profielen abrupt een einde gekomen. De geschiedenisdocenten hebben nu te maken met grotere groepen leerlingen, die minder vrij voor hun vak gekozen hebben. Voeg daarbij de enorme terugloop in het aantal contacturen en het veel strakker voorgeschreven programma, en de frictie wordt duidelijk.

Examenmakers hebben de opdracht de stofomschrijving en de eindtermen uit domein A te toetsen. De stofomschrijving gaat uit van ongeveer 27 sluis per onderwerp, het niveau van de examens mag door de tweede fase niet achteruitgaan. Voor de spelers op dit veld is het een hele kunst al deze ballen in de lucht te houden. Dat we daar in één jaar niet in slagen ligt voor de hand. Dat we daaraan moeten blijven werken is duidelijk.

Wat te doen?

Hoe de omstandigheden ook zijn veranderd, er moet toch gewerkt wor-

den aan een goede voorbereiding op het centrale eindexamen.

In het PTA kan iedere school al aanpassen opnamen. Maak daarvoor gebruik van de Adelmund-maatregel die de mogelijkheid biedt een historisch overzicht als vervanging van drie domeinen op te voeren. In dit historisch overzicht kan aangesloten worden bij de examenonderwerpen, waardoor de leerlingen wat meer achtergrondkennis kunnen opdoen. Voor het komend examen (Indonesië en de Sovjet-Unie) zal dat overigens niet eenvoudig zijn; deze onderwerpen liggen wel erg ver uit elkaar in tijd en plaats. Probeer in het PTA mogelijkheden voor toetsing van de examenstof op te nemen. Via bijvoorbeeld het voorschrijven van de onderwerpen van de praktische opdrachten of in het oefenen met stelopdrachten zijn mogelijkheden te creëren. Houd minimaal twee maal twintig lesuren vrij voor de examenonderwerpen; het is werkelijk van belang dat de leerlingen de stof van de onderwerpen goed beheersen.

Het bekendmaken van de interne afspraak rondom de casus in de septembermededeling van de Cevo verschaft enige ruimte. Het oefenen met stelopdrachten kan in die ruimte gebeuren, en dat blijkt vruchten af te werpen. Het laten maken van oude examenopgaven (zeker als het dezelfde onderwerpen betreft) ook. Hulp daarbij kan gevonden worden in de oude 'Wieg-brief' met gehanteerde vraagformuleringen.⁷ Op grond van deze afspraken formuleren examenmakers hun vragen. Het verdient aanbeveling als docenten in hun eigen toetsen ook gebruik maken van deze formuleringen. Leerlingen weten daardoor beter wat er van hen verwacht wordt.

De meest gebruikte vormen in de examens zijn *Noem ...* en *Geef ...* (met een kort antwoord) en *Leg uit ...* (wat een langer antwoord inhoudt). Voor het havo-examen maken we zeer spaarzaam gebruik van formuleringen als *Ontleen ...*, *Vergelijk ...*, *Beredeneer ...* of *Hypothese*. Als het wel gebruikt wordt in het havo-examen is de vraag ingebed in meer ingevulde instructies in intro of vraag.

Overigens wil dit niet zeggen dat uitsluitend deze vraagformuleringen worden gebruikt. In de stelopdracht past bijvoorbeeld de instructie *Beschrijf*, meestal gevolgd door een nadere precisering in de vorm van *Doe dat door ...*

Een vraag als: *Wanneer vond dit plaats?* of iets dergelijks blijft uiteraard mogelijk.

Tenslotte

De examens worden gemaakt door de Citogroep in opdracht van de Cevo (waarvan de vaksectie de examens vaststelt). Het Cito maakt daarbij gebruik van een constructiegroep, docenten die werkzaam zijn in de eind-examenklassen van het voortgezet onderwijs. De *'knappe koppen van de Cevo'* maken dus niet gebruik van *'negentiende-eeuwse onderwijzers met hoofddact'* waarvan afgevraagd moet worden of zij *'ooit een havo-leerling in het echt gezien, gesproken, onderwezen, getoetst'* hebben om een examen in elkaar te zetten, om her en der wat (tendentieus) te citeren uit de VGN-site.

Uitgangspunt van de spreiding van de vragen over de stof is de stofomschrijving; die bepaalt wat de inhoud van de antwoorden moet zijn en waar de accenten moeten worden gelegd. De hoofdvraag van de stofomschrijving bepaalt de thematiek van het examen, de hoeveelheid regels die besteed worden aan een aspect, bepaalt de zwaarte van dit aspect. De inhoudsvaliditeit van het examen wordt op deze manier gewaarborgd.

Nog één keer terug naar de VGN-site en de reacties van de geschiedenisdocenten. *'Vorig jaar was ik ziedend van woede. Dit jaar is het echter al berusting'.* *'Een examen als het havo-examen 2001 vermoordt ons geschiedenisonderwijs', 'prutswork', 'druk de leerlingen op het hart zoveel mogelijk van buiten te leren'.*

Dit soort opmerkingen laten de betrokkenheid van geschiedenisdocenten bij hun leerlingen en hun vak zien. Die betrokkenheid is er ook bij Cevo, Cito en de constructiegroep. Het is een zware opgave in de snel veranderende dynamiek van het voortgezet onderwijs een voor iedereen aanvaardbaar centraal examen te maken, passend in de stringente eisen van stofomschrijving en eindtermen. In het verbeteren van de communicatie tussen leerkrachten en examenmakers kan nog veel winst geboekt worden. Meer transparantie in de oplossingen die examenmakers kiezen voor de problemen die de eisen van het examenprogramma bij de toetsconstructie opleveren, meer uitleg van de overwegingen van de examenmakers daarbij, meer discussie in bij-

Geef, noem, noteer:
Leg uit, toon aan:

kort-antwoord vragen
lang-antwoord vragen

Afweging

De verschillende aspecten van een (complex) historisch probleem worden vergeleken waarna op basis van argumenten beredeneerd gewicht moet worden toegekend aan die aspecten of argumenten. Bij dit type vraag moeten in het antwoord de stappen van een redenering zichtbaar zijn. Hierbij moet een bepaalde bewering met argumenten ondersteund worden. Het aantal te vermelden argumenten wordt aangegeven.

Beredeneer

Beargumenteer

Bewering of Stelling

Een in vrijwel ieder examen toegepaste vraagvorm waarbij de in de intro opgeworpen (soms aanvechtbare) historische uitspraak door de kandidaat met argumenten moet worden bevestigd, genuanceerd of ontkracht.

Hypothese

De kandidaat wordt geacht een historisch aanvaardbare hypothese op te stellen die ondubbelzinnig voortvloeit uit de aangereikte informatie (en eventueel toegevoegde eigen kennis).

Verandering

In het antwoord moet de situatie voor en de situatie na de verandering aangegeven worden.

Vergelijking

Ook hier dienen (meerdere) aspecten naast elkaar in het antwoord te worden opgenomen.

Maak aannemelijk

Bij deze vraagvorm moet de kandidaat met één of meerdere argument(en) aangeven dat sprake is van een hoge mate van historische waarschijnlijkheid.

Vraagformuleringen

voorbeeld *Kleio*, kan daartoe bijdragen. Dit artikel gaat daarmee verder op de vorig jaar ingeslagen weg.

Eventueel kunnen we allemaal terugval-len op tip nummer 7 van reactie 28 op de VGN-site: *Geef nu uiteindelijk maar toe dat u het ook allemaal niet meer weet maar dat er gelukkig mensen zijn die dat nog wel doen. Laat die nou toevallig net allemaal bij het Cito werken.* En zo is het maar net ... of toch niet?

Noten

- 1 Voor uitleg van dit begrip: zie www.cito-groep.nl, kenniscentrum, toetstechnische begrippenlijst. Voor het gemak: de p-waarde is de gemiddelde score op een vraag gedeeld door de maximaal haalbare score op die vraag. Een p-waarde van 10 betekent een (zeer) makkelijke, een p-waarde van 90 een (erg) moeilijke vraag. Gestreefd wordt naar een p'-waarde waarbij $40 < p' < 70$. De notatie dient eigenlijk te zijn .10, maar dat is minder leesbaar.
- 2 Uiteraard is het niet juist als docenten op grond van kritiek op bepaalde vraag- en antwoorden 'alles' van hun leerlingen goed rekenen. Daarmee zouden hun leerlingen dubbel worden beloond, zowel door hun docent als door de normstelling.
- 3 De docent die opmerkte dat er maar liefst 14 vragen zijn waar pure kennis voor nodig

is, heeft op twee manieren geen juiste analyse gemaakt: het gaat om veel minder reproductieve vragen en er worden in de door hem als reproductief bestempelde vragen ook vaardigheden getoetst.

- 4 Een brongerelateerde vraag is hier een vraag naar *aanleiding* van een bron. Een voorbeeld: een tekst over de revolutie in 1848 in Wenen en daarbij de vraag: wat gebeurde er nog meer in dat jaar? Het gaat bij dit soort vragen meer om associaties op dan om een directe analyse van de bron.
- 5 Leerkrachten met zowel havo- als vwo-examengroepen merkten de afgelopen jaren herhaaldelijk op: "Ik wou dat mijn havisten het vwo-betoog hadden gehad, dan was het veel beter gegaan!"
- 6 Moelijk, maar niet per se onmogelijk, is beantwoorden aan de 'roep uit het veld' om meer 'eigen mening' of standpuntvragen in het examen op te nemen. De stelopdracht biedt daarvoor weliswaar de ruimte, maar vooral het ontwikkelen van werkbare correctievoorschriften is daarbij geen sinecure.
- 7 Deze formuleringen zijn afkomstig uit de brief van de WIEG van 1 februari 1997 en de "oude" WIEG-brief van februari 1994; de formulering 'In hoeverre' is daaruit –na kritiek in onder ander *Kleio*– geschrapt.