Examenverslag 2004

havo/vwo

R.V.J. Alberts

Cito groep
Voortgezet onderwijs

Examenverslag
havo/vwo 2004

R.V.J. Alberts

Citogroep
Arnhem, oktober 2004

Art. nr. 58854

© Citogroep Arnhem 2004. Auteursrecht voorbehouden.
Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van de Citogroep worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie/reprografie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Woord vooraf

Het examenverslag geeft u een indruk van de wijze van produceren van de examens en van de resultaten in 2004. In opdracht van de CEVO (Centrale Examencommissie Vaststelling Opgaven) produceert de Citogroep jaarlijks de examens. De evaluatie, waarvan dit verslag een onderdeel is, behoort ook bij deze activiteiten.

Examens krijgen elk jaar in de examenperiode veel aandacht. Men is benieuwd naar de effecten van de veranderingen in het onderwijs. Daarnaast gaan examens steeds meer een ijkpunt vormen, niet alleen voor de niveaubepaling en certificering van de leerling, maar ook voor de resultaten van de school. Omdat examenresultaten onderdeel van de kwaliteitskaart uitmaken, wordt ook buiten de examenperiode over examens en examenresultaten van scholen gesproken. Centrale examens maken het mogelijk vergelijkingen met landelijke gemiddelden en vergelijkingen tussen scholen onderling te maken. Daarbij dient echter te worden opgemerkt dat centrale examens zich beperken tot een deel van het examen-programma. Alleen voor dat deel vormen de examens een ijkpunt.

Bij de samenstelling van de examens is, naast het leerstofinhoudelijke aspect, de moeilijkheidsgraad steeds van groot belang. Gelijkwaardigheid qua eisen is een van de billijkheidseisen waaraan centrale examens dienen te voldoen. Procedures van normvergelijking, waarvan in dit verslag ook gegevens opgenomen zijn, dragen ertoe bij dat de centrale examens een echt ijkpunt kunnen zijn en maken het mogelijk om prestaties van een gelijk niveau over verschillende jaren en tijdvakken heen met een zelfde cijfer te waarderen.
Het accent op de inhoudelijke ontwikkeling en de aandacht voor de niveaubepaling heeft een belangrijke rol gespeeld bij de ontwikkeling van de nieuwe examens voor de Tweede Fase havo/vwo.

Ik hoop dat dit verslag u een goed beeld geeft van het verloop en de resultaten van de centrale examens havo/vwo in 2004.
De gegevens waarop dit verslag gebaseerd is, ontvangen wij van de scholen. Zonder de medewerking van vele docenten en schoolleiders was dit verslag niet mogelijk. Wij willen dit woord vooraf besluiten met een woord van dank aan allen die aan de totstandkoming van dit verslag meegewerkt hebben.

drs. E.J.J. Kremers
programmamanager
centrale examens havo/vwo

Inhoud

Woord vooraf 3

Inleiding 7

1 Algemeen 9

1.1 De eindexamenkandidaten 9
1.2 Eindexamenvakken en vraagvormen 10
1.3 De constructie van opgaven en correctievoorschriften 10
1.4 De afname van de examens 11
1.5 De verzameling van afnamegegevens 11
1.6 Fouten in examenopgaven 13
1.7 Normering 13
1.8 Normvergelijking 18
1.9 Vergelijking van deelvak- en totaalvak-examens en populaties 21
1.10 Vergelijking van reguliere en experimentele examens en -populaties 21
1.11 Verschillen tussen de resultaten van jongens en meisjes 22
1.12 De examens tweede tijdvak 23

2 Nieuwe examenvormen 27

2.1 Nieuwe examenprogramma's 27
2.2 Computers en examens 27
2.2.1 Constructie van IMEX-examens 29
2.2.2 Technische aspecten 29
2.2.3 Conclusies 31
2.2.4 Beschrijving van de IMEX-onderdelen 32
2.3 Vooruitblik 2005 36

3 Examengegevens per vak 37

3.1 Cito-vakmedewerkers 37
3.2 Tabellen 38

Bijlage 49

1 Enkele begrippen 50

Inleiding

De examens 2004 stonden evenals andere jaren volop in de publiciteit. Toch kenden ze over het algemeen een rustig verloop. De reacties op de examens waren zoals dat gebruikelijk is: wisselend kritisch en positief.
Het LAKS (landelijk aktie komitee scholieren) verzamelt en ordent al sinds enkele jaren op systematische wijze de klachten van examenkandidaten. Het aantal klachten steeg tot nu toe van jaar tot jaar. Dat wordt vermoedelijk mede veroorzaakt door het gemak waarmee steeds meer kandidaten gebruik maken van internet.
In het algemeen was er sprake van een zekere mate van overeenstemming tussen de inhoudelijke reacties die het LAKS mocht ontvangen en de analyses van de examen-resultaten. De examens waarover in relatie tot het aantal kandidaten de meeste klachten binnen kwamen, bleken vaak bij nadere analyse van de resultaten ook de examens te zijn die een hogere moeilijkheidsgraad hadden dan was beoogd.
Het door het LAKS aangelegde bestand over de examens 2002 is door de Citogroep aan een nadere inhoudelijke analyse onderworpen. Uit die analyse blijkt geenszins dat de kwaliteit van de examens achteruit is gegaan. Het merendeel van de klachten was onvoldoende gespecificeerd om daar richtlijnen uit te destilleren voor de constructie van toekomstige examens. De verzameling van klachten door het LAKS is in 2004 verder aangescherpt en de klagers worden nu gevraagd hun klacht inhoudelijk te omschrijven. Dat biedt mogelijk meer aangrijpingspunten voor een inhoudelijke evaluatie van de examens.

Dit examenverslag geeft voor havo en vwo inzicht in de landelijke resultaten van de centrale examens 2004. In het eerste hoofdstuk wordt een aantal globale cijfermatige gegevens verstrekt en toegelicht. Tevens wordt nader ingegaan op de werkwijze bij de examen-productie, de verwerking van de gegevens en de normering. Speciale aandachtspunten zijn normvergelijking, niveauverschillen, verschillen tussen de resultaten van jongens en meisjes en de resultaten van het tweede tijdvak.
In het tweede hoofdstuk komen enkele nieuwe of bijzondere examenvormen aan de orde. Er wordt een samenvatting gegeven van de gang van zaken bij de afname van de experimentele examens waarbij de computer gebruikt werd. Vervolgens wordt per computer-examen een korte beschrijving gegeven van de toegepaste computerapplicaties. Ook wordt in dit hoofdstuk een korte vooruitblik op 2005 gegeven.
Het derde en laatste hoofdstuk bevat de cijfermatige gegevens van alle examens afzonderlijk. De gegevens over de behaalde resultaten zijn afkomstig van de informatie die de scholen verstrekt hebben. Het betreft hier bij het merendeel van de vakken de scores van de (alfabetisch) eerste vijf kandidaten per vak, na eerste correctie.
Sinds 1998 is er uit oogpunt van vergelijkbaarheid voor gekozen om bij de normering alleen te kijken naar de afnameresultaten van de kandidaten uit het reguliere VO. Bij de presentatie van de afnameresultaten in hoofdstuk 3 zijn in de steekproef dan ook geen kandidaten uit het VAVO opgenomen.

1 Algemeen

1.1 De eindexamenkandidaten

In 2004 namen er ongeveer 45.000 havo- en 32.000-vwo-kandidaten deel aan het Centrale Examen, eerste tijdvak, dat in mei afgenomen werd.
Tabel 1 geeft een overzicht van de aantallen eindexamenkandidaten die stonden ingeschreven voor het eerste tijdvak van de Centrale Examens 2004.
Ter voorkoming van misverstanden het volgende: de vermelde aantallen zijn de aantallen ingeschreven kandidaten bij het begin van het cursusjaar, zoals verstrekt door de Informatie Beheer Groep te Groningen. Deze aantallen zijn in de regel een overschatting van de aantallen kandidaten die daadwerkelijk aan het examen deelnamen.

Tabel 1 Aantal examenkandidaten[1)]

	havo	vwo
Nederlands (leesvaardigheid)	45.176	31.574
Fries leesvaardigheid	50	9
Latijn		5.698
Grieks		2.086
Frans 1,2	8.733	7.698
Duits 1,2	11.588	6.911
Engels	45.792	31.967
Spaans (1,2)	605	773
Russisch (1,2)	3	13
Turks (1,2)	24	5
Arabisch (1,2)	9	12
Wiskunde A1		5.692
Wiskunde A1,2	19.956	11.610
Wiskunde B1	7.842	9.396
Wiskunde B1,2	6.033	6.592
Natuurkunde 1	7.729	9.306
Natuurkunde 1,2	5.545	5.660
Scheikunde 1		8.338
Scheikunde (1,2)	12.973	6.970
Biologie (1,2)	16.535	11.893
Economie 1	18.048	6.798
Economie 1,2	19.718	11.987
Management & Organisatie	11.883	7.432
Aardrijkskunde	22.079	13.523
Geschiedenis en Staatsinrichting	33.164	17.698
Maatschappijleer	2.214	1.315
Filosofie	469	2.427
Muziek	1.159	687
Tehatex	5.812	2.675
CKV 2	5.305	2.480

[1)] In deze aantallen zijn ook de examenkandidaten van de Antillen opgenomen. De cijfers voor Nederland liggen derhalve iets lager. Daarnaast ligt het feitelijk aantal kandidaten gemiddeld enkele procenten lager dan het opgegeven aantal, omdat scholen een zekere veiligheidsmarge in hun bestellingen inbouwen.

1.2 Eindexamenvakken en vraagvormen

In tabel 1 zijn de algemene eindexamenvakken genoemd die in 2004 in het havo en vwo centraal geëxamineerd zijn en waarvoor de Citogroep de productie en verwerking verricht. De specifiek oude stijl examens voor vwo die er in 2004 voor het laatst waren zijn hier buiten beschouwing gelaten.

Sinds 2001 kennen de algemene vakken geen examens meer die volledig uit machinaal scoorbare vragen bestaan. De antwoorden op de open vragen of opdrachten worden beoordeeld door docenten aan de hand van door de CEVO vastgestelde correctie-voorschriften. De examens die zowel open als meerkeuzevragen bevatten, worden eveneens geheel door docenten gescoord. De meerkeuzevragen in deze examens worden gescoord aan de hand van de in het correctievoorschrift opgenomen antwoordsleutels. De scoring van examens wordt tweemaal uitgevoerd: eerst door de eigen docent en vervolgens door een docent van een andere school, de tweede corrector.
De examens voor de moderne vreemde talen bestaan voor ongeveer 60% uit meerkeuze-vragen. Daarnaast bestaan ook de examens biologie en Nederlands havo en vwo voor een gedeelte uit open vragen en voor een gedeelte uit meerkeuzevragen. De examens beeldende vakken, Latijn en muziek bevatten een beperkt aantal meerkeuzevragen naast overwegend of (soms uitsluitend) open vragen. Tabel 2 geeft een overzicht van de gebruikte vraagvormen in de examens.

Tabel 2 Overzicht van toegepaste vraagvormen in de centrale examens

	havo	vwo
Latijn		□
Grieks		◪
Nederlands	◪	◪
Fries	◪	◪
Frans 1,2	◪	◪
Duits 1,2	◪	◪
Engels	◪	◪
Spaans (1,2)	◪	◪
Russisch (1,2)	◪	◪
Turks (1,2)	◪	◪
Arabisch (1,2)	◪	◪
Wiskunde	□	□
Natuurkunde	□	□
Scheikunde	□	□
Biologie (1,2)	◪	◪
Economie	□	□
Management & Organisatie	□	□
Aardrijkskunde	□	□
Geschiedenis / staatsinrichting	□	□
Filosofie	□	□
Maatschappijleer	□	□
Muziek	◪	◪
Tehatex	◪	□
CKV 2	□	□

□ open vragen
◪ open en gesloten vragen

1.3 De constructie van opgaven en correctievoorschriften

Bij de constructie van opgaven en correctievoorschriften is een groot aantal docenten uit het voortgezet onderwijs betrokken. De verantwoordelijkheid voor de vaststelling van opgaven en correctievoorschriften ligt bij de CEVO (Centrale Examencommissie Vaststelling Opgaven). In het algemeen bestuur van de CEVO zijn de algemene onderwijsorganisaties en het

vervolgonderwijs vertegenwoordigd. De CEVO kent voor de verschillende vakken vaksecties. Het voorzitterschap van de CEVO-vaksecties wordt vervuld door een vakdeskundige uit het vervolgonderwijs. Naast de voorzitter hebben twee of drie docenten zitting in de vaksectie. Deze docenten worden voorgedragen door de diverse algemene en vakinhoudelijke onderwijsorganisaties.
Het eigenlijke constructiewerk wordt verricht door constructiegroepen waarvan docenten met ervaring in examenklassen deel uitmaken. Per examen worden door een drietal vakdocenten onder leiding van een Cito-medewerker conceptopgaven geschreven en correctie-voorschriften opgesteld.

Bij de constructie van examenopgaven en correctievoorschriften wordt gewerkt met examen-modellen. In een examenmodel zijn de productspecificaties ten aanzien van het uiteindelijke examen opgenomen, zoals:
– de verdeling van de vragen over de examenstof;
– de mate waarin het examen vragen met een reproductief en/of productief karakter moet bevatten;
– de keuze van teksten en contextmateriaal;
– de toe te passen vraagvormen en vaardigheidsvragen;
– de wijze waarop de correctievoorschriften opgesteld dienen te zijn.
De examenmodellen worden door de CEVO vastgesteld en fungeren voor de Citogroep en de constructiegroepen als opdracht voor de constructie. Door het gebruik van een examen-model wordt bevorderd dat elk examen een evenwichtige representatie van het examen-programma is en dat examens van verschillende tijdvakken en verschillende jaren inhoudelijk als gelijkwaardig beschouwd kunnen worden.

De conceptexamens worden van commentaar voorzien en beoordeeld door de vaksecties van de CEVO. Aan de commentaarrondes wordt onder verplichting van strikte geheimhouding ook deelgenomen door niet-schoolgebonden deskundigen, zoals universitaire deskundigen voor speciale onderwerpen en voor de moderne vreemde talen deskundigen die de desbetreffende taal als moedertaal spreken.
De CEVO-vaksectie stelt ten slotte de opgaven en het correctievoorschrift in hun definitieve vorm vast.
De constructie van opgaven en correctievoorschriften betreft elk jaar drie tijdvakken: het eerste tijdvak, dat in de tweede helft van mei afgenomen wordt, het tweede tijdvak (de herkansing), dat in juni wordt afgenomen, en het derde tijdvak (afname in augustus) dat onder meer bedoeld is voor kandidaten die – om geldige redenen – niet aan de eerste twee tijdvakken konden deelnemen.
In totaal zijn er dit jaar voor de algemene vakken havo/vwo 191 examens geconstrueerd, 73 voor het eerste tijdvak, 61 voor het tweede tijdvak en 57 voor het derde tijdvak.

1.4 De afname van de examens

De periodes waarin de Centrale Examens afgenomen worden, zijn ruim voor het begin van het cursusjaar aan de scholen bekendgemaakt. Het examenrooster 2004 vermeldde 18 mei tot en met 4 juni als de periode van het eerste tijdvak en 23 juni als de belangrijkste dag van het tweede tijdvak. Voor enkele weinig gekozen vakken wordt de afname van het tweede tijdvak door de staatsexamencommissie georganiseerd. Deze afnames vonden plaats op vrijdag 25 juni. De examens van deze aangewezen vakken worden niet gepubliceerd. Het derde tijdvak begon op 16 augustus.
De organisatie van de afname van de eerste- en het merendeel van de tweede-tijdvak-examens is in handen van de schoolleiding. Het correctiewerk wordt verricht door aan de school verbonden docenten/examinatoren. De correctieperiode loopt door tot enkele dagen vóór de datum waarop de uitslag wordt meegedeeld.
In dezelfde periode komen bij de Citogroep dagelijks duizenden optisch leesbare formulieren binnen om machinaal verwerkt te worden. Daarover meer in de volgende paragraaf.

1.5 De verzameling van afnamegegevens

De activiteiten voor de verzameling van afnamegegevens ten behoeve van de normering beginnen lang voor de examenperiode. In november vindt de inschrijving van kandidaten bij

de Productgroep examendiensten van de Informatie Beheer Groep in Groningen plaats. Deze gegevens worden ook door de Citogroep gebruikt. Op basis van deze inschrijf-gegevens worden scholen door de Citogroep aangeschreven en wordt de keuze voorgelegd op welke wijze scholen de afnamegegevens wensen te verstrekken.
Dit jaar is het merendeel van de afnamegegevens via elektronische weg verzameld. Iets minder dan de helft van de gegevens is via optisch leesbare formulieren verzameld. Beide werkwijzen worden hieronder kort aan de orde gesteld.

Gegevensverzameling via optisch leesbare formulieren
Begin april zijn de pakketten met formulieren voor de correctoren aan de scholen toegestuurd. Op deze formulieren worden de gegevens van vijf kandidaten (of voor enkele vakken tien) per school verzameld. De gegevens uit deze steekproef worden op de normeringsvergaderingen gebruikt om na te gaan hoe de examens door de kandidaten gemaakt zijn.
Voor de specifieke oude-stijl-vwo-examens zijn dit jaar geen gegevens meer opgevraagd. Voor het tweede tijdvak werden, net als vorig jaar, op formulieren voor de corrector de gegevens van vijf kandidaten of minder per school verzameld. Tevens werd gevraagd het cijfer voor het eerste tijdvak aan te strepen. Deze gegevens werden gebruikt voor een vergelijking tussen eerste en tweede tijdvak.
Op het formulier voor de corrector diende per open vraag de behaalde score aangestreept te worden en per gesloten vraag het gekozen alternatief. Dit laatste gegeven is essentieel om de kwaliteit van de afleiders (de onjuiste alternatieven) te kunnen beoordelen. Sinds 2001 bestaat de mogelijkheid om het vraagnummer door te strepen in geval een kandidaat de vraag in het geheel niet beantwoord heeft. De mate waarin met name vragen aan het eind van het examen door kandidaten niet beantwoord zijn, kan geïnterpreteerd worden als een indicatie voor tijdnood en kan er op wijzen dat een examen als geheel te bewerkelijk is geweest.

Gegevensverzameling via elektronische weg
Er zijn technisch gesproken ook andere mogelijkheden beschikbaar voor het verzamelen van afnamegegevens. Een belangrijke impuls voor vernieuwing ging uit van het afschaffen van de antwoordbladen in 2001. Antwoordbladen waren optisch leesbare formulieren die de leerlingen zelf invulden bij de 100%-meerkeuze-examens.
De verzameling van afnamegegevens via optisch leesbare formulieren is niet alleen kostbaar, bij sommige vakken treedt ook uitval van gegevens op omdat bij het invullen van de formulieren door de correctoren fouten worden gemaakt.
Door elektronische invoer en uitwisseling van afnamegegevens wordt het mogelijk om invoerfouten te minimaliseren en de snelheid van de gegevensverzameling en -verwerking te verhogen. Sinds 2001 wordt ervaring opgedaan met het electronisch verzamelen van afnamegegevens. Scholen nemen op basis van vrijwilligheid deel aan deze nieuwe werkwijze. Het aantal scholen breidt zich elk jaar verder uit.
Dit jaar hebben ongeveer de helft van de havo/vwo-scholen meegedaan aan het experiment WOLF2004. Wolf staat voor: Windows Optisch Leesbaar Formulier. Dat betekent dat de docenten van examenklassen na het examen de scoregegevens van de alfabetisch eerste vijf (of tien) leerlingen niet op papieren (optisch leesbare) formulieren aanstrepen, maar via een computerprogramma hebben ingevoerd. De school- en examengegevens zijn via een pincode te downloaden van internet en de ingevulde scores worden weer via het web teruggestuurd naar de Citogroep. WOLF2004 neemt de gebruiker bij de hand; het is te vergelijken met het elektronisch aangifteformulier van de belastingdienst. Wordt er iets fout ingevuld of overgeslagen, dan meldt het programma dat direct.
Het voordeel voor de docent is dat hij met behulp van WOLF2004 tegelijkertijd een score-overzicht voor de tweede corrector kan samenstellen. Er hoeven dus maar één keer gegevens te worden ingevoerd. Ook voor de Citogroep zal (op termijn) het programma zijn vruchten afwerpen: de data die we op deze manier binnenkrijgen zullen eerder beschikbaar zijn en gebleken is dat er geen fouten meer in de gegevens zitten.
Helaas bleek er dit jaar tijdens de examenperiode een ernstige onvolkomenheid in het programma te zitten. Als gevolg daarvan zou het kunnen voorkomen dat onder bepaalde omstandigheden aan kandidaten een verkeerde totaalscore wordt toegekend. Omdat de consequenties daarvan zo ernstig zijn, is de fout in het programma onmiddellijk onder de aandacht van alle scholen gebracht en werden de scholen verzocht om in die gevallen waarin de scores met behulp van WOLF zijn vastgesteld nog een extra controle uit te voeren. In overleg met de CEVO zal bekeken worden of in de campagne 2005 op grotere schaal gebruik gemaakt gaat worden van de elektronische gegevensverzameling. Op termijn wordt gestreefd naar invoering bij alle scholen.

Het aantal formulieren (inclusief de elektronische) dat in de analyses verwerkt werd, bedroeg voor havo/vwo voor het eerste tijdvak ruim 75.000 en voor het tweede tijdvak 10.000.

1.6 Fouten in examenopgaven

Ondanks het feit dat elke examenopgave door verschillende groepen deskundigen onafhankelijk van elkaar bekeken wordt en dat drukproefcorrecties door verschillende betrokkenen uitgevoerd worden, is het nooit helemaal uitgesloten dat er in een of enkele examens een storende fout over het hoofd wordt gezien. Wanneer deze fout nog bij de laatste controle, enkele weken voor het examen, geconstateerd wordt, ontvangen de scholen een erratumblad. Een fout in de opgaven moet bij het begin van het examen worden voorgelezen en is dus zeer storend voor de kandidaten. Een fout in het correctievoorschrift moet ter kennis worden gebracht van de examinatoren/correctoren. Wanneer de keuze-mogelijkheid aanwezig is, wordt de voorkeur gegeven aan een erratumblad voor het correctievoorschrift boven een erratumblad voor de opgaven.
Dit jaar moesten er voor de eerste-tijdvak-havo- en -vwo-examens vijf erratumbladen voor de opgaven en zes erratumbladen voor de correctievoorschriften verzonden worden. Voor het tweede tijdvak moesten er één erratumblad voor de opgaven en vijf voor de correctie-voorschriften verzonden worden.
Scholen krijgen van de Informatie Beheer Groep een overzicht van de examens waarvoor een erratumblad verschenen is. Dit wordt ook vermeld op internet.
Blijkt er bij de afname desondanks nog een fout te staan in een examen of in een correctie-voorschrift, dan publiceert de CEVO op het “Examenblad” een tussentijds bericht over de wijze waarop deze vraag beoordeeld moet worden en distribueert dit bericht op elektronische wijze. Dit jaar was dat bij 7 examens in het eerste tijdvak nodig, en bij één examen in het tweede tijdvak. Een tussentijds bericht kan alleen gepubliceerd worden wanneer de fout kort na afname van het examen ontdekt is. Wordt tijdens de normeringsvergadering besloten dat een vraag op vakinhoudelijke gronden toch minder gelukkig was, dan rest de mogelijkheid dit in de normeringsterm te verdisconteren. We spreken dan van een incidentele aanpassing. Wanneer een bepaalde druk- of zetfout géén probleem oplevert bij de beantwoording van de vraag, is dat uiteraard niet aan de orde.
In 2004 lag het aantal errata op iets meer dan 0,6% van het aantal vragen. Voor de Citogroep geldt echter dat elke fout er een te veel is en wordt voortdurend getracht het aantal fouten te verminderen.

1.7 Normering

De normeringsvergaderingen voor de eerste-tijdvak-examens vonden dit jaar plaats op 16 juni. De normeringsbeslissingen werden een dag later bekend gemaakt via de websites www.eindexamen.nl en www.citogroep.nl. De verzending van de normeringstermen op papier vond voor het eerste tijdvak op donderdag 17 juni plaats.

Onder normering van examens verstaan we in het algemeen de procedure die gevolgd wordt bij de omzetting van scores in cijfers. Sinds 2000 wordt genormeerd volgens de ‘CEVO-methode’. Deze methode is (m.u.v. Nederlands stelopdracht) voor alle examens toepasbaar ongeacht de gehanteerde vraagvormen of de lengte van de scoreschaal. De vaststelling van de normeringsterm is de belangrijkste beslissing.
De normeringsterm is een ijkvariabele die zo wordt vastgesteld dat in verschillende jaren en tijdvakken even zware eisen aan kandidaten gesteld kunnen worden om een voldoende voor het centrale examen te halen. De normeringsterm kan variëren van 0,0 tot en met 2,0.

In de besluitvorming ten aanzien van de normering zijn globaal genomen drie stappen te onderscheiden. De eerste stap betreft de advisering door de Cito(vak)medewerker op basis van de geanalyseerde afnamegegevens. De Cito(vak)medewerker stelt het advies op voorafgaand aan de normeringsvergadering van de vaksectie.

De tweede stap betreft de advisering door de vaksectie aan het dagelijks bestuur van de CEVO. Op basis van het Cito-advies alsook een vakinhoudelijke beoordeling van het examen en weging van binnengekomen reacties en commentaren stelt de CEVO-vaksectie een normeringsadvies op voor het dagelijks bestuur van de CEVO.

Bij de derde stap inventariseert het dagelijks bestuur van de CEVO alle adviezen en beoordeelt deze o.a. op consistentie over vakken en schooltypen en stelt vervolgens de definitieve normering vast.
De advisering door de Cito(vak)medewerker berust voornamelijk op de interpretatie van de analyses van de afnamegegevens. Het advies van de Cito(vak)medewerker is in eerste instantie vooral een technisch advies. Het is technisch in die zin dat op basis van de verzamelde afnamegegevens een zodanige normering wordt voorgesteld dat over jaren en tijdvakken heen gelijke eisen aan kandidaten worden gesteld. De norm die als uitgangspunt wordt genomen is door de CEVO bepaald door voor elk vak een referentie-examen aan te wijzen met een daarbij toe te passen referentienorm. De Cito(vak)medewerker adviseert voor het nieuwe examen een normering die equivalent is aan de referentienorm op het referentie-examen.
De gegevensverzameling en de daarop aansluitende normeringsprocedures zijn niet voor alle vakken volledig identiek. Voor sommige examens zijn er meer en andersoortige gegevens beschikbaar om de normeringsbeslissing te ondersteunen. Voor die examens waarbij de gegevensverzameling minder volledig is moeten noodzakelijkerwijze enkele aannamen worden gedaan. De waarde van het Cito-advies is afhankelijk van de juistheid van die aannamen. Maar gegeven het uitgangspunt is de redeneertrant transparant. Het Cito-advies heeft geen ‘absolute’ pretenties, maar als de geëxpliciteerde aannamen juist zijn dan zou ook het resultaat juist moeten zijn. We zullen dat hieronder nader toelichten.

Bij de gegevensverzameling worden globaal de volgende varianten onderscheiden:
1. de standaardprocedure van gegevensverzameling die voor (nagenoeg) alle examens geldt. Dit houdt in dat er bij een steekproef van ongeveer 2000 kandidaten deelscores per vraag worden verzameld;
2. sommige examens met overwegend de gesloten vraagvorm zijn naast de standaard-procedure ook nog opgenomen in de procedure ‘normvergelijking achteraf’ (dit zijn de examens Frans, Duits en Engels voor zowel havo als vwo;
3. sommige examens met uitsluitend of overwegend open vragen zijn naast de standaard-procedure ook nog opgenomen in de procedure ‘normvergelijking vooraf’ (dit zijn naast de exacte vakken op havo en vwo-niveau ook Nederlands, economie en biologie);
4. tot slot zijn er de examens die naast de standaardprocedure ook nog overlap vertonen met een ander examen en waarbij die examens in relatie tot elkaar worden genormeerd. Dit geldt met name voor experimentele examens die altijd een zekere mate van overlap vertonen met het reguliere examen. Op basis van overlap tussen beide examens worden aanvullende analyses uitgevoerd.

Ad 1) De standaardprocedure is als volgt. Het definitieve correctievoorschrift wordt vóór afname van het examen vastgesteld. Tegelijk met de opgaven wordt dit definitieve voorschrift naar de scholen gestuurd. Bij de constructie van het merendeel van de examens is een gemiddelde moeilijkheidsgraad nagestreefd zodat in de regel een normeringsterm tussen 0,5 en 1,3 kan gelden[1]. De docent/examinator kan onmiddellijk na afname van het examen met de correctie beginnen en op grond van het correctievoorschrift een score bepalen.
De normeringsterm wordt definitief vastgesteld op basis van een steekproef van de scores van kandidaten na eerste correctie – de alfabetisch eerste 5 kandidaten per vak – en tevens op basis van reacties van docenten. Bij het opstellen van het Cito-advies wordt aangenomen dat de populaties van jaar tot jaar een vergelijkbaar prestatieniveau hebben. Een verschil in de gemiddelde score met het referentie-examen wordt derhalve gezien als een gevolg van een verschil in moeilijkheidsgraad met het referentie-examen. Aan de hand van de analyse-resultaten wordt vervolgens nagegaan met welke normeringsterm het verschil in moeilijkheidsgraad adequaat gecompenseerd kan worden.

Ad 2) De procedure voor examens die zijn opgenomen in de ‘normvergelijking achteraf’ verloopt als volgt. Bij de samenstelling en de vaststelling van de examens wordt een zelfde moeilijkheidsgraad nagestreefd als die van een vooraf aangewezen referentie-examen met de daarbij behorende referentienorm. Na afname van de examens wordt de feitelijke moeilijkheidsgraad van het nieuwe examen vergeleken met die van het referentie-examen en wordt de norm van het referentie-examen overgebracht op het nieuwe examen. Dit gebeurt op basis van de normvergelijkingsgegevens. Bij deze werkwijze wordt er niet voorshands van uitgegaan dat populaties over jaren heen een zelfde prestatieniveau hebben.

[1] De ondergrenzen die voor 2004 golden voor het bepalen van de N-term zijn gepubliceerd in de maartmededeling 2004.

Ad 3) De gegevens van de ‘normvergelijking vooraf’ worden verzameld door middel van een pretest. Deze gegevens zijn reeds verwerkt bij de samenstelling van de definitieve examens en het opstellen van de correctievoorschriften. Deze examens zijn meestal met een normeringsterm van 1,0 equivalent aan het aangewezen referentie-examen. Ook voor deze examens wordt niet voorshands aangenomen dat populaties over jaren heen een zelfde prestatieniveau hebben. Alleen wanneer na de pretest nog ingrijpende veranderingen in het examen zijn aangebracht geldt een vergelijkbare werkwijze als hierboven beschreven onder Ad 1.

Ad 4) Bij de experimentele examens (zie ook § 1.10) speelt de overlap met de reguliere examens een belangrijke rol. Hier wordt op basis van de overlap getracht de beide examens zo te normeren dat aan beide groepen kandidaten vergelijkbare eisen worden gesteld.
Een belangrijk uitgangspunt bij de definitieve vaststelling van de normeringsterm is dat de kandidaten die aan het experimentele examen deelnemen niet gedupeerd mogen worden door het experimentele karakter van het examen.

tweede-tijdvak-normering
In principe heeft het tweede tijdvak dezelfde normeringsterm als het eerste tijdvak. Wanneer bij de vaststelling van de normeringsterm voor het eerste tijdvak het feit van één of meer onjuiste vragen in dat examen is meegewogen, dan werkt dit niet door naar het tweede tijdvak. De ‘voorlopige’ normeringsterm voor het tweede tijdvak is in dat geval de normeringsterm die vastgesteld zou zijn als er geen onjuiste vraag in het eerste tijdvak had gezeten.
De basale aanname daarbij is dat de eerste- en tweede-tijdvak-examens, die gelijktijdig geconstrueerd en vastgesteld worden, gelijkwaardige examens zijn.
De voorlopige normeringstermen voor het tweede tijdvak zijn voorlopig in die zin dat ze nooit ten nadele van de kandidaten zullen worden bijgesteld. De normeringstermen voor het tweede tijdvak zullen nooit lager worden vastgesteld. Ze worden alleen gewijzigd wanneer uit de vergelijking van het eerste en tweede tijdvak onomstotelijk blijkt, dat het tweede tijdvak moeilijker is dan het eerste tijdvak. De vergelijking van het eerste en tweede tijdvak vindt plaats aan de hand van de scores van het tweede tijdvak en het cijfer voor het eerste tijdvak. Deze worden direct na de correctie op optisch leesbare formulieren naar de Citogroep opgestuurd.

Score-cijfer-transformatie
Behalve de normeringsterm is ook de wijze van omzetten van score naar cijfer bij de normering van belang. Sinds 2000 geldt ook hiervoor de ‘CEVO-methode’ gevolgd.
Uitgangspunten bij de omzetting van score naar cijfer zijn:
– elk volgend gescoord punt leidt altijd tot een hoger examencijfer (afgezien van afrondingen);
– een score van 0% correspondeert altijd met het examencijfer 1;
– een score van 100% correspondeert altijd met het examencijfer 10;
– over een zo breed mogelijk centraal interval van de scoreschaal is sprake van een constante stijging van het cijfer per extra scorepunt (onafhankelijk van de normering).

De omzetting van scores naar cijfers dient op basis van door de CEVO vastgestelde formules te worden uitgevoerd (zie Uitleg Gele Katern nr. 8 van 22 maart 2000).
Voor elk examen zijn na de normering omzettingstabellen gegenereerd. Deze zijn per e-mail verzonden aan die scholen die zich daarvoor hebben aangemeld. Daarnaast werden de omzettingstabellen gepubliceerd op de sites www.eindexamen.nl en www.citogroep.nl.
Deze omzettingstabellen zijn berekend volgens de vastgestelde formules en de vastgestelde normeringstermen.

In de hierna volgende tabellen wordt per schooltype en vak een overzicht gepresenteerd van de normeringsbeslissing en het daarmee corresponderende percentage onvoldoendes.

Havo
De afgelopen jaren (2002 en 2003) leverden met name de wiskunde-examens gezien de relatief hoge percentages onvoldoendes de meeste moeilijkheden op voor de kandidaten. Daarnaast gaven ook de economische vakken en geschiedenis en aardrijkskunde relatief hoge percentages onvoldoendes te zien.

Voor de andere vakken leek de afstemming van de moeilijkheidsgraad van de nieuwe examens op het prestatieniveau van de leerlingenpopulatie goed geslaagd. In enkele gevallen waren de examens iets aan de makkelijke kant.
In 2004 is het beeld iets gematigder maar niet wezenlijk anders. Wiskunde en economie laten in 2004 nog steeds bovengemiddelde percentages onvoldoendes zien.Geschiedenis vertoont echter een duidelijk trendbreuk doordat het examen 2004 relatief een stuk makkelijker is.

Enkele examens bleken dit jaar voor de derde keer op rij aan de makkelijke kant te zijn. Dat geldt o.a. voor de vakken natuurkunde 1,2 en scheikunde. Het laagste percentage onvoldoende treffen we aan bij natuurkunde 1,2 compex. Het gaat hier echter om een relatief kleine groep die het in verhouding tot de reguliere groep relatief goed doet.

In tabel 3 worden de resultaten voor havo vermeld. Gemiddeld genomen werd voor de eerste tijdvak examens een N-term van 1,0 vastgesteld. Het gemiddelde van de percentages onvoldoende bedraagt voor havo 24%.

Tabel 3 Normeringstermen en percentages onvoldoendes examens havo, 1e tijdvak

vak	L (lengte van de scoreschaal)	N (normerings-term)	% onvoldoendes	incidentele aanpassing[2)]
Nederlands	50	1,4	30	
Frans 1,2	45	0,6	34	–
Duits 1,2	47	0,6	23	–
Engels	50	0,6	22	–
Spaans 1,2	46	0,8	29	–
Russisch 1,2	48	0,7	–	–
Turks 1,2	48	0,7	–	–
Arabisch 1,2	47	0,7	–	–
Wiskunde A 1,2	83	1,1	29	0,1
Wiskunde B 1	81	0,6	32	–
Wiskunde B 1,2	82	1,1	28	–
Natuurkunde 1	81	0,8	24	–
Natuurkunde 1,2	81	0,5	16	–
Natuurkunde 1,2 compex	80	0,5	7	–
Scheikunde	81	0,7	25	–
Biologie	87	1,4	25	0,1
Biologie compex	84	1,5	23	–
Economie 1	58	1,3	26	–
Economie 1 compex	59	1,4	24	–
Economie 1,2	59	1,2	33	–
Management & organisatie	69	1	28	–
Aardrijkskunde	60	1,4	21	0,2
Aardrijkskunde compex	58	1,8	14	–
Geschiedenis	81	0,5	21	–
Maatschappijleer	88	0,8	29	–
Filosofie	48	0,9	19	–
Muziek	90	1,3	19	–
Tehatex	74	1,4	23	–
CKV 2 [1)]	70	1,0	19	–

[1)] Geen Centraal Examen

[2)] De N-term is in verband met een minder geslaagde vraag met genoemde factor aangepast.

Vwo-examens nieuwe stijl
Bij de vwo-examens bleken enkele examens achteraf toch iets moeilijker te zijn dan beoogd. Dit betrof de examens Latijn, Grieks en wiskunde A1 en A1,2.
Ook vorig jaar bleken enkele van deze vakken moeilijk voor de kandidaten. Het percentage onvoldoenden bij deze vakken is ondanks een relatief hoge N-term aan de hoge kant. Globaal genomen zijn de percentages onvoldoendes in 2004 redelijk vergelijkbaar met die van 2003.
Enkele examens zoals Frans, Spaans, Scheikunde 1, Geschiedenis, Management en organisatie en filosofie bleken relatief aan de makkelijke kant.
Het laagste percentage onvoldoende treffen we aan bij natuurkunde 1,2 compex. Het gaat hier echter net als bij havo natuurkunde compex om een relatief kleine groep die in verhouding tot de reguliere groep relatief goed presteert.

In tabel 4 worden de resultaten voor vwo vermeld. Gemiddeld genomen werd voor de eerste tijdvak examens een N-term van 0,9 vastgesteld. Het gemiddelde van de percentages onvoldoende bedraagt voor vwo 21%.

Tabel 4 Normeringstermen en percentages onvoldoendes examens vwo nieuwe stijl, 1e tijdvak

vak	L (lengte van de scoreschaal)	N (normerings-term)	% onvoldoendes	incidentele aanpassing[2)]
Nederlands	48	1,1	19	–
Latijn	86	1,5	26	–
Grieks	86	1,6	29	
Frans 1,2	46	0,0	27	–
Duits 1,2	47	0,4	23	–
Engels	51	1,1	25	–
Spaans	50	0,2	20	–
Russisch	48	0,7	–	–
Turks	50	0,7	–	–
Arabisch	46	0,7	–	–
Wiskunde A 1	83	1,4	24	0,1
Wiskunde A 1 compex	87	1,4	17	–
Wiskunde A 1,2	87	1,5	32	0,1
Wiskunde A 1,2 compex	90	1,6	28	–
Wiskunde B 1	86	0,8	29	–
Wiskunde B 1,2	86	1,2	26	–
Natuurkunde 1	79	1,3	18	0,2
Natuurkunde 1,2	79	0,8	14	–
Moderne Natuurkunde	82	0,9	14	–
Natuurkunde 1,2 compex	80	0,8	10	–
Scheikunde 1	67	0,5	16	–
Scheikunde 1,2	71	0,7	17	–
Biologie 1,2	77	1,3	17	–
Biologie 1,2 compex	71	1,3	10	–
Economische wetenschappen 1	56	0,9	27	–
Economische wetenschappen 1,2	56	0,7	31	0,1
Management & organisatie	70	0,5	26	–
Aardrijkskunde	72	1,0	27	
Geschiedenis	84	0,5	13	–
Maatschappijleer	87	0,7	17	–
Filosofie	44	0,5	19	–
Muziek	90	1,3	18	–
Tekenen, handvaardigh., text. werkv.	73	1,2	18	–
CKV 2 [1)]	67	1,0	17	–

[1)] Geen Centraal Examen
[2)] De N-term is in verband met een minder geslaagde vraag met genoemde factor aangepast.

1.8 Normvergelijking

In de vorige paragraaf is aangegeven dat de gegevensverzameling niet voor alle examens identiek is. Sommige examens zijn opgenomen in procedures van normvergelijking. Daarbij wordt onderscheid gemaakt tussen 'normvergelijking achteraf' en 'normvergelijking vooraf'.
Normvergelijkingsprocedures 'achteraf' zijn in 2004 toegepast bij de eerste-tijdvak-examens Frans, Duits en Engels zowel voor havo als vwo.
Normvergelijkingsprocedures 'vooraf' zijn voor de examens 2004-1 toegepast bij

- wiskunde A en B, natuurkunde, scheikunde (zowel voor havo als vwo);
- economie 1 en 1,2 en Nederlands leesvaardigheid (zowel voor havo als vwo);
- biologie (vwo).

In deze procedures worden onderdelen van de nieuwe examens gecombineerd met onderdelen uit een referentie-examen afgenomen. Het referentie-examen is tevoren door de CEVO uitgekozen en voorzien van een referentienorm die men als een redelijke eis beschouwt bij het betreffende referentie-examen.
De afnamen maken uitspraken mogelijk over de moeilijkheidsgraad van het examen 2004 in vergelijking met het referentie-examen. Ook kunnen uitspraken worden gedaan over de populatie 2004 in vergelijking met de populatie van het referentiejaar. De CEVO gebruikt deze gegevens bij de vaststelling van het examen of bij de vaststelling van de normerings-term na de afname van het examen.

Normvergelijking 'achteraf'
Voor de talenexamens havo werd de normvergelijking reeds voor het elfde achtereenvolgende jaar uitgevoerd. Voor de vwo-examens is deze normvergelijking dit jaar voor de achtste maal uitgevoerd. Voor Frans zijn echter onvoldoende gegevens ontvangen om een zinvolle vergelijking uit te voeren.
Tabel 5 geeft een overzicht van de verschillen in moeilijkheidsgraad tussen de examens van dit jaar en de referentie-examens.
Omdat de lengte van de scoreschaal geen vast gegeven meer is, worden verschillen uitgedrukt in een procentuele score (ook wel p-waarde genoemd; zie bijlage 1).
Een procentuele score van 50 wil zeggen dat de kandidaten gemiddeld genomen de helft van het maximum hebben behaald. Wanneer het nieuwe examen 2% makkelijker is wil dat zeggen dat de referentiepopulatie op dat examen een geschatte procentuele score van 52 zou hebben gehaald.

Tabel 5 Vergelijking tussen de referentie-examens en de examens 2004

Vak	Referentie-examen	verschil t.o.v. referentie-examen
Duits 1,2 havo	2001	2% makkelijker
Engels havo	2001	1% makkelijker
Frans 1,2 havo	2001	1% makkelijker
Duits 1,2 vwo	2001	6% makkelijker
Engels vwo	2001	10% moeilijker

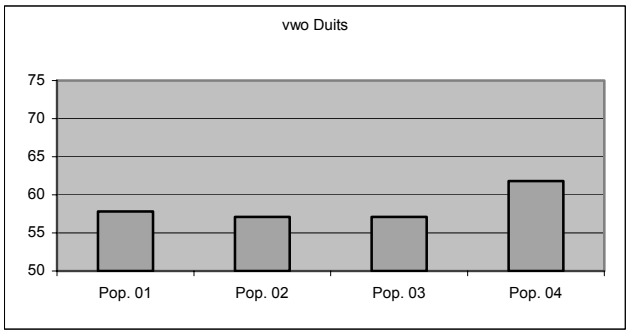
De normvergelijking vindt bij deze examens plaats direct na de afname van het examen. Dit betekent dat eventuele verschillen in moeilijkheidsgraad niet vooraf gereduceerd worden, maar pas achteraf door vaststelling van een juiste normeringsterm gecompenseerd kunnen worden.
Tabel 5 laat zien in hoeverre zich verschillen in moeilijkheidsgraad voordeden tussen het examen 2004 en het referentie-examen. In de regel wordt bij de constructie gemikt op een gelijke moeilijkheidsgraad als die van het referentie-examen.
In 2002 is besloten om de in 2001 afgenomen examens als referentie-examen te kiezen. Deze keuze hield mede verband met de veranderingen die sinds 2000 en 2001 in de MVT-examens zijn doorgevoerd zoals het opnemen van eindterm 1 en het gebruik van een woordenboek als hulpmiddel.

De Cito-vakmedewerkers adviseren op basis van de uitgevoerde vergelijking de normeringstermen bij deze 2004-examens zodanig vast te stellen dat deze equivalent zijn aan de in 2001 gestelde eisen.

Het systeem van normvergelijking maakt het – onder bepaalde aannamen – mogelijk om uitspraken te doen over ontwikkelingen in de populaties over de jaren heen.
De staafdiagrammen hieronder brengen de ontwikkeling in de populatie in beeld. Op de verticale as zijn de gemiddelde procentuele scores afgezet. De kolommen voor de verschillende jaren geven de gemiddelde procentuele scores aan die de opeenvolgende populaties op hetzelfde referentie-examen behaald (zouden) hebben.
In de diagrammen voor de moderne vreemde talen zijn de gegevens van havo en vwo opgenomen.
Het zij uitdrukkelijk opgemerkt dat deze ontwikkeling slechts geconstateerd wordt en dat de verzamelde informatie niets zegt over de mogelijke oorzaken die daaraan ten grondslag liggen. Veranderingen die van jaar tot jaar worden doorgevoerd compliceren de interpretatie van de gegevens. De vergelijking van het prestatieniveau van de populaties van de elkaar opvolgende jaren is dan ook mede door deze complicaties enigszins speculatief van aard. De vergelijking van het prestatieniveau van de tien elkaar opvolgende populaties zou dan ook mede door deze complicaties speculatief van aard zijn. Mede als gevolg van deze veranderingen heeft de CEVO in 2002 nieuwe referentie-examens aangewezen en beperkt deze rapportage zich tot de vernieuwde tweede fase.

Geringe schommelingen in het prestatieniveau van jaar tot jaar kunnen zowel door verschillen in prestatieniveau als door de meetprocedure veroorzaakt zijn.
Onderstaande staafdiagrammen brengen de ontwikkeling in de populatie in beeld.
Op de verticale as zijn de gemiddelde procentuele scores afgezet. De kolommen voor de verschillende jaren geven de geschatte gemiddelde procentuele scores aan die de opeenvolgende populaties op hetzelfde oude referentie examen behaald zouden hebben.

Diagram 1 Prestatieniveau van de opeenvolgende populaties Duits uitgedrukt in een score op de referentietoets



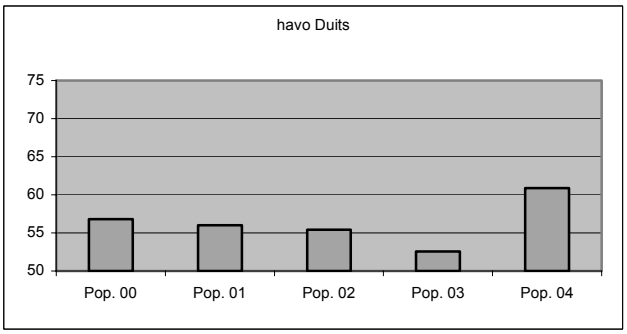
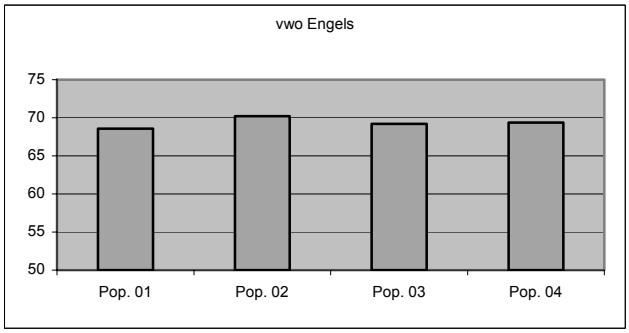


Diagram 2 Prestatieniveau van de opeenvolgende populaties Engels uitgedrukt in een score op de referentietoets



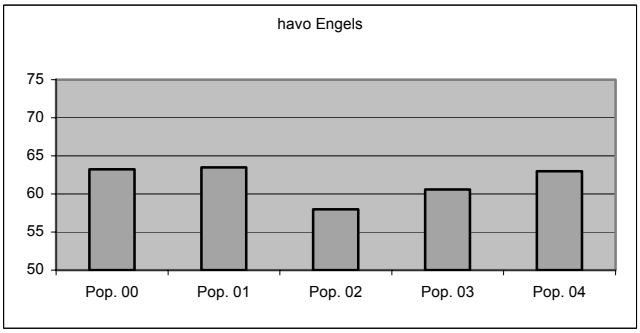
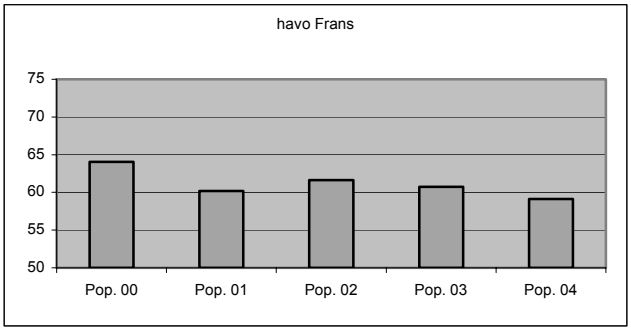


Diagram 3 Prestatieniveau van de opeenvolgende populaties Frans uitgedrukt in een score op de referentietoets



Normvergelijking 'vooraf'

Voor havo en vwo zijn in 2004 de examens wiskunde A en B, natuurkunde, scheikunde, economie 1 en 1,2 en biologie vwo, en Nederlands havo vastgesteld met gebruikmaking van normvergelijkingsgegevens. Dit betekent dat er bij de samenstelling van de examens naar gestreefd is dat ze grotendeels dezelfde moeilijkheidsgraad zouden hebben als een door de CEVO gekozen referentie-examen met een door de CEVO vastgestelde referentienorm.
Als de procedure voor het gehele examen is toegepast, d.w.z. dat alle opgaven gepretest zijn en het examen precies de juiste moeilijkheidsgraad heeft, kan voor deze examens de normeringsterm van 1,0 aangehouden worden. Bij het merendeel van de gepreteste examens was dit echter niet het geval. In tabel 6 zijn de resultaten van 2004 voor deze vakken bij elkaar gezet. Het gemiddelde van de vastgestelde N-termen is weliswaar 1,0 maar de bedoeling is dat daar nauwelijks nog van wordt afgeweken.
Bij de invoering van de vernieuwde tweede fase heeft een uitgebreide oriëntatie plaatsgevonden m.b.t. de consequenties daarvan voor de continuering van de normvergelijkingsprocedures. Daarbij deed zich de paradox voor dat juist in een complexe en nieuwe situatie referentiegegevens over nieuwe examens zeer wenselijk zijn maar dat vanwege de vele veranderingen in het onderwijs het maken van valide vergelijkingen buitengewoon gecompliceerd is. Bij het continueren van de procedures is het uitgangspunt geweest dat er weliswaar minder strikte vergelijkingen gemaakt zouden kunnen worden maar dat getracht zou moeten worden zo relevant mogelijke gegevens te verzamelen over het niveau en de maakbaarheid van de toekomstige examens.
De verwachting is dat de invoering van de vernieuwde tweede fase inmiddels zover uitgekristalliseerd is dat er op basis van de pretest betere voorspellingen omtrent de toekomstige moeilijkheidsgraad kunnen worden gedaan.

Tabel 6 Aantal kandidaten, N-termen, gemiddelde cijfers en percentages onvoldoendes examens 2004

vak	aantal kandidaten	N-term	gemiddeld cijfer	percentage onvoldoendes
Nederlands havo	1.822	1,4	6,0	30
Nederlands vwo	2.035	1,1	6,6	19
Wiskunde A 1,2 havo	2.125	1,1	6,2	29
Wiskunde A 1,2 vwo	1.937	1,5	6,1	32
Wiskunde B 1,2 havo	1.813	1,1	6,2	28
Wiskunde B 1,2 vwo	2.027	1,2	6,5	26
Natuurkunde 1,2 havo	1.788	0,5	6,6	16
Natuurkunde 1,2 vwo	1.779	0,8	6,8	14
Scheikunde havo	2.053	0,7	6,2	25
Scheikunde 1,2 vwo	1.811	0,7	6,8	17
Biologie 1,2 vwo	1.814	1,3	6,3	17
Economie 1 havo	1.807	1,3	6,2	26
Economie 1,2 havo	1.869	1,2	5,9	33
Economie 1,2 vwo	1.844	0,7	6,0	31

1.9 Vergelijking van deelvak- en totaalvak-examens en -populaties

In tabel 7 worden de vergelijkingen van deelvak- en totaalvakpopulaties voor de vakken wiskunde A en B, natuurkunde en scheikunde weergegeven. In het algemeen geldt dat de deelvakpopulaties op de overlap (aanmerkelijk) lager scoren. Alleen bij scheikunde blijken de populaties relatief dicht bij elkaar te zitten. Bij natuurkunde vwo is het verschil tussen de heel- en deelvakpopulatie dit jaar uitzonderlijk gering.
Bij de verschillen tussen heel- en deelvakpopulatie dient te worden aangetekend dat deelvak en totaalvak gebaseerd zijn op verschillende programma's, verschil in studielast kennen en op verschillende vervolgopleidingen voorbereiden.

Tabel 7 Vergelijking van totaalvak- en deelvakpopulaties, totaalvak- en deelvakexamens en de normering

vak	schooltype	Verschil in populaties	Verschil in moeilijkheidsgraad	Afstand na normering in cijferpunten
Wiskunde A 1,2 / Wiskunde A 1	vwo	8	11	0,9
Wiskunde B 1,2 / Wiskunde B 1	vwo	11	10	0,5
Natuurkunde 1,2 / Natuurkunde 1	vwo	2	12	0,7
Scheikunde 1,2 / Scheikunde 1	vwo	3	2	0,1
Wiskunde B 1,2 / Wiskunde B 1	havo	6	13	0,6
Natuurkunde 1,2 / Natuurkunde 1	havo	8	2	0,5

Alle heelvakexamens blijken moeilijker te zijn dan het deelvak examen. Bij scheikunde vwo en natuurkunde havo blijkt het heelvakexamen ietsje moeilijker te zijn. Bij de andere vakken is het verschil in moeilijkheidsgraad substantieel.
Kijken we naar het resultaat na de normering dan worden voor alle onderhavige combinaties van vakken aan de heelvakkers de hoogste eisen gesteld. Bij scheikunde vwo zijn de prestatie-eisen om aan een voldoende te komen voor heelvakkers slechts ietsje hoger dan voor deelvakkers.
Bovenstaande gegevens zijn uitsluitend beschrijvend van aard en zijn pas achteraf vastgesteld. Ze spelen met andere woorden geen rol bij de normering van de heel- en deelvakexamens.

1.10 Vergelijking van reguliere en experimentele examens en -populaties

Voor de vakken aardrijkskunde, economie 1 op havo en biologie, natuurkunde 1,2 op havo en vwo en wiskunde A1 en A1,2 op vwo zijn naast de reguliere examens ook op beperkte schaal experimentele examens afgenomen. In § 2.2 wordt de inhoudelijke kant daarvan nader belicht. Hier wordt ingegaan op de resultaten.
In alle gevallen bestond er een aanzienlijke overlap in vragen tussen het reguliere en experimentele examen (± 65%). Op basis van deze overlap worden de populaties, de examens en de normering op dezelfde wijze als in de beide voorgaande paragrafen vergeleken.
Het enige wezenlijke verschil is dat we hier i.t.t. bij heel- en deelvak geen of slechts geringe verschillen in moeilijkheidsgraad zouden willen aantreffen omdat er geen redenen zijn om aan de experimentele groep andere eisen te stellen dan aan de reguliere populatie.
Tabel 8 geeft de resultaten van deze vergelijking tussen de experimentele en reguliere examens.

Tabel 8 Vergelijking van populaties, examens en normering regulier / experimenteel

vak	schooltype	Verschil in populaties	Verschil in moeilijkheidsgraad	Afstand na normering in cijferpunten
Aardrijkskunde compex	havo	3	-4	0,0
Biologie compex	havo	-2	-1	0,0
Economie 1 compex	havo	0	1	0,0
Natuurkunde 1,2 compex	havo	2	0	0,0
Biologie 1,2 compex	vwo	3	1	0,1
Natuurkunde 1,2 (proj.mod.nat.)	vwo	-1	0	0,1
Natuurkunde 1,2 compex	vwo	2	2	0,0
Wiskunde A1 compex	vwo	-2	2	0,2
Wiskunde A1,2 compex	vwo	3	-1	0,0

De experimentele groepen blijken soms iets beter en soms iets zwakker te scoren op de overlap. Gezien het beperkt aantal scholen (en kandidaten) dat aan het experiment heeft deelgenomen zijn deze verschillen te verwachten. De examens zouden bij voorkeur van gelijke moeilijkheidsgraad moeten zijn. Sommige experimentele examens zijn echter ietsje moeilijker dan de reguliere examens en bij andere is dat weer omgekeerd. Gemiddeld genomen blijken experimentele en reguliere examens even moeilijk te zijn.
Het is uitdrukkelijk niet de bedoeling dat er bij de experimentele examens hogere eisen worden gesteld dan bij de reguliere examens. Daarom wordt er voor een eventueel verschil in moeilijkheidsgraad bij de normering gecompenseerd.
In de laatste kolom wordt de afstand na normering gerapporteerd. Voor het merendeel van de examens blijkt er een identiek prestatieniveau gevraagd te worden. Voor de overige zijn de verschillen gering.

1.11 Verschillen tussen de resultaten van jongens en meisjes

Jongens en meisjes presteren op de meeste examens verschillend. De gemiddelden van de jongens bleken de achterliggende jaren hoger dan die van meisjes. De examens Nederlands, filosofie en kunstbeschouwing werden meestal beter gemaakt door meisjes. De examens moderne vreemde talen en de examens exacte vakken laten vaak een hoger gemiddeld cijfer voor jongens zien. In tabel 10 zijn alle significante verschillen[2] tussen de gemiddelden van jongens en meisjes opgenomen, uitgedrukt in een percentage van de maximumscore die op elk examen te behalen was. De streepjes in de tabel geven examens aan waarbij zich geen significant verschil voordeed. De gearceerde vlakken geven aan dat een vak niet in dat bepaalde schooltype wordt geëxamineerd.

De verschillen zijn het grootst bij Turks vwo en Spaans havo. Het gaat in beide gevallen om een groot verschil maar vastgesteld op een gering aantal jongens en meisjes. De examens wiskunde, natuurkunde, scheikunde, biologie en economie gaven in het verleden bijna altijd wel voor een of meer van de schooltypen significante verschillen in het voordeel van de jongens te zien. Dat patroon was tot en met 2001 altijd redelijk stabiel. In 2002 werd dat patroon voor het eerst doorbroken. Op de examens wiskunde A1,2 en B1 en scheikunde 1 scoorden meisjes toen significant hoger. Ook in 2003 scoorden de vwo-meisjes hoger op enkele van die traditionele ‘jongensvakken’ namelijk op scheikunde 1,2 en wiskunde B1,2. In 2004 scoren de vwo-meisjes opnieuw significant beter op scheikunde 1 en 1,2 en wiskunde B1,2.
Over de oorzaken van de verschillen valt weinig met zekerheid te zeggen. Uit onderzoek blijkt dat meisjes gemiddeld jonger zijn wanneer ze examen doen; ze lopen minder studievertraging op en jongens verlaten de school vaker vóór het eindexamen dan meisjes. Meisjes worden vaker naar ‘hogere’ schooltypen doorverwezen dan jongens. Bij een gelijke score op de eindtoets basisonderwijs worden meisjes meer toegelaten tot moeilijker geachte schooltypen dan jongens.

[2] Als criterium is hier een overschrijdingskans van genomen van $\alpha < 0,05$. Deze toetsing kan ook uitgevoerd worden bij geringe aantallen kandidaten. Het verschil zal bij geringe aantallen kandidaten wel substantieel moeten zijn om het niveau van significantie te bereiken.

Tabel 9 Significante verschillen in p'-waarde tussen jongens en meisjes

	vwo		havo	
	meisjes beter	jongens beter	meisjes beter	jongens beter
Nederlands (leesv.)	4	–	3	
Frans 1,2		3	–	–
Duits 1,2		2		2
Engels		6		2
Spaans (1,2)	–	–		13
Russisch (1,2)	–	–	–	–
Turks (1,2)		11	–	–
Arabisch(1,2)	–	–	–	–
Latijn	6			
Grieks	4			
Wiskunde A1		1		
Wiskunde A1,2	–	–	–	–
Wiskunde B1	–	–	–	–
Wiskunde B1,2	2		–	–
Natuurkunde 1	–	–		2
Natuurkunde 1,2	–	–		4
Scheikunde (1)	1			2
Scheikunde 1,2	2			
Biologie (1,2)		2		4
Economie 1		5		6
Economie 1,2	–	–		5
Management & Organisatie		1		2
Gesch.en staatsinr.	–	–	–	–
Aardrijkskunde		3		3
Maatschappijleer	–	–	–	–
Filosofie	5		–	–
Muziek	–	–		6
Tehatex	–	–	–	–
CKV 2	–	–	–	–

Behalve deze mogelijke oorzaken, die met de schoolloopbaan te maken hebben, kunnen oorzaken misschien ook gevonden worden in kenmerken van examens zelf. Meisjes scoren hoger op traditioneel vrouwelijke onderwerpen, zoals huishouden, opvoeding en onderwerpen als literatuur, kunst, intermenselijke relaties en gevoelens. Jongens scoren hoger op 'typische mannenonderwerpen' zoals oorlog, misdaad, politiek, sport en op economische, financiële en natuurwetenschappelijke onderwerpen.
Overigens moet benadrukt worden dat als het te toetsen leerdoel aan het examen-programma ontleend is, de toetsing zonder meer gerechtvaardigd is. Voor alle kandidaten geldt immers hetzelfde examenprogramma. Van de examenconstructeurs mag gevraagd worden rekening te houden met de verschillende interesses van jongens en meisjes als het factoren betreft die niet essentieel zijn voor het te toetsen leerdoel, zoals de teksten bij talenexamens en het uitgangs- of contextmateriaal in overige examens. Hoewel bij de selectie van teksten en contexten hier expliciet rekening mee wordt gehouden, bleken de verschillen in resultaten van jongens en meisjes over een reeks van jaren toch redelijk consistent te zijn. Van de 29 gesignaleerde significante verschillen in tabel bleken er in 2003 maar liefst 23 eveneens significant en vergelijkbaar verschil te laten zien. Zelfs de verschuivingen die sinds 2002 zijn opgetreden blijken redelijk consistent te zijn.
Dit jaar signaleren we bij het vwo voor de derde maal op rij dat meisjes het beter doen op enkele examens waarop tot 2001 de jongens het nagenoeg altijd beter deden.

1.12 De examens tweede tijdvak

Op 23 juni 2004 (en voor enkele vakken op 25 juni) werd het tweede tijdvak afgenomen.
De deelname aan het tweede tijdvak staat voor alle kandidaten voor één vak open. Een ieder

die meent dat hij of zij voor één vak een beter cijfer zal kunnen behalen, mag aan het tweede tijdvak deelnemen.
Alleen kandidaten die tijdens het eerste tijdvak ziek waren, mogen in twee vakken examen doen. Wie in meer dan twee vakken nog examen moet doen, kan in augustus, in het derde tijdvak, het examen afronden.
De scholen is gevraagd voor elk examen maximaal van vijf kandidaten (de in alfabetische volgorde eerste vijf) de scores van het tweede tijdvak op speciaal daarvoor toegezonden optisch leesbare formulieren aan te strepen, samen met het cijfer dat de kandidaat voor het eerste-tijdvak-examen behaalde, en deze formulieren op te sturen naar de Citogroep.
Voor de scholen die aan het experiment WOLF2004 meededen gold een vergelijkbare verplichting. De verwerkingstijd is zowel voor docenten als voor de Citogroep bijzonder kort omdat de resultaten nog voor de datum van de uitslag – zes dagen na de afname – gebruikt moeten kunnen worden. De scholen is gevraagd de formulieren twee dagen na de afname – op vrijdag 25 juni – al op te sturen. In de tabellen met resultaten is te zien dat deze snelle inzending en verwerking met succes zijn verlopen.
De gegevens werden verzameld om een vergelijking tussen het eerste en tweede tijdvak mogelijk te maken.

De resultaten van kandidaten die een onvoldoende op het eerste tijdvak behaald hebben, worden afgezet tegen hun resultaten op het tweede tijdvak. De aanname is dat het gemiddelde cijfer van deze kandidaten in het tweede tijdvak hoger moet zijn dan in het eerste tijdvak. Er zijn twee redenen om dit aan te nemen. In de eerste plaats zullen kandidaten zich nog speciaal kunnen richten op één vak waardoor leerwinst kan optreden. In de tweede plaats heeft de regressie naar het gemiddelde, die optreedt bij een herhaalde meting bij kandidaten die op het eerste tijdvak een lage score hebben behaald, een positief effect op de hoogte van de score. Dit regressie-effect kan ook begrepen worden als het effect dat veroorzaakt wordt doordat met name die kandidaten herkansen die vinden dat ze pech hebben gehad of, met andere woorden, dat door toevalsfluctuaties hun score op het eerste tijdvak een onderwaardering van hun niveau weergeeft. Is de gemiddelde score van deze groep kandidaten op het tweede tijdvak niet hoger dan die op het eerste tijdvak, dan moet dit veroorzaakt worden doordat het tweede tijdvak een hogere moeilijkheidsgraad heeft dan het eerste tijdvak. Een andere normeringsterm wordt dan overwogen.

Bij de tweede-tijdvak-examens van 2004 wijkt de definitieve normeringsterm van het tweede tijdvak voor de volgende examens af van de ‘voorlopige’ normeringsterm:

vak	afwijking	N-term
Wiskunde B1,2 vwo	0,6	1,8
Management & organisatie	0,8	1,3

De verzameling van de gegevens betekent ook dat de resultaten van het tweede tijdvak voor een groot aantal examens beschikbaar zijn. In de tabellen 11 en 12 worden deze resultaten gepresenteerd. Als toelichting bij deze tabellen het volgende:
– Niet van alle examens worden gegevens gepresenteerd. Bij een aantal vakken zijn geen gegevens verzameld, omdat vooraf reeds te voorzien was dat er te weinig kandidaten zouden zijn voor een zinvolle gegevensverzameling.
– In de eerste kolom wordt het vak/examen vermeld.
– De tweede kolom betreft alle voor dat vak door de Citogroep verwerkte formulieren, hetgeen overeenkomt met het aantal kandidaten in de steekproef.
– In de derde kolom wordt het gemiddelde cijfer vermeld gebaseerd op de gehele steekproef van tweede-tijdvak-kandidaten.
– In de vierde kolom wordt vermeld welk percentage van de kandidaten in de steekproef op het tweede tijdvak een onvoldoende haalt. Over het algemeen scoort de populatie die deelneemt aan het tweede tijdvak gemiddeld lager dan de populatie die deelneemt aan het eerste tijdvak. Dit wordt veroorzaakt doordat in de populatie van het tweede tijdvak relatief meer zwakke kandidaten zitten.

Zoals te verwachten liggen de gemiddelde cijfers van het tweede tijdvak lager dan die van het eerste tijdvak. Een groot deel van de tweede-tijdvakpopulatie bestaat immers uit herkansers die op het eerste tijdvak een onvoldoende resultaat hadden behaald. Van deze groep haalt (gemiddeld over vakken) ruim 40% alsnog een voldoende in het tweede tijdvak.

Tabel 10 Gemiddeld cijfer, aantallen kandidaten examens havo 2e tijdvak

vak	steekproefgrootte	gemiddeld cijfer	% onvoldoendes
Nederlands	580	5,5	47
Frans 1,2	233	5,0	69
Duits 1,2	188	5,2	62
Engels	482	5,5	51
Spaans 1,2	–	–	–
Russisch 1,2	–	–	–
Turks 1,2	–	–	–
Arabisch 1,2	–	–	–
Wiskunde A 1,2	350	5,2	56
Wiskunde B 1	220	4,9	64
Wiskunde B 1,2	184	5,1	58
Natuurkunde 1	167	6,2	18
Natuurkunde 1,2	111	6,3	23
Scheikunde	325	5,9	29
Biologie	180	5,6	45
Economie 1	335	6,5	13
Economie 1,2	638	5,2	57
Management & organisatie	94	5,8	35
Aardrijkskunde	212	5,4	48
Geschiedenis	341	5,6	37
Muziek	–	–	–
Tekenen/handenarb./text.werkv.	–	–	–
Filosofie	–	–	–
Maatschappijleer	–	–	–
CKV 2 [1)]	–	–	–

[1)] Geen Centraal Examen

Tabel 11 Gemiddeld cijfer, aantallen kandidaten examens vwo 2e tijdvak

vak	steekproefgrootte	gemiddeld cijfer	% onvoldoendes
Nederlands	466	5,3	57
Latijn	–	–	–
Grieks	–	–	–
Frans 1,2	194	5,5	55
Duits 1,2	131	5,5	52
Engels	713	5,8	44
Spaans	–	–	–
Russisch	–	–	–
Turks	–	–	–
Arabisch	–	–	–
Wiskunde A 1	73	4,8	71
Wiskunde A 1,2	322	5,6	45
Wiskunde B 1	341	5,3	54
Wiskunde B 1,2	285	5	65
Natuurkunde 1	183	5,7	39
Natuurkunde 1,2	129	5,6	40
Scheikunde 1	164	5,1	62
Scheikunde 1,2	138	6,1	28
Biologie 1,2	282	6,2	28
Economie 1	81	5,5	49
Economie 1,2	395	5	68
Management & organisatie	86	4,9	72
Aardrijkskunde	167	5,4	48
Geschiedenis en staatsinrichting	101	5,3	57
Maatschappijleer	–	–	–
Muziek	–	–	–
Tekenen/handenarbeid/text.werkv.	–	–	–
Filosofie	–	–	–
CKV 2 [1)]	–	–	–

[1)] Geen Centraal Examen

2 Nieuwe examenvormen

In dit hoofdstuk wordt aandacht besteed aan veranderingen in de examens. Het gaat daarbij zowel om zaken die betrekking hebben op veranderingen die pas in de toekomst hun beslag zullen krijgen als om veranderingen die dit jaar voor het eerst of betrekkelijk recent gerealiseerd werden. In § 2.1 wordt kort ingegaan op de ontwikkelingen m.b.t. de examenprogramma's en de consequenties die daaruit voortvloeien voor de toekomstige examens.
In § 2.2 wordt ingegaan op de inzet van computers bij het afnemen van de examens. Het gaat daarbij om de toepassing van specifieke computer applicaties in experimentele examens. In § 2.3 wordt tot slot een korte vooruitblik gegeven op 2005.

2.1 Nieuwe examenprogramma's

In 2007 worden een aantal aanpassingen in de opzet van de tweede fase doorgevoerd. Deze aanpassingen worden voor het eerst zichtbaar in de havo-examens van 2009 en de vwo-examens van 2010. Zo zal bij de exacte en economische vakken het onderscheid tussen heel- en deelvak komen te vervallen. De examenprogramma's van deze vakken zullen daartoe aangepast worden. Voor de meeste andere examens zullen de aanpassingen in de examenprogramma's minder ingrijpend zijn. De aanpassingen zullen betrekking hebben op evidente beperkte feilen in de eindtermen voor het huidige centraal examen. In de eerste helft 2005 vindt de besluitvorming over deze aanpassingen plaats. Los van bovenstaande aanpassingen zijn voor een aantal vakken ingrijpende wijzingen in de examenprogramma's in voorbereiding. Die zullen echter pas in de examens na 2010 geëffectueerd gaan worden. Het betreft de vakken geschiedenis, economie en scheikunde.

2.2 Computers en examens

In de tweede fase van het havo/vwo wordt in de examenprogramma's expliciet het aspect van ICT-vaardigheden genoemd. Bij de CEVO leefde de wens in 2003 de eerdere experimenten met het gebruik van de computer bij de afname van centrale examens te intensiveren en te gebruiken als een start voor een beleidsplan voor de jaren 2004 tot en met 2005.
Als uitwerking van deze CEVO-plannen is binnen de unit VO van de Citogroep het examen-innovatieproject 'Compex3' van start gegaan. De Citogroep voert het project in overleg met de CEVO en in samenwerking met projectscholen uit onder de werknaam 'Compex' (Computers en Examen).
Het project heeft tot doel:

- het opdoen van ervaring met de productie van opgaven en de eisen die aan zulke opgaven moeten worden gesteld;
- het opdoen van ervaring met de afname van opgaven waarbij binnen de setting van een centraal examen gebruik moet worden gemaakt van de computer;
- het in kaart brengen van de organisatorische, juridische, technische en didactische consequenties van het invoeren van het gebruik van ICT bij de centrale examens;
- het opdoen van ervaring met de mogelijkheden voor automatische rapportage van de resultaten van de kandidaten

Het project Compex3 is een vervolg op de projecten 'Compex' en 'Compex2' uit de jaren 2000 en 2002.

ICT kan op diverse manieren worden toegepast in de examens. In het kader van de Compex-projecten worden twee vormen nader verkend, te weten 'IMEX' (ICT en multimedia in het examen) en 'CBT' (Computer Based Test).

- In geval van een 'IMEX' gaat het om de inzet van de computer als hulpmiddel bij examens op papier, bijv. voor het afspelen van multimediabestanden en/of het werken met vakspecifieke software. Bij slechts een deel van de opgaven van het examen moet

de leerling gebruik maken van de computer. Het overige deel van het examen is gelijk aan het regulier centraal schriftelijk examen.

– In geval van een ‘CBT’ gaat het om een examen dat geheel via het beeldscherm aan de leerling wordt aangeboden. De computer wordt hier dus ingezet ter vervanging van het papieren examen.

Het streven is dat voor elk profiel (vwo/havo), en voor elke sector in elke leerweg (vmbo) minstens één examen beschikbaar is waarbij de computer gebruikt moet worden.
Op (vak-)inhoudelijke gronden worden twee typen computerexamens onderscheiden.
1. Examens met voornamelijk open vragen of praktijkexamens lenen zich niet voor geheel gecomputeriseerde afname, maar wel kan in die examens standaardprogrammatuur gebruikt worden of kan uitgangsmateriaal in de vorm van (kleur)plaatjes of audio en video op het beeldscherm aangeboden worden.
2. Examens met alleen gesloten vragen zijn geschikt om geheel op de computer afgenomen te worden en ook door de computer gescoord te worden. De mogelijkheid van het aanbieden van uitgangsmateriaal in de vorm van (kleur)plaatjes of audio en video op het beeldscherm is ook hierbij aanwezig. Deze zogenaamde CBT-examens komen in havo/vwo overigens niet voor.

Gebruik van standaardprogrammatuur en multimedia (IMEX):
De examens worden zo opgesteld dat de leerling bij de beantwoording van vragen of de uitvoering van opdrachten gebruik moet maken van bekende (standaard)programmatuur, audiovisuele of andere informatiebronnen. De kandidaten krijgen de vragen of opdrachten op papier aangeboden of op papier en op beeldscherm. De antwoorden moeten op papier gegeven worden, behoudens sommige opdrachten die in een programma moeten worden uitgevoerd en waarvan de bestanden moeten worden opgeslagen. Voorbeelden hiervan zijn het gebruik van de belastingdiskette bij het vmbo economie, het gebruik van een boekhoudprogramma bij administratie of het gebruik van een CAD (programma voor technisch tekenen) bij bouwtechniek, het gebruik van IPCoach (een programma waarmee een natuurkundig model kan worden gesimuleerd) bij een examen natuurkunde havo en het gebruik van een tekstverwerker voor schrijfvaardigheid bij Nederlands vmbo. Deze examens worden met de werknaam IMEX aangeduid (Ict en Multimedia in Examens). Andere voorbeelden van dergelijke examens zijn het al een aantal jaren succesvol afgenomen examen voor Culturele en Kunstzinnige Vorming 2 (CKV2) in havo/vwo en het examen Dans en Drama voor de gemengde en theoretische leerweg in het vmbo.

Complete beeldschermexamens (CBT):
Het examen wordt geheel op het scherm aangeboden. De leerlingen voeren de antwoorden in. De computer beoordeelt de antwoorden voor zover dat mogelijk is. Deze vorm van examinering lijkt op dit moment alleen toepasbaar en haalbaar bij examens met louter gesloten vragen. De Citogroep is nu bezig voor dit soort examens een afnameomgeving (= toetsprogramma) te maken waarmee het mogelijk is meerkeuzevragen op betrouwbare en succesvolle wijze te toetsen en te scoren. In 2004 is in het vmbo bij diverse beroepsgerichte vakken in het CSPE deze afnameomgeving ingezet. De planning is om in 2005 alle vakken in het CSPE en alle algemene vakken op het niveau van de basisberoepsgerichte leerweg met deze afnameomgeving de examens te laten maken. We gaan hier niet verder in op dit type examen.

De vakken in havo/vwo waarbij de computer bij de examens ingezet wordt, zijn gekozen vanuit diverse overwegingen:
1. een vak kan alleen maar geëxamineerd worden door van de computer gebruik te maken;
2. welk vak biedt de beste mogelijkheden voor ICT-toepassingen;
3. welk vak is een profiel- of sectorvak waaraan niet alle kandidaten deelnemen (dit in verband met de omvang van de organisatie);
4. welke leerlingen zijn het meest gediend met computertoetsen.

Er zijn voor havo en vwo net als in 2003 acht IMEX-examens ontwikkeld. Hieronder wordt aangegeven voor welke vakken en schooltypen.

Profiel	havo	vwo
Cultuur en Maatschappij	Economie 1	Wiskunde A1
Economie en Maatschappij	Aardrijkskunde	Wiskunde A1,2
Natuur en Gezondheid	Biologie	Biologie 1,2
Natuur en Techniek	Natuurkunde 1,2	Natuurkunde 1,2

2.2.1 Constructie van IMEX-examens

Voor het construeren van computerexamens is naast veel creativiteit ook ervaring met toetsconstructie vereist. Op welke manier formuleer je een vraag zodanig dat deze de beoogde doelstelling toetst en daarnaast slechts voor één uitleg vatbaar is. En hoe zorg je dat het examen als geheel niet te moeilijk is. Aspecten die daarbij een rol spelen zijn:

- de constructeur moet overzicht hebben over wat er mogelijk is met de programmatuur,
- de selectie van materiaal (video, websites, animaties, programmatuur) vereist dat de constructeur zelf intensief met de mogelijke programmatuur en uitgangs-materialen bekend is,
- de vragen/opdrachten moeten de vakdoelstellingen toetsen op het juiste niveau,
- voor een opgave waarbij de computer wordt gebruikt geldt net zo goed dat er een efficient gebruik van de examentijd moet worden gemaakt,

In het Compex-project blijkt dat deze kennis opgebouwd moet worden. Examens aanbieden, reacties van leerlingen en docenten verzamelen, leerlingen observeren bij het maken van een examen zijn methodes om deze kennis te verzamelen. Docenten verwachten inhoudelijke meerwaarde wanneer de computer betrokken wordt bij de examinering. De ene constructiegroep slaagt er de eerste keer ook beter in om meerwaarde te bieden dan de andere. Het opbouwen van nieuwe ervaring kost tijd.
Zo'n examen wordt door docenten en leerlingen natuurlijk kritisch bekeken. De volgende reacties zijn verzameld:

- Het lezen vanaf het scherm van grotere teksten waarbij gescrold moet worden, wordt niet door iedereen als prettig ervaren.
- In het algemeen blijkt dat docenten en leerlingen computergebruik bij examens pas zinvol vinden wanneer exameninhouden duidelijk op een andere manier aangeboden worden, dat kan betrekking hebben op het gebruik van uitdagende programmatuur; andere manieren om eindtermen te toetsen of door examens beter op de dagelijkse realiteit te laten aansluiten.

Veel leerlingen vinden een examen op de computer aantrekkelijk. De sfeer is vaak iets meer ontspannen tijdens het examen. De verwachtingen ten aanzien van computergebruik zijn echter vaak hooggespannen.

2.2.2 Technische aspecten

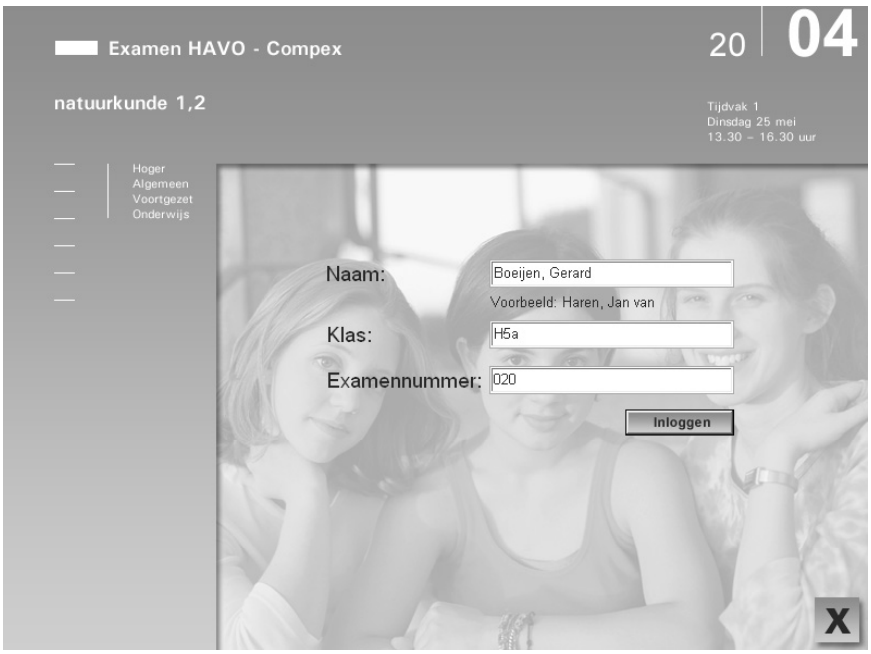
De software en de hardware moeten tijdens de afname van het examen probleemloos functioneren. Ook op dit terrein moet ervaring opgebouwd worden, zowel bij de leveranciers van de examens als bij de scholen. Het gebruik van algemeen gebruikte standaard programmatuur als Word, Excel, Internet Explorer roept weinig problemen op. Wanneer deze programmatuur in het gewone gebruik goed functioneert, moet dat ook tijdens examens kunnen. Het gebruik van programma's die van verschillende leveranciers komen, moet goed uitgetest worden, op verschillende configuraties en in alle omstandigheden die zich tijdens een examen kunnen voordoen. De Citogroep heeft hiervoor een speciaal testlaboratorium ingericht, waar diverse configuraties opgesteld zijn. Het uitvoerig uittesten buiten de examensituatie voorafgaande aan gebruik in de examens is een absolute must.
De Citogroep hanteert inmiddels een vaste procedure waarbij eerst een interne technische test wordt uitgevoerd en daarna pas een test op verschillende scholen buiten de examen-situatie. Het samenspel met Mediaplayers is niet altijd probleemloos. Hetzelfde geldt voor de opslag in databases. Om het aantal testsituaties en het aantal condities waaraan programmatuur moet voldoen enigszins in te perken heeft de Citogroep in overleg met de Cevo de volgende minimum systeemeisen geformuleerd.

werkstation Kandidaat (minimale configuratie)		netwerk infrastructuur (minimale configuratie)	
snelheid PC:	450 Mhz	snelheid netwerkverbinding:	10 Mbit/s
beschikbaar werkgeheugen:	64 Mb	snelheid netwerkserver:	Pentium III 500 Mhz
vrije ruimte op harddisk	1 Gb per examen	beschikbaar werkgeheugen netwerk server:	128 Mb
Cd-rom speler:	24 x of sneller	vrije ruimte op harddisk:	1 Gb per examen
Geluidskaart :	stereo, 16-bits, 44 kHz (met koptel.-aansluiting)		
Koptelefoon:	geluidsniveauregeling		
Videokaart:	resolutie: 800x600 pixels, kleuren: 24-bits (RGB) geheugen: 8 Mb of meer		
Besturingssysteem:	Windows 9x / 2000/XP		
Overige software:	MS IE 5.5 Mediaplayer 6.4		

Om het voor de leerlingen zo makkelijk mogelijk te maken en de kans op problemen tijdens het examen te verkleinen worden de toepassingen menugestuurd aangeboden. Vanuit dat menu worden de programma's gestart volgens de aanwijzingen in het examen.
Een voorbeeld van het getoonde openingsscherm voor het vak natuurkunde 1,2 havo in het examen 2004 is te zien in onderstaande figuur.

Daarnaast is het mogelijkheid om van tevoren de programma's te (laten) installeren op het systeem en voor de examensituatie een aantal opties te laten instellen door de systeem-beheerder.

Deze instellingen hebben tot doel er voor te zorgen dat als er examenresultaten moeten worden opgeslagen, deze op een veilige manier op het netwerk worden weggeschreven en er tevens een back-up wordt gemaakt.
Dit jaar is hierbij voor het eerst gebruik gemaakt van door de Citogroep ontwikkelde software voor een veilige opslag van resultaten en de back-up daarvan op het netwerk.
Het openingsscherm bij die examens waarbij resultaten moeten worden opgeslagen is dan ook een inlogscherm. De door de leerling ingevoerde gegevens worden door het programma uitgelezen en gebruikt om de opgeslagen resultaten van een logbestand en een identificatie te voorzien. Daardoor wordt optimale veiligheid ten aanzien van het behoud van de bestanden gegarandeerd.



Examens worden afgenomen op scholen. De inrichting van netwerken en PC's is een zaak van de school zelf. De technische kennis ten aanzien van de computerinfrastructuur is aanwezig bij systeembeheerders en ICT-coördinatoren. De feitelijke situatie op de meeste scholen is dat er verschillende lokalen met PC's zijn die op verschillende momenten aangeschaft zijn en derhalve ook een verschillende inrichting kennen. De examens maken veel gebruik van multimedia. Derhalve is er voor gekozen om te werken met een lokale installatie in een cliënt–server omgeving. Dit betekent dat deze examens vooralsnog niet kunnen worden afgenomen in een terminal–server omgeving. De scholen moeten zich bij de aanschaf en inrichting van hun netwerkconfiguratie terdege realiseren dat een terminal-server-omgeving niet geschikt is om rijke multimedia toepassingen te hanteren en dus ook niet geschikt is voor afname van compex-examens. Een school (systeembeheerder, ICT-coördinator, examensecretaris en vakdocenten) moet vooraf goed nadenken over hoe ze de afname van computerexamens willen organiseren.

2.2.3 Conclusies

De computeropgaven maakten gebruik van standaard programmatuur, zoals Excel, Website simulaties op Cd-rom, IPCoach, de module remote sensing van de Grote Bosatlas Extra, verschillende Flash animaties etc., de antwoorden werden merendeels op papier opgeschreven. Daar waar de resultaten van het gebruik van programmatuur moest worden opgeslagen is gebruik gemaakt van de door de Citogroep ontworpen software. Reacties van de scholen zijn over het algemeen positief. Niettemin is er een aantal belangrijke knelpunten dat genoemd moet worden en waaraan voor het examenjaar 2005 gewerkt zal worden. De volgende knelpunten verdienen aandacht of vragen om een oplossing of een verbetering:

- Voor de vakken met veel kandidaten is de organisatie van het examen een knelpunt. Voor elke leerling moet er een werkende computer zijn.

- Er moet voldoende ruimte zijn rond de computer omdat er ook nog moet worden geschreven.
- De surveillance moet worden uitgebreid met een terzake kundig persoon (ICT-coördinator liefst systeembeheerder).
- In geval van calamiteiten moet snel worden gehandeld.
- De ruimte waarin de computers staan moet geschikt gemaakt worden voor de examensituatie. Dat kan betekenen dat er schotten tussen de computers moeten worden geplaatst, of iets dergelijks. Kortom deze examens kosten, zeker in het begin, meer werk dan de reguliere aula of gymzaal examens.Toch is men over het algemeen van mening dat de organisatie (ook met grotere groepen) te doen is.
- De installatie van de software die nodig is om met het oefenmateriaal te kunnen werken moet ruim van tevoren plaats vinden. Alle systeembeheerders zijn in staat gebleken met de diversiteit aan software om te gaan. Er zal gewerkt worden aan een verbetering van de installatieprocedures voor oefenmateriaal en examen.
- De opslag en back-up procedure is nog niet helemaal waterdicht gebleken. Onderzocht zal worden wat daarvan de oorzaak is.

In het voorgenomen invoeringstraject is voorzien dat het examenjaar 2005 het laatste jaar is waarin de IMEX-examens als experimentele examens zullen worden aangeboden. In 2006 zullen de IMEX-examens eerste tijdvak via de IB-groep worden aangeboden aan alle scholen als alternatief voor het reguliere schriftelijke CE. Ook de logistiek rondom de uitlevering zal door de IB-groep worden verzorgd. Het ministerie onderzoekt de mogelijkheden om in 2007 te beginnen met de invoering van deze ICT aspecten in het reguliere examen (eerste en tweede tijdvak) voor het vak natuurkunde 1,2 op havo/vwo. Een geleidelijke invoering in de reguliere examens voor andere vakken zal in de jaren daarna volgen.

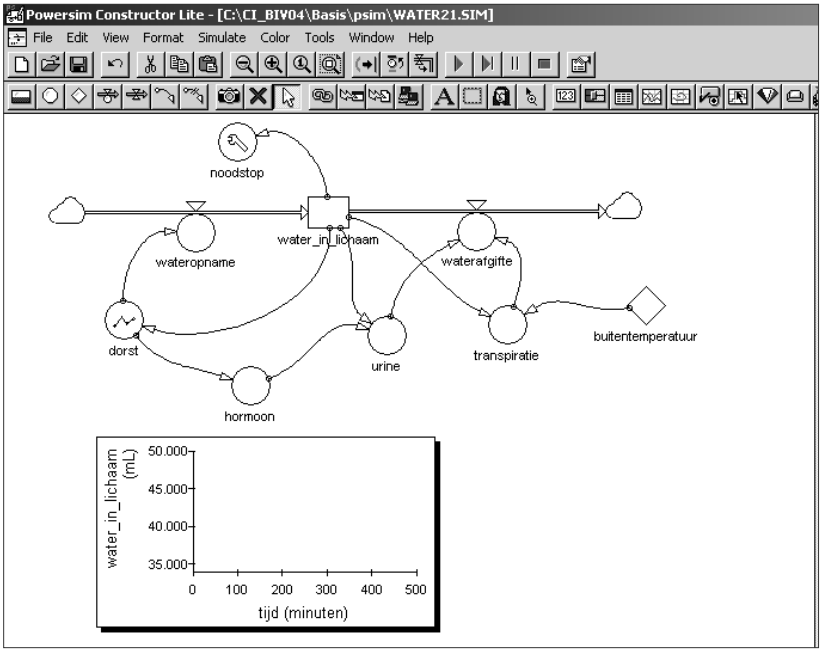
2.2.4 Beschrijving van de IMEX-onderdelen

De computeronderdelen maakten in de regel zo'n 30 tot 40 % van het examen uit.
De overige vragen waren gelijk aan vragen uit het reguliere examen. Hieronder zal per compex-examen een korte typering worden gegeven van de wijze waarop de computer bij het oplossen van vakinhoudelijke problemen werd ingezet.

biologie havo
Voor het compexdeel wordt gebruik gemaakt van de volgende simulatieprogramma's
- het simulatieprogramma 'Meer' (ecologisch aquatisch simulatieprogramma) waarin effecten van maatregelen, ook op langere termijn, zichtbaar worden;
- het programma Powersim, een modelleringprogramma over de urineproductie van het menselijk lichaam onder verschillende omstandigheden.
- het programma Fly Lab, een simulatieprogramma over erfelijkheid. Door het uitvoeren van verschillende kruisingen met fruitvliegjes worden erfelijke eigenschappen onderzocht, waaronder geslachtsgebonden en letale eigenschappen

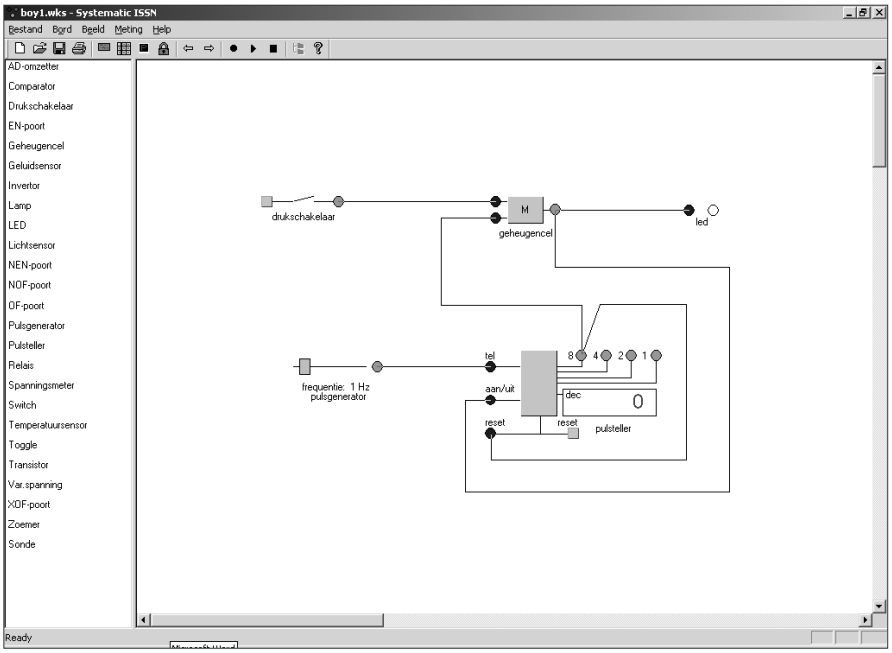
Voor een deel zijn deze simulatieprogramma's gekoppeld aan videofragmenten. De volgende videofragmenten zijn in het examen verwerkt:
- een videofragment over het gedrag van Brasem (gekoppeld aan het programma Meer);
- drie videofragmenten over 'Afvalwaterzuivering';
- een videofragment over de mestkever en zijn bijzondere manier van het leggen van eieren in mest.

biologie 1,2 vwo
Het computergedeelte bestaat uit vier verschillende clusters.
1. Evolutie: Bij deze opdracht wordt in een virtueel laboratorium (Evolution Lab) een aantal experimenten over evolutie gedaan. De evolutie van zaadetende vinken kan op deze manier nader worden onderzocht. Bepalende factoren die kunnen worden ingesteld zijn o.a. de grootte van de zaden en door het experiment te 'runnen' kan onderzocht worden wat de gevolgen zijn voor de snavelgrootte en de populatie op de verschillende eilanden.
2. Poetsvissen: na het bekijken van een video over poetsvissen werd een aantal vragen over gedrag gesteld.
3. In het cluster 'Dorst tijdens de vierdaagse' wordt met behulp van het grafisch simulatieprogramma Powersim een model geboden voor de waterhuishouding van het menselijk lichaam tijdens een vierdaagse wandeltocht. Zie figuur.



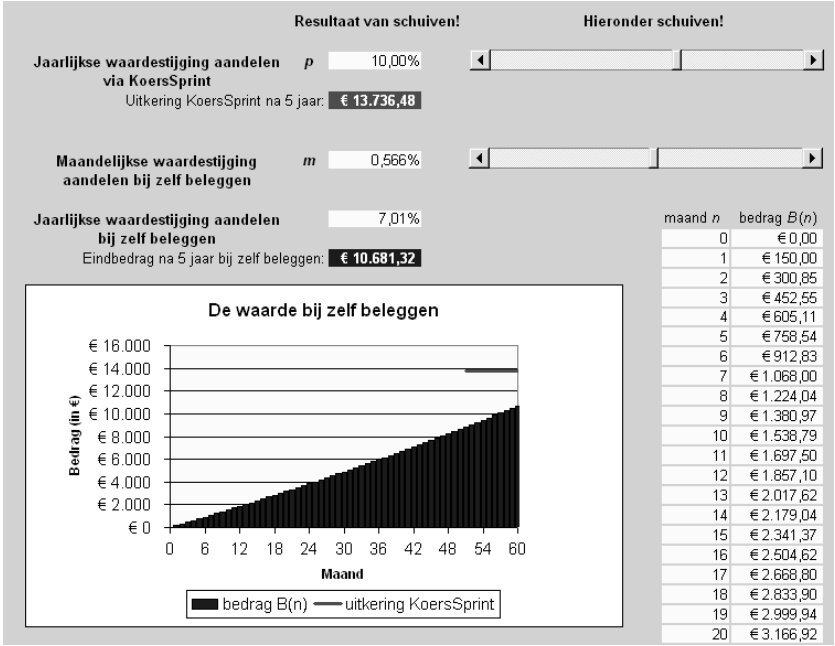
Tot slot wordt met behulp van een Flash animatie de werking van het hart bekeken en afgevraagd.

natuurkunde 1,2 havo
Het computergedeelte van het examen bestaat uit twee opgaven, Hellingbaan en Broodrooster.
Door op ‘Hellingbaan’ te klikken in het openingsscherm komt de leerling in Coach Videometen waarin een paar filmpjes van een pretpark attractie zijn opgenomen. In een van de filmpjes is het omhoog takelen van een treintje van wagentjes te zien. Na het starten van het tweede filmpje wordt de kandidaten onder andere gevraagd een geschikt (x,t) en (y,t)-diagram te maken en daarin de beweging van de treintjes vast te leggen. Vervolgens moet uit de diagrammen de snelheid worden bepaald of berekend.
Een derde filmpje laat de beweging van het treintje zien als het wordt losgelaten. Hieraan is al een videometing gedaan en in een (x,t)- en (y,t)-diagram klaargezet. Aan de hand van deze resultaten moeten de afstand die het treintje heeft afgelegd en de versnelling van het treintje worden bepaald met behulp van de faciliteiten van het programma.
Bij de opgave ‘Broodrooster’ krijgen de leerlingen de opdracht met behulp van het computerprogramma ‘Systematic’ een automatisch systeem te ontwerpen. Dit is in het examenprogramma opgenomen als het onderwerp signaalverwerking. Zie figuur.



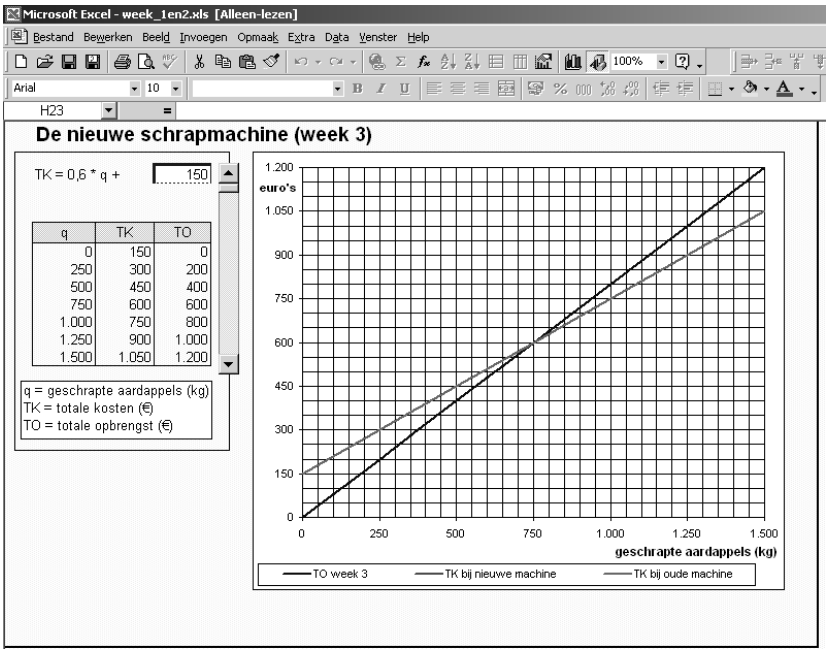
natuurkunde 1,2 vwo
Het computergedeelte van het examen bestaat uit één opgave, ‘Sahara’.
De zon komt op boven de Sahara. Dat betekent het einde van een ijskoude nacht en het begin van moordend hete dag. Als de zon naar het hoogste punt klimt, neemt de stralingswarmte toe en stijgt de temperatuur. Hoe heter het zand en de lucht erboven, hoe meer de aarde ook van deze warmte terugstraalt: niet in het zichtbare licht, maar in het infrarode golflengtegebied.
In deze opgave wordt met een computermodel gesimuleerd hoe de temperatuur in de Sahara over een etmaal verloopt. De opgave bestaat uit drie delen: in deel I wordt het effect bestudeerd van de zonnestraling alleen, in deel II het temperatuurverloop uitsluitend tengevolge van de uitstraling door de aarde en in deel III worden beide effecten tot één totaal model gecombineerd. Het programma waarmee deze modellen kunnen worden gegeven, gemaakt of gewijzigd is IPCoach 5 van de Universiteit van Amsterdam.

wiskunde A1 vwo en A1,2 vwo
Voor de computeropgave moesten de kandidaten met het softwareprogramma Excel kunnen werken.
De wiskunde A1 en A1,2 examens bestonden dit jaar uit de computeropgave ‘Koerssprint’. Deze opgave is gebaseerd op het nog steeds actuele thema van het leasen van aandelen. Het onderzoeksthema was: “Wanneer is het voordeliger aandelen te leasen, wanneer zelf beleggen?” Na enkele theoretische vragen over de wiskundige achtergrond (exponentiële De wiskunde groei of meetkundige rijen) werden vragen gesteld voer de beste keus bij een bepaalde rentestand: “Zelf beleggen of laten beleggen via leasen”. Daartoe kon men in een spreadsheet het groeipercentage van de aandelenwaarde instellen als men zou gaan leasen. Daarnaast kon men het groeipercentage instellen waarmee de waarde van de eigen aandelen zou toenemen, die men maandelijks zou kunnen aankopen voor het geld dat men anders aan de kosten voor het leasen kwijt was. Zie figuur.



economie 1 havo

In de opgave ‘Arme scholier bestaat niet meer’ wordt een vergelijking gemaakt tussen de scholiereninkomsten van Nederland en die van een economiegroep in een klas op een school. De inkomensgegevens worden aangeboden in de vorm van een rapport van het NIBUD en bij de vergelijking moet actief gebruik gemaakt worden van Excel spreadsheets. Bij de tweede opgave wordt een klassiek economieprobleem op een interactieve manier aangeboden via een Excel spreadsheet. Zie figuur.



Het gaat hier om een optimaliseringprobleem ten aanzien kosten en opbrengsten. De leerling kan door middel van het veranderen van waarden het break-even punt op een actieve manier vinden.

In de laatste opgave wordt gebruik gemaakt van een videofragment om een economisch probleem te schetsen. Het onderwerp is werkloosheid in Nederland. Bovendien wordt in deze opgave van de leerlingen verwacht dat zij met Excel en een gegevensreeks een grafiek kunnen maken waarin de ontwikkeling van de werkloosheid in Nederland zichtbaar wordt.

aardrijkskunde havo
Het computerdeel van het examen bestaat uit het onderdeel Remote Sensing. Via reeds geïmporteerde Remote Sensing kleurenbanden moeten leerlingen een zogenaamd combinatiebeeld maken van een gegeven gebied, waarbij ze dienen te weten welke kleur ze in welk kleurkanaal moeten stoppen om een combinatiebeeld te maken dat vervolgens de basis wordt van een later te maken kaart. Dit combinatiebeeld wordt opgeslagen en is het eerste digitale product van de leerling waar een score voor wordt toegekend. Vervolgens gaat de leerling het gebied van het combinatiebeeld bemonsteren: hij/zij neemt monsters van een zo homogeen mogelijke aard van de verschillende soorten grondgebruik en water-kwaliteit. Hoe homogener de monsters en hoe beter ze zijn gespreid over het gebied, hoe beter de kaart wordt die hij/zij gaat maken. Het grondgebruik kan de leerling controleren via een bijgeleverde topografische zwart-witkaart van het gebied. De leerling maakt zelf een legenda, kiest een cartografisch verantwoorde kleur en betekenis, zet hier en daar een speciaal symbool in de kaart, zet er een titel boven en plaatst een schaalaanduiding op de kaart. Kortom: de leerling produceert zelf een kaart.

2.3 Vooruitblik 2005

Examens met ICT
De inzet van computers bij de examinering is voor havo/vwo vastgelegd in een beleidsnotitie van de CEVO voor de komende jaren. In 2005 wordt voor elk profiel bij één van de profielvakken een examen ontwikkeld waarbij de computer bij de examenafname zal worden ingezet.
In 2004 is gestart met de constructie van nieuwe opgaven en zijn de scholen aangezocht die aan de afname van de ICT-examens zullen deelnemen. Gemiddeld zal per examen bij 30% van de vragen gebruik moeten worden gemaakt van de computer om de vragen te beantwoorden. De overige vragen zijn identiek aan die in de reguliere examens voor die vakken.
In 2005 zullen deze examens worden afgenomen op een veel groter aantal scholen en bij een veel groter aantal leerlingen dan in de jaren 2003 en 2004. Het aantal deelnemende vwo-scholen zal net als het aantal havo-scholen uitgebreid worden tot ongeveer 40. Door het uitbreiden van het aantal deelnemende leerlingen per vak is het ook mogelijk om meer psychometrische zekerheid te krijgen ten aanzien van de vergelijking van de resultaten op deze examens ten opzichte van het landelijk reguliere examen.
Met ingang van 2006 zal de deelname aan de ICT-examens worden opengesteld voor alle scholen. In de tweede helft van 2004 is besluitvorming voorzien over de inzet van computers bij de examinering ná 2006, en de mogelijkheid om deze in de plaats te laten komen van de reguliere geheel schriftelijke, examens

Normvergelijking
De normvergelijkingsprocedures zullen in 2005 gecontinueerd worden voor de examens Frans, Duits en Engels op alle niveaus. Voor havo/vwo vindt normvergelijking ook plaats voor de examens wiskunde A, wiskunde B, natuurkunde en scheikunde, Nederlands en economie 1,2. Daarnaast voor biologie vwo en economie 1 voor havo.
Ook zal in 2005 de vergelijking tussen de resultaten van het eerste en tweede tijdvak weer uitgevoerd worden.

3 Examengegevens per vak

3.1 Cito-vakmedewerkers

Bij de samenstelling van de examenopgaven 2004 werkten de verschillende constructiegroepen onder leiding van de volgende Cito-medewerkers:

Talen		
Latijn	vwo	S. Jeurissen
Grieks	vwo	S. Jeurissen
Nederlands	havo/vwo	A. v.d. Kerkhof
Frans	havo/vwo	C. Peer
Duits	havo/vwo	O. de Vries, L. Melse
Engels	havo/vwo	N. v. Zuijlen, D. Samson
Spaans	havo/vwo	F. Agerkop
Russisch	havo/vwo	O. Petri
Turks	havo/vwo	T. Duindam
Arabisch	havo/vwo	A. de Graaf
Exacte vakken		
Wiskunde A	havo/vwo	K. Lagerwaard, G. Limpens, H. Boertien
Wiskunde B	havo/vwo	E. van Kervel, G. Stroomer, H. Boertien
Natuurkunde	havo/vwo	H. Joosten, B. Kneepkens, G. Boeijen
Scheikunde	havo/vwo	K. Beers, D. Witte
Biologie	havo/vwo	J. Brûens, M. Lieverse
M&M-vakken		
Economische vakken	havo/vwo	K. Blokker, F. Denie, N. Dieteren, L. Knoben
Aardrijkskunde	havo/vwo	B. v. Erp Taalman Kip
Geschiedenis	havo/vwo	S. Boom
Filosofie	havo/vwo	H. Wessels
Maatschappijleer	havo/vwo	V. Gijselhart
Kunst- en cultuurvakken		
Muziek	havo/vwo	T. Doevendans
Tehatex	havo/vwo	M. Wensing, M. Knüppe
CKV 2	havo/vwo	T. Doevendans, M. Wensing

3.2 Tabellen

In de hiernavolgende tabellen volgen per vak de algemene gegevens van de examens eerste tijdvak, verdeeld over de verschillende schooltypen. Nieuw hierbij is dat dit jaar de betrouwbaarheid op een andere wijze wordt geschat.
Tot 2003 werd voor de schatting van de interne consistentie de grootheid Cronbach's alpha gehanteerd. Er zijn echter betere schatters beschikbaar. Om die reden wordt in dit verslag de asymptotische GLB[3] gerapporteerd. Ook deze grootheid is net als Cronbach's alpha nog steeds een onderschatting van de werkelijke betrouwbaarheid, maar hij komt dichter in de buurt van de werkelijke betrouwbaarheid. Om deze betrouwbaarheidsmaat te kunnen berekenen dient het aantal kandidaten in de analyse ten minste een viervoud van het aantal vragen te zijn. Een toelichting op een aantal psychometrische begrippen staat in de bijlage.
De gegevens over de examens zijn bij alle vakken ontleend aan de verwerking van de afnameresultaten van een steekproef van kandidaten. In deze steekproef zijn uitsluitend kandidaten opgenomen uit het reguliere VO.
De genoemde aantallen van de kandidaten zijn gebaseerd op de gegevens van de inschrijving. Deze aantallen kunnen een overschatting van enkele honderden kandidaten te zien geven (vgl par 1.1).
Doorgaans worden per examen de gegevens gerapporteerd uit de toets- en itemanalyse. Wanneer deze is uitgevoerd op een aantal leerlingen kleiner dan 25 worden gegevens als gemiddeld cijfer, percentage onvoldoende, betrouwbaarheid en standaardmeetfout niet vermeld.

Tabel 12 Algemene gegevens van het CE Nederlands 2004

	havo	vwo
Totaal aantal kandidaten	45.176	31.967
Steekproefgrootte	1.822	2.035
Aantal vragen*	29	27
P'-waarde van het examen jongens + meisjes	0,51	0,61
P'-waarde, alleen jongens	0,50	0,58
P'-waarde, alleen meisjes	0,52	0,62
Maximumscore	50	48
Gemiddelde score	26	29
Normeringsterm	1,4	1,1
Gemiddeld cijfer	6	6,6
Percentage onvoldoendes (<5.5)	30	19
Standaardafwijking	5,6	6,2
Betrouwbaarheid (asymp. GLB)	0,53	0,53
Standaardmeetfout	3,9	4,2

*aantal vragen inclusief beoordelingscomponenten van de samenvattingsopdracht

[3] De afkorting GLB staat voor Greatest lower bound wat zo veel wil zeggen als de hoogste ondergrens.

Tabel 13 Algemene gegevens van het CE Latijn en Grieks 2004

	Latijn	Grieks
Totaal aantal kandidaten	5.698	2.086
Steekproefgrootte	1.220	739
Aantal vragen	47	47
P'-waarde van het examen jongens + meisjes	0,55	0,53
P'-waarde, alleen jongens	0,52	0,51
P'-waarde, alleen meisjes	0,57	0,54
Maximumscore	86	86
Gemiddelde score	48	46
Normeringsterm	1,5	1,6
Gemiddeld cijfer	6,5	6,4
Percentage onvoldoendes (<5.5)	26	29
Standaardafwijking	14,2	14,7
Betrouwbaarheid (asymp. GLB)	0,91	0,93
Standaardmeetfout	4,3	3,9

Tabel 14 Algemene gegevens van het CE Frans 1,2 2004

	havo	vwo
Totaal aantal kandidaten	8.733	7.698
Steekproefgrootte	1.875	1.696
Aantal vragen	41	44
P'-waarde van het examen jongens + meisjes	0,60	0,68
P'-waarde, alleen jongens	0,61	0,70
P'-waarde, alleen meisjes	0,60	0,68
Maximumscore	45	46
Gemiddelde score	27	31
Normeringsterm	0,6	0
Gemiddeld cijfer	6,0	6,2
Percentage onvoldoendes (<5.5)	34	27
Standaardafwijking	5,6	6,4
Betrouwbaarheid (asymp. GLB)	0,76	0,83
Standaardmeetfout	2,7	2,6

Tabel 15 Algemene gegevens van het CE Duits 1,2 2004

	havo	vwo
Totaal aantal kandidaten	11.588	6.911
Steekproefgrootte	2.007	1.777
Aantal vragen	45	41
P'-waarde van het examen jongens + meisjes	0,63	0,67
P'-waarde, alleen jongens	0,64	0,69
P'-waarde, alleen meisjes	0,62	0,67
Maximumscore	47	47
Gemiddelde score	29	32
Normeringsterm	0,6	0,4
Gemiddeld cijfer	6,2	6,5
Percentage onvoldoendes (<5.5)	23	23
Standaardafwijking	5,4	6,5
Betrouwbaarheid (asymp. GLB)	0,76	0,83
Standaardmeetfout	2,6	2,7

Tabel 16 Algemene gegevens van het CE Engels 2004

	havo	vwo
Totaal aantal kandidaten	45.792	31.967
Steekproefgrootte	2.036	1.858
Aantal vragen	42	44
P'-waarde van het examen jongens + meisjes	0,64	0,59
P'-waarde, alleen jongens	0,65	0,62
P'-waarde, alleen meisjes	0,63	0,57
Maximumscore	50	51
Gemiddelde score	32	30
Normeringsterm	0,6	1,1
Gemiddeld cijfer	6,4	6,4
Percentage onvoldoendes (<5.5)	22	25
Standaardafwijking	7,0	7,8
Betrouwbaarheid (asymp. GLB)	0,82	0,86
Standaardmeetfout	3,0	2,9

Tabel 17 Algemene gegevens van het CE Spaans (1,2) 2004

	havo	vwo
Totaal aantal kandidaten	605	773
Steekproefgrootte	80	122
Aantal vragen	41	45
P'-waarde van het examen jongens + meisjes	0,61	0,70
P'-waarde, alleen jongens	0,70	0,73
P'-waarde, alleen meisjes	0,57	0,69
Maximumscore	46	50
Gemiddelde score	28	35
Normeringsterm	0,8	0,2
Gemiddeld cijfer	6,3	6,5
Percentage onvoldoendes (<5.5)	29	20
Standaardafwijking	7,4	6,5
Betrouwbaarheid (asymp. GLB)*	*	*
Standaardmeetfout*	*	*

* niet berekend wegens gering aantal kandidaten in verhouding tot het aantal vragen

Tabel 18 Algemene gegevens van het CE Russisch (1,2) 2004

	havo	vwo
Totaal aantal kandidaten	3	13
Steekproefgrootte	–	4
Aantal vragen	44	44
P'-waarde van het examen jongens + meisjes	–	–
P'-waarde, alleen jongens	–	–
P'-waarde, alleen meisjes	–	–
Maximumscore	48	48
Gemiddelde score	–	–
Normeringsterm	0,7	0,7
Gemiddeld cijfer	–	–
Percentage onvoldoendes (<5.5)	–	–
Standaardafwijking	–	–
Betrouwbaarheid (asymp. GLB)	–	–
Standaardmeetfout	–	–

Tabel 19 Algemene gegevens van het CE Turks (1,2) 2004

	havo	vwo
Totaal aantal kandidaten	24	5
Steekproefgrootte	–	–
Aantal vragen	42	44
P'-waarde van het examen jongens + meisjes	–	–
P'-waarde, alleen jongens	–	–
P'-waarde, alleen meisjes	–	–
Maximumscore	48	50
Gemiddelde score	–	–
Normeringsterm	0,7	0,7
Gemiddeld cijfer	–	–
Percentage onvoldoendes (<5.5)	–	–
Standaardafwijking	–	–
Betrouwbaarheid (asymp. GLB)	–	–
Standaardmeetfout	–	–

Tabel 20 Algemene gegevens van het CE Arabisch (1,2) 2004

	havo	vwo
Totaal aantal kandidaten	9	12
Steekproefgrootte	–	–
Aantal vragen	41	43
P'-waarde van het examen jongens + meisjes	–	–
P'-waarde, alleen jongens	–	–
P'-waarde, alleen meisjes	–	–
Maximumscore	47	46
Gemiddelde score	–	–
Normeringsterm	0,7	0,7
Gemiddeld cijfer	–	–
Percentage onvoldoendes (<5.5)	–	–
Standaardafwijking	–	–
Betrouwbaarheid (asymp. GLB)	–	–
Standaardmeetfout	–	–

Tabel 21 Algemene gegevens van het CE Wiskunde A 2004

	havo Wisk. A1,2	vwo Wisk. A1	vwo Wisk. A1 compex	vwo Wisk. A1,2	vwo Wisk. A1,2 compex
Totaal aantal kandidaten	19.956	5.673	74	11.568	160
Steekproefgrootte	2.125	1.738	64	1.937	90
Aantal vragen	21	21	24	21	22
P'-waarde van het examen jongens + meisjes	0,57	0,55	0,56	0,51	0,53
P'-waarde, alleen jongens	0,57	0,57	0,57	0,51	0,52
P'-waarde, alleen meisjes	0,56	0,55	0,56	0,50	0,53
Maximumscore	83	83	87	87	90
Gemiddelde score	47	46	49	44	47
Normeringsterm	1,1	1,4	1,4	1,5	1,6
Gemiddeld cijfer	6,2	6,4	6,5	6,1	6,3
Percentage onvoldoendes (<5.5)	29	24	17	32	28
Standaardafwijking	12,2	12,2	10,6	11,9	12,5
Betrouwbaarheid (asymp. GLB)	0,8	0,8	*	0,77	0,91
Standaardmeetfout	5,5	5,4	*	5,7	3,7

* niet berekend wegens gering aantal kandidaten in verhouding tot het aantal vragen

Tabel 22 Algemene gegevens van het CE Wiskunde B 2004

	havo Wisk. B1	havo Wisk. B1,2	vwo Wisk. B1	vwo Wisk. B1,2
Totaal aantal kandidaten	7.842	6.033	9.396	6.592
Steekproefgrootte	1.869	1.813	2.081	2.027
Aantal vragen	21	21	19	18
P'-waarde van het examen jongens + meisjes	0,60	0,57	0,59	0,59
P'-waarde, alleen jongens	0,60	0,57	0,59	0,58
P'-waarde, alleen meisjes	0,60	0,58	0,59	0,60
Maximumscore	81	82	86	86
Gemiddelde score	49	47	51	50
Normeringsterm	0,6	1,1	0,8	1,2
Gemiddeld cijfer	6,0	6,2	6,1	6,5
Percentage onvoldoendes (<5.5)	32	28	29	26
Standaardafwijking	11,4	11,8	11,7	14,3
Betrouwbaarheid (asymp. GLB)	0,77	0,79	0,79	0,82
Standaardmeetfout	5,5	5,4	5,4	6,1

Tabel 23 Algemene gegevens van het CE Natuurkunde HAVO 2004

	havo N1	havo N1,2	havo compex
Totaal aantal kandidaten	7.729	5.532	221
Steekproefgrootte	1.964	1.788	190
Aantal vragen	25	27	26
P'-waarde van het examen jongens + meisjes	0,61	0,67	0,70
P'-waarde, alleen jongens	0,62	0,68	0,70
P'-waarde, alleen meisjes	0,60	0,64	0,63
Maximumscore	81	81	80
Gemiddelde score	49	55	56
Normeringsterm	0,8	0,5	0,5
Gemiddeld cijfer	6,3	6,6	6,8
Percentage onvoldoendes (<5.5)	24	16	7
Standaardafwijking	10,3	9,8	9,0
Betrouwbaarheid (asymp. GLB)	0,77	0,75	0,77
Standaardmeetfout	4,9	4,9	4,3

Tabel 24 Algemene gegevens van het CE Natuurkunde VWO 2004

	vwo N1	vwo N1,2	vwo moderne natuurkunde	vwo N1,2 compex
Totaal aantal kandidaten	9.306	5.647	302	152
Steekproefgrootte	1.891	1.779	239	83
Aantal vragen	25	23	24	26
P'-waarde van het examen jongens + meisjes	0,58	0,67	0,66	0,69
P'-waarde, alleen jongens	0,59	0,67	0,66	0,70
P'-waarde, alleen meisjes	0,57	0,66	0,65	0,62
Maximumscore	79	79	82	80
Gemiddelde score	46	53	54	55
Normeringsterm	1,3	0,8	0,9	0,8
Gemiddeld cijfer	6,5	6,8	6,8	7,0
Percentage onvoldoendes (<5.5)	18	14	14	10
Standaardafwijking	9,9	11,2	11,5	11,0
Betrouwbaarheid (asymp. GLB)	0,79	0,83	0,87	*
Standaardmeetfout	4,6	4,6	4,2	*

* niet berekend wegens gering aantal kandidaten in verhouding tot het aantal vragen

Tabel 25 Algemene gegevens van het CE Scheikunde 2004

	havo	vwo Sk 1	vwo Sk 1, 2
Totaal aantal kandidaten	12.973	8.338	6.970
Steekproefgrootte	2.053	1.810	1.811
Aantal vragen	37	24	27
P'-waarde van het examen jongens + meisjes	0,61	0,68	0,68
P'-waarde, alleen jongens	0,62	0,67	0,67
P'-waarde, alleen meisjes	0,59	0,69	0,69
Maximumscore	81	67	71
Gemiddelde score	49	46	48
Normeringsterm	0,7	0,5	0,7
Gemiddeld cijfer	6,2	6,6	6,8
Percentage onvoldoendes (<5.5)	25	16	17
Standaardafwijking	10,6	8,8	10,1
Betrouwbaarheid (asymp. GLB)	0,83	0,80	0,84
Standaardmeetfout	4,4	3,9	4,0

Tabel 26 Algemene gegevens van het CE Biologie (1,2) 2004

	havo	havo compex	vwo	vwo compex
Totaal aantal kandidaten	16.502	504	11.846	206
Steekproefgrootte	1.766	370	1.814	165
Aantal vragen	49	49	41	37
P'-waarde van het examen jongens + meisjes	0,52	0,50	0,56	0,59
P'-waarde, alleen jongens	0,55	0,53	0,57	0,62
P'-waarde, alleen meisjes	0,51	0,49	0,55	0,58
Maximumscore	87	84	77	71
Gemiddelde score	45	42	43	42
Normeringsterm	1,4	1,5	1,3	1,3
Gemiddeld cijfer	6,1	6,0	6,3	6,7
Percentage onvoldoendes (<5.5)	25	23	17	10
Standaardafwijking	9,0	8,5	8,4	7,3
Betrouwbaarheid (asymp. GLB)	0,74	0,78	0,70	0,77
Standaardmeetfout	4,6	4,0	4,6	3,5

Tabel 27 Algemene gegevens van het CE Economische vakken havo 2004

	havo Ec1	havo Ec1 compex	havo Ec1,2	havo M&O
Totaal aantal kandidaten	18.001	555	19.718	11.883
Steekproefgrootte	1.807	238	1.869	1.745
Aantal vragen	31	31	27	33
P'-waarde van het examen jongens + meisjes	0,54	0,54	0,53	0,58
P'-waarde, alleen jongens	0,59	0,61	0,54	0,58
P'-waarde, alleen meisjes	0,53	0,51	0,50	0,56
Maximumscore	58	59	59	69
Gemiddelde score	31	32	31	40
Normeringsterm	1,3	1,4	1,2	1
Gemiddeld cijfer	6,2	6,2	5,9	6,2
Percentage onvoldoendes (<5.5)	26	24	33	28
Standaardafwijking	7,6	7,0	8,2	9,3
Betrouwbaarheid (asymp. GLB)	0,77	0,82	0,77	0,83
Standaardmeetfout	3,6	3,0	4,0	3,8

Tabel 28 Algemene gegevens van het CE Economische vakken vwo 2004

	vwo Ec 1	vwo Ec 1, 2	vwo M&O
Totaal aantal kandidaten	6.798	11.987	7.432
Steekproefgrootte	1.584	1.844	1.880
Aantal vragen	28	25	32
P'-waarde van het examen jongens + meisjes	0,60	0,59	0,64
P'-waarde, alleen jongens	0,62	0,59	0,64
P'-waarde, alleen meisjes	0,57	0,59	0,63
Maximumscore	56	56	70
Gemiddelde score	33	33	45
Normeringsterm	0,9	0,7	0,5
Gemiddeld cijfer	6,3	6,0	6,2
Percentage onvoldoendes (<5.5)	27	31	26
Standaardafwijking	7,8	7,3	9,1
Betrouwbaarheid (asymp. GLB)	0,76	0,73	0,84
Standaardmeetfout	3,8	3,8	3,6

Tabel 29 Algemene gegevens van het CE Aardrijkskunde 2004

	havo	havo compex	vwo
Totaal aantal kandidaten	22.079	365	13.523
Steekproefgrootte	1.977	280	1.881
Aantal vragen	32	26	40
P'-waarde van het examen jongens + meisjes	0,52	0,52	0,56
P'-waarde, alleen jongens	0,54	0,53	0,58
P'-waarde, alleen meisjes	0,51	0,49	0,54
Maximumscore	60	58	72
Gemiddelde score	31	30	40
Normeringsterm	1,4	1,8	1,0
Gemiddeld cijfer	6,1	6,5	6,0
Percentage onvoldoendes (<5.5)	21	14	27
Standaardafwijking	6,3	6,0	7,3
Betrouwbaarheid (asymp. GLB)	0,66	0,73	0,78
Standaardmeetfout	3,7	3,1	3,4

Tabel 30 Algemene gegevens van het CE Geschiedenis en staatsinrichting 2004

	havo	vwo
Totaal aantal kandidaten	33.164	17.698
Steekproefgrootte	2.033	1.932
Aantal vragen	27	25
P'-waarde van het examen jongens + meisjes	0,63	0,67
P'-waarde, alleen jongens	0,63	0,66
P'-waarde, alleen meisjes	0,63	0,67
Maximumscore	81	84
Gemiddelde score	51	56
Normeringsterm	0,5	0,5
Gemiddeld cijfer	6,2	6,5
Percentage onvoldoendes (<5.5)	21	13
Standaardafwijking	8,4	8,8
Betrouwbaarheid (asymp. GLB)	0,73	0,71
Standaardmeetfout	4,4	4,7

Tabel 31 Algemene gegevens van het CE Maatschappijleer 2004

	havo	vwo
Totaal aantal kandidaten	2.214	1.315
Steekproefgrootte	407	350
Aantal vragen	40	35
P'-waarde van het examen jongens + meisjes	0,57	0,63
P'-waarde, alleen jongens	0,56	0,63
P'-waarde, alleen meisjes	0,58	0,64
Maximumscore	88	87
Gemiddelde score	50	55
Normeringsterm	0,8	0,7
Gemiddeld cijfer	5,9	6,4
Percentage onvoldoendes (<5.5)	29	17
Standaardafwijking	10,7	11,1
Betrouwbaarheid (asymp. GLB)	0,87	0,87
Standaardmeetfout	3,9	4,0

Tabel 32 Algemene gegevens van het CE Filosofie 2004

	havo	vwo
Totaal aantal kandidaten	469	2.427
Steekproefgrootte	186	759
Aantal vragen	16	15
P'-waarde van het examen jongens + meisjes	0,63	0,68
P'-waarde, alleen jongens	0,62	0,64
P'-waarde, alleen meisjes	0,64	0,69
Maximumscore	48	44
Gemiddelde score	30	30
Normeringsterm	0,9	0,5
Gemiddeld cijfer	6,6	6,6
Percentage onvoldoendes (<5.5)	19	19
Standaardafwijking	6,3	6,2
Betrouwbaarheid (asymp. GLB)	0,72	0,75
Standaardmeetfout	3,3	3,1

Tabel 33 Algemene gegevens van het CE Muziek 2004

	havo	vwo
Totaal aantal kandidaten	1.159	687
Steekproefgrootte	469	373
Aantal vragen	49	50
P'-waarde van het examen jongens + meisjes	0,58	0,58
P'-waarde, alleen jongens	0,63	0,58
P'-waarde, alleen meisjes	0,57	0,57
Maximumscore	90	90
Gemiddelde score	52	52
Normeringsterm	1,3	1,3
Gemiddeld cijfer	6,5	6,5
Percentage onvoldoendes (<5.5)	19	18
Standaardafwijking	11,6	10,6
Betrouwbaarheid (asymp. GLB)	0,89	0,90
Standaardmeetfout	3,9	3,3

Tabel 34 Algemene gegevens van het CE Tekenen, handvaardigheid en textiele werkvormen 2004

	havo tehatex	vwo tehatex
Totaal aantal kandidaten	5.812	2.675
Steekproefgrootte	1.415	1.017
Aantal vragen	37	38
P'-waarde van het examen jongens + meisjes	0,53	0,58
P'-waarde, alleen jongens	0,53	0,56
P'-waarde, alleen meisjes	0,53	0,58
Maximumscore	74	73
Gemiddelde score	39	42
Normeringsterm	1,4	1,2
Gemiddeld cijfer	6,2	6,4
Percentage onvoldoendes (<5.5)	23	18
Standaardafwijking	7,5	8,6
Betrouwbaarheid (asymp. GLB)	0,76	0,83
Standaardmeetfout	3,7	3,5

Tabel 35 Algemene gegevens Cultureel Kunstzinnige Vorming 2 (CKV 2) 2004

	havo	vwo
Totaal aantal kandidaten	5305	2480
Steekproefgrootte	756	666
Aantal vragen	38	35
P'-waarde van het examen jongens + meisjes	0,59	0,60
P'-waarde, alleen jongens	0,57	0,58
P'-waarde, alleen meisjes	0,59	0,61
Maximumscore	70	67
Gemiddelde score	41	41
Normeringsterm	1,0	1,0
Gemiddeld cijfer	6,3	6,4
Percentage onvoldoendes (<5.5)	19	17
Standaardafwijking	7,6	7,3
Betrouwbaarheid (asymp. GLB)	0,80	0,78
Standaardmeetfout	3,4	3,4

Bijlage

Bijlage 1 Enkele begrippen

betrouwbaarheid

De betrouwbaarheid is de mate waarin men staat kan maken op meetresultaten, dat wil zeggen de mate waarin de scores consistent, nauwkeurig en reproduceerbaar zijn, kortom vrij van meetfouten.

Bij een betrouwbare meting zal het resultaat dus niet beïnvloed mogen zijn door storende factoren met betrekking tot de toets, kandidaat en beoordelaar zoals het tijdstip van de toetsafname, de specifieke vormgeving van de toets, of pech of geluk met de opgaven. De mate waarin een meting ongevoelig is voor bepaalde, verstorende factoren, kan worden geschat door berekening van een betrouwbaarheidscoëfficiënt. De waarde van een betrouwbaarheidscoëfficiënt ligt tussen 0 en 1. De coëfficiënt kan op verschillende manieren worden vastgesteld; elke manier richt zich op bepaalde storende factoren:

a via de test-hertest-methode;
b via de parallelle-vormen-methode;
c via een generaliseerbaarheidscoëfficiënt;
d via een berekening van de interne consistentie

De berekening van de interne consistentie wordt in de praktijk het meest gebruikt omdat deze uitgevoerd kan worden op één toetsafname. De interne consistentie geeft vooral aan in hoeverre van item tot item hetzelfde wordt gemeten. Dit gegeven wordt gebruikt om te schatten in hoeverre iemand dezelfde score zou behalen bij het maken van een denkbeeldige, andere vergelijkbare toets over dezelfde leerstof. Tot 2003 werd voor de schatting van de interne consistentie de grootheid Cronbach's alpha gehanteerd. Er zijn echter betere schatters beschikbaar. Om die reden wordt in dit verslag de asymptotische GLB gerapporteerd. Ook deze grootheid is net als Cronbach's alpha nog steeds een onderschatting van de werkelijke betrouwbaarheid, maar hij komt dichter in de buurt van de werkelijke betrouwbaarheid.

cesuur

De cesuur is de grens tussen de hoogste toetsscore waaraan een onvoldoende en de laagste toetsscore waaraan een voldoende wordt toegekend. Wanneer bijvoorbeeld aan een score van 30 punten een voldoende wordt toegekend en aan 29 punten een onvoldoende dan ligt de cesuur tussen 29 en 30. Sinds 2000 wordt de cesuur niet meer expliciet vastgesteld, maar wordt het examen genormeerd door de vaststelling van een normeringsterm. Na vaststelling van de normeringsterm kan men vervolgens de cesuur berekenen door na te gaan welke score het cijfer 5,5 oplevert.

normeringsterm (N-term)

De normeringsterm is een ijkvariabele die na de afname van het examen zodanig wordt vastgesteld dat het examen vergelijkbare eisen stelt als in eerdere jaren. De normeringsterm kan variëren tussen 0,0 en 2,0.

p'-waarde
open vraag

De p'-waarde is een getal tussen 0 en 1 dat de moeilijkheidsgraad van een opgave weergeeft. De p'-waarde wordt berekend door de gemiddelde score op een opgave te delen door de maximaal haalbare score op die opgave.

Een opgave met een p'-waarde van .10 is erg moeilijk; een opgave met een p'-waarde van .90 is erg gemakkelijk.

p-waarde
gesloten vraag

De p-waarde van een meerkeuzevraag is de proportie kandidaten die het goede antwoord heeft gekozen. Met dit getal wordt de moeilijkheidsgraad van een item weergegeven.

p- of p'-waarde
van het examen

Zie: gemiddelde p-waarde ($\overline{P}$).

gemiddelde p-waarde ($\overline{P}$)

De gemiddelde p-waarde ($\overline{P}$) is het gemiddelde van een reeks p-waarden, verkregen door alle p-waarden op te tellen en de som te delen door het aantal p-waarden of, wat nauwkeuriger is, de gemiddelde score ($\overline{X}$) te delen door het aantal items van een toets.

Bij gewogen scores, die bij open vragen kunnen voorkomen, wordt de p'-waarde van de gehele toets berekend door de gemiddelde toetsscore te delen door het maximaal haalbare aantal punten.

standaardafwijking

De standaardafwijking (S_x) is een maat voor de spreiding van getallen rondom hun gemiddelde. In het geval van een toets gaat het om de spreiding van toetsscores rondom de gemiddelde score.

De formule voor het berekenen van de standaardafwijking is:

$$S_x \quad \sqrt{\frac{(X_i \quad \overline{X})^2}{N}}$$

S_x = standaardafwijking

X_i = toetsscore van kandidaat i

$\overline{X}$ = gemiddelde toetsscore

N = aantal kandidaten

standaardmeetfout

De standaardmeetfout (S_e) is een indicatie voor de onnauwkeurigheid van een meting.
De grootte van de standaardmeetfout van een toets hangt af van de betrouwbaarheid en de standaardafwijking van de toetsscores.

De formule voor het berekenen van de standaardmeetfout is:

$$S_e = S_x \sqrt{(1-r)}$$

S_e = standaardmeetfout
S_x = standaardafwijking toetsscores
r = toetsbetrouwbaarheid