

Het schoolexamen in het voortgezet onderwijs

**Verslag van een onderzoek naar de kwaliteit van het schoolexamen
bij de vakken Engels, Nederlands, biologie en wiskunde**

Eindversie 3 december 2008



zeker weten

Samenvatting

In opdracht van de Inspectie van het Onderwijs heeft het Cito een onderzoek uitgevoerd naar de vakinhoudelijke en toetstechnische kwaliteit van de schoolexamens (SE) in het voortgezet onderwijs. Het onderzoek is uitgevoerd bij de vakken Engels en biologie in het havo en Nederlands en wiskunde in het vmbo.

De kwaliteit van het SE is onder leiding van toetsdeskundigen van het Cito vastgesteld door in totaal twaalf beoordelaars (drie per vak). Allen waren ervaren en volledig bevoegde docenten die daarnaast betrokken waren bij de constructie van de centrale examens. Na een training in het beoordelen hebben de beoordelaars de kwaliteit van het SE op negentien indicatoren beoordeeld. Belangrijke aandachtspunten waren daarbij de inhoudelijke dekking van het eindexamenprogramma, de toetstechnische kwaliteit van de opgaven, de moeilijkheid van de toetsen, de strengheid/soepelheid van de beoordelingsmiddelen en de becijfering, de volledigheid en relevantie van de beoordelingsaspecten, het vooraf vastleggen en vroegtijdig meedelen van de scorepunten, het aangeven van het gewicht bij de bepaling van het cijfer, het moment van toetsing en de mate waarin docenten de toetsen, beoordeling en normering in onderling overleg vaststellen.

De belangrijkste conclusie uit het onderzoek is dat de vakinhoudelijke en toetstechnische kwaliteit van het SE naar het oordeel van de beoordelaars op de overgrote meerderheid van de onderzochte scholen in orde is. Althans, naar de mening van de beoordelaars waren de geconstateerde tekortkomingen vrijwel nooit ernstig genoeg om de kwalificatie 'onvoldoende' te rechtvaardigen. De percentages scholen met een SE van voldoende kwaliteit bedragen volgens de gehanteerde criteria voor de vakken Engels, Nederlands, biologie en wiskunde respectievelijk 78%, 100%, 96% en 94%.

Een tweede conclusie is dat er volgens de beoordelaars aanzienlijke verschillen tussen scholen bestaan in de inhoudelijke en toetstechnische kwaliteit van het SE. Zo gaven de beoordelaars de school met het "slechtste" SE Engels op vijf indicatoren een voldoende; de dertien scholen met het "beste" SE Engels hadden hun schoolexamen op alle indicatoren in orde. Ook bij de vakken biologie en wiskunde verschilden scholen in de kwaliteit van het SE, zij het dat de verschillen minder uitgesproken waren dan bij Engels. Het vak Nederlands vormt hierop een uitzondering in de zin dat de beoordelaars de kwaliteit van het SE Nederlands volgens de gehanteerde criteria van geen enkele school als onvoldoende aanmerkten (al constateerden zij wel kwaliteitsverschillen).

Een derde conclusie is dat de kwaliteit van het schoolexamen naar de mening van de beoordelaars niet voor alle indicatoren even hoog is. Zo waren de verplichte onderdelen van het SE Engels op een derde van de scholen in onvoldoende mate in het examen vertegenwoordigd; daarentegen maakten nagenoeg alle scholen in voldoende mate gebruik van antwoord- of beoordelingsmodellen. Bij biologie werden bijvoorbeeld de toetsen en opdrachten op een op de zes scholen als te moeilijk of te gemakkelijk aangemerkt; daarentegen voldeden de beoordelingsaspecten op nagenoeg alle scholen in redelijke mate aan de eis van relevantie. Bij wiskunde werden bijvoorbeeld de scorepunten naar het oordeel van de beoordelaars op ruim een derde van de scholen niet of te laat aan de kandidaten meegedeeld; daarentegen namen alle scholen de toetsen in voldoende mate af op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt konden hebben. Wederom vormt het vak Nederlands een uitzondering: de beoordelaars waren de mening toegedaan dat alle onderzochte scholen aan alle indicatoren van de kwaliteit van het SE voldeden.

Een vierde conclusie is dat de kwaliteit van het schoolexamen niet voor alle vakonderdelen gelijk is. Zo bleken de beoordelaars Engels de kwaliteit van het onderdeel luistervaardigheid veel hoger te waarderen dan van het onderdeel schrijfvaardigheid. Bij wiskunde werd de kwaliteit van de vakonderdelen Algebraïsche verbanden, Rekenen, meten en schatten en Meetkunde op nagenoeg alle scholen als voldoende aangemerkt, maar van het onderdeel Statistiek was de kwaliteit op ruim een vijfde van de scholen niet in orde. Het vak Nederlands vormt hierop een uitzondering omdat de beoordelaars hier van mening waren dat de kwaliteit van alle beoordeelde vakonderdelen voldoende was. Over de kwaliteit van de onderdelen van het vak biologie kan vanwege het grote percentage ontbrekende waarnemingen geen uitspraak worden gedaan.

Al met al zijn de beoordelaars van mening dat de schoolexamens in het voortgezet onderwijs op de overgrote meerderheid van de scholen in voldoende mate voldoen aan redelijke eisen ten aanzien van de inhoudelijke dekking van het eindexamenprogramma, het moment van toetsing, de toetstechnische kwaliteit van de opgaven, de moeilijkheid van de toetsen, de strengheid/soepelheid van de beoordelingsmiddelen en de becijfering, de volledigheid en relevantie van de beoordelingsaspecten, het vooraf vastleggen en

vroegtijdig meedelen van de scorepunten, het aangeven van het gewicht bij de bepaling van het cijfer en de samenwerking tussen docenten bij de toetsontwikkeling, beoordeling en normering.

Inhoudsopgave

Samenvatting

1 Inleiding 7

2 Methode van onderzoek 9

- 2.1 Examenvakken, steekproef en response 9
- 2.2 Toetsingskader 9
- 2.3 Geraadpleegde bronnen 13
- 2.4 Beoordelingsmethode 13
- 2.5 De uitvoering van het onderzoek 16
- 2.6 De vragenlijst Inventarisatie Examenpraktijk 17
- 2.7 Statistische analyse 18
- 2.8 Beoordelaarsovereenstemming 19
- 2.9 Ontbrekende waarnemingen 23

3 Resultaten 27

- 3.1 De kwaliteit van het SE Engels 27
 - 3.1.1 Verschillen tussen scholen 27
 - 3.1.2 Algemene kwaliteit per indicator 27
 - 3.1.3 Kwaliteit per vakonderdeel 28
 - 3.1.4 Voorbeelden examenpraktijk 31
 - 3.1.5 Conclusie en aanbevelingen 37
- 3.2 De kwaliteit van het SE Nederlands 40
 - 3.2.1 Verschillen tussen scholen 40
 - 3.2.2 Algemene kwaliteit per indicator 40
 - 3.2.3 Kwaliteit per vakonderdeel 40
 - 3.2.4 Voorbeelden examenpraktijk 41
 - 3.2.5 Conclusie en aanbevelingen 46
- 3.3 De kwaliteit van het SE biologie 48
 - 3.3.1 Verschillen tussen scholen 48
 - 3.3.2 Algemene kwaliteit per indicator 48
 - 3.3.3 Kwaliteit per vakonderdeel 49
 - 3.3.4 Voorbeelden examenpraktijk 51
 - 3.3.5 Conclusies en aanbevelingen 56
- 3.4 De kwaliteit van het SE wiskunde 57
 - 3.4.1 Verschillen tussen scholen 57
 - 3.4.2 Algemene kwaliteit per indicator 57
 - 3.4.3 Kwaliteit per vakonderdeel 58
 - 3.4.4 Voorbeelden examenpraktijk 59
 - 3.4.5 Conclusies en aanbevelingen 61

4 Conclusies, discussie en aanbevelingen 62

- 4.1 Conclusies 62
- 4.2 Discussie 67
 - 4.2.1 Vakinhoudelijke kanttekeningen 68
 - 4.2.2 Organisatorische en beoordelingstechnische kanttekeningen 69
- 4.3 Aanbevelingen 70
 - 4.3.1 Regelgeving 70
 - 4.3.2 Het eindexamenprogramma 70
 - 4.3.3 De praktijk van het schoolexamen 70
 - 4.3.4 Vervolgonderzoek 71

Literatuur 73

Bijlagen 75

- 1 Toetsingskader Engels 77
- 2 Toetsingskader Nederlands 82
- 3 Toetsingskader biologie 88
- 4 Toetsingskader wiskunde 92
- 5 Vragenlijst Inventarisatie Examenpraktijk 98
- 6 Vakinhoudelijke en andere kanttekeningen bij de resultaten Nederlands 110

1 Inleiding

In Nederland is de school verantwoordelijk voor het samenstellen, afnemen en beoordelen van de schoolexamens. Anders dan in sommige andere landen wordt er op de kwaliteit van de schoolexamens geen interne of externe moderatie toegepast (Kuhlemeier, Van Weeren & Van der Werf, 2006). Het toezicht op de kwaliteit van de schoolexamens is in handen van de Inspectie die daartoe de scholen regelmatig bezoekt. Volgens een recent Onderwijsverslag is de kwaliteit van het SE op driekwart van de scholen voor voortgezet onderwijs in orde (Inspectie van het Onderwijs, 2007c). Op deze scholen aanvaardt de inspectie de verklaring van de schoolleiding dat deze de kwaliteit van de schoolexamens waarborgt, na controle bij leraren en leerlingen en bij gebrek aan contra-indicaties. In het toezicht op de schoolexamens kijkt de Inspectie vooral naar de mate waarin de school de wettelijke bepalingen naleeft. Belangrijke aandachtspunten zijn daarbij het Examenreglement, het Programma van Toetsing en Afsluiting en de documentatie van schoolinterne afspraken rond de afname, correctie en communicatie. In het Onderwijsverslag 2005-2006 constateert de Inspectie dat de mate waarin scholen de kwaliteit van de toetsing van het schoolexamen waarborgen, moet verbeteren (Inspectie van het Onderwijs, 2007c).

De Inspectie heeft al enige tijd geen onderzoek meer gepubliceerd naar de vakinhoudelijke en toetstechnische kwaliteit van de schoolexamens. Het laatste onderzoek is van bijna tien jaar geleden (Inspectie van het onderwijs, 1995, 1996, 1998, 1999). Een van de conclusies was dat de kwaliteit van het SE sterk van school tot school verschilde. Het meest recent gepubliceerde onderzoek was voor het vak Latijn (Inspectie van het onderwijs, 1999). In dat onderzoek voldeed 70% van de scholen niet aan alle inhoudelijke eisen van het eindexamenprogramma. De Inspectie constateerde verder grote verschillen in de vorm en inhoud van de toetsen en in de spreiding en omvang van de toetsen die maakten dat de slaagkansen van kandidaten sterk uiteenliepen. Van alle schoolexamens was 40% (veel) te gemakkelijk in vergelijking met het centrale examen. De Inspectie concludeerde dat het merendeel van de schoolexamens Latijn niet voldeed aan redelijke eisen ten aanzien van de inhoud, de vorm, de omvang, de constructie, de normering en het niveau. Ter verklaring voerde de Inspectie aan dat veel docenten zich onvoldoende bewust waren van de formele eisen in het eindexamenprogramma en onvoldoende professionele aandacht of discipline hadden om het niveau van de toetsen adequaat aan te scherpen. Hoe het anno 2008 met de vakinhoudelijke en toetstechnische kwaliteit van de schoolexamens gesteld is, weten we niet precies. Wel heeft de Inspectie op grond van de schoolbezoeken het vermoeden dat de kwaliteitsverschillen tussen scholen nog steeds groot zijn (Inspectie van het Onderwijs, 2006). Volgens de Inspectie manifesteren deze kwaliteitsverschillen zich vooral in de aantallen toetsen, de moeilijkheidsgraad van de toetsen en de kwaliteit van de vormgeving en afnamecondities van de toetsen (Inspectie van het Onderwijs, 2006).

De laatste jaren is het aandeel van de schoolexamens in de afsluiting van het voortgezet onderwijs steeds groter geworden. Tegelijkertijd is ook het percentage geslaagden toegenomen. Een van de vele mogelijke oorzaken heeft mogelijk te maken met de schoolexamens (Kremers, Kuhlemeier & Wiegers, 2006). Volgens de Inspectie van het Onderwijs komt de toename van het percentage geslaagden vooral door de stijging van de cijfers voor de schoolexamens (Inspectie van het Onderwijs, 2006). Voor vakken met alleen een schoolexamen halen maar weinig leerlingen een onvoldoende. Er zijn scholen waar leerlingen het schoolexamen net zo vaak mogen herkansen tot zij (bijna) allen een voldoende hebben gehaald. Onderzoekers signaleren vooral bij sommige particuliere, vrije en zwarte scholen een toegeeflijke cijfercultuur (De Lange & Dronkers, 2006; Rekers-Mombargs & Harms, 2008; Inspectie van het onderwijs, 2007a,b). Op respectievelijk 13, 10 en 28 procent van de reguliere scholen voor vmbo-gt, havo en vwo krijgen kandidaten gemiddeld op het schoolexamen tussen een half en een heel punt meer dan op het centrale examen (Inspectie van het Onderwijs, 2007a,b). Daardoor zou de slaagkans – gegeven gelijke kennis en vaardigheden - op de ene school groter zijn dan op de andere school. Volgens sommigen tast dit gebrek aan landelijke vergelijkbaarheid het civiel effect van het diploma voortgezet onderwijs aan.

Voor de Inspectie was het geconstateerde gebrek aan recente informatie over de vakinhoudelijke en toetstechnische kwaliteit van de schoolexamens aanleiding om een nieuw onderzoek naar de schoolexamens te initiëren. De centrale vraag van dit onderzoek is in hoeverre de schoolexamens in het voortgezet onderwijs voldoen aan redelijke eisen ten aanzien van de inhoud, de wijze van beoordelen en normeren en de organisatie. In dit rapport wordt verslag gedaan van een onderzoek dat het Cito in opdracht van de Inspectie heeft uitgevoerd. Daarbij is ook nagegaan in hoeverre de moeilijkheid van de toetsen en opdrachten uit het schoolexamen in overeenstemming is met hetgeen men van kandidaten mag eisen.

2 Methode van onderzoek

In opdracht van de Inspectie van het Onderwijs heeft het Cito een onderzoek uitgevoerd naar de kwaliteit van de schoolexamens in het voortgezet onderwijs. In dit hoofdstuk wordt uiteengezet hoe het onderzoek is opgezet en uitgevoerd.

2.1 Vakken, steekproef en response

De analyse van de kwaliteit van het schoolexamen is uitgevoerd voor de vakken Engels en biologie op havo-niveau en Nederlands en wiskunde op vmbo-niveau. Het onderzoek vormt het vervolg op een in het schooljaar 2001-2002 uitgevoerde pilot waarbij de vakken Engels en biologie betrokken waren en waarvoor allerlei materiaal ontwikkeld was zoals toetsingskaders en de vragenlijst *Evaluatie Schoolexamen schooljaar 2007-2008*. Zowel voor vmbo als havo is een taal en een exact vak gekozen. Bovendien is gekozen voor kernvakken die veel leerlingen volgen.

Voor het onderzoek is een representatieve steekproef getrokken van 74 havo-scholen en 84 vmbo-scholen (vmbo-breed). De havo-scholen is gevraagd al het schriftelijke leerlingenwerk voor de vierde en vijfde klas op te sturen. De vmbo-scholen is gevraagd van iedere op de school aanwezige leerweg het materiaal van een derde klas en een vierde klas op te sturen. Hierbij werden de gemengde en theoretische leerweg als één leerweg beschouwd.

Het door de scholen ingezonden materiaal bleek niet in alle gevallen volledig. Zo bleek ook na rappelling van veel scholen het PTA te ontbreken. Bij het vak Nederlands is het PTA niet in de beoordeling betrokken, maar bij de overige vakken is dit wel gebeurd.

De beschikbare tijd en menskracht lieten het niet toe om de kwaliteit van het SE voor elke leerweg van een vmbo-school apart te beoordelen (d.w.z. basisberoepsgerichte leerweg, kaderberoepsgerichte leerweg en de gemengde/theoretische leerweg). Besloten is om voor de vakken Nederlands en wiskunde per school in beginsel slechts één leerweg in de beoordeling te betrekken. De verdeling van de steekproef naar leerweg was als volgt: 26 scholen met een basisberoepsgerichte leerweg, 26 scholen met een kaderberoepsgerichte leerweg en 45 scholen met een gemengde/theoretische leerweg.

2.2 Toetsingskader

Ten behoeve van de analyse van de kwaliteit van de schoolexamens is voor elk vak een conceptueel kader opgesteld. Het zogeheten toetsingskader bevatte onder meer een beschrijving van de standaarden en indicatoren aan de hand waarvan de beoordelaars de algemene kwaliteit van het SE en de specifieke kwaliteit per vakonderdeel beoordeelden. Voor elk vak zijn dezelfde standaarden en indicatoren gehanteerd (waarvan de uitwerking uiteraard per vak verschilde). De zes standaarden en de negentien indicatoren zijn samen met een korte omschrijving weergegeven in onderstaand overzicht. In de omschrijving is geabstraheerd over de vier vakken. Voor de vakinhoudelijke uitwerking van de standaarden en indicatoren wordt verwezen naar Bijlage 1 tot en met 4 waarin de toetsingskaders voor de vier vakken integraal zijn opgenomen.

Standaard	Indicator	Omschrijving
1 Inhoud	1	Verplichte onderdelen SE Het SE voldoet aan de inhoudelijke specificaties uit het eindexamenprogramma
2 Niveau	2	Conform niveauspecificaties Het SE voldoet aan de niveauspecificaties uit het eindexamenprogramma
	3	Afnamemoment gerelateerd aan niveau Het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereiken kunnen hebben
	4	Moeilijkheid toetsen en opdrachten De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (in vergelijking met het door de eindtermen geïmpliceerde niveau en de moeilijkheid van de toetsen en opdrachten uit het centraal examen)
	5	Soepelheid/strengheid beoordelingsmiddelen De beoordelingsmiddelen zijn noch te streng noch te soepel
3 Toetsconstructie	6	Becijfering kandidaten De toegekende cijfers zijn noch te hoog noch te laag
	7	Conform producteisen Het SE is zorgvuldig samengesteld en voldoet aan relevante producteisen
	8	Vergelijkbaarheid opgaven Als met toetsvarianten wordt gewerkt, is de moeilijkheidsgraad ervan vergelijkbaar
4 Beoordeling	9	Verplichte toetsinhoud in overleg De vakcollega's stellen de inhoud van het SE in overleg vast
	10	Beoordelingsaspecten relevant Alle beoordelingsaspecten zijn relevant
	11	Beoordelingsaspecten volledig Alle relevante beoordelingsaspecten worden in de beoordeling betrokken
	12	Gebruik antwoord- c.q. beoordelingsmodel De beoordeling vindt plaats met een antwoord-c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan
	13	Cijfers voor toetsen en opdrachten Voor de toetsen en handelingsdelen worden cijfers respectievelijk de kwalificaties "naar behoren" of "onvoldoende" toegekend
	14	Beoordeling in overleg De beoordeling geschiedt door ten minste twee docenten
5 Normering	15	Scorepunten en/of cesuur vooraf vastgelegd De scorepunten en indien van toepassing ook de cesuur worden vóór de afname van het toetsmateriaal schriftelijk vastgesteld
	16	Scorepunten vroegtijdig meegedeeld De scorepunten worden vóór de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt
	17	Bijstelling scorepunten/cesuur beargumenteerd De scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd
	18	Normering in overleg De vakcollega's bepalen de normering in onderling overleg
6 Weging	19	Gewicht Het gewicht bij de bepaling van het SE cijfer is voor elke toets aangegeven

Hieronder lichten we de indicatoren exemplarisch toe aan de hand van het vak Engels.

1 Verplichte onderdelen SE

Het SE moet voldoen aan de inhoudelijke specificaties uit het eindexamenprogramma. Voor het vak Engels moet het examendossier in havo 5 bijvoorbeeld toetsen bevatten voor luistervaardigheid, gespreksvaardigheid en schrijfvaardigheid. Daarnaast moet het handelingsdeel een aantal onderdelen bevatten, te weten leesvaardigheid (extensief lezen, samenvatten als strategie en informatie verwerven met behulp van ICT), luistervaardigheid (extensief luisteren en maken van aantekeningen), gespreksvaardigheid (reële communicatieve situaties en presentaties), schrijfvaardigheid (correspondentie mede met behulp van ICT) en letterkunde (samenstelling leesdossier en verwerkingsopdrachten). Omdat de wet scholen niet tot een handelingsdeel verplicht, is besloten de inhoud van het handelingsdeel niet in de beoordeling te betrekken.

- 2 Conform niveauspecificaties
Het SE moet voldoen aan de niveauspecificaties uit het eindexamenprogramma. Voor het vak Engels betreft het de specificaties van niveau 3 zoals verwoord in bijlage 19 van het eindexamenprogramma. Bij het vakonderdeel Leesvaardigheid gaat het om eisen aan de te lezen teksten. Die moeten bijvoorbeeld bestemd zijn voor een redelijk ontwikkeld publiek en een breed scala aan onderwerpen beslaan. Bij het vakonderdeel Luistervaardigheid betreft het eisen aan de te beluisteren teksten die bijvoorbeeld een transparante opbouw en een normaal spreektempo moeten hebben. Bij het vakonderdeel gespreksvaardigheid gaat het om een nadere specificatie van de spreekhandelingen en eisen waaraan de spreekprestatie van de kandidaat moet voldoen, zoals zich begrijpelijk en verstaanbaar uitdrukken en strategieën toepassen om tekorten in zijn kennis te compenseren. Bij het vakonderdeel Schrijfvaardigheid gaat het om een nadere specificatie van de schrijfhandelingen, de te schrijven teksten, de eisen waar de prestatie van de kandidaat aan moet voldoen (zoals teksten logisch opbouwen en zich duidelijk uitdrukken) en toegestaan naslagmateriaal en hulpmiddelen.
- 3 Afnamemoment gerelateerd aan niveau
Het toetsmateriaal moet worden afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben. Als een school bijvoorbeeld een vakonderdeel al in de vierde klas afsluit, is het de vraag of alle kandidaten voldoende de gelegenheid hebben gehad om zich het vereiste niveau eigen te maken.
- 4 Moeilijkheid toetsen en opdrachten
De toetsen en opdrachten mogen noch te moeilijk noch te gemakkelijk zijn. Als het toetsmateriaal steevast te moeilijk of te gemakkelijk is en docenten dat niet compenseren door een soepele respectievelijk strenge beoordeling (zie indicator 5 en 6), kunnen kandidaten in vergelijking met leeftijdsgenoten op andere scholen worden benadeeld respectievelijk bevoordeeld.
- 5 Soepelheid/strengheid beoordelingsmiddelen
De beoordelingsmiddelen moeten noch te streng noch te soepel zijn. Onder beoordelingsmiddelen verstaan we hier bijvoorbeeld een antwoordmodel voor het scoren van open vragen of een beoordelingsschaal voor het beoordelen van gespreksvaardigheid. De scorings- en beoordelingsmiddelen laten zien hoe de school scores (punten) toekent aan zwakke, middelmatige en goede prestaties. Verschillen tussen scholen in de strengheid van de scorings- en beoordelingsmiddelen kunnen ertoe leiden dat een kandidaat met een vergelijkbaar prestatieniveau op de ene school een grotere kans heeft om te slagen dan op de andere school.
- 6 Becijfering kandidaten
De toegekende cijfers moeten noch te hoog noch te laag zijn. Samen met de toetsen en opdrachten en de scorings- en beoordelingsmiddelen geven de scores en cijfers die de examinator op de leerlingenwerken heeft genoteerd inzicht in de billijkheid van de beoordeling. Bij het toekennen van cijfers kunnen docenten compenseren voor te moeilijke of te gemakkelijke toetsen of voor te strenge of te soepele beoordelingsmiddelen. Examinatoren kunnen kandidaten uiteindelijk strenger of soepeler beoordelen dan men op grond van de toetsen en opdrachten en de gehanteerde beoordelingsmiddelen zou verwachten. Dit is alleen na te gaan aan de hand van de uiteindelijk toegekende scores en cijfers.
- 7 Conform producteisen
Het SE moet zorgvuldig zijn samengesteld en voldoen aan relevante producteisen. Het gaat hier om elementaire eisen van algemene kwaliteit, validiteit, betrouwbaarheid en efficiëntie. Bij de algemene kwaliteit gaat het om toetstechnische eisen zoals de aansluiting bij de eindtermen, de eenduidigheid van de vragen en opdrachten en de vakinhoudelijke correctheid. Met betrekking tot de validiteit van de examinering van het onderdeel gespreksvaardigheid Engels moet deze vaardigheid bijvoorbeeld met een mondelinge toets worden geëxamineerd. Ten aanzien van de betrouwbaarheid van het SE moeten de toetsen voldoende lang zijn om een betrouwbare toetsing mogelijk te maken. Een schrijfopdracht van een kwartier kan bijvoorbeeld niet bijzonder betrouwbaar zijn. De producteis van efficiëntie verlangt bijvoorbeeld dat de toetsen niet langer zijn dan voor een betrouwbare meting strikt noodzakelijk is.
- 8 Vergelijkbaarheid opgaven
Als met toetsvarianten wordt gewerkt, bijvoorbeeld bij inhaaltoetsen of herkansingen, moet de moeilijkheidsgraad ervan vergelijkbaar zijn. Scholen werken met toetsvarianten als zij grote aantallen kandidaten moeten examineren en er vanwege de geheimhouding niet steeds dezelfde toets of opdracht kan worden gebruikt. Ook bij de herkansing dienen dezelfde toetsen en opdrachten te worden gebruikt. Omwille van een rechtvaardige examinering is het gewenst dat de varianten van een vergelijkbare moeilijkheidsgraad zijn.

- 9 **Verplichte toetsinhoud in overleg**
De vakcollega's moeten de inhoud van het SE in overleg vaststellen. In de situatie dat aan een kleine school voor het desbetreffende examenvak slechts één vakdocent verbonden is, moet deze de inhoud van het SE vaststellen met ten minste één vakcollega uit de onderbouw, met een vakdocent van een verwant vak of met een vakdocent van een andere school. Dit betekent overigens niet dat de collega zelf actief betrokken moet zijn bij het ontwikkelen van het examen. Hij of zij moet echter wel de gelegenheid hebben gehad om het conceptmateriaal van commentaar te voorzien.
- 10 **Beoordelingsaspecten relevant**
Alle beoordelingsaspecten moeten relevant zijn. Bij de beoordeling van de kandidaten mogen alleen aspecten worden betrokken die relevant zijn voor de getoetste kennis en vaardigheden. In het Eindexamenprogramma Engels havo 5 is voor twee vaardigheden aangegeven welke aspecten examinatoren in de beoordeling moeten betrekken. Bij gespreksvaardigheid moeten zij letten op de inhoud van het taalgebruik, de vlotheid en de uitspraak. Bij schrijfvaardigheid zijn het de inhoud, opbouw en correctheid van het taalgebruik. Voor luistervaardigheid noemt het examenprogramma geen beoordelingsaspecten, maar in de beoordeling van deze vaardigheid mag de stel- en spelvaardigheid van de kandidaat bijvoorbeeld geen rol spelen.
- 11 **Beoordelingsaspecten volledig**
Alle relevante beoordelingsaspecten moeten in de beoordeling worden betrokken. Zo moeten er bij schrijfvaardigheid Engels ook andere aspecten worden beoordeeld dan uitsluitend grammatica (zoals doel- en publiekgerichtheid).
- 12 **Gebruik antwoord- c.q. beoordelingsmodel**
De beoordeling moet plaatsvinden met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan. Als de school bijvoorbeeld de luister- of schrijftoetsen Engels van het Cito gebruikt, kan men ervan uitgaan dat voor deze vakonderdelen aan deze eis voldaan is.
- 13 **Cijfers voor toetsen en opdrachten**
Voor de toetsen Engels Havo 5 moeten cijfers worden gegeven en voor het handelingsdeel mogen alleen de kwalificaties "naar behoren" of "onvoldoende" worden toegekend.
- 14 **Beoordeling in overleg**
De beoordeling geschiedt door minimaal twee docenten. Bij de mondelinge toetsing van gespreksvaardigheid Engels kan men bijvoorbeeld eisen dat de score in onderling overleg tussen twee examinatoren wordt vastgesteld. Is er in de bovenbouw slechts één vakdocent Engels werkzaam, dan moet deze een vakcollega uit de onderbouw inschakelen of bijvoorbeeld een collega Duits of Frans of een docent Engels van een andere school.
- 15 **Scorepunten en/of cesuur vooraf vastgelegd**
De scorepunten en indien van toepassing ook de cesuur dienen vóór de afname van het toetsmateriaal schriftelijk te worden vastgelegd. Bij schrijfvaardigheid Engels is er bijvoorbeeld van uitgegaan dat aan deze indicator is voldaan als in de toets is aangegeven hoeveel punten de kandidaat met de hele opgave of met elke deelopgave kan verdienen.
- 16 **Scorepunten vroegtijdig meegedeeld**
De scorepunten worden vóór of tijdens de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt. Bij schrijfvaardigheid Engels moet de kandidaat bijvoorbeeld vroegtijdig geïnformeerd zijn over de te behalen maximumscore en het gewicht van de beoordelingsaspecten in het cijfer en de cesuur.
- 17 **Bijstelling scorepunten/cesuur beargumenteerd**
De scorepunten en indien van toepassing de cesuur dienen te worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd.
- 18 **Normering in overleg**
De vakcollega's dienen de normering in onderling overleg te bepalen. Is er in de bovenbouw slechts één vakdocent Engels werkzaam, dan moet deze een vakcollega uit de onderbouw inschakelen of een docent van een andere school.
- 19 **Gewicht**
Het gewicht bij de bepaling van het SE cijfer dient voor elke toets te zijn aangegeven. De kandidaat moet niet alleen geïnformeerd zijn over de te behalen maximumscore en de puntentoekenning per vraag maar ook over het gewicht van de toetsen in het eindcijfer voor het SE.

Bij havo Engels en biologie was er sprake van twee verschillende examenprogramma's. Terwijl het SE in havo 5 moest voldoen aan de eisen in het examenprogramma van 1998 was in havo 4 een nieuw programma van kracht. Het nieuwe examenprogramma laat de scholen zeer veel vrijheid bij de invulling van het SE. Veel indicatoren uit het toetsingskader waren daarom niet relevant bij dit nieuwe programma. Scholen konden met betrekking tot dit nieuwe programma alleen maar opgaven voor het vierde leerjaar insturen – als de school ten minste al in de vierde klas met toetsing van deze SE-onderdelen begon. Bij het vak Engels is de beoordeling daarom beperkt gebleven tot de SE-praktijk met betrekking tot het vorige examenprogramma – en dus havo 5. Afgesproken is met de inspectie dat het handelingsdeel voor Engels en biologie niet beoordeeld wordt, omdat het niet verplicht is. Wel hebben de beoordelaars Engels aan de hand van de door docenten ingevulde vragenlijst *Evaluatie Schoolexamen schooljaar 2007-2008* bekeken of er cijfers voor onderdelen van het handelingsdeel meetelden voor het SE-cijfer (zie indicator 13).

Bij biologie is voor de kwaliteit van het SE in havo 4 een algemene indicator aan het toetsingskader toegevoegd. De verdeling van de afnames over havo 4 en 5 verschilde van school tot school. Er waren scholen die alleen maar examenwerk van havo 5 inleverden (en dus kennelijk geen SE in havo 4 organiseerden). Sommige scholen die wel in Havo 4 examineerden, beperkten zich in dat leerjaar tot praktische opdrachten.

2.3 Geraadpleegde bronnen

Ten behoeve van de beoordeling van de kwaliteit van het SE is de scholen verzocht het volgende bronnenmateriaal ter beschikking te stellen:

- Het Programma van Toetsing en Afsluiting (PTA).
- De vragenlijst *Evaluatie Schoolexamen schooljaar 2007-2008* die was ingevuld door een van de docenten die les gaf aan de desbetreffende eindexamenkandidaten;
- De toetsen en opdrachten van het SE, de bijbehorende scorings- en beoordelingsmiddelen en van twee klassen al het leerlingenwerk (per leerjaar per leerweg per vakonderdeel) met daarop vermeld de cijfers.

De vragenlijst *Evaluatie Schoolexamen schooljaar 2007-2008* bood de beoordelaars informatie over onder meer de getoetste onderdelen van het SE, het afnamemoment, de relevantie en volledigheid van de beoordelingsaspecten en het gewicht van de onderdelen van het SE in het eindcijfer.

De door de school opgestuurde toetsen en opdrachten gaven in het bijzonder informatie over de vraag of het SE voldoet aan de niveauspecificaties uit het eindexamenprogramma, de moeilijkheidsgraad van het toetsmateriaal, de toetstechnische kwaliteit ervan (producteisen) en de vergelijkbaarheid van varianten van toetsen en opdrachten.

De meegestuurde scorings- en beoordelingsmiddelen verschaften informatie over de vraag of er was voldaan aan de niveauspecificaties uit het examenprogramma, de soepelheid/strengheid van de scorings- en beoordelingsmiddelen, de relevantie en volledigheid van de beoordeelde aspecten, het gebruik van antwoord- c.q. beoordelingsmodellen, de toekenning van cijfers voor de toetsen en de kwalificaties "naar behoren" of "onvoldoende" voor de handelingsdelen en de vraag of de scorepunten en cesuur voorafgaand aan het SE waren vastgelegd.

De leerlingenwerken van de beide klassen boden de beoordelaars informatie over onder meer de billijkheid van de cijfers van de kandidaten en of er was voldaan aan de niveauspecificaties uit het examenprogramma.

Het PTA verschaftte de beoordelaars onder meer informatie over de verplichte onderdelen van het SE, of het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben, het vooraf vastleggen en vroegtijdig meedelen van de scorepunten en de cesuur en het gewicht van de toets of opdracht bij het bepalen van het cijfer voor het SE.

2.4 Beoordelingsmethode

Beoordelaars

De beoordeling van de kwaliteit van het schoolexamen heeft plaatsgevonden door in totaal twaalf beoordelaars. Allen waren volledig bevoegd in het desbetreffende examenvak en beschikten over ruime onderwijservaring. Daarnaast waren alle beoordelaars actief betrokken (geweest) bij de constructie of vaststelling van de centrale examens en als zodanig goed op de hoogte van de vakinhoudelijke eisen, de toetstechnische eisen en het niveau dat men van examenkandidaten mag verwachten.

Beoordelaarstraining

Ter voorbereiding op de beoordelingstaak ontvingen de beoordelaars onder leiding van een toetsdeskundige van het Cito een training. Na een inleiding over het doel van onderzoek, het toetsingskader met de standaarden en indicatoren, het bronnenmateriaal en de regels voor het beoordelen werd het materiaal van enkele scholen gezamenlijk beoordeeld waarbij discrepanties tussen beoordelaars werden doorgesproken.

Kwaliteitsoordelen

De beoordelaars moesten de kwaliteit van het SE op negentien indicatoren beoordelen. Zij konden daarbij kiezen uit de oordelen *voldoende*, *onvoldoende* of *niet beoordeelbaar*. Bij sommige indicatoren moesten zij simpelweg kiezen tussen Ja (dus voldoende) of Nee (dus onvoldoende). Het oordeel *niet beoordeelbaar* moest men geven als het benodigde bronnenmateriaal wel voorhanden was maar geen uitsluitel bood. Was het voor de beoordeling van een indicator benodigde bronnenmateriaal niet voorhanden, dan hoefde men het desbetreffende veld op het beoordelingsformulier niet in te vullen. Bij de indicatoren 4, 5 en 6 voor het niveau van het SE waren de antwoordmogelijkheden enigszins afwijkend. Bij indicator 4 voor de moeilijkheid van de toetsen en opdrachten kon men kiezen uit *juiste moeilijkheidsgraad (voldoende)*, *te gemakkelijk* en *te moeilijk* (afgezien van de antwoordmogelijkheid niet beoordeelbaar). Bij indicator 5 voor de soepelheid/stringheid van de scorings- en beoordelingsmiddelen waren de keuzemogelijkheden *noch te soepel noch te streng (voldoende)*, *te soepel* en *te streng*. Bij indicator 6 voor de billijkheid van de becijfering waren de alternatieven *noch te hoog noch te laag (voldoende)*, *te hoog* en *te laag*. In uitzonderingsgevallen konden de beoordelaars bij deze drie niveau-indicatoren het alternatief *onvoldoende* kiezen. Hiervoor moest worden gekozen als er evenveel indicaties waren voor een te soepele als voor een te strenge beoordeling. Bij indicator 4 is hiervan bijvoorbeeld sprake als de toetsen en opdrachten voor de ene helft van de vakonderdelen te moeilijk zijn en voor de andere helft te gemakkelijk.

Vakonderdelen

De beoordelaars gaven voor elk van de negentien indicatoren een algemeen oordeel over de kwaliteit van het schoolexamen. Twee overige indicatoren - te weten indicator 1 (Verplichte onderdelen SE) en indicator 19 (Weging) - werden alleen op algemene kwaliteit beoordeeld. Voor deze twee indicatoren was de informatie uit de vragenlijsten te halen, evenals voor indicator 3 (Afnamemoment gerelateerd aan niveau), 9 (Verplichte toetsinhoud in overleg), 13 (Cijfers voor toetsen en opdrachten) en 14 (Beoordeling in overleg). Voor de beoordeling van de overige indicatoren waren de vragenlijsten en/of de concrete opgaven relevant.

Daarnaast spraken zij in bepaalde gevallen ook een oordeel uit over de kwaliteit per vakonderdeel. De vakonderdelen waren uiteraard niet voor al de vier vakken dezelfde.

De drie vakonderdelen bij Engels waren Luistervaardigheid, Schrijfvaardigheid en gespreksvaardigheid.

Bij het vak Nederlands werden zeven vakonderdelen onderscheiden, te weten Luister- en kijkvaardigheid, Schrijfvaardigheid, Fictie, Gedocumenteerd schrijven, Basale taalvaardigheden (spelling, grammatica, woordkennis e.d.), Mondelinge vaardigheden en Leesvaardigheid.

Bij biologie werden naast de negentien algemene oordelen drie aanvullende oordelen gegeven, te weten Toetsing vaardigheden op niveau, Nieuwe examenprogramma in havo 4 en Algemene natuurwetenschappen (ANW). Voor de kwaliteit van het SE voor deze drie "vakonderdelen" werd telkens één totaaloordeel gegeven (dus niet per indicator). Het vakonderdeel ANW behoeft enige toelichting. ANW is als apart vak op het havo verdwenen. Als oplossing heeft men de eindtermen verdeeld over de zogeheten monovakken. Zo heeft ook biologie er een aantal eindtermen over het onderdeel gezondheid bij gekregen die meestal bij het onderdeel menskunde aan bod komen. Het is derhalve moeilijk vast te stellen of er expliciet gewerkt is aan de eindtermen die vroeger bij ANW thuishoorden. Daarnaast is de kwaliteit van de schriftelijke toetsen en de praktische opdrachten beoordeeld (telkens op alle indicatoren met uitzondering van indicator 1 en 19).

Bij wiskunde werden vier vakonderdelen onderscheiden, namelijk a) Algebraïsche verbanden, b) Rekenen, meten en schatten, c) Meetkunde en Informatieverwerking en d) Statistiek.

De beoordelingsmethode

De beoordelaars beoordeelden volgens de methode van progressieve focusering. Dit wil zeggen dat zij eerst een *quick scan* uitvoerden en vervolgens, als daar aanleiding toe was, een *deep scan*. De reden voor deze getrapte procedure is dat de beschikbare tijd en menskracht er niet in voorzagen om voor alle scholen zowel een oordeel over de algemene kwaliteit uit te spreken als afzonderlijke oordelen per vakonderdeel.

Tijdens de *quick scan* werd het materiaal van de school globaal doorgenomen. Was er geen reden om aan de algemene kwaliteit van het SE te twijfelen, dan werd de algemene kwaliteit voor de desbetreffende indicator als voldoende beschouwd (bij gebrek aan contra-indicaties). Leek de kwaliteit twijfelachtig of onvoldoende, dan werd het oordeel over de kwaliteit opgeschort tot na de *deep scan*.

Tijdens de *deep scan* werd al het materiaal van de school nauwkeurig bestudeerd. Daarbij moesten de beoordelaars voor elk vakonderdeel een afzonderlijk kwaliteitsoordeel geven. Naar aanleiding van de *deep scan* kon voor de algemene kwaliteit van het SE alsnog de kwalificatie voldoende worden toegekend, maar het was ook mogelijk dat de algemene kwaliteit op de desbetreffende indicator een onvoldoende kreeg. In overeenstemming met de methode van progressieve focusering zouden de beoordelaars altijd een oordeel over de algemene kwaliteit geven maar de kwaliteit per vakonderdeel alleen beoordeelden in het geval van twijfel.

Opgemerkt zij dat de methode van progressieve focusering alleen toepasbaar is als het materiaal van een school compleet is en er voor alle vakonderdelen beoordeelbaar materiaal beschikbaar is. Ontbrak het materiaal van een of meer vakonderdelen, is de beoordelaars gevraagd de vakonderdelen en leerwegen waarvoor al wel materiaal binnen was alvast te beoordelen (waardoor zij van progressieve focusering moesten afzien), hun oordeel over de algemene kwaliteit op te schorten tot al het materiaal beschikbaar was en op het moment dat het materiaal compleet was een algemeen oordeel te geven op basis van de oordelen over de vakonderdelen.

De teams van beoordelaars hebben regelmatig van de methode van progressieve focusering moeten afwijken (d.w.z. eerst een *quick scan* en vervolgens bij twijfel aan de algemene kwaliteit een *deep scan* per vakonderdeel). Een aannemelijke verklaring is het veelvuldig ontbreken van materiaal voor bepaalde vakonderdelen waardoor de progressieve focusering van het geheel naar de vakonderdelen niet uitvoerbaar was.

Beoordelingsmethode Engels

De beoordeling van Engels is vrijwel onmiddellijk na de eerste zending van het materiaal van start gegaan. Een probleem was dat het materiaal van de scholen verre van compleet was. Zo ontbrak van veel scholen het materiaal voor gespreksvaardigheid. Een eerste reden is dat de meeste scholen dit vakonderdeel pas in de vijfde klas examineren. Een tweede reden is dat scholen spreekvaardigheid mondeling examineren zonder daarbij van scorings- of beoordelingsmiddelen gebruik te maken (en er dus geen examenwerken of ander materiaal opgestuurd kon worden). Vanwege het ontbreken van materiaal was de methode van progressieve focusering vaak niet toepasbaar. Als er materiaal van een school bleek te ontbreken, zijn beoordelaars alvast begonnen met het beoordelen van de kwaliteit per vakonderdeel (en moesten zij vooralsnog afzien van de methode van progressieve focusering). De beoordelaars kregen de opdracht om de kolom "Algemeen" (voor de algemene kwaliteit) op het beoordelingsformulier alleen in te vullen als het materiaal compleet was en de oordelen over de drie vakonderdelen overeenstemden – dus of alle positief of alle drie negatief waren. Alleen in het geval van compleet en volledig beoordeelbaar materiaal gaven de beoordelaars dus zowel een oordeel over de algemene kwaliteit als over de kwaliteit voor de drie vakonderdelen. Ontbrak er voor een of twee vakonderdeel materiaal of bleek de desbetreffende indicator voor een of twee vakonderdelen niet te beoordelen, dan werden alleen de aanwezige en beoordeelbare vakonderdelen beoordeeld (en gaven de beoordelaars dus geen algemeen oordeel). In de statistische analyse zijn ontbrekende algemene oordelen gesubstitueerd op basis van de oordelen per vakonderdeel (zie paragraaf 2.7).

Beoordelingsmethode Nederlands

Bij het vak Nederlands kon pas tegen de zomervakantie met de beoordeling worden begonnen. Het voordeel was dat de derde zending inmiddels binnen was en de beoordelaars het materiaal in nagenoeg complete staat aantreffen. Bij Nederlands is de beoordeling grondiger uitgevoerd dan oorspronkelijk was afgesproken. Bij de *quick scan* is de algemene indruk van de kwaliteit van het SE namelijk gevormd aan de hand van een bestudering van het materiaal van alle zeven afzonderlijke vakonderdelen. Het ging daarbij om een *quick scan* per vakonderdeel per leerweg. Er is dus niet alleen bij de *deep scan* naar de afzonderlijke vakonderdelen gekeken. Tijdens de *quick scan* bekeken de beoordelaars voor ieder vakonderdeel de toetsen en/of opdrachten waarbij er slechts per vakonderdeel gerapporteerd werd als één van de eerste bekeken toetsen van een school tot twijfel aan de kwaliteit leidde.

Beoordelingsmethode biologie

Bij biologie waren er naast de negentien indicatoren drie aanvullende indicatoren, te weten Toetsing vaardigheden op niveau, Nieuwe examenprogramma in havo 4 en Algemene natuurwetenschappen (ANW). Voor de kwaliteit van het SE voor deze drie "vakonderdelen" werd telkens één totaaloordeel gegeven (dus niet per indicator).

Naast de algemene oordelen en de vier aanvullende oordelen hebben de beoordelaars biologie twee extra vakonderdelen beoordeeld, namelijk de kwaliteit van de schriftelijke toetsen en de praktische opdrachten.

Beoordelingsmethode wiskunde

De beoordelaars wiskunde hebben naast de negentien algemene kwaliteitskenmerken ook de kwaliteit van vier vakonderdelen beoordeeld, te weten a) Algebraïsche verbanden, b) Rekenen, meten en schatten, c) Meetkunde en d) Statistiek. De beoordelaars hebben de kwaliteit van de vier vakonderdelen altijd beoordeeld (voor zover het benodigde materiaal aanwezig en de indicator beoordeelbaar was). Dus ook als de *quick scan* geen aanleiding gaf tot twijfel aan de algemene kwaliteit.

2.5 De uitvoering van het onderzoek

De uitvoering van het onderzoek stelde het projectteam voor een aantal organisatorische uitdagingen.

Aanlevering van het materiaal

Al voor het begin van het nieuwe schooljaar heeft de Inspectie de scholen erover geïnformeerd dat zij in de steekproef zaten en het materiaal van het schoolexamen moesten bewaren. De havo-scholen is gevraagd al het schriftelijke leerlingenwerk voor de vierde en vijfde klas op de sturen. De vmbo-scholen moesten voor iedere op de school aanwezige leerweg het materiaal van een derde klas en een vierde klas opsturen (waarbij de gemengde en theoretische leerweg als één leerweg werden beschouwd). Daarnaast vereiste de beoordeling de aanwezigheid van de vragenlijst *Evaluatie Schoolexamen schooljaar 2007-2008* en het Programma van Toetsing en Afsluiting (PTA). Vanwege de periodisering van het schoolexamen en de planning van het onderzoek heeft de Inspectie de scholen gevraagd het leerlingenwerk in drie zendingen aan te leveren: een eerste zending per 1 februari 2008 van alle schoolexamens van de eerste helft van het schooljaar, een tweede zending per 1 juni 2008 van alle schoolexamens van de tweede helft en een nazending per 28 juni 2008 voor het geval de school nog schoolexamens na 1 juni 2008 zou afnemen. Nadat de eerste zending op het Cito arriveerde, bleek het materiaal veel volumineuzer was dan oorspronkelijk gedacht was. Dit gold vooral voor het vmbo, dat meer toetsmomenten en meer toetsen en opdrachten kent dan het havo. Uit een eerste inspectie van het omvangrijke materiaal bleek dat het door de scholen ingezonden materiaal maar zelden compleet was. Dit betrof niet alleen de PTA's en de vragenlijsten. Zeer vaak ontbrak toetsmateriaal van een of meer klassen, leerwegen en/of vakonderdelen. De controle op volledigheid werd verder bemoeilijkt doordat scholen bepaalde vakonderdelen later in het schooljaar nog konden afnemen en alsnog bij de tweede of derde zending zouden kunnen opsturen. Het was dus niet duidelijk in hoeverre het ontbreken van voor de beoordeling benodigd materiaal was toe te schrijven aan onzorgvuldigheid van de scholen dan wel aan een afwijkende programmering van de examinering van de vakonderdelen. Tegelijkertijd hadden drie van de vier toetsdeskundigen met de beoordelaars vaste afspraken gemaakt over de periode waarin het materiaal beoordeeld zou worden (beginnen in maart en afronden vóór de zomervakantie). De controle op volledigheid was vanwege de onverwacht grote hoeveelheid materiaal zeer arbeidsintensief en kostbaar. In overleg met de Inspectie is besloten het overzicht van het ingezonden materiaal niet aan de scholen toe te zenden en de scholen niet te rappelleren (overigens werden er wel pogingen gedaan om de PTA's compleet te krijgen).

Beoordeling van niet compleet materiaal

Vanwege de tijdsdruk zijn de meeste beoordelaars op een tijdstip tussen de eerste en tweede zending met beoordelen begonnen. Niet zelden moesten de beoordelaars constateren dat er materiaal ontbrak. Regelmatig ontbrak van een school het materiaal van een bepaald leerjaar of van een bepaalde leerweg. In andere gevallen hadden de scholen voor een of meer vakonderdelen geen materiaal meegestuurd (zonder dat duidelijk was of er hierbij sprake was van nalatigheid of dat het desbetreffende vakonderdeel op een later tijdstip alsnog getoetst zou worden en de school het materiaal bij een latere zending nog zou opsturen). Bij het beoordelen zagen de beoordelaars zich telkenmale geconfronteerd met ontbrekend materiaal waarvan zij de kwaliteit nog niet konden beoordelen (zoals PTA, vragenlijst, antwoordmodel, correctievoorschrift, leerlingenwerk). Daarnaast was het materiaal van de verschillende zendingen van een school vaak over verschillende dozen verdeeld, wat veel zoekwerk met zich meebracht. Tegelijkertijd stapelden de dozen zich bij de beoordelaars thuis op. Sommige beoordelaars deden hierover hun beklag. Direct na de derde fase bleek uit berichten van beoordelaars dat het materiaal nog steeds niet compleet was. Er is toen besloten tot een vierde fase van gegevensverzameling waarvan de overigens geringe opbrengst vrij snel na de zomervakantie op het Cito arriveerde. Sommige beoordelaars hadden hun formulieren toen al ingeleverd. Dit gold overigens niet voor het vak Nederlands waar de beoordelaars pas vlak voor de zomervakantie op basis van nagenoeg compleet en netjes op school geordend materiaal met de beoordeling begonnen.

Ontwikkeling toetsingskaders

Het onderzoek vormt het vervolg op een in het schooljaar 2001-2002 uitgevoerde pilot waarin onder meer toetsingskaders ontwikkeld zijn voor de vakken Engels en biologie. Bij de aanvang van het project zijn de toetsingskaders voor de vakken Nederlands en biologie opgesteld waarbij de toetsingskaders voor Engels en biologie als voorbeeld dienden. Op het moment dat de beoordelaars met hun werkzaamheden zouden beginnen bleken de toetsingskaders voor de vier vakken te verschillen in onder meer het aantal, de aard en de formulering van de indicatoren. Het overleg over het op elkaar afstemmen van de vier toetsingskaders en het ontwikkelen van een voor alle vier vakken bruikbaar scanbaar beoordelingsformulier nam meer tijd in beslag dan oorspronkelijk was voorzien, zodat de beoordelaars pas laat over de formulieren konden beschikken en met het beoordelen konden beginnen. Door deze tegenvaller resteerde weinig tijd voor achtereenvolgens de training van de beoordelaars, het onderzoek naar de beoordelaaroverstemming, de aanscherping van de beoordelingsregels, de eigenlijke beoordeling van de kwaliteit van de schoolexamens en de controle op de kwaliteit van de beoordeling. Zo kon de training van de beoordelaars op zijn vroegst pas in mei plaatsvinden in plaats van zoals gepland in april.

Uitvoering onderzoek beoordelaarsovereenstemming

In overeenstemming met de planning zouden de resultaten van het onderzoek naar de overeenstemming tussen de beoordelaars vrij snel na de start van de beoordeling ter beschikking gesteld worden (om zo nodig nog tijdig te kunnen bijsturen). Vanwege de hoeveelheid werk en de kosten werd besloten het omvangrijke materiaal van de vijf scholen niet in drievoud te vermenigvuldigen (zodat iedere beoordelaar direct met zijn "eigen" kopie kon beginnen). In plaats van het te vermenigvuldigen werd het materiaal van de vijf scholen over de drie beoordelaars verdeeld en werd ieder van hen gevraagd het hem of haar toegewezen materiaal na beoordeling aan een van de andere beoordelaars via de post te doen toekomen. In de praktijk bleek het rouleren van het materiaal zoveel tijd in beslag te nemen dat de gegevens pas rond de zomervakantie ingevoerd en geanalyseerd konden worden en dus niet meer bruikbaar waren voor verbetering van het beoordelingsproces.

2.6 De vragenlijst Inventarisatie Examenpraktijk

Tijdens de eerste statistische analyses bleek dat de beoordelaars Nederlands de kwaliteit van het SE voor hun vak duidelijk vaker als voldoende hadden aangemerkt dan de beoordelaars Engels (de gegevens voor biologie en wiskunde waren toen nog niet beschikbaar). Het grote verschil tussen beide vakken leek te groot om uitsluitend aan het vak te kunnen toeschrijven en gaf aanleiding tot twijfel aan de validiteit van de beoordeling. Om meer te weten te komen over waar de beoordelaars bij het beoordelen vooral op gelet hadden en hoe het beoordelingsproces in zijn werk was gegaan is toen besloten een vragenlijst te ontwikkelen. In de vragenlijst *Inventarisatie Examenpraktijk* is de beoordelaars en toetsdeskundigen gevraagd om per indicator een aantal voorbeelden te geven van goede en minder goede examenpraktijken. De informatie over de examenpraktijk werd nodig geacht om de cijfermatige gegevens te kunnen interpreteren en aanbevelingen te kunnen doen voor de verbetering van het SE.

In de *Instructie voor het beantwoorden van de vragen* is de toetsdeskundigen en beoordelaars gevraagd de meest frequente afwijkingen te beschrijven. In de vragenlijst is hiervoor onder het kopje "Welke afwijkingen heb je geconstateerd?" ruimte gereserveerd. Daarnaast is hun per indicator gevraagd enkele voorbeelden van voldoende of goede examenpraktijken te geven, dit wil zeggen voorbeelden of kenmerken van een voldoende, goed of uitstekend verzorgd SE. De beoordelaars noteerden deze voorbeelden of kenmerken onder het kopje "Voorbeelden van voldoende of goede examenpraktijken". De beoordelaars is nadrukkelijk op gewezen dat het geven van positieve voorbeelden ook van belang is bij de indicatoren waarbij de kwaliteit van het SE op alle of bijna alle scholen als voldoende werd beoordeeld (zoals bij het vak Nederlands het geval was). Zij moesten in dat geval afwijkingen beschrijven waar zij naar gezocht hadden, maar die zij (vrijwel) niet constateerden omdat deze zich (bijna) niet voordeden. De respondenten is verzocht hun antwoorden zodanig te formuleren dat de beschrijving voor derden volledig begrijpelijk zou zijn. Er was geen onder- of bovengrens voor de lengte van de antwoorden. Als bepaalde belangrijke afwijkingen of voorbeelden van voldoende of goede examenpraktijken maar heel weinig voorkwamen, is de beoordelaars gevraagd deze voorbeelden toch te beschrijven. Als een indicator niet of moeilijk te beoordelen was, konden de beoordelaars volstaan met de opmerking dat zij deze indicator niet of onvoldoende hadden kunnen beoordelen met vermelding van de reden of oorzaak. De beoordelaars is erop gewezen dat het niet de bedoeling was dat zij het materiaal nogmaals zouden bestuderen. Hun is verzocht de vragen "geheel vanuit het geheugen" te beantwoorden. Ook is de beoordelaars gevraagd eventuele verbeteringssuggesties op het formulier te vermelden (bij de indicator in kwestie).

De instructie bevatte de volgende voorbeelden waaruit de toetsdeskundigen en beoordelaars konden opmaken wat van hen verwacht werd:

- Overeenkomstig indicator 1 moet het SE voldoen aan de inhoudelijke specificaties uit het eindexamenprogramma. Noteer bij vraag 1 s.v.p. de vijf meest frequente afwijkingen van de inhoudelijke specificaties die je geconstateerd hebt, zo mogelijk op volgorde van afnemende frequentie.
- Overeenkomstig indicator 2 moet het SE voldoen aan de niveauspecificaties uit het eindexamenprogramma. Noteer bij vraag 2 s.v.p. de vijf meest frequente afwijkingen van de niveauspecificaties die je geconstateerd hebt, zo mogelijk op volgorde van afnemende frequentie.
- Overeenkomstig indicator 4 mogen de toetsen en opdrachten niet te moeilijk en niet te gemakkelijk zijn (gegeven het door de eindtermen geïmpliceerde niveau en in vergelijking met de moeilijkheid van het centraal examen). Geef bij vraag 4 vijf voorbeelden van te moeilijke of te gemakkelijke toetsen en opdrachten, zo mogelijk op volgorde van afnemende frequentie.
- Overeenkomstig indicator 7 moet het SE zorgvuldig zijn samengesteld en moeten de toetsen en opdrachten voldoen aan relevante producteisen. Geef bij vraag 7 vijf voorbeelden van een onzorgvuldige samenstelling en geef voorbeelden van toetsen op opdrachten die NIET aan de toetstechnische kwaliteitseisen voldoen (zo mogelijk op volgorde van afnemende frequentie). Bij Engels is bijvoorbeeld een schrijftoets met een afnameduur van een kwartier als onvoldoende beoordeeld omdat schrijfvaardigheid volgens de beoordelaars in een zo korte afnameduur niet voldoende betrouwbaar getoetst kan worden.

Bij drie van de vier vakken is de vragenlijst rondgestuurd *nadat* de beoordeling voltooid was en het materiaal al was teruggezonden (bij wiskunde hadden we te kampen met tegenslag en was de beoordeling nog in volle gang toen de vragenlijst werd rondgestuurd). Het was voor sommige beoordelaars vrij lastig om voorbeelden te geven als zij het materiaal er niet meer bij hadden, zelfs al hadden zij die beoordeling heel recent afgerond. Ten slotte wordt opgemerkt dat lang niet alle indicatoren zich lenen voor het geven van voorbeelden van goede en minder goede examenpraktijken. Het geven van voorbeelden is bijvoorbeeld goed mogelijk als alle docenten alle verplichte onderdelen toetsen en als alle sectieleden bij de verplichte toetsen samenwerken in de constructie (omdat ze allemaal een toets van het Cito afnemen of een batterij aan schoolboekentoetsen).

De vragenlijst is integraal opgenomen in Bijlage 5. Alle vijftien potentiële respondenten hebben de vragenlijst ingevuld en geretourneerd.

2.7 Statistische analyse

Statistische analyse

De statistische analyse is uitgevoerd met het programma SPSS.

Behandeling van ontbrekende waarnemingen

Bij het vak Engels ontbraken relatief veel waarnemingen (zie paragraaf 2.9). Waar het algemene oordeel ontbrak en er voor een, twee of drie vakonderdelen oordelen beschikbaar waren, is het ontbrekende algemene oordeel in de statistische analyse gesubstitueerd (d.w.z. geïmputeerd) op basis van de oordelen per vakonderdeel. Hierbij zijn de volgende regels gehanteerd. In het geval er slechts voor één vakonderdeel een oordeel beschikbaar was, werd het algemene oordeel aan dat ene oordeel gelijk gesteld. Waren er voor twee onderdelen oordelen beschikbaar, dan werd de algemene kwaliteit als voldoende beoordeeld als beide oordelen voldoende waren en als onvoldoende als ten minste één van beide onvoldoende was. Bij drie beschikbare oordelen per vakonderdeel werd de algemene kwaliteit als voldoende beschouwd als er ten minste twee vakonderdelen als voldoende waren beoordeeld.

Bij de indicatoren 4, 5 en 6 waarbij de beoordelaars twee soorten onvoldoendes konden geven – namelijk te gemakkelijk of te moeilijk (indicator 4), te soepel of te streng (indicator 5) en te hoog of te laag (indicator 6) – werden ontbrekende oordelen voor de algemene kwaliteit op vergelijkbare wijze gesubstitueerd op basis van de onderdeel-specifieke oordelen (voor zover deze beschikbaar waren). Waar de beoordeling van de vakonderdelen evenveel evidentie opleverde voor te gemakkelijke als voor te moeilijke toetsen en opdrachten (of te soepele of te strenge scorings- en beoordelingmiddelen of te hoge of te lage cijfers) werd voor het ontbrekende oordeel over de algemene kwaliteit het oordeel onvoldoende gesubstitueerd. Deze situatie deed zich overigens alleen voor in het geval er voor twee van de drie vakonderdelen oordelen beschikbaar waren en de toetsen en opdrachten bij het ene vakonderdeel te moeilijk waren en bij het andere vakonderdeel te gemakkelijk (indicator 4), de scorings- en beoordelingsmiddelen bij het ene onderdeel te soepel en bij het andere onderdeel te streng waren

(indicator 5) of de cijfers voor de leerlingenwerken bij het ene onderdeel te hoog waren en bij het andere onderdeel te laag waren (indicator 6).

Behandeling ontbrekende waarnemingen in analyse verschillen tussen scholen

Een belangrijk doel van het onderzoek is meer te weten te komen over de mate waarin scholen verschillen in de kwaliteit van het SE. Daartoe is voor elk vak een somscore over de indicatoren van de algemene kwaliteit van het SE berekend. Vanwege de vaak hoge percentages ontbrekende waarnemingen werd berekening van een somscore over alle 19 indicatoren niet zinvol geacht. De somscore is alleen berekend over de indicatoren met een percentage ontbrekende waarnemingen van hooguit 20%. Na verwijdering van indicatoren met meer dan twintig procent ontbrekende waarnemingen resteerden bij het vak Engels zestien bruikbare indicatoren (1 t/m 15 en 19), bij biologie tien indicatoren (namelijk: 3, 4, 5, 6, 8, 10, 11, 12, 13 en 15) en bij wiskunde elf indicatoren (namelijk: 3, 4, 5, 6, 7, 10, 11, 12, 13, 15 en 16). Bij Nederlands was rapportage van verschillen tussen scholen niet zinvol omdat alle scholen aan alle indicatoren voldeden (zie paragraaf 3.2.1).

Na selectie van de voor analyse bruikbare indicatoren is bij de constructie van de somscore voor ontbrekende waarnemingen per school een waarde gesubstitueerd op basis van de wel beschikbare oordelen. Een voorbeeld moge dit verduidelijken. Stel een school uit de responsgroep wiskunde heeft voor de elf indicatoren negen voldoende (score 1), één onvoldoende (score 0) en één ontbrekend oordeel (missing). In de analyse is dan voor het ontbrekend oordeel de score .9 gesubstitueerd, waardoor de somscore uitkomt op 9.9 (waarbij deze waarde in de frequentieverdeling van de somscore over de scholen is afgerond op een geheel getal).

Beoordelaarsovereenstemming

Ter bepaling van de overeenstemming tussen de beoordelaars is voor elk van de vier vakken een aselechte steekproef van vijf scholen door alle beoordelaars beoordeeld. De beoordelaars beoordeelden geheel onafhankelijk van elkaar. De kwaliteit van de beoordeling is bepaald aan de hand van het percentage volledig identieke oordelen en *Cohen's multi-rater Kappa*, een maat die corrigeert voor overeenstemming op basis van toeval en die niet eist dat de beoordelaars precies dezelfde categorieën gebruiken. *Kappa* is 1 als de beoordelaars perfect overeenstemmen en 0 als de overeenstemming niet groter is dan men op basis van toeval mag verwachten. *Kappa* is berekend met het SPSS-macro "mKappasc.sps" (Nichols, 1997). Landis en Koch (1997) geven de volgende vuistregel voor het interpreteren van de hoogte van *Kappa*: .00 tot .20 is *slight*, .21 tot .40 is *fair*, .41 tot .60 is *moderate*, .61 tot .80 is *substantial* en .81 tot 1.00 is *almost perfect*.

2.8 Beoordelaarsovereenstemming

Beoordelaarsovereenstemming Engels

Ter bepaling van de overeenstemming tussen de beoordelaars is een aselechte steekproef van vijf scholen door alle drie beoordelaars beoordeeld. De beoordelaars beoordeelden geheel onafhankelijk van elkaar. Aan het onderzoek nam behalve de drie beoordelaars ook de toetsdeskundige Engels die de training verzorgde deel. Van de in totaal 4 (beoordelaars) * 5 (scholen) * 19 (indicatoren) = 380 oordelen over de algemene kwaliteit van het SE bleken er in totaal 276 (73%) te ontbreken (waarvan de beoordelaars in 228 gevallen geen algemeen oordeel hadden gegeven en in 48 gevallen de algemene kwaliteit expliciet als "niet beoordeelbaar" hadden beoordeeld). Waar er onderdeel-specifieke oordelen beschikbaar waren, zijn voor ontbrekende algemene oordelen waarden gesubstitueerd op basis van de oordelen voor de onderdeel-specifieke indicatoren (zie paragraaf 2.7). Hiermee daalde het aantal ontbrekende algemene oordelen van 276 naar 45. In het onderzoek naar de beoordelaarsovereenstemming zijn deze 45 ontbrekende algemene oordelen als legitieme scores beschouwd (die bijdragen dan wel afbreuk doen aan de overeenstemming). In Tabel 1 is aan de hand van indicator 14 weergegeven hoe deze substitutie heeft plaatsgevonden. Te zien is onder meer dat geheel verschillende oordelen over de kwaliteit per vakonderdeel kunnen leiden tot eenzelfde oordeel over de algemene kwaliteit van het SE.

Het percentage volledig identieke oordelen bedraagt 73% (waarbij de codes 2 en 3 van de indicatoren 4, 5 en 6 zijn gehercodeerd naar 0 en de codes voor ontbrekend materiaal of niet beoordeelbaar tot één code zijn samengevoegd).

Cohen's Kappa voor de overeenstemming tussen de vier beoordelaars bedraagt .36, hetgeen volgens de vuistregel van Landis en Koch (1997) als *fair* beoordeeld moet worden.

In Tabel 2 zijn de gegevens uitgesplitst naar beoordelaar-paar. Te zien is onder meer dat beoordelaar 1 en 3 het sterker met elkaar eens waren dan beoordelaar 2 en 4.

De verdeling van de codes per beoordelaar is weergegeven in Tabel 3 (vóór hercodering van 2 en 3 naar 0 en samenvoeging van de codes voor ontbrekend of niet beoordeelbaar tot één code). Daaruit valt op te maken dat de vier beoordelaars in verschillende mate van de beoordelingscategorieën gebruik maakten. Beoordelaar 1 en 3 gebruikten ze alle zes en beoordelaars 2 en 4 slechts vier. Beoordelaar 2 lijkt de vijf scholen duidelijk soepeler te hebben beoordeeld dan beoordelaar 1 en 3. Het percentage als voldoende beoordeelde indicatoren voor de algemene kwaliteit van het SE voor beoordelaars 1, 2, 3 en 4 bedroeg respectievelijk 68%, 87%, 68% en 67%.

Tabel 1 Voorbeeld van het substitueren van ontbrekende algemene oordelen over de kwaliteit van het SE Engels vanuit de onderdeel-specifieke indicatoren (indicator 14)

School	Beoordelaar	Scores luisteren	Scores schrijven	Scores spreken	Scores algemene kwaliteit vóór substitutie	Scores algemene kwaliteit na substitutie
1	1	99	0	0	99	0
2	1	1	1	0	99	1
3	1	1	0	0	99	0
4	1	1	1	1	99	1
5	1	1	0	0	99	0
1	2	1	1	1	1	1
2	2	99	99	99	1	1
3	2	1	1	1	1	1
4	2	1	0	0	99	0
5	2	1	1	0	99	1
1	3	1	0	0	99	0
2	3	1	1	1	99	1
3	3	1	1	9	99	1
4	3	1	1	9	99	1
5	3	1	0	9	99	0
1	4	1	0	0	99	0
2	4	1	1	0	99	1
3	4	99	0	0	99	0
4	4	99	99	99	1	1
5	4	99	99	99	0	0

Noot. De betekenis van de scores is: 0 = kwaliteit onvoldoende; 1 = kwaliteit voldoende; 9 = indicator niet beoordeelbaar; 99 = indicator is niet beoordeeld.

Tabel 2 Overeenstemming tussen de vier beoordelaars per beoordelaarspaar Engels

Beoordelaarspaar	Percentage identieke oordelen	Cohen's Kappa
1 – 2	71	.20
1 – 3	75	.47
1 – 4	69	.37
2 – 3	75	.31
2 – 4	69	.19
3 – 4	75	.48

Tabel 3 Gebruik van de beoordelingscategorieën per beoordelaar Engels

Oordeel	Beoordelaar 1	Beoordelaar 2	Beoordelaar 3	Beoordelaar 4
0 Onvoldoende	16	7	18	11
1 Voldoende	65	83	65	64
2 Te gemakkelijk of te soepel	2		1	
3 Te moeilijk of te streng	2		1	
9 Niet beoordeelbaar	10	5	10	20
Totaal	95	95	95	95

Beoordelaarsovereenstemming Nederlands

Het percentage volledig identieke oordelen in het onderzoek naar de overeenstemming tussen de drie beoordelaars Nederlands bedraagt 88% (waarbij de codes voor ontbrekend en materiaal en onbeoordeelbaar tot één code zijn toegevoegd).

Cohen's multi-rater Kappa voor de overeenstemming tussen de drie beoordelaars bedraagt .13, hetgeen volgens de vuistregel van Landis en Koch (1997) als *slight* beoordeeld moet worden. Een mogelijke verklaring voor de afwijkende oordelen is dat de indicatoren 15 t/m 18 niet of nauwelijks te beoordelen bleken. Berekenen we *Cohen's multi rater Kappa* zonder indicator 15 t/m 18, dan blijkt de waarde te zijn gestegen tot .24 hetgeen als *fair* beoordeeld kan worden.

In Tabel 4 zijn de gegevens uitgesplitst naar beoordelaarpaar. Te zien is dat beoordelaar 2 en 3 het sterker met elkaar eens zijn dan beoordelaar 1 en 2 en beoordelaar 1 en 3. De overeenstemming tussen beoordelaar 2 en 3 kan als *moderate* worden gewaardeerd. De waarde van *Kappa* voor de overeenstemming van beoordelaar 1 met beoordelaar 2 en 3 is respectievelijk -.09 en -.06 en ligt in de buurt van wat men op basis van toeval zou mogen verwachten.

Inspectie van Tabel 5 brengt aan het licht dat geen van de beoordelaars de code 0 heeft toegekend en dat alleen beoordelaar 1 de codes "Te moeilijk of te streng" en "Materiaal ontbreekt" gebruikte (zij het in beperkte mate). Navraag bij de toetsdeskundige Nederlands leerde dat de beoordelaars bij het coderen van niet beoordeelbare indicatoren waarschijnlijk onvoldoende consequent te werk zijn gegaan en feitelijk eerder en consequenter een 9 (Niet beoordeelbaar) hadden moeten invullen als er geen keiharde aanwijzingen of bewijzen waren.

Tabel 4 Overeenstemming tussen de drie beoordelaars per beoordelaarpaar Nederlands

Beoordelaarpaar	Percentage identieke oordelen	Cohen's Kappa
1 – 2	83	-.09
1 – 3	88	-.06
2 – 3	93	.49

Tabel 5 Gebruik van de beoordelingscategorieën per beoordelaar Nederlands

Oordeel	Beoordelaar 1	Beoordelaar 2	Beoordelaar 3
1 Voldoende	89	85	90
3 Te moeilijk of te streng	1		
8 Materiaal ontbreekt	2		
9 Niet beoordeelbaar	3	10	5
Totaal	95	95	95

Beoordelaarsovereenstemming biologie

Het percentage volledig identieke oordelen over de kwaliteit van het SE op de vijf scholen die zijn onderzocht in het onderzoek naar de overeenstemming tussen de drie beoordelaars biologie bedraagt 69%

(waarbij de codes 2 en 3 van de indicatoren 4, 5 en 6 zijn gehercodeerd naar 0 en de codes voor ontbrekend of niet beoordeelbaar tot één code zijn samengevoegd).

Cohen's multi-rater Kappa voor de overeenstemming tussen de drie beoordelaars bedraagt .31, hetgeen volgens de vuistregel van Landis en Koch (1997) als *fair* beoordeeld moet worden.

In Tabel 6 zijn de gegevens uitgesplitst naar beoordelaarpaar waarbij opvalt dat *Kappa* voor beoordelaarpaar 2-3 hoger is dan voor de paren 1-2 en 1-3.

Inspectie van de toegekende scores in Tabel 7 laat zien dat beoordelaar 1 geen enkele keer het oordeel "Onvoldoende" uitsprak en dat beoordelaar 3 de kwaliteit van het SE van de school 38 keer als "Niet beoordeelbaar" beschouwde.

Tabel 6 Overeenstemming tussen de drie beoordelaars per beoordelaarpaar biologie

Beoordelaarpaar	Percentage identieke oordelen	Cohen's Kappa
1 – 2	71	.28
1 – 3	65	.20
2 – 3	72	.42

Tabel 7 Gebruik van de beoordelingscategorieën per beoordelaar biologie

Oordeel	Beoordelaar 1	Beoordelaar 2	Beoordelaar 3
0 Onvoldoende		5	1
1 Voldoende	75	65	56
2 Te gemakkelijk of te soepel		2	
3 Te moeilijk of te streng	1		
8 Materiaal ontbreekt	1	2	
9 Niet beoordeelbaar	18	21	38
Totaal	95	95	95

Beoordelaarsovereenstemming wiskunde

Het percentage volledig identieke oordelen in het onderzoek naar de overeenstemming tussen de drie beoordelaars wiskunde bedraagt 71% (waarbij de code 2 bij de indicatoren 4, 5 en 6 is gehercodeerd naar 0 en de codes voor ontbrekend en onbeoordeelbaar tot één code zijn samengevoegd).

Cohen's multi-rater Kappa voor de overeenstemming tussen de drie beoordelaars bedraagt .43, hetgeen volgens de vuistregel van Landis en Koch (1997) als *moderate* beoordeeld kan worden.

Tabel 8 toont de overeenstemming per beoordelaarpaar waarbij opvalt dat beoordelaar 1 en 2 sterker overeenstemmen zijn de paren 1-3 en 2-3.

De beoordelaars verschillen ook in het aantal gebruikte beoordelingscategorieën (zie Tabel 9): beoordelaars 1 en 3 gebruiken vier van de zes categorieën en beoordelaar 2 slechts twee. Tegelijkertijd valt op dat beoordelaars 1 en 2 de kwaliteit van het SE van de scholen vaker als "Niet beoordeelbaar" kwalificeerden dan beoordelaar 3. Beoordelaar 3 heeft relatief streng beoordeeld.

Tabel 8 Overeenstemming tussen de drie beoordelaars per beoordelaarpaar wiskunde

Beoordelaarpaar	Percentage identieke oordelen	Cohen's Kappa
1 – 2	74	.41
1 – 3	68	.42
2 – 3	71	.39

Tabel 9 Gebruik van de beoordelingscategorieën per beoordelaar wiskunde

Oordeel	Beoordelaar 1	Beoordelaar 2	Beoordelaar 3
0 Onvoldoende	7		12
1 Voldoende	52	64	63
2 Te gemakkelijk of te soepel			3
8 Materiaal ontbreekt	1		
9 Niet beoordeelbaar	35	31	17

2.9 Ontbrekende waarnemingen

Het door de scholen ingezonden materiaal bleek niet in alle gevallen volledig. Niet altijd waren alle toetsen en opdrachten, scorings- en beoordelingsmiddelen en leerlingenwerken voor alle klassen, leerwegen en vakonderdelen voorhanden (zie paragraaf 2.7). Sommige respondenten hadden de vragenlijst *Evaluatie Schoolexamen schooljaar 2007-2008* onvolledig ingevuld. Van relatief veel scholen ontbrak het PTA. Bij het vak Nederlands is het PTA in het geheel niet in de beoordeling betrokken en bij de overige vakken is dat alleen gebeurd voor zover dit document voorhanden was. In deze paragraaf wordt verslag gedaan van de ontbrekende waarnemingen. Hoe hiermee in de statistische analyse is omgegaan is uiteengezet in paragraaf 2.7.

Tabel 10 Enkele beschrijvende statistische gegevens voor de 19 indicatoren voor de algemene kwaliteit van het SE Engels

	Aantal scholen	Laagste score	Hoogste score	Gemiddelde	Standaard deviatie	% missings
1 Verplichte onderdelen SE	67	0	1	.66	.48	6
2 Conform niveauspecificaties	71	0	1	.68	.47	0
3 Afnamemoment gerelateerd aan Niveau	68	0	1	.96	.21	4
4 Moeilijkheid toetsen en opdrachten	71	1	3*	1.18	.54	0
5 Soepelheid strengheid beoordelingsmiddelen	70	1	3*	1.17	.45	1
6 Becijfering kandidaten	69	1	3*	1.16	.44	3
7 Conform producteisen	70	0	1	.70	.46	1
8 Vergelijkbaarheid opgaven	64	0	1	.98	.13	10
9 Verplichte toetsinhoud in overleg	68	0	1	.93	.26	4
10 Beoordelingsaspecten relevant	71	0	1	.70	.46	0
11 Beoordelingsaspecten volledig	71	0	1	.66	.48	0
12 Gebruik antwoord- c.q. beoordelingsmodel	71	0	1	.99	.12	0
13 Cijfers voor toetsen en opdrachten	69	0	1	.84	.37	3
14 Beoordeling in overleg	68	0	1	.56	.50	4
15 Scorepunten en/of cesuur vooraf Vastgelegd	62	0	1	.97	.18	13
16 Scorepunten vroegtijdig meegedeeld	50	0	1	.98	.14	30
17 Bijstelling scorepunten cesuur beargumenteerd	9	1	1	1.00	.00	87
18 Normering in overleg	9	0	1	.89	.33	87
19 Gewicht	67	0	1	.67	.47	6

* De "niveau-indicatoren" zijn op een andere schaal beoordeeld, bijvoorbeeld 1 = juiste moeilijkheidsgraad (voldoende); 2 = te gemakkelijk; 3 = te moeilijk (zie paragraaf 2.4).

Ontbrekende waarnemingen Engels

Van 71 (96%) van de in totaal 74 getrokken havo-scholen is de kwaliteit van het SE Engels beoordeeld. De beoordelaars hadden de mogelijkheid om over de kwaliteit van het SE geen oordeel uit te spreken, bijvoorbeeld omdat er materiaal ontbrak of omdat het beschikbare materiaal het vellen van een valide oordeel niet toeliet. In de analyse is voor ontbrekende oordelen over de algemene kwaliteit van het SE een score gesubstitueerd op basis van de oordelen voor de afzonderlijke vakonderdelen (zie voor meer informatie paragraaf 2.7). Enkele beschrijvende statistische gegevens van de 19 oordelen over de algemene kwaliteit van het SE zijn opgenomen in Tabel 10. Weergegeven zijn onder meer het aantal beoordeelde scholen, het gemiddelde oordeel, de standaarddeviatie en het percentage ontbrekende waarnemingen. Te zien is dat het percentage ontbrekende waarnemingen bij de meeste indicatoren redelijk is. Duidelijke uitzonderingen zijn indicator 17 en 18 (beide 87% missings).

Ontbrekende waarnemingen Nederlands

De steekproef bestond uit 84 scholen. Om tijd en kosten te besparen, zijn niet alle leerwegen van deze scholen in het onderzoek betrokken (zie paragraaf 2.1). Er is een steekproef getrokken van 26 scholen met een leerweg van de basisberoepsgerichte leerweg, 26 scholen met een leerweg van de kaderberoepsgerichte leerweg en 45 scholen met een leerweg van de gemengde/theoretische leerwegen.

Van de 84 scholen uit de steekproef Nederlands blijken er 82 voor een of meerdere leerwegen te zijn beoordeeld. Tegelijkertijd bevat het responsebestand twee particuliere scholen beoordeeld die niet in het steekproefbestand voorkwamen en die daar naderhand aan zijn toegevoegd.

Bij Nederlands zijn alle 82 scholen in de responsegroep voor alle aanwezige leerwegen beoordeeld (als er voor de desbetreffende leerweg materiaal voorhanden was) en niet alleen de getrokken leerwegen. De verdeling van de beoordeelde leerwegen over de 82 scholen was 38 voor de basisberoepsgerichte leerweg, 35 voor de kaderberoepsgerichte leerweg en 74 voor de gemengde/theoretische leerweg.

Tabel 11 Enkele beschrijvende statistische gegevens voor de 19 indicatoren van de algemene kwaliteit van het SE Nederlands

	Aantal scholen	Laagste score	Hoogste score	Gemiddelde	Standaard deviatie	% missings
1 Verplichte onderdelen SE	73	1	1	1.00	.00	13
2 Conform niveauspecificaties	77	1	1	1.00	.00	8
3 Afnamemoment gerelateerd aan niveau	75	1	1	1.00	.00	11
4 Moeilijkheid toetsen en opdrachten	77	1	3*	1.03	.23	8
5 Soepelheid strengheid beoordelingsmiddelen	76	1	1*	1.00	.00	10
6 Becijfering kandidaten	77	1	1*	1.00	.00	8
7 Conform producteisen	77	1	1	1.00	.00	8
8 Vergelijkbaarheid opgaven	76	1	1	1.00	.00	10
9 Verplichte toetsinhoud in overleg	74	1	1	1.00	.00	12
10 Beoordelingsaspecten relevant	77	1	1	1.00	.00	8
11 Beoordelingsaspecten volledig	75	1	1	1.00	.00	11
12 Gebruik antwoord- c.q. beoordelingsmodel	77	1	1	1.00	.00	8
13 Cijfers voor toetsen en opdrachten	76	1	1	1.00	.00	10
14 Beoordeling in overleg	75	1	1	1.00	.00	11
15 Scorepunten en/of cesuur vooraf vastgelegd	72	1	1	1.00	.00	14
16 Scorepunten vroegtijdig meegedeeld	70	1	1	1.00	.00	17
17 Bijstelling scorepunten cesuur beargumenteerd	21	1	1	1.00	.00	75
18 Normering in overleg	73	1	1	1.00	.00	13
19 Gewicht	71	1	1	1.00	.00	13

* De "niveau-indicatoren" zijn op een andere schaal beoordeeld, bijvoorbeeld 1 = juiste moeilijkheidsgraad (voldoende); 2 = te gemakkelijk; 3 = te moeilijk (zie paragraaf 2.4).

De beoordelaars hadden de mogelijkheid om over de kwaliteit van het SE geen oordeel uit te spreken, bijvoorbeeld omdat er materiaal ontbrak of omdat het beschikbare materiaal het niet toeliet om een valide oordeel te vellen. Tabel 11 laat zien dat het percentage ontbrekende waarnemingen bij de meeste indicatoren relatief hoog is. Dit geldt vooral voor indicator 17 met 75% ontbrekende waarnemingen.

Ontbrekende waarnemingen biologie

Van 71 (96%) van de in totaal 74 getrokken havo-scholen is de kwaliteit van het SE Biologie beoordeeld. De beoordelaars hadden de mogelijkheid om over de kwaliteit van het SE geen oordeel uit te spreken, bijvoorbeeld omdat er materiaal ontbrak of omdat het beschikbare materiaal het niet toeliet om een valide oordeel te vellen. Enkele beschrijvende statistische gegevens van de 19 oordelen over de algemene kwaliteit van het SE zijn opgenomen in Tabel 12. Weergegeven zijn onder meer het aantal beoordeelde scholen, het gemiddelde oordeel, de standaarddeviatie en het percentage ontbrekende waarnemingen. Te zien is dat het percentage ontbrekende waarnemingen bij een aantal indicatoren zeer hoog is (vooral bij de indicatoren 2, 14, 17 en 18 die veel vaker niet dan wel beoordeeld zijn).

Tabel 12 Enkele beschrijvende statistische gegevens voor de 19 indicatoren van de algemene kwaliteit van het SE biologie

	Aantal scholen	Laagste score	Hoogste score	Gemiddelde	Standaard deviatie	% missings
1 Verplichte onderdelen SE	48	0	1	.96	.20	32
2 Conform niveauspecificaties	2	1	1	1.00	.00	97
3 Afnamemoment gerelateerd aan niveau	67	0	1	.99	.12	6
4 Moeilijkheid toetsen en opdrachten	66	1	2*	1.17	.38	7
5 Soepelheid strengheid beoordelingsmiddelen	68	1	2*	1.09	.29	4
6 Becijfering kandidaten	69	1	2*	1.09	.28	3
7 Conform producteisen	45	0	1	.93	.25	37
8 Vergelijkbaarheid opgaven	63	0	1	.98	.13	11
9 Verplichte toetsinhoud in overleg	50	0	1	.94	.24	30
10 Beoordelingsaspecten relevant	68	0	1	.99	.12	4
11 Beoordelingsaspecten volledig	66	0	1	.98	.12	7
12 Gebruik antwoord- c.q. beoordelingsmodel	67	1	1	1.00	.00	6
13 Cijfers voor toetsen en opdrachten	68	1	1	1.00	.00	4
14 Beoordeling in overleg	3	1	1	1.00	.00	96
15 Scorepunten en/of cesuur vooraf vastgelegd	60	1	1	1.00	.00	15
16 Scorepunten vroegtijdig meegedeeld	39	0	1	.64	.49	45
17 Bijstelling scorepunten cesuur beargumenteerd	7	1	1	1.00	.00	90
18 Normering in overleg	10	1	1	1.00	.00	86
19 Gewicht	44	1	1	1.00	.00	38

* De "niveau-indicatoren" zijn op een andere schaal beoordeeld, bijvoorbeeld 1 = juiste moeilijkheidsgraad (voldoende); 2 = te gemakkelijk; 3 = te moeilijk (zie paragraaf 2.4).

Ontbrekende waarnemingen wiskunde

De steekproef bestond uit 84 scholen voor vmbo. Om tijd en kosten te besparen, zijn niet alle leerwegen van deze scholen getrokken (zie paragraaf 2.1). Er is een steekproef getrokken van 26 scholen met een leerweg van de basisberoepsgerichte leerweg, 26 scholen met een leerweg van de kaderberoepsgerichte leerweg en 45 scholen met een leerweg van de gemengde/theoretische leerwegen.

Van de 84 scholen uit de steekproef wiskunde blijken er 69 (82%) voor een of meerdere leerwegen te zijn beoordeeld. Van vijf scholen ontbreken waarnemingen voor alle negentien algemene indicatoren. Deze scholen zijn tot de non-response gerekend. Niet alle beoordelaars hebben zich aan de afspraken voor de

per school te beoordelen leerwegen gehouden. De verdeling van de leerwegen over de 69 beoordeelde scholen was 11 voor de basisberoepsgerichte leerweg, 25 voor de kaderberoepsgerichte leerweg en 39 voor de gemengde/theoretische leerweg (in plaats van respectievelijk 26, 26 en 45; zie paragraaf 2.1). Bij navraag bleek dat een beoordelaar zich gezien zijn opleiding en onderwijservaring niet bekwaam achtte om de kwaliteit van het SE in de basisberoepsgerichte leerweg te beoordelen. De beoordelaars waren unaniem van mening dat de kwaliteitsverschillen tussen verschillende leerwegen van dezelfde school miniem waren en dat het maken van een onderscheid tussen leerwegen derhalve niet zinvol was. Als de kwaliteit van het SE van de ene leerweg van een school goed of slecht was, gold dat vrijwel altijd ook voor de andere leerweg(en) van diezelfde school. Om deze reden is afgezien van de voorgenomen analyse en rapportage voor elke leerweg afzonderlijk.

De beoordelaars hadden de mogelijkheid om over de kwaliteit van het SE geen oordeel uit te spreken, bijvoorbeeld omdat er materiaal ontbrak of omdat het beschikbare materiaal het niet toeliet om een valide oordeel te vellen. Tabel 13 laat zien dat het percentage ontbrekende waarnemingen bij de meeste indicatoren doorgaans hoog tot zeer hoog is. Dit geldt vooral voor de indicatoren 1, 9, 14, 17 en 18 die vaker wel dan niet beoordeeld zijn.

Tabel 13 Enkele beschrijvende statistische gegevens voor de 19 indicatoren van de algemene kwaliteit van het SE wiskunde

	Aantal scholen	Laagste score	Hoogste score	Gemiddelde	Standaard deviatie	% missings
1 Verplichte onderdelen SE	26	1	1	1.00	.00	62
2 Conform niveauspecificaties	53	0	1	.98	.14	23
3 Afnamemoment gerelateerd aan niveau	67	1	1	1.00	.00	3
4 Moeilijkheid toetsen en opdrachten	68	1	2*	1.04	.21	1
5 Soepelheid strengheid beoordelingsmiddelen	68	1	2*	1.03	.17	1
6 Becijfering kandidaten	65	1	2*	1.03	.17	6
7 Conform producteisen	63	0	1	.97	.18	9
8 Vergelijkbaarheid opgaven	45	1	1	1.00	.00	35
9 Verplichte toetsinhoud in overleg	30	0	1	.50	.51	57
10 Beoordelingsaspecten relevant	67	0	1	.99	.12	3
11 Beoordelingsaspecten volledig	67	0	1	.99	.12	3
12 Gebruik antwoord- c.q. beoordelingsmodel	68	0	1	.96	.21	1
13 Cijfers voor toetsen en opdrachten	68	0	1	.96	.21	1
14 Beoordeling in overleg	25	0	1	.36	.49	64
15 Scorepunten en/of cesuur vooraf vastgelegd	61	0	1	.97	.18	12
16 Scorepunten vroegtijdig meegedeeld	66	0	1	.64	.48	4
17 Bijstelling scorepunten cesuur beargumenteerd	10	0	1	.60	.52	86
18 Normering in overleg	26	0	1	.38	.50	62
19 Gewicht	46	0	1	.85	.36	33

* De "niveau-indicatoren" zijn op een andere schaal beoordeeld, bijvoorbeeld 1 = juiste moeilijkheidsgraad (voldoende); 2 = te gemakkelijk; 3 = te moeilijk (zie paragraaf 2.4).

Behandeling van ontbrekende waarnemingen

De hiervoor gerapporteerde analyses laten zien dat het percentage ontbrekende waarnemingen vaak erg hoog is. De analyse van de verschillen tussen scholen en de algemene kwaliteit van het SE is daarom alleen uitgevoerd op die indicatoren waarvan het responsepercentage tachtig procent of hoger was. Voor Engels, biologie en wiskunde resteerden respectievelijk zestien, tien en elf indicatoren.

3 Resultaten

3.1 De algemene kwaliteit van het SE Engels

3.1.1 Verschillen tussen scholen

Allereerst is nagegaan in welke mate scholen verschillen in de algemene kwaliteit van het SE Engels. Vanwege een percentage ontbrekende waarnemingen groter dan 20% zijn de indicatoren 16, 17 en 18 in deze analyse buiten beschouwing gelaten. De analyse is dus gebaseerd op zestien indicatoren. Het gemiddelde aantal als voldoende beoordeelde indicatoren bedraagt 12.91 (78%) met een standaarddeviatie van 2.74. De verdeling van de somscore is weergegeven in Tabel 14. De school met het "slechtste" SE Engels scoorde op slechts vijf indicatoren een voldoende en van de dertien scholen met het "beste" SE waren alle zestien indicatoren in orde.

Tabel 14 Verdeling van de scholen naar de algemene kwaliteit van het SE Engels

Aantal indicatoren voldoende	Aantal scholen	Percentage
5	1	1
6	1	1
7	3	4
8	1	1
9	2	3
10	4	6
11	6	8
12	8	11
13	9	13
14	11	15
15	12	17
16	13	18
Totaal	71	100

3.1.2 Algemene kwaliteit per indicator

Het percentage onvoldoende en voldoende oordelen is per indicator gepresenteerd in Tabel 15¹. In deze tabel is het aantal valide oordelen als percentagebasis gehanteerd (dit wil zeggen dat de ontbrekende waarnemingen buiten de berekening zijn gehouden). De vijf indicatoren waarvoor de kwaliteit van het SE Engels het vaakst als voldoende beoordeeld werd en die derhalve het minst voor verbetering in aanmerking komen, zijn (waarbij indicatoren met meer dan 20% ontbrekende waarnemingen buiten beschouwing zijn gelaten):

- Gebruik antwoord- c.q. beoordelingsmodel (1% onvoldoende)
- Vergelijkbaarheid opgaven (2% onvoldoende)
- Scorepunten en/of cesuur vooraf vastgelegd (3% onvoldoende)
- Afnamemoment gerelateerd aan niveau (4% onvoldoende)
- Verplichte toetsinhoud in overleg (7% onvoldoende)

¹ Bij de interpretatie van de laagste en hoogste score bij de niveau-indicatoren 3, 4 en 5 zij opgemerkt dat de scores 2 en 3 voor te gemakkelijk of te moeilijk toetsmateriaal of voor een te strenge of te soepele beoordeling zijn gecodeerd naar 0 (voor onvoldoende).

De vijf indicatoren waarvoor de kwaliteit van het SE Engels het vaakst als onvoldoende beoordeeld werd en die derhalve het meest voor verbetering in aanmerking komen, zijn (waarbij indicatoren met meer dan 20% ontbrekende waarnemingen buiten beschouwing zijn gelaten):

- Conform niveauspecificaties (32% onvoldoende)
- Gewicht (33% onvoldoende)
- Verplichte onderdelen SE (34% onvoldoende)
- Beoordelingsaspecten volledig (34% onvoldoende)
- Beoordeling in overleg (44% onvoldoende)

Tabel 15 Percentage onvoldoende en voldoende oordelen per indicator (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen Engels

Indicator	Onvoldoende	Voldoende	Te gemakkelijk of te soepel	Te moeilijk of te streng	% missing
1 Verplichte onderdelen SE	34	66			6
2 Conform niveauspecificaties	32	68			0
3 Afnamemoment gerelateerd aan niveau	4	96			4
4 Moeilijkheid toetsen en opdrachten		89	4	7	0
5 Soepelheid strengheid beoordelingsmiddelen		86	11	3	1
6 Becijfering kandidaten		87	10	3	3
7 Conform producteisen	30	70			1
8 Vergelijkbaarheid opgaven	2	98			10
9 Verplichte toetsinhoud in overleg	7	93			4
10 Beoordelingsaspecten relevant	30	70			0
11 Beoordelingsaspecten volledig	34	66			0
12 Gebruik antwoord- c.q. beoordelingsmodel	1	99			0
13 Cijfers voor toetsen en opdrachten	16	84			3
14 Beoordeling in overleg	44	56			4
15 Scorepunten en/of cesuur vooraf vastgelegd	3	97			13
16 Scorepunten vroegtijdig meegedeeld	2	98			30
17 Bijstelling scorepunten cesuur beargumenteerd		100			87
18 Normering in overleg	11	89			87
19 Gewicht	33	67			6

3.1.3 Kwaliteit per vakonderdeel

De oordelen van de beoordelaars over de kwaliteit van het SE voor de vakonderdelen luisteren, schrijven en spreken zijn samengevat in Tabel 18, 19 en 20. De meest rechter kolom vermeldt telkens het percentage ontbrekende waarnemingen. De percentages in de overige kolommen zijn gebaseerd op de beschikbare valide beoordelingen (en tellen per rij op tot honderd procent). Het percentage ontbrekende beoordelingen blijkt sterk van indicator tot indicator en van vakonderdeel tot vakonderdeel te verschillen. De indicator "Conform niveauspecificaties" bij de onderdelen luisteren en schrijven is op vrijwel alle scholen beoordeeld. Van slechts weinig scholen zijn oordelen over de kwaliteit van het SE gespreksvaardigheid voorhanden.

Luistervaardigheid

Twaalf indicatoren voor de kwaliteit van het SE Engels blijken voor rapportage bruikbaar te zijn, dit wil zeggen hooguit 20% ontbrekende waarnemingen te hebben. De beoordelaars hebben de kwaliteit van het SE luisteren vrijwel zonder uitzondering als voldoende beoordeeld (zie Tabel 16). De drie uitzonderingen

zijn Scorepunten en/of cesuur vooraf vastgelegd (2% onvoldoende), Verplichte toetsinhoud in overleg (3% onvoldoende) en Beoordeling in overleg (3% onvoldoende). Hierbij zij vermeld dat de meeste scholen de luistertoetsen van het Cito afnemen (waarbij de cesuur niet van tevoren wordt vastgelegd en de indicator Beoordeling in overleg niet van toepassing is).

Tabel 16 Oordelen en percentage ontbrekende waarnemingen voor het vakonderdeel luisteren Engels per indicator per vakonderdeel (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen

Indicator	Onvoldoende	Voldoende	Te gemakkelijk	Te moeilijk	% missing
1 Verplichte onderdelen SE					100
2 Conform niveauspecificaties		100			6
3 Afnamemoment gerelateerd aan niveau	2	98			42
4 Moeilijkheid toetsen en opdrachten		100			8
5 Soepelheid strengheid beoordelingsmiddelen		100			7
6 Becijfering kandidaten		100			14
7 Conform producteisen		100			8
8 Vergelijkbaarheid opgaven		100			11
9 Verplichte toetsinhoud in overleg	3	97			8
10 Beoordelingsaspecten relevant		100			10
11 Beoordelingsaspecten volledig		100			8
12 Gebruik antwoord- c.q. beoordelingsmodel		100			8
13 Cijfers voor toetsen en opdrachten		100			75
14 Beoordeling in overleg	3	97			8
15 Scorepunten en/of cesuur vooraf vastgelegd	2	98			18
16 Scorepunten vroegtijdig meegedeeld		100			31
17 Bijstelling scorepunten cesuur beargumenteerd		100			97
18 Normering in overleg		100			93
19 Gewicht					100

Schrijfvaardigheid

Op basis van de zeven indicatoren waarvoor van ten minste 80% van de scholen bruikbare gegevens voorhanden waren, kan worden geconcludeerd dat de kwaliteit van het SE schrijfvaardigheid lang niet op alle scholen in orde is (zie Tabel 17). Het meest voor verbetering vatbaar zijn de volledigheid van de beoordelingsaspecten (69% onvoldoende), de beoordeling in overleg (63% onvoldoende), de relevantie van de beoordelingsaspecten (46% onvoldoende), conform producteisen (42% onvoldoende) en conform niveauspecificaties (34% onvoldoende). Indicatoren waarop vrijwel alle scholen een voldoende scoren zijn de verplichte toetsinhoud in overleg (8% onvoldoende) en de moeilijkheid van de toetsen en opdrachten (13% onvoldoende).

Tabel 17 Oordelen en percentage ontbrekende waarnemingen voor het vakonderdeel schrijven Engels per indicator per vakonderdeel (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen

Indicator	Onvoldoende	Voldoende	Te		% missing
			gemakkelijk	Te moeilijk	
1 Verplichte onderdelen SE					100
2 Conform niveauspecificaties	34	66			1
3 Afnamemoment gerelateerd aan niveau	13	87			45
4 Moeilijkheid toetsen en opdrachten		87	5	8	15
5 Soepelheid strengheid beoordelingsmiddelen		82	14	4	20
6 Becijfering kandidaten		82	14	4	28
7 Conform producteisen	42	58			8
8 Vergelijkbaarheid opgaven		100			48
9 Verplichte toetsinhoud in overleg	8	92			13
10 Beoordelingsaspecten relevant	46	54			14
11 Beoordelingsaspecten volledig	69	31			14
12 Gebruik antwoord- c.q. beoordelingsmodel	4	96			23
13 Cijfers voor toetsen en opdrachten	7	93			80
14 Beoordeling in overleg	63	38			10
15 Scorepunten en/of cesuur vooraf vastgelegd	2	98			42
16 Scorepunten vroegtijdig meegedeeld	18	82			85
17 Bijstelling scorepunten cesuur beargumenteerd		100			94
18 Normering in overleg	33	67			96
19 Gewicht					100

Gespreksvaardigheid

Bij het onderdeel gespreksvaardigheid Engels laten de zeer hoge percentages ontbrekende waarnemingen het doen van uitspraken over de kwaliteit van het SE slechts bij twee indicatoren toe (zie Tabel 18). Het betreft Verplichte toetsinhoud in overleg (8% onvoldoende) en Beoordeling in overleg (67% onvoldoende).

Tabel 18 Oordelen en percentage ontbrekende waarnemingen voor het vakonderdeel spreken per indicator per vakonderdeel (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen

Indicator	Onvoldoende	Voldoende	Te		% missing
			gemakkelijk	Te moeilijk	
1 Verplichte onderdelen SE					100
2 Conform niveauspecificaties	32	68			73
3 Afnamemoment gerelateerd aan niveau		100			77
4 Moeilijkheid toetsen en opdrachten		90		10	86
5 Soepelheid strengheid beoordelingsmiddelen		80		20	93
6 Becijfering kandidaten		100			90
7 Conform producteisen	24	76			52
8 Vergelijkbaarheid opgaven	13	88			89
9 Verplichte toetsinhoud in overleg	8	92			13
10 Beoordelingsaspecten relevant	26	74			45
11 Beoordelingsaspecten volledig	20	80			44
12 Gebruik antwoord- c.q. beoordelingsmodel		100			63
13 Cijfers voor toetsen en opdrachten	11	89			87
14 Beoordeling in overleg	67	33			11
15 Scorepunten en/of cesuur vooraf vastgelegd		100			90
16 Scorepunten vroegtijdig meegedeeld		100			93
17 Bijstelling scorepunten cesuur beargumenteerd		100			97
18 Normering in overleg	100				99
19 Gewicht					100

3.1.4 Voorbeelden examenpraktijk

In het voorgaande is volstaan met een cijfermatige weergave van de belangrijkste bevindingen. De tabellen laten niet zien welke minder acceptabele examenpraktijken de beoordelaars nu precies geconstateerd hebben. Evenmin verschaffen de tabellen inzicht in hetgeen wel in orde is. Vandaar dat wij de beoordelaars gevraagd hebben de meest frequente afwijkingen van de wettelijke voorschriften en de algemene kwaliteitseisen die zij geconstateerd hebben per indicator te omschrijven. Daarnaast is de beoordelaars gevraagd om enkele voorbeelden van voldoende of goede examenpraktijken te geven. Het onderliggende idee is dat scholen niet alleen kunnen leren van wat nog niet helemaal in orde is, maar ook van wat er goed of zeer goed gaat. In deze paragraaf presenteren we de belangrijkste bevindingen van de schriftelijke bevraging van de beoordelaars Engels (waarbij we het woordgebruik van de respondenten zoveel mogelijk intact hebben gelaten). Opgemerkt wordt dat het noemen van voorbeelden van goede en minder goede examenpraktijken niet voor alle indicatoren mogelijk was.

1 Verplichte onderdelen SE: Het SE voldoet aan de inhoudelijke specificaties uit het eindexamenprogramma

Voorbeelden minder goede examenpraktijk

- Schrijfvaardigheid toetste inhoudelijke kennis van behandelde teksten of leesvaardigheid.
- Vertalen wordt getoetst in plaats van schrijven.
- Idioom toetsen/voortgangstoetsen bepaalden voor een groot deel het schrijfvaardigheidscijfer.
- Schrijfvaardigheid was soms alleen in naam schrijfvaardigheid. De toets bestond dan b.v. alleen uit invuloefeningen met woordjes (de vocabulaire toetsen van Wasp/reporter werden gebruikt). Ook grammatica invuloefeningen als schrijftoets kwamen voor.
- Sommige scholen tellen in havo 5 toetsen voor woordenschat en/of grammatica mee voor het schoolexamen, wat volgens het "oude" programma niet is toegestaan.
- Spreekvaardigheid toetst literatuurkennis.

- In plaats van een gesprek moet er een presentatie worden gegeven.
- Leesvaardigheid werd beoordeeld en meegeteld als onderdeel van het SE.
- Literatuur werd beoordeeld en meegeteld als onderdeel van SE Engels.
- Handelingsdelen werden hier en daar beoordeeld met een cijfer dat meetelde voor het SE cijfer.

Voorbeelden van voldoende of goede examenpraktijk

- Schrijfvaardigheid was in de vorm van een briefopdracht
- Spreekvaardigheid gaf leerling de kans zijn mondelinge vaardigheid te etaleren
- Scholen gebruikten bijna allen de Cito-luistertoets in combinatie met de norm
- De meeste scholen toetsten de drie onderdelen (schrijfvaardigheid, luistervaardigheid en spreekvaardigheid) in de een of andere vorm

2 Conform niveauspecificaties: Het SE voldoet aan de niveauspecificaties uit het eindexamenprogramma

Voorbeelden minder goede examenpraktijk

- In een enkel geval had spelling een zodanig grote invloed dat een schrijftoets meer een spellingstoets leek.
- Schrijfvaardigheid betrof geen correspondentie/schrijfofdracht, maar geïsoleerde grammatica oefeningen en vragen over behandelde teksten en idoomtoetsen
- Schrijfvaardigheid blijkt een ruim begrip. Woordjes invultoetsen kwamen voor (de wasp/reporter vocabulaire toetsen bleken populair) als schrijftoets. Zo ook grammatica toetsen. Soms was een van deze dingen de complete toets. Soms kwamen ze samen voor of in combinatie met een echte schrijfofdracht zoals een brief.
- Spreekvaardigheid werd gebruikt om literatuur te toetsen
- Spreekvaardigheid was vrijwel nergens inhoudelijk te beoordelen. Het materiaal was er niet.
- Bij de zes scholen die ik heb beoordeeld, heb ik geen afwijkingen kunnen vinden.

Voorbeelden van voldoende of goede examenpraktijk

- Een briefopdracht voor schrijfvaardigheid
- Een flink aantal scholen gebruikt de door Cito geproduceerde schrijfvaardigheidstoetsen en beoordelingsnormen. Vaak worden op deze toetsen gebaseerde maar zelf gemaakte opdrachten gebruikt en of een variatie op de beoordelingsnormen. Deze leken vaak zeer doordacht.
- Voor luistervaardigheid wordt meestal de cito kijk en luistertoets gebruikt met de daarbij geadviseerde beoordeling en normering. Dit heeft als extra voordeel de vergelijkbaarheid tussen leerlingen van verschillende scholen en de objectiviteit.
- Een gesprek i.p.v. toets over boeken bij spreekvaardigheid
- Een enkele school vermeldde ook voor spreekvaardigheid de door Cito vervaardigde toetsen en beoordelingen te gebruiken. Ze waren echter niet bijgesloten, dus moest ik het maar voor waar aannemen.

3 Afnamemoment gerelateerd aan niveau: Het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben

Voorbeelden minder goede examenpraktijk

- Geen tot weinig
- Dat scholen luisteren/spreken/schrijven (weet ik niet meer) al afsloten in het 4e leerjaar
- Ook hier was natuurlijk de spreekvaardigheid moeilijk te beoordelen. Waar ik naar gezocht heb, was, bijvoorbeeld, een toets schrijfvaardigheid of luistervaardigheid in de vierde klas die in de vijfde niet herhaald werd. Ik ging ervan uit dat, tenzij men modulair werkt, leerlingen pas in de loop van havo 5 het vereiste niveau hebben om een eindtoets te maken. Een (opvolg) toets in Havo 5 leek me dus in alle gevallen noodzakelijk. Eigenlijk was dat overal in orde.

Voorbeelden van voldoende of goede examenpraktijk

- Meestal deed men het op het gewenste tijdstip
- Alles scholen leken te toetsen op een moment dat je redelijkerwijs mag verwachten dat leerlingen een Havo 5 niveau bereikt hebben.

4 Moeilijkheid toetsen en opdrachten: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Voorbeelden minder goede examenpraktijk

- Inhoudelijk veel te ambitieuze/moeilijke opdracht voor schrijfvaardigheid, ingaan op politiek/filosofisch probleem
- Ingewikkelde en abstracte schrijfoopdrachten waar Havo 5 eigenlijk niet toe in staat hoeven zijn (meer V6 niveau). Dus te moeilijk
- te moeilijk, een vwo schrijfoopdracht of een presentatie die moest worden gegeven
- te gemakkelijk, een vmbo schrijfoopdracht
- Alleen woordkennis toetsen (te gemakkelijk)
- Ook hier was mondeling moeilijk of niet te beoordelen. Daar waar de procedure beschreven was, leken me de opdrachten en de te volgen procedure erg veel van de leerling te eisen.

Voorbeelden van voldoende of goede examenpraktijk

- Schrijfvaardigheidopdrachten die de kandidaat vroegen een brief te schrijven passend bij wat van hen verwacht kon worden (presenteren van eigen situatie/informeren naar vrijetijdsbesteding/stageplek/opleiding)
- Daar waar de vorm de juiste was, bij voorbeeld voor schrijfvaardigheid was meestal ook de moeilijkheidsgraad in orde.
- De meeste scholen volgden voor luistervaardigheid de cito kijk en luistertoets. Dat is een prettig ijkpunt.

5 Soepelheid strengheid beoordelingsmiddelen: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Voorbeelden minder goede examenpraktijk

- Combinatie van ingewikkelde en te moeilijke opdracht met daarnaast een veel te soepele norm, of juist ook erg ingewikkelde/streng nakijkmethodiek, met veel nadruk op spelling
- Als de toets al niet conform de specificaties is (bijv. Grammatica toets) valt niet te achterhalen of de beoordeling billijk is.
- Voortgang/huiswerkcontrole e.d. bepalen mede het cijfer voor 1 van de vaardigheidtoetsen
- Soms veel te streng
- Soms helemaal niet streng
- Bij schrijfvaardigheid werd vaak alleen grammaticale correctheid vrij streng beoordeeld en werden opbouw en effectieve communicatie buiten beschouwing gelaten.
- Bij een enkele school werd erg veel nadruk gelegd op spelling en werd dit dus erg streng beoordeeld
- Ook hier was het onderdeel spreekvaardigheid niet te beoordelen wegens gebrek aan materiaal.

Voorbeelden van voldoende of goede examenpraktijk

- Schrijfvaardigheid- en spreekvaardigheidbeoordeling voorziet ook in communicatieve efficiëntie, is niet alleen het gevolg van aftrekken van fouten.
- Vaak ging dit goed
- Vele secties volgen de toetsen en beoordelingen van Cito die zijn objectief en laten ook bij schrijfvaardigheid een evenwichtige beoordeling toe
- Beoordelingsverschillen tussen verschillende leerlingen binnen een school heb ik gelukkig niet geconstateerd.

**6 Becijfering kandidaten: De toegekende cijfers zijn noch te hoog noch te laag
Scholen volgen de normering die bij de proefwerken wordt geleverd.**

Voorbeelden van goede of minder goede examenpraktijk

- Hangt samen met eventuele afwijkingen m.b.t. de opdracht
- Soms veel te hoog
- Soms veel te laag
- Verbazend genoeg kwamen vrijwel overal redelijke cijfers uit de bus. Ook daar waar spelling veel betekenis kreeg, waren de cijfers gemiddeld vrij goed. Daar waar grammatica slechts aandacht kreeg waren de cijfers gemiddeld ongeveer gelijk aan gemiddelden van scholen die cito toetsen en

beoordeling gebruiken. De beoordeling samen met de normering blijken uiteindelijk op een voor de meesten acceptabel gemiddelde uit te komen.

7 Conform producteisen: Het SE is zorgvuldig samengesteld en voldoet aan relevante producteisen

Voorbeelden minder goede examenpraktijk

- Schoolexamen door 1 persoon, zonder overleg, gemaakt
- (Te) vaak schrijfvaardigheid toetsen die slechts bestonden uit vocabulaire toetsen
- Toetsen soms geconstrueerd door een docent terwijl sectie uit meerdere docenten bestond.

Voorbeelden van voldoende of goede examenpraktijk

- Merendeel der secties werkt samen bij het construeren/samenstellen van de verschillende schoolexamens
- De meeste docenten stellen toetsen in overleg samen, soms ook met collega's uit aanpalende secties.
- Ondanks dat je mag aannemen dat overleg noodzakelijk is, was een van de best doordachte examens die ik onder ogen kreeg, het werk van een docente zonder vakcollega's binnen haar school.
- Er wordt standaard gebruikt gemaakt van Cito-toetsen luistervaardigheid
- Cito toetsen schrijfvaardigheid zijn vaak aan de basis van de schrijfvaardigheid opdrachten

8 Vergelijkbaarheid opgaven: Als met toetsvarianten wordt gewerkt, is de moeilijkheidsgraad ervan vergelijkbaar

Voorbeelden minder goede examenpraktijk

- Dit was niet te constateren, want in 99% van de gevallen stuurden scholen geen vergelijkbare opgaven op weinig herkansingsopdrachten gezien
- Als scholen al toetsvarianten van bepaalde toetsen opstuurden betrof het veelal idioomtoetsen, die feitelijk geen onderdeel van het schoolexamen zouden moeten uitmaken.
- Ik zou verwachten dat dit zou kunnen spelen bij spreekvaardigheid maar dat heb ik niet kunnen beoordelen wegens gebrek aan materiaal.
- Verder werden op alle scholen toetsen aan het gehele leerjaar tegelijk gegeven
- Wegens ontbreken van informatie onbeoordeelbaar.

Voorbeelden van voldoende of goede examenpraktijk

- De (weinige) scholen die toetsvarianten opstuurden, hadden vergelijkbare toetsen
- Op vrijwel alle scholen werd een SE toets (schrijfvaardigheid of luistervaardigheid) op hetzelfde moment aan het hele cohort gegeven. Dan zijn er dus geen problemen met vergelijkbaarheid.
- Eenmaal mochten leerlingen zelf een keuze maken uit drie opdrachten. Die opdrachten waren wellicht niet alle drie gelijk maar elke leerling had wel dezelfde keuze.

9 Verplichte toetsinhoud in overleg: De vakcollega's stellen de inhoud van het SE in overleg vast

Voorbeelden minder goede examenpraktijk

- Open deur: er zijn scholen waar docenten geen overleg met collega's (lijken te) plegen, soms zelfs als er wel degelijk vakcollega's op school zijn volgens de docentenvragenlijst.

Voorbeelden minder goede examenpraktijk

- Alle scholen met meer dan een sectie Engels gaven aan de inhoud in overleg vast te stellen. Daar moeten we dan maar vanuit gaan.
- Heel vaak gebeurde het wel

10 Beoordelingsaspecten relevant: Alle beoordelingsaspecten zijn relevant

Voorbeelden minder goede examenpraktijk

- Opbouwen en/of (brief)conventies wordt niet meebeoordeeld bij schrijfvaardigheid.
- In geval van grammatica/idioom/tekstkennistoetsen voor schrijfvaardigheid is alle relevantie verdwenen
- Teveel nadruk op wat hoogstens een aspect van schrijfvaardigheid mag zijn. (bijvoorbeeld spelling of grammatica)
- Er zijn scholen waarbij grammatica- en spelfouten erg zwaar wegen ten opzichte van inhoudelijke factoren.
- Literatuur bepaalt mede spreekvaardigheidcijfer
- Ten overvloede: voor spreekvaardigheid was dit niet te beoordelen.
- Soms was beoordeling ondoorzichtig omdat geen correctiemodel aanwezig was, of omdat ik het correctiemodel niet kon begrijpen.

Voorbeelden van voldoende of goede examenpraktijk

- Heldere algemene scoreformulieren voor de toetsen
- Zie bij 7. Weliswaar passen niet alle scholen de beoordelingsmodellen van Cito onverkort toe, maar dan blijken ze in ieder geval nagedacht te hebben over de systematiek die ze wel willen toepassen.

11 Beoordelingsaspecten volledig: Alle relevante beoordelingsaspecten worden in de beoordeling betrokken

Voorbeelden minder goede examenpraktijk

- Vooral bij spreken en schrijven werden niet altijd alle vereiste beoordelingsaspecten betrokken
- Vaak was van schrijfvaardigheid in de schrijfvaardigheidstoets geheel geen sprake
- Opbouw en effectiviteit van communicatie werden vaak niet in beschouwing genomen.

Voorbeelden van voldoende of goede examenpraktijk

- Luistertoets (vaak CITO) ging goed
- Hoewel hier duidelijk geen consensus over bestaat, waren er aardig wat scholen met een doordacht en evenwichtig correctiemodel voor schrijfvaardigheid

12 Gebruik antwoord- c.q. beoordelingsmodel: De beoordeling vindt plaats met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan

Voorbeelden minder goede examenpraktijk

- Dat het niet werd gebruikt
- Erg ingewikkelde en onduidelijke methodiek voor schrijfvaardigheid
- Bij sommige scholen blijkt op de antwoordbladen dat gescoord wordt op verscheidene deelaspecten van schrijfvaardigheid (al dan niet met verschillend onderling gewicht), terwijl er geen beoordelingsmodel is bijgeleverd. Er lijken dus afspraken te zijn, maar het is niet duidelijk of deze zijn vastgelegd.
- Beoordelingsmodel voor spreekvaardigheid ontbrak vaak
- Antwoordmodel en of criteria waren onvolledig of afwezig

Voorbeelden van voldoende of goede examenpraktijk

- Soms stonden criteria en gewicht voor de duidelijkheid op de opgaven voor de leerlingen vermeld.
- Veel scholen gebruiken voor schrijfvaardigheid de Citotoetsen en/of beoordelingsmodellen of toetsen en/of beoordelingsmodellen die daarop geïnspireerd zijn.

13 Cijfers voor toetsen en opdrachten: Voor de toetsen en handlingsdelen worden cijfers respectievelijk de kwalificaties “naar behoren” of “onvoldoende” toegekend

Voorbeelden minder goede examenpraktijk

- Dat literatuur soms werd meegenomen in spreekvaardigheid en dus een cijfer kreeg
- Literatuuropdrachten werden soms becijferd
- Voortgang/idioomtoetsen bepaalden soms mede het cijfer

Voorbeelden van voldoende of goede examenpraktijk

- Eigenlijk was dit meestal in orde. Slechts in een enkel geval werden handelingdelen met een cijfer beoordeeld dat meetelde voor het totaal gemiddelde. Toetsen kregen altijd cijfers.

14 Beoordeling in overleg: De beoordeling geschiedt door ten minste twee docenten

Voorbeelden minder goede examenpraktijk

- Bij spreekvaardigheid vaak niet het geval
- Dit was alleen te beoordelen voor schrijfvaardigheid. Voor Spreekvaardigheid was geen materiaal en luistervaardigheid gaat meestal volgens Cito. Ik moet hier voornamelijk afgaan op de antwoorden uit de vragenlijst
- Vaak beoordeelt alleen de eigen docent het werk van de leerlingen. Aan het leerling werk was nooit te zien dat een tweede docent er naar gekeken had, hoewel dat volgens de vragenlijst wel het geval zou zijn.
- Hier en daar geven docenten aan bij twijfel een collega in te schakelen.

Voorbeelden van voldoende of goede examenpraktijk

- Bij schrijfvaardigheid werd hier bij een behoorlijk aantal scholen getracht meerdere docenten bij te betrekken, soms heel summier (alleen twijfelgevallen)
- Twee docenten bij de spreekvaardigheid

15 Scorepunten en/of cesuur vooraf vastgelegd: De scorepunten en indien van toepassing ook de cesuur worden vóór de afname van het toetsmateriaal schriftelijk vastgesteld

Voorbeelden minder goede examenpraktijk

- Dat het niet gebeurde en niet stond vermeld op de opgave
- Was slecht te beoordelen
- Dit was eigenlijk niet te beoordelen. Tenzij op de aan de leerling uitgereikte toets de scorepunten en de cesuur vermeld stonden, was voor mij niet na te gaan wanneer deze vastgesteld waren. Zelfs voor de cito kijk en luister toets komt de cesuur pas achteraf.
- Soms zat bij een zending helemaal niets waaruit scorepunten/of cesuur bleken.

Voorbeelden van voldoende of goede examenpraktijk

- Dat het vaak wel gebeurde

16 Scorepunten vroegtijdig meegedeeld: De scorepunten worden vóór de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt

Voorbeelden minder goede examenpraktijk

- Dat het niet gebeurde
- Vaak nergens terug te vinden /slecht te beoordelen
- Het is niet na te gaan wat kandidaten van te voren te horen hebben gekregen. Vraag 27 van de vragenlijst was wel door alle scholen ingevuld. Daaruit zou volgen dat alle leerlingen van te voren op de hoogte zijn gesteld van de "beoordelingssystematiek". Ik vind dat wat te onduidelijk.

Voorbeelden van voldoende of goede examenpraktijk

- Bij luistertoetsen CITO wel
- Bij een deel van de scholen is bij schrijfvaardigheid de methode van beoordelen (beoordelingsaspecten en onderling gewicht, maximumscore) en scoren helder vermeld op de opgaven.

17 Bijstelling scorepunten cesuur beargumenteerd: De scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd

Voorbeelden van goede of minder goede examenpraktijk

- Wegens ontbreken van informatie onbeoordeelbaar.

18 Normering in overleg: De vakcollega's bepalen de normering in onderling overleg

Voorbeelden minder goede examenpraktijk

- Dat een docent het alleen deed
- Wegens ontbreken van informatie onbeoordeelbaar.

Voorbeelden van voldoende of goede examenpraktijk

- Heel vaak gebeurde dit wel in onderling overleg
- Dit was op grond van het werk en het vragenformulier niet na te gaan

19 Gewicht: Het gewicht bij de bepaling van het SE cijfer is voor elke toets aangegeven

Voorbeelden minder goede examenpraktijk

- Dat men zich niet hield aan de verdeling 1/3, 1/3, 1/3
- 1 school kwam op 110 procent uit
- Weging klopt niet of zat erg ingewikkeld in elkaar
- Als per onderdeel meerdere toetsen werden gegeven, was niet altijd duidelijk hoe deze zich verhielden.
- Wegens ontbreken van informatie onbeoordeelbaar. Daarom hebben we deze indicator opgevat als: De school past de gewichtsverdeling toe die het examenprogramma voorschrijft, waarbij we kleine afwijkingen hebben geaccepteerd. Hierbij konden we gebruik maken van de antwoorden op vraag 16 en 17 van de docentenvragenlijst.

Voorbeelden van voldoende of goede examenpraktijk

- Dat men zich wel hield aan de verdeling 1/3, 1/3, 1/3
- In het algemeen was de verdeling (1/3 schrijfv., 1/3 spreekv., 1/3 luisterv.) in orde.
- Helder overzicht op het PTA, bijv.
- Bij een enkele school stond de weging op de toets vermeld.

3.1.5 Conclusies en aanbevelingen

Conclusies

Een eerste conclusie is dat de beoordelaars de kwaliteit van het SE Engels gemiddeld bij 78% van de beoordelingen als voldoende hebben aangemerkt (waarbij drie indicatoren vanwege te veel ontbrekende oordelen buiten de analyse zijn gelaten).

Een tweede conclusie is dat er volgens de beoordelaars aanzienlijke verschillen tussen scholen bestaan in de kwaliteit van het SE Engels. Zo scoorde de school met het "slechtste" SE Engels op vijf van de zestien bruikbare indicatoren een voldoende; de dertien scholen met het "beste" SE Engels hadden hun schoolexamen op alle zestien indicatoren in orde.

Een derde conclusie is dat de kwaliteit van het SE Engels naar de mening van de beoordelaars niet voor alle indicatoren even hoog is. De vijf "beste" indicatoren waren Gebruik antwoord- c.q. beoordelingsmodel (1% onvoldoende), Vergelijkbaarheid opgaven (2% onvoldoende), Scorepunten en/of cesuur vooraf vastgelegd (3% onvoldoende), Afnamemoment gerelateerd aan niveau (4% onvoldoende) en Verplichte toetsinhoud in overleg (7% onvoldoende). De vijf indicatoren die mogelijk het meest voor verbetering in aanmerking komen, zijn Conform niveauspecificaties (32% onvoldoende), Gewicht (33% onvoldoende), Verplichte onderdelen SE (34% onvoldoende), Beoordelingsaspecten volledig (34% onvoldoende) en Beoordeling in overleg (44% onvoldoende).

Een vierde conclusie is dat de kwaliteit van het SE Engels niet voor elk vakonderdeel gelijk is. De kwaliteit van het SE luisteren is vrijwel zonder uitzondering als voldoende beoordeeld (hetgeen weinig verbazing wekt als men bedenkt dat vrijwel alle scholen de luistertoetsen van het Cito gebruiken). Het SE schrijfvaardigheid is naar het oordeel van de beoordelaars nog lang niet op alle scholen in orde. Het meest voor verbetering vatbaar zijn de volledigheid van de beoordelingsaspecten (69% onvoldoende), de beoordeling in overleg (63% onvoldoende), de relevantie van de beoordelingsaspecten (46% onvoldoende), conform producteisen (42% onvoldoende) en conform niveauspecificaties (34% onvoldoende). Indicatoren van de kwaliteit van het vakonderdeel schrijfvaardigheid waarop vrijwel alle scholen een voldoende scoren, zijn de verplichte toetsinhoud in overleg (8% onvoldoende) en de moeilijkheid van de toetsen en opdrachten (13% onvoldoende). Over de kwaliteit van het derde en laatste vakonderdeel gespreksvaardigheid kan vanwege het grote aantal ontbrekende waarnemingen geen conclusie worden getrokken.

Kanttekeningen

Bij Engels havo bleek de kwaliteit van het SE gespreksvaardigheid vrijwel niet te beoordelen (zie paragraaf 3.1.3). Het SE luistervaardigheid werd in feite alleen beoordeeld als de school voor dit vakonderdeel ander examenmateriaal dan de kijk-luistertoetsen van het Cito gebruikte. De oordelen over de algemene kwaliteit zijn daarom voor een belangrijk deel gebaseerd op de beoordelingen van het SE schrijfvaardigheid. Ook voor de toetsing van schrijfvaardigheid gebruiken veel scholen toetsen en beoordelingsmiddelen van het Cito of volgens het model van het Cito. Dit zal eveneens in veel gevallen hebben geleid tot een voldoende beoordeling. Speciaal bij de toetsing en beoordeling van gespreksvaardigheid echter zijn de risico's op niet-valide toetsing en onbevredigende beoordelingspraktijken groot. Dat van dit laatste sprake is, blijkt uit een van de weinige indicatoren met betrekking tot gespreksvaardigheid waarover dit onderzoek gegevens heeft opgeleverd. Maar liefst 67% van de respondenten blijkt onvoldoende te scoren op indicator 14 (Beoordeling in overleg). Dit is des te kwalijker omdat de toetsing van gespreksvaardigheid een uitermate vluchtige aangelegenheid is. Tenzij er bandopnames van de afnames worden gemaakt, moeten de taaluitingen van de kandidaten onmiddellijk beoordeeld worden. Maar hoe vaak worden bandopnames gemaakt? De vragenlijst *Evaluatie schoolexamen Engels havo schooljaar 2007-2008* heeft hierover informatie opgeleverd. Van 60 scholen zijn ingevulde vragenlijsten op dit punt geïventariseerd. De vraag of en, zo ja, hoe de verrichtingen van de leerlingen op het schoolexamenonderdeel gespreksvaardigheid worden geregistreerd, is door 58 scholen beantwoord, als volgt. 39 scholen (67%) registreren alleen schriftelijk. Zeven scholen (12%) registreren alleen door middel van audio-opname, waarbij één school aangeeft dat deze opname 48 uur wordt bewaard met het oog op eventueel beroep door de kandidaten. Zes scholen (10%) combineren schriftelijke registratie en audio-opname; één school (2%) combineert schriftelijke registratie en video-opname. Tot slot geven vijf scholen (9%) aan helemaal geen registratie te voeren. Kortom, ruim drie kwart van de scholen registreert alleen schriftelijk of helemaal niet. En wat het schriftelijk registreren betreft: in dit onderzoek is slechts bij een enkele school duidelijk geworden hoe docenten daarbij te werk gaan en daarbij lijkt niet altijd sprake te zijn van heldere beoordelingscriteria. Wat betreft de inhoud van de toetsing levert dit onderzoek één geruststellend gegeven op: bij de overgrote meerderheid van de scholen (92%) wordt de verplichte toetsinhoud in overleg vastgesteld (zie indicator 9 in Tabel 17 in paragraaf 3.1.3). Helaas is uit dit onderzoek nauwelijks duidelijk geworden waar die toetsinhoud dan uit bestaat. Echter, uitgaande van de hoge percentages onvoldoende op de indicatoren 2 (Conform niveauspecificaties; 34%), 7 (Conform producteisen; 42%), 10 (Beoordelingsaspecten relevant; 46%) en 11 (Beoordelingsaspecten volledig; 69%) voor het SE schrijfvaardigheid is er alle reden tot zorg ten aanzien van inhoud en beoordeling van het onderdeel gespreksvaardigheid. Maar ook de beoordelingspraktijken die in de vorige alinea werden beschreven, geven aanleiding om te veronderstellen dat minder scholen voor Engels voldoende zouden zijn beoordeeld, als genoeg informatie over de beoordeling van het SE gespreksvaardigheid voorhanden was geweest.

Aanbevelingen

De resultaten van dit onderzoek zijn voor Engels alleen gebaseerd op de schoolexamenpraktijk in havo 5 op grond van het examenprogramma van 1998. In 2007 is een nieuw examenprogramma van kracht geworden dat op een aantal punten sterk afwijkt van het vorige. Vooral de volgende punten zijn in het licht van dit onderzoek van belang:

- Volgens het nieuwe programma mag het bevoegd gezag ervoor kiezen vakonderdelen toe te voegen die niet worden omschreven in het programma.
- Over het onderlinge gewicht van verschillende vakonderdelen van het schoolexamen bevat het nieuwe examenprogramma geen voorschriften of aanwijzingen.
- Het nieuwe programma bevat geen niveauspecificaties.

Het eerste punt kan ertoe leiden dat de toetsing van deelaspecten van taalvaardigheid zoals woordkennis en grammaticakennis mee gaat tellen voor het schoolexamen. Uit dit onderzoek blijkt dat zulke praktijken nu al voorkomen in havo 5 – waar het volgens de regelgeving bij het oude examenprogramma niet is toegestaan. Dit is misschien al ingegeven door de mogelijkheden die het nieuwe programma biedt. Maar het kan ook heel goed zijn dat docenten/scholen onvoldoende op de hoogte zijn van de regelgeving of zich niet daaraan wensen te houden.

Docenten/scholen kunnen goede argumenten hebben om bepaalde deelaspecten van taalvaardigheid te laten meetellen voor het schoolexamen. Op die manier hebben ze bijvoorbeeld een stok achter de deur om leerlingen ertoe te brengen om woorden te leren. De inrichting van de centrale examens – waarbij het gebruik van een woordenboek is toegestaan – stimuleert daar niet toe. Niettemin is uit de literatuur bekend (zie bijvoorbeeld Schouten-Van Parreren (1985, 44) en Schmitt (2000, 19) dat woordkennis een cruciale rol speelt in de beheersing van een taal. Het is dan ook niet bezwaarlijk, als scholen ervoor kiezen deelaspecten van taalvaardigheid apart te toetsen en de resultaten mee te laten tellen voor het schoolexamen. Het zou echter wel bezwaarlijk zijn, als zulke deelaspecten een te groot gewicht in de schaal zouden leggen bij de bepaling van het eindcijfer voor een bepaalde vaardigheid. Ook in het nieuwe

examenprogramma is het er immers om begonnen dat leerlingen kunnen aantonen dat zij de taalgebruikswijzen (bijvoorbeeld gespreksvaardigheid en Schrijfvaardigheid) beheersen als geïntegreerde instrumenten om bepaalde doelen te bereiken, zoals “adequaat reageren in sociale contacten met anderen” en “informatie vragen en verstrekken” – hoe dan ook, om communicatie tot stand te brengen. Woord- en grammaticakennis zijn daarvoor slechts voorwaardenscheppend, wat tot uitdrukking zou moeten komen in een gering gewicht van deze vakonderdelen bij de berekening van het schoolexamencijfer. Daarbij is tevens van belang dat alle betrokkenen goed duidelijk is hoe het eindcijfer voor elk onderdeel van het schoolexamen tot stand komt. Eenzelfde transparantie moet uiteraard worden geboden ten aanzien van het onderlinge gewicht van de schoolexamenonderdelen bij de berekening van het uiteindelijke schoolexamen voor het hele vak. Dit lijkt des te belangrijker, omdat de grotere vrijheid die het nieuwe examenprogramma biedt, zou kunnen leiden tot forse verschillen in de manier waarop scholen de schoolexamens inrichten. Er is een risico dat sommige scholen relatief veel gewicht zullen toekennen aan de resultaten van leerlingen met betrekking tot deelaspecten, terwijl andere deze niet apart zullen toetsen. Of dat bepaalde scholen bijvoorbeeld de cijfers van de – receptieve – luistervaardigheid lichter zullen laten wegen dan die van de – productieve – gespreks- of schrijfvaardigheid. Dit is in lijn met de politieke trend om scholen meer zeggenschap te geven over de inrichting van hun onderwijs, maar staat op gespannen voet met de verantwoordelijkheid van de overheid om een bepaald onderwijspeil te garanderen. Een factor die dit risico ernstig versterkt, is het bovengenoemde ontbreken van niveauspecificaties in het nieuwe examenprogramma. Veel moeilijker dan voorheen zullen scholen eraan gehouden kunnen worden een bepaald onderwijsniveau te bereiken, want er is geen instrument om dat niveau objectief aan te toetsen.

Het is frappant dat de bovenstaande factoren nauw samenhangen met drie indicatoren waarop veel scholen in dit onderzoek al onvoldoende scores, namelijk: Verplichte onderdelen SE, Gewicht en Niveauspecificaties. Ook de bevindingen wat betreft toetsconstructie en de beoordeling van schrijfvaardigheidopdrachten zijn zorgelijk. Het is mogelijk dat de toetsing van gespreksvaardigheid lijdt aan dezelfde manco's – de beoordeling vermoedelijk zelfs in sterkere mate, gezien de vluchtigheid van het medium. Met betrekking tot dit onderdeel van het SE heeft dit onderzoek echter onvoldoende resultaten opgeleverd.

Een belangrijke aanbeveling is het herintroduceren van niveaueisen in het examenprogramma, bijvoorbeeld gestoeld op het Europees Referentiekader (zie bijvoorbeeld Meijer en Noijons, 2008).

Los daarvan kunnen scholen worden verplicht of gestimuleerd tot transparantie over hun praktijken. Daarnaast kan op verschillende manieren worden gewerkt aan bewustwording bij scholen en docenten ten aanzien van de inrichting en opzet van de schoolexamens.

Dit leidt tot de volgende aanbevelingen ter verbetering en/of handhaving van de kwaliteit van het SE Engels:

- Verplicht scholen om voor alle betrokkenen (dus ook de Inspectie) gedetailleerd te omschrijven welke vakonderdelen meetellen voor het SE en met welk gewicht.
- Verplicht scholen om voor alle betrokkenen te verantwoorden waarom zij de bovenstaande keuze hebben gemaakt.
- Verplicht scholen om voor alle betrokkenen duidelijk te maken welk(e) eindniveau(s) zij nastreven voor de verschillende onderdelen van het SE. Als leidraad kan daarbij de *Handreiking schoolexamen moderne vreemde talen* (Meijer & Fasoglio, 2006) dienen.
- Controleer jaarlijks een steekproef van de bovengenoemde documenten en wijs scholen zo nodig terecht.
- Stimuleer deskundigheidsbevordering bij scholen en docenten ten aanzien van de valide, betrouwbare en efficiënte toetsing en beoordeling van de verschillende onderdelen van het SE.

Zoals boven opgemerkt, heeft dit onderzoek onvoldoende resultaten opgeleverd wat betreft de toetsing van gespreksvaardigheid. Ook op andere punten konden onvoldoende gegevens worden verzameld (zie paragraaf 2.9), met name indicator 17 (Bijstelling scorepunten cesuur beargumenteerd) en 18 (Normering in overleg).

3.2 De kwaliteit van het SE Nederlands

3.2.1 Verschillen tussen scholen

Nagegaan is in welke mate scholen verschillen in de algemene kwaliteit van het SE Nederlands. Van de gemiddelde school uit het onderzoek is de kwaliteit van het SE Nederlands op gemiddeld 19 van de 19 indicatoren als voldoende beoordeeld (met een standaard deviatie van 0). Slechts bij één van de maximaal 22344² toe te kennen beoordelingen is het oordeel “onvoldoende” gerapporteerd. Het betrof de indicator “Moeilijkheid van de toetsen en opdrachten” waar de desbetreffende beoordelaar het oordeel “te moeilijk” rapporteerde.

Voor een samenvatting van de belangrijkste vakinhoudelijke, organisatorische en beoordelingstechnische kanttekeningen bij de onderzoeksresultaten wordt verwezen naar de algemene discussie in paragraaf 4.2. Een uitgebreide toelichting op de onderzoeksresultaten is opgenomen in Bijlage 6.

3.2.2 Algemene kwaliteit per indicator

Omdat de algemene kwaliteit van het SE op alle scholen voor alle negentien indicatoren op één uitzondering na als voldoende beoordeeld is, is cijfermatige rapportage van de kwaliteit per indicator voor het vak Nederlands niet zinvol. Voor een kwalitatieve bespreking wordt verwezen naar paragraaf 3.2.4 en 3.2.5.

3.2.3 Kwaliteit per vakonderdeel

Bij het vak Nederlands zijn de beoordelaars te werk gegaan volgens de methode van progressieve focusering. Dit wil zeggen dat zij eerst een *quick scan* uitvoerden en vervolgens, als daar aanleiding toe was, een *deep scan*. Tijdens de *quick scan* werd het materiaal van de school doorgenomen waarbij alle zeven vakonderdelen en alle leerwegen bekeken werden. Was er geen reden om aan de algemene kwaliteit van het SE te twifelen, dan werd de algemene kwaliteit voor de desbetreffende indicator als voldoende beoordeeld (bij gebrek aan contra-indicaties). Leek de kwaliteit twijfelachtig of onvoldoende, dan werd het oordeel over de kwaliteit opgeschort tot na de *deep scan*. Tijdens de *deep scan* werd al het materiaal van de school nauwkeurig bestudeerd waarbij de beoordelaars voor elk vakonderdeel een afzonderlijk kwaliteitsoordeel moesten geven. De vakonderdelen bij Nederlands waren: Luister- en kijkvaardigheid, Schrijfvaardigheid, Fictie, Gedocumenteerd schrijven, Basale taalvaardigheden (spelling, grammatica, woordkennis e.d), Mondelinge vaardigheden en Leesvaardigheid. Naar aanleiding van de *deep scan* kon voor de algemene kwaliteit van het SE alsnog de kwalificatie voldoende worden toegekend, maar het was ook mogelijk dat de algemene kwaliteit op de desbetreffende indicator een onvoldoende kreeg. De methode van progressieve focusering brengt met zich mee dat de beoordelaars altijd een oordeel gaven over de algemene kwaliteit maar de kwaliteit per vakonderdeel alleen beoordeelden in het geval van twijfel.

Een hier niet gerapporteerde analyse laat zien dat het percentage ontbrekende waarnemingen voor de oordelen over de zeven vakonderdelen 99.35% bedraagt.³ In de sporadische gevallen (.65%) dat een beoordelaar wel een oordeel over de kwaliteit per vakonderdeel uitsprak, bleek dat in geen enkel geval te leiden tot het oordeel “onvoldoende” voor de algemene kwaliteit van het SE. Kennelijk gaf de *quick scan* (waarbij alle vakonderdelen betrokken werden) slechts zeer zelden aanleiding tot twijfel aan de kwaliteit van het SE en tot nadere bestudering van de kwaliteit per vakonderdeel. En als er twijfel was, dan werd er in alle gevallen voor de algemene kwaliteit van het SE alsnog het oordeel “voldoende” uitgesproken. Wij zien dan ook af van verdere rapportage van de kwaliteit per vakonderdeel. Voor een kwalitatieve bespreking wordt verwezen naar paragraaf 3.2.4 en 3.2.5.

² Het maximaal aantal van 22344 toe te kennen oordelen is als volgt berekend: (38 bb + 35 kb + 74 gtl) * 19 indicatoren * 8 (1 algemeen oordeel en 7 oordelen per vakonderdeel)

³ Het grote aantal ontbrekende waarnemingen heeft ook te maken met de bij Nederlands toegepaste beoordelingsmethode. Tijdens de *quick scan* bekeken de beoordelaars voor ieder vakonderdeel de toetsen en/of opdrachten. Er werd echter slechts per vakonderdeel gerapporteerd als één van de eerste bekeken toetsen van een school tot twijfel aan de kwaliteit leidde.

3.2.4 Voorbeelden examenpraktijk

De vragenlijst *Inventarisatie Examenpraktijk* is ook voorgelegd aan de beoordelaars Nederlands. In deze vragenlijst is hun gevraagd zowel voorbeelden te geven van goede als minder goede examenpraktijken. Onderstaand overzicht bevat een bloemlezing uit hun antwoorden. Daarbij zijn de voorbeelden zoveel mogelijk in hun oorspronkelijke bewoording weergegeven. Opgemerkt wordt dat de respondenten niet alleen voorbeelden van goede en minder goede examenpraktijken gegeven hebben, maar ook kanttekeningen van meer algemene aard. Daarnaast zijn er niet voor alle indicatoren voorbeelden gegeven.

1 Verplichte onderdelen SE: Het SE voldoet aan de inhoudelijke specificaties uit het eindexamenprogramma

Voorbeelden van minder goede examenpraktijk

- BB, KB en GT geen. Opgemerkt moet worden dat Luistervaardigheid, Schrijfvaardigheid en Leesvaardigheid geen verplichte onderdelen zijn in het SE, zij zitten in het CSE. Resteert spreek- en gespreksvaardigheid (dat zich goeddeels aan de beoordeling in het kader van dit onderzoek onttrokken heeft), Fictie (dat op zeer gevarieerde wijzen getoetst wordt, zowel mondeling als schriftelijk, als ook door middel van een combinatie ervan, al dan niet aan de hand van zogenaamde fictiedossiers) en de Basisvaardigheden (veelal vormgegeven in een batterij kleine (combinatie-)toetsen, met spelling, woordenschat, zinsbouw).
- Bij GT zijn wel schrijfopdrachten aangetroffen in de SE's, maar deze hadden toch veelal het karakter van de opgaven zoals we die op het Centraal Examen aantreffen en die toch zelden het karakter hebben van 'gedocumenteerd schrijven' (zoals het althans door de opstellers van de eindtermen bedoeld is). Zie elders in het verslag opmerkingen over 'Gedocumenteerd schrijven'
- Het schrijven op basis van documentatie is vaak niet compleet of ontbreekt volledig. Scholen lijken niet te weten dat dit onderdeel verplicht is. Mogelijke oorzaak: De omschrijving van dit examenonderdeel is erg ruim en vaag. Docenten hebben te weinig houvast.
- Het toetsen van Basisvaardigheden en Leervaardigheden gebeurt op grote schaal. Scholen hechten er veel waarde aan. Vaak worden de methodegebonden toetsen gebruikt. Er is een groot verschil in niveau.
- Scholen weten over het algemeen niet hoe ze om moeten gaan met het onderdeel gedocumenteerd schrijven. Vaak geven scholen aan het onderdeel te toetsen bij het sectorwerkstuk. Het sectorwerkstuk wordt niet beoordeeld met een cijfer. Het onderdeel gedocumenteerd schrijven dus ook niet. Omdat het hier mijns inziens gaat om een van de belangrijkste onderdelen van het examen, dient er voor de scholen een duidelijker instructie te komen hoe om te gaan met dit onderdeel. Een uitgewerkt voorbeeld (met beoordelingscriteria) lijkt mij onontbeerlijk.
- Uit de ingevulde enquêtes blijkt dat veel docenten niet wisten wat ze moesten invullen bij 'basis' of 'leer'. Vaak werd aangegeven dat het onderdeel 'basis' niet werd getoetst, terwijl uit de meegestuurde toetsen bleek, dat dit onderdeel juist prominent aanwezig was. Het lijkt mij gewenst de (te) algemene termen 'basisvaardigheden' en 'leervaardigheden' op een begrijpelijke en eenduidige manier te definiëren.

Voorbeelden van voldoende of goede examenpraktijk

- Feitelijk zou, gelet op wat verplicht is in het SE, voor BB en KB alleen gespreksvaardigheid en fictie in aanmerking komen (en het diffuse, want weinig gepreciseerde Basis- en Leervaardigheden); voor GT komt daar luistervaardigheid bij. Aangezien de feitelijke schrijf- en luisteronderdelen in het bestaande centrale examen BB slechts in beperkte mate als afzonderlijke toets voor deze domeinen kunnen gelden (iets waarvan verondersteld mag worden dat het op scholen onvoldoende bekend is), doen scholen er, willen zij hun leerlingen goed voorbereiden voor vervolgstudie en verdere beroepsopleiding, goed aan ook schrijfvaardigheidsonderdelen en luistervaardigheidsonderdelen in hun schoolexamen te stoppen. Het centrale examen BB bevat immers slechts één of twee kleinere schrijfopdrachten en het luistergedeelte is zeer beperkt van omvang. Momenteel is een pilot digitaal examens KB in voorbereiding en daarbij dient zich het zelfde vraagstuk aan: in welke mate is het aangeboden examen voldoende dekkend voor luisteren en schrijven, in welke mate zou het schoolexamen complementair op deze onderdelen moeten werken (nb. iets soortgelijks geldt voor de moderne vreemde talen vmbo ten aanzien van de luistervaardigheid). Het huidige examenprogramma wekt ten onrechte de indruk dat in het centraal examens luister- en schrijfvaardigheid volledig worden afgedekt, een precisering tussen wat op het schoolexamen en wat op het centraal examen zou moeten worden getoetst om de eindtermen volledig te dekken, heeft nog onvoldoende plaatsgevonden).

- Bij BB en KB zijn er scholen die ook (niet-verplichte) schrijfvaardigheidopdrachten en leesvaardigheidopdrachten in hun schoolexamens aanbieden. Daarbij moet worden opgemerkt dat in het programma zelf nu nog onvoldoende beschreven is welke specifieke eindtermen ten aanzien van luisteren en schrijven zich lenen voor de schoolexamens en welke voor de centrale examens, waardoor het maken van keuzes door de docenten niet vergemakkelijkt wordt. Om uiteenlopende redenen zijn de (functionele) schrijfopgaven op het centrale examen strak aangestuurd, de eenduidigheid in de correctie is er de belangrijkste van. Veel van wat in de schoolexamens werd aangetroffen is er simpelweg een doublure van, wat gezien mag worden als een veilige keuze. Wat daarbij een stevig accent krijgt zijn schrijfconventies, de vaardigheden die onderbelicht blijven zijn het publiekgericht schrijven (veelal eenzijdig gericht op ‘formele instanties’, zoals toekomstige werkgevers bij sollicitatiebrieven) en het zelf verzamelen van documentatie. De vaardigheid in het reviseren van schrijfproducten wordt vrijwel nooit expliciet getoetst.
- Voorbeelden van *good practise* voor alle leerwegen zijn, zuiver vanuit toetstechnisch oogpunt, toetsen die zich richten op één domein, zodat er zuivere (eventueel ook diagnosticerende) uitspraken gedaan kunnen worden over de beheersing ervan. Zo is het beter het toetsen van de fictionele vaardigheden gescheiden te houden van de spreek- en gespreksvaardigheden (veel scholen combineren deze domeinen: het scheelt de docenten tijd, en geeft de onderwerpen waarover gesproken wordt een betekenisvolle inhoud). Veelal bieden de scholen de leerlingen van alle leerwegen voldoende mogelijkheden om creatief met fictievaardigheden om te gaan. Voorbeelden van *good practise* zijn die vormen van verwerking van fictie waarbij de leerlingen, behalve het beschrijven situaties en het benoemen van relaties, ook worden aangestuurd tot het geven van een beargumenteerd persoonlijk oordeel. Waar de toetsing schriftelijk plaats vindt, kon dit in de beoordeling worden meegenomen, waar de toetsing mondeling plaatsvindt, kon niet worden waargenomen of en in welke mate aan het laatste werd voldaan.

2 Conform niveauspecificaties: Het SE voldoet aan de niveauspecificaties uit het eindexamenprogramma

Voorbeelden van goede of minder goede examenpraktijk

- Opgemerkt moet worden dat in het programma zelf weinig of geen niveauspecificaties voorkomen. Docenten zullen dus veelal houvast zoeken bij wat de methodes aanbieden en, voor zover de schoolexamens en de centrale examens een overlap vertonen (geldt vooral voor leesvaardigheid en schrijfvaardigheid – overigens facultatieve onderdeel in het SE), zo dicht mogelijk de niveaus benaderen die in het CE verondersteld worden.
- Er is sprake van niveauverschil tussen verschillende toetsen. Doordat de PTA-programma's ontbreken is moeilijk te beoordelen of het niveau van de toets past bij het niveau van de leerlingen op dat moment.
- Dit criterium is moeilijk meetbaar, aangezien de niveauspecificaties dermate vaag zijn, dat bijna alles eronder kan vallen. Mijns inziens dient duidelijker aangegeven te worden (met voorbeelden, ook van beoordeling) wat van welk niveau verwacht mag worden. Het ontbreken van heldere criteria leidt aantoonbaar tot grote verschillen in de moeilijkheidsgraad van de schoolexamens.

3 Afnamemoment gerelateerd aan niveau: Het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben

Voorbeelden van goede of minder goede examenpraktijk

- Opgemerkt moet worden dat er geen dwingende volgorde in het aanbieden van de domeinen (die veelal een cyclisch karakter hebben) is voorgeschreven. Leesvaardigheid, schrijfvaardigheid, luistervaardigheid en gespreksvaardigheid worden zonder vastgestelde uitzondering in het laatste examenjaar aangeboden. Veel wordt bij herhaling getoetst. De toetsen die betrekking hebben op deelvaardigheden (spelling, woordkennis e.d.) worden verspreid over de examenjaren aangeboden. Veel van deze (soms erg kleine) toetsen verdienen eerder de kwalificatie van voortgangstoetsing. Fictie kent een variëteit aan schriftelijke toetsvormen, verspreid over twee leerjaren (soms parallel lopend met het ‘afwerken’ van een leeslijst). De mondelinge examens Fictie vinden, voor zover ze plaatsvinden, veelal in het laatste examenjaar plaats.
- Er is sprake van niveauverschil tussen verschillende toetsen. Doordat de PTA's ontbreken is moeilijk te beoordelen of het niveau van de toets past bij het niveau van de leerlingen op dat moment.
- Opnieuw een vaag criterium. De vaardigheden Nederlands kunnen op ieder moment afgenomen worden. Alleen de beoordelingscriteria zullen verschillen.

4 Moeilijkheid toetsen en opdrachten: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Voorbeelden van goede of minder goede examenpraktijk

- De schrijfvaardigheid- en leesvaardigheidstoetsen zijn dikwijls derivaten van wat in het centraal examen wordt aangeboden (en worden door de docenten als van het relevante niveau beschouwd). M.m. geldt hetzelfde voor de luistertoetsen (dikwijls door derden aangeleverd). De kleinere toetsen (spelling, grammatica, woordenschat) bevatten veel herhalings-elementen (bijvoorbeeld de spelling van werkwoorden). Veel taalregels (zoals die m.b.t. spelling) bekliven bij de zwakkere taalgebruikers slechts door voortdurende herhaling, en gelet op de maatschappelijke aandacht voor de matige schrijfvaardigheid der Nederlandse burgers (spelling), lijkt deze aandacht momenteel gerechtvaardigd. Sommige opdrachten (bijvoorbeeld die voor het maken van een folder in het kader van schrijfvaardigheid) worden door de leerlingen met plezier ter hand genomen, maar zijn niet altijd uitdagend te noemen. Ook verwerkingsopgaven voor literatuur vergen soms wel veel tijd, maar niet altijd veel inspanning. Het gaat dan duidelijk om stimuli ter vergroting van het leesplezier, zij leiden niet altijd per se tot dieper inzicht of kennisvergroting. Veelal gaat het echter om meer dan één opdracht, waarbij altijd wel, conform de eindtermen, op enigerlei moment situaties, handelingen en relaties getypeerd moeten worden.
- De methodegebonden toetsen zijn meestal voldoende. Een aantal toetsen had een te hoog niveau voor vmbo-leerlingen. Een school stuurde een se literatuur in, waaraan havo-leerlingen nog wel eens een grote kluit zouden kunnen hebben. Een andere school neemt een schoolexamen af waarin de verschillende moeilijke bijzinnen getoetst worden.
- In enkele gevallen waren de toetsen beslist te moeilijk, in enkele gevallen waren de toetsen beslist te gemakkelijk. Te moeilijk waren vooral bepaalde dictees en grammaticatoetsen (samengestelde zin en naamwoordelijk gezegde kwamen voor).
- Docenten zijn geen toetsmakers. De zelf geformuleerde vragen bij leesteksten en de formuleringen bij schrijfopdrachten zorgden herhaaldelijk voor onbegrijpelijke opdrachten en vragen. Hierdoor werden de toetsen soms moeilijker dan nodig en bedoeld was.

5 Soepelheid strengheid beoordelingsmiddelen: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Voorbeelden van goede of minder goede examenpraktijk

- Duidelijk te soepele of te strenge beoordelingsmiddelen zijn vrijwel niet aangetroffen, De weging der afzonderlijke beoordelingscomponenten bij schrijfvaardigheid is veelal ontleend aan wat op het centraal examen gebruikelijk is, m.m. geldt hetzelfde voor leesvaardigheidstoetsen.
- Dit onderdeel is lastig te beoordelen. Niet alle scholen hebben hun beoordeling van te voren vastgesteld. Onderdelen als spreken en fictie (mondeling) lijken meer op gevoel te worden beoordeeld.
- Over het algemeen is de beoordeling in orde. Opvallend is, dat de gemiddelde resultaten per klas en per school niet al te ver uiteenlopen. Te moeilijke toetsen worden dus met een zekere soepelheid beoordeeld. De uiteindelijke resultaten van de verschillende scholen verschillen daardoor niet al te zeer.
- De beoordeling van de spreekvaardigheid blijft ondoorzichtig. Dit kan vooral problematisch zijn, wanneer een school erg veel gewicht geeft aan het onderdeel, zoals enkele keren voorkomt.
- De beoordelingscriteria voor fictie verschillen per school dermate, dat enige sturing wenselijk lijkt. Binnen de vage criteria die in de wet gesteld worden is zo ongeveer iedere toetsing toegestaan. In de praktijk betekende dit, dat op sommige scholen pure verhaaltheorie getoetst wordt, op andere scholen dient het literatuurtentamen voornamelijk om te komen tot een cijfer voor spreekvaardigheid.

6 Becijfering kandidaten: De toegekende cijfers zijn noch te hoog noch te laag

Voorbeelden van goede of minder goede examenpraktijk

- Scholen volgen de normering die bij de proefwerken wordt geleverd.
- Vrijwel alle toetsen laten een normaal beeld zien, met een duidelijke spreiding van de cijfers en een te verwachten verhouding tussen voldoende en onvoldoende. Het onderdeel fictie kent doorgaans een wat mildere beoordeling dan de andere, er worden doorgaans weinig diepe onvoldoendes voor gegeven (maar ook niet vaak hele hoge).
- Over het algemeen is de beoordeling in orde. Opvallend is, dat de gemiddelde resultaten per klas en per school niet al te ver uiteenlopen. Te moeilijke toetsen worden dus met een zekere

soepelheid beoordeeld. De uiteindelijke resultaten van de verschillende scholen verschillen daardoor niet al te zeer.

- De beoordeling van de spreekvaardigheid blijft ondoorzichtig. Dit kan vooral problematisch zijn, wanneer een school erg veel gewicht geeft aan het onderdeel, zoals enkele keren voorkomt.
- De beoordelingscriteria voor fictie verschillen per school dermate, dat enige sturing wenselijk lijkt. Binnen de vage criteria die in de wet gesteld worden is zo ongeveer iedere toetsing toegestaan. In de praktijk betekende dit, dat op sommige scholen pure verhaalthorie getoetst wordt, op andere scholen dient het literatuurtentamen voornamelijk om te komen tot een cijfer voor spreekvaardigheid.

7 Conform producteisen: Het SE is zorgvuldig samengesteld en voldoet aan relevante producteisen

Voorbeelden van goede of minder goede examenpraktijk

- Veel (kleinere) toetsen vormen een mêlee van (sub)vaardigheden: spelling, interpunctie, woordenschat, een enkele schrijfopdracht, wat kennisvragen aangaande fictie, soms nog met een enkele leesttekstje. Dit soort toetsen is te typeren als een algemene taalvaardigheidstoets en het is lastig om ze onder te brengen bij één van de domeinen schrijven of lezen. Zij kunnen zeer wel functioneren als voortgangstoetsen of diagnostische toetsen (basale vaardigheden) en hebben ook de duidelijke functie van de leerlingen aan het werk te houden. Spelling en formuleervaardigheden zouden echter ook als (instrumentele, functionele) subvaardigheden bij schrijfvaardigheid kunnen worden getoetst, maar schrijfvaardigheid zit niet verplicht in het SE.
- Door het ontbreken van de PTA's is niet te constateren of scholen de se's gelijkwaardig hebben verdeeld of dat zij een onderdeel zwaarder laten wegen dan een ander.
- In de samenstelling van de SE's zitten enorme verschillen. Dat wil niet zeggen dat ze niet zorgvuldig zijn samengesteld. Het geeft wel aan, dat scholen blijkbaar alle ruimte hebben om het SE in te richten naar eigen goeddunken. Veel scholen leggen de nadruk op toetsen algemene vaardigheden uit schoolboeken.
- De weging van de verschillende onderdelen verschilt enorm. Sommige scholen leggen de nadruk op algemene vaardigheden, andere scholen waarderen fictie bovenmate en weer andere scholen spreek- en luistervaardigheid. Ik denk dat het samenstellen van een PTA voor veel (de meeste) scholen gemakkelijker zou worden, wanneer de weging van de verplichte onderdelen voor alle scholen hetzelfde zou zijn. Met andere woorden: het PTA wordt voor alle scholen identiek, alleen de invulling en beoordeling van de verschillende onderdelen gebeurt per school.
- Het onderscheid 'verplicht in SE' en 'mag in SE' werkt verwarrend en zorgt voor een onevenwichtige samenstelling van SE's in den lande. Eenduidigheid hieromtrent zou de schoolexamens gelijkwaardiger maken.

8 Vergelijkbaarheid opgaven: Als met toetsvarianten wordt gewerkt, is de moeilijkheidsgraad ervan vergelijkbaar

Voorbeelden van goede of minder goede examenpraktijk

- Toetsvarianten komen zelden voor, zelfs niet daar waar ze verwacht zouden kunnen worden: bij schrijfvaardigheid. Bij fictie hebben leerlingen vrijheid om te kiezen uit een reeks van boeken (al dan niet verfilmd), die beslist niet allemaal een zelfde moeilijkheidsgraad hebben, maar doorgaans wel aansluiten op de belevingswereld van de leerlingen. De opdrachten zijn doorgaans onafhankelijk van het gelezen, bekeken fictionele werk.
- Geen afwijkingen gevonden

9 Verplichte toetsinhoud in overleg: De vakcollega's stellen de inhoud van het SE in overleg vast

Voorbeelden van goede of minder goede examenpraktijk

- Docenten blijken volgens de enquêtes zelden als solist te (mogen) opereren. Het is veelal de gezamenlijke sectie (of althans de subsectie van de eindexamendocenten) die de inhoud en de frequentie van de SE-toetsen bepaalt. Onderzoek naar het toetsbeleid van scholen op het punt van toetsfrequentie en gewicht van toetsen zou hier gewenst zijn.
- Volgens de enquêtes is dit het geval. Uiteraard is dit gegeven niet op te maken uit de toegestuurde opgaven.

10 Beoordelingsaspecten relevant: Alle beoordelingsaspecten zijn relevant

Voorbeelden van goede of minder goede examenpraktijk

- Blijkens de enquête zijn de relevante beoordelingscriteria voor spreekvaardigheid en schrijfvaardigheid landelijk gangbaar. Niet uitgesloten mag worden dat de voorgelegde criteria in de enquête (bij fictie, luistervaardigheid, spreekvaardigheid en schrijfvaardigheid) tot sociale wenselijke invulling hebben geleid. Om te onderzoeken wat op het vlak van spreekvaardigheid (inclusief het mondeling 'fictie') daadwerkelijk aan criteria wordt gehanteerd, zou er op de scholen geobserveerd moeten worden.
- Dit criterium is te vaag om goed te kunnen beantwoorden.

11 Beoordelingsaspecten volledig: Alle relevante beoordelingsaspecten worden in de beoordeling betrokken

Voorbeelden van goede of minder goede examenpraktijk

- Blijkens de enquête zijn de relevante beoordelingscriteria voor spreekvaardigheid en schrijfvaardigheid landelijk gangbaar. Niet uitgesloten mag worden dat de voorgelegde criteria in de enquête (bij fictie, luistervaardigheid, spreekvaardigheid en schrijfvaardigheid) tot sociale wenselijke invulling hebben geleid. Om te onderzoeken wat op het vlak van spreekvaardigheid (inclusief het mondeling 'fictie') daadwerkelijk aan criteria wordt gehanteerd, zou er op de scholen geobserveerd moeten worden.
- Opnieuw te vaag om te kunnen beoordelen.

12 Gebruik antwoord- c.q. beoordelingsmodel: De beoordeling vindt plaats met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan

Voorbeelden van goede of minder goede examenpraktijk

- Voor de onderdelen leesvaardigheid (ontleend aan centraal examen), schrijfvaardigheid (ontleend aan centraal examen), luistervaardigheid (Citotoetsen) is dit vrijwel altijd op orde. Voor wat betreft spreek- en gespreksvaardigheid ook, al zijn niet altijd de gewichten aangegeven. Bij fictie zijn op onderdelen leerlingenwerken aangetroffen met globale beoordelingen, zonder dat duidelijk was welke criteria werden gehanteerd, laat staan dat het gewicht ervan bekend was.
- Er is verschil in beoordelen tussen de verschillende toetsen. Voor bepaalde onderdelen, zoals spelling en lezen is het vastleggen van de beoordelingscriteria gemakkelijker dan voor onderdelen als spreken, fictie en schrijven.
- Over het algemeen klopt dit. Criteria en gewicht kunnen per school/docent echter verschillen.

13 Cijfers voor toetsen en opdrachten: Voor de toetsen en handelingsdelen worden cijfers respectievelijk de kwalificaties "naar behoren" of "onvoldoende" toegekend

14 Beoordeling in overleg: De beoordeling geschiedt door tenminste twee docenten

Voorbeelden van goede of minder goede examenpraktijk

- De algemene taalvaardigheidstoetsen worden doorgaans door één docent beoordeeld (inclusief de kleinere schrijfoopdrachten, die in de meerderheid zijn). Grotere schrijfoopdrachten ('gedocumenteerd schrijven' zou er daarvan één kunnen zijn) worden, blijkens de enquête, ook wel door andere vakdocenten beoordeeld, maar worden juist binnen het vak zelf weinig aangeboden (en al helemaal niet op BB- en KB-niveau). De kleine schrijfoopdrachten spitsen zich toe op schrijfconventies (het gaat om vrijwel 'gesloten' opgaven met een strak beoordelingsmodel, waardoor (beoordelingstechnisch gezien) met één beoordelaar volstaan kan worden. M.m. geldt hetzelfde voor leestoetsen (ontleend aan oude examens) en de vele kleinere taaltoetsen. Het beoordelen van spreek- en gespreksvaardigheid (dikwijls gecombineerd met fictie), vergt ten minste één tweede beoordelaar, maar aangezien dit onderdeel zich aan observaties in het kader van dit onderzoek heeft onttrokken, valt niet te melden of docenten / scholen zich hier in de praktijk werkelijk aan houden. Verwacht mag worden dat veel scholen om schoolorganisatorische redenen (lesuitval, tijdgebrek) zullen volstaan met een beoordeling door één docent.
- In de praktijk wordt er niet altijd door twee docenten beoordeeld. Het afnemen van een mondeling literatuur met een 'bijzitter' is vaak al een probleem; het kost te veel uren!)

- Uit de enquêtes blijkt niet dat ten minste twee docenten beoordelen. Volgens mij is dat (behalve bij het sectorwerkstuk) ook geen eis.

15 Scorepunten en/of cesuur vooraf vastgelegd: De scorepunten en indien van toepassing ook de cesuur worden vóór de afname van het toetsmateriaal schriftelijk vastgesteld

Voorbeelden van goede of minder goede examenpraktijk

- Bij de beoordeling van dit criterium kan slechts gebruik gemaakt worden van de antwoorden uit de enquête. Niet echt goed te beoordelen dus.

16 Scorepunten vroegtijdig meegedeeld: De scorepunten worden vóór de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt

Voorbeelden van goede of minder goede examenpraktijk

- Kon niet worden vastgesteld.
- Voor veel docenten is reeds bij voorbaat voldaan aan dit criterium, doordat de leerlingen beschikken over een PTA.
- Uit de examenopgaven zelf blijkt overigens vaak niet hoeveel punten een leerling kan krijgen per vraag. Ook is zelden op het examen vermeld wat de weging van het betreffende SE is binnen het PTA.

17 Bijstelling scorepunten cesuur beargumenteerd: De scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd

Voorbeelden van goede of minder goede examenpraktijk

- Kon niet worden vastgesteld, maar mag wel bij gemeenschappelijke toetsen met antwoordmodel en scorevoorschrift worden verondersteld.
- Niet te beoordelen.

18 Normering in overleg: De vakcollega's bepalen de normering in onderling overleg

Voorbeelden van goede of minder goede examenpraktijk

- Uit de enquêtes mag worden afgeleid dat dit doorgaans het geval is voor de schriftelijke toetsen (de algemene taalvaardigheidstoetsen, de luistertoetsen, de leesvaardigheidstoetsen, de schrijftoetsen).
- Ook hier kan alleen beoordeeld worden op grond van de invulling van de enquête.

19 Gewicht: Het gewicht bij de bepaling van het SE cijfer is voor elke toets aangegeven

Voorbeelden van goede of minder goede examenpraktijk

- Of dit daadwerkelijk in de PTA's is vastgelegd, is niet nagegaan.
- Door het ontbreken van de PTA-regelingen is dit niet te beoordelen.
- Dit geldt alleen voor het PTA. Opgemerkt dient te worden dat uit het PTA soms moeilijk valt te herleiden welk gewicht een enkele toets heeft.

3.2.5 Conclusie en aanbevelingen

Conclusie

De belangrijkste conclusie is dat de beoordelaars van mening waren dat het SE Nederlands op alle onderzochte scholen voor alle indicatoren volgens de gehanteerde criteria van voldoende kwaliteit is. Dat betekent niet dat er geen fouten of tekortkomingen zijn aangetroffen of dat scholen niet verschillen in de kwaliteit van het SE Nederlands, maar per saldo bleek de kwaliteit steeds voldoende. Voor een nadere toelichting op deze conclusie en een kwalitatieve bespreking van de geconstateerde fouten en tekortkomingen wordt verwezen naar Bijlage 6 (die tevens aanleiding gaven tot de hieronder besproken aanbevelingen).

Aanbevelingen

Op grond van deze studie kunnen voor het vak Nederlands aanbevelingen worden gedaan ten aanzien van de regelgeving, het eindexamenprogramma, de examenpraktijk en vervolgonderzoek.

Regelgeving

- Facultatieve onderdelen in SE beperken, dat wil zeggen: luistervaardigheid en schrijfvaardigheid in alle leerwegen op alle niveaus in de SE's verplicht stellen (met daarbij een toewijzing van de eindtermen per domein naar SE en CSE) (NB het huidige CSE dekt de domeinen schrijfvaardigheid en luistervaardigheid onvoldoende).
- Het aantal SE's limiteren (zowel voor klas drie als klas vier); voorts een minimum per domein voorschrijven.
- Heldere voorschriften omtrent de noodzaak van een tweede correctie/beoordeling per type SE (bij Nederlands o.i. alleen noodzakelijk bij mondelinge SE's en grotere schrijfopdrachten).
- Heldere regelbepaling omtrent de gecombineerde toetsing van spreek- en gespreksvaardigheid en fictie (NB verschillende bewindspersonen hebben zich hier in het verleden op verschillende wijze over deze vermeende ontlasting bij gecombineerde toetsing van de docent/scholier uitgelaten).
- Het gewicht van de onderdelen in de bepaling van het SE-cijfer nader regelen.

Eindexamenprogramma Nederlands

- De Basisvaardigheden preciseren en herformuleren tot duidelijke leerdoelen.
- Nauwkeurig preciseren wat er op het vlak van spelling en grammatica beheerst moet worden (een taalbeheersingscanon opstellen).
- De V-leerdoelen ('gedocumenteerd schrijven') preciseren en duidelijk maken welke minimale rol hier precies is vastgelegd voor het vak Nederlands; aanwijzingen voor toetsing geven.
- Eindtermen bij fictie kwantificeren: omvang leeslijst/fictielijst vaststellen (vgl. havo en vwo).
- Eindtermen bij fictie kwalificeren: een algemene canon voor (jeugd)literatuur opstellen.
- Preciseren notie 'publiekgericht schrijven'.
- Het gewicht der onderdelen in het programma opnemen.
- Bij elk domein niveau-indicatoren opnemen en/of taalgebruikerscontexten vermelden (zie hoe dat in het ERK gebeurt).

Examenpraktijk Nederlands

- Het aantal toetsen limiteren.
- Het aantal combinatietoetsen beperken, dan wel de resultaten ervan (dus op de kleinere toetsen) steeds overhevelen naar een resultatenoverzicht van de afzonderlijke vaardigheden (en dit overzicht ook rapporteren aan de leerlingen); de kleinere toetsen die bedoeld zijn om de leerlingen te activeren plaatsen onder de voortgangstoetsing.
- In de SE's schrijfvaardigheid ook andere, bredere toetsen voor schrijfvaardigheid aanbieden dan op het CSE.
- Bij fictie ook toewerken naar het zelfstandig oordelen van de leerlingen (kennis dienstig maken aan beleving).
- Bij spreek- en gespreksvaardigheid meer dan één beoordelaar inschakelen.

Vervolgonderzoek Nederlands

- Onderzoek naar de praktijk van het fictie-onderwijs en de toetsing van dit onderdeel.
- Onderzoek naar spreek- en gespreksvaardigheden en de toetsing ervan (al dan niet in combinatie met fictie).
- Onderzoek naar de gewenste niveaus aangaande basale taalvaardigheden (spelling, woordenschat, grammatica, formuleervaardigheden).
- Onderzoek naar de mate waarin docenten nog zelfstandig toetsen (durven) maken en de mate waarin zij daarin scholing behoeven.
- Onderzoek naar het effect van toetsfrequentie op het leerlingenresultaat.
- Onderzoek naar een juist evenwicht van de verdeling van de examenstof over CSE en SE.
- Onderzoek naar nauwkeurige niveaubeschrijvingen van de afzonderlijke vaardigheden.
- Onderzoek naar de toetsing van onderbelichte subvaardigheden als reviseren bij schrijven en het gebruik van ICT.
- Onderzoek naar een gewenste canon voor taalvaardigheden per leerweg (dus per niveau)
- Detailstudie naar de kwaliteit van toetsen van gerenommeerde methoden (veel docenten hebben een blindelings vertrouwen in de kwaliteit van deze toetsen).

3.3 De kwaliteit van het SE biologie

3.3.1 Verschillen tussen scholen

In hoeverre verschillen scholen in de kwaliteit van het SE biologie? Het antwoord op deze vraag wordt bemoeilijkt door het grote aantal indicatoren met zeer veel ontbrekende waarnemingen. Rapportage van de verschillen tussen scholen op basis van de somscore over alle 19 indicatoren is niet zinvol. Als we het criterium arbitrair op hooguit 20% ontbrekende waarnemingen stellen, resteren voor het berekenen van een somscore tien bruikbare indicatoren (d.w.z. indicator 3, 4, 5, 6, 8, 10, 11, 12, 13 en 15). Van een school ontbraken alle tien oordelen en van een andere school ontbraken er acht. Deze twee scholen zijn niet in de berekening van de somscore meegenomen. Van de resterende 69 scholen waren voor zes of meer indicatoren oordelen beschikbaar.

De kwaliteit van het SE biologie is op gemiddeld 9.58 (96%) van de tien indicatoren als voldoende beoordeeld (met een standaard deviatie van .97). De verdeling van de oordelen is weergegeven in Tabel 19. De school met het "slechtste" SE kreeg voor vier indicatoren een voldoende; de 54 (78%) scholen met het "beste" SE behaalden de maximaal mogelijke score van tien.

Tabel 19 Verdeling van de scholen naar de algemene kwaliteit van het SE biologie

Aantal indicatoren voldoende	Aantal scholen	Percentage	Valide percentage
4	1	1	1
7	1	1	1
8	6	8	9
9	7	10	10
10	54	76	78
Valide	69	97	100
Missing	2	3	
Totaal	71	100	

3.3.2 Algemene kwaliteit per indicator

Het percentage onvoldoende en voldoende oordelen is per indicator gepresenteerd in Tabel 20. In deze tabel is het aantal valide oordelen als percentagebasis gehanteerd (dit wil zeggen dat de ontbrekende waarnemingen buiten de berekening zijn gehouden). De vijf indicatoren waarvoor de kwaliteit van het SE biologie het vaakst als voldoende beoordeeld werd en die derhalve het minst voor verbetering in aanmerking komen, zijn (waarbij indicatoren met meer dan 20% ontbrekende waarnemingen buiten beschouwing zijn gelaten):

- Cijfers voor toetsen en opdrachten (0% onvoldoende)
- Gebruik antwoord- c.q. beoordelingsmodel (0% onvoldoende)
- Scorepunten en/of cesuur vooraf vastgelegd 0% onvoldoende)
- Beoordelingsaspecten relevant (1% onvoldoende)
- Afnamemoment gerelateerd aan niveau (1% onvoldoende)

De vijf indicatoren waarvoor de kwaliteit van het SE biologie het vaakst als onvoldoende beoordeeld werd zijn en die mogelijk het meest voor verbetering in aanmerking komen, zijn (waarbij indicatoren met meer dan 20% ontbrekende waarnemingen buiten beschouwing zijn gelaten):

- Beoordelingsaspecten volledig (2% onvoldoende)
- Vergelijkbaarheid opgaven (2% onvoldoende)
- Becijfering kandidaten (9% onvoldoende)
- Soepelheid strengheid beoordelingsmiddelen (9% onvoldoende)
- Moeilijkheid toetsen en opdrachten (17% onvoldoende)

Tabel 20 Percentage onvoldoende en voldoende oordelen per indicator (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen voor biologie

Indicator	Onvoldoende	Voldoende	Te gemakkelijk of te soepel	Te moeilijk of te streng	% missing
1 Verplichte onderdelen SE	4	96			32
2 Conform niveauspecificaties		100			97
3 Afnamemoment gerelateerd aan niveau	1	99			6
4 Moeilijkheid toetsen en opdrachten		83	17		7
5 Soepelheid strengheid beoordelingsmiddelen		91	9		4
6 Becijfering kandidaten		91	9		3
7 Conform producteisen	7	93			37
8 Vergelijkbaarheid opgaven	2	98			11
9 Verplichte toetsinhoud in overleg	6	94			30
10 Beoordelingsaspecten relevant	1	99			4
11 Beoordelingsaspecten volledig	2	98			7
12 Gebruik antwoord- c.q. beoordelingsmodel		100			6
13 Cijfers voor toetsen en opdrachten		100			4
14 Beoordeling in overleg		100			96
15 Scorepunten en/of cesuur vooraf vastgelegd		100			15
16 Scorepunten vroegtijdig meegedeeld	36	64			45
17 Bijstelling scorepunten cesuur beargumenteerd		100			90
18 Normering in overleg		100			86
19 Gewicht		100			38

3.3.3 Kwaliteit per vakonderdeel

Aanvullende indicatoren

Bij het vak biologie zijn behalve de negentien algemene indicatoren nog drie indicatoren beoordeeld, te weten of de vaardigheden op niveau getoetst worden, of het SE voldoet aan de eisen van het nieuwe examenprogramma in havo 4 en of het SE-onderdeel Algemene Natuur Wetenschappen (ANW) van voldoende kwaliteit is. De oordelen van de beoordelaars zijn samengevat in Tabel 21. De percentages ontbrekende waarneming zijn in alle drie gevallen groter dan 20%, zodat we moeten afzien van verdere rapportage. Het percentage ontbrekende waarnemingen is zo hoog omdat veel scholen geen praktische opdrachten hebben opgestuurd en de kerndoelen van ANW naar het oordeel van de beoordelaars op vrijwel geen enkele school expliciet en afzonderlijk getoetst worden.

Tabel 21 *Kwaliteitsoordelen voor drie aanvullende indicatoren biologie (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen*

Indicator	Onvoldoende	Voldoende	Te gemakkelijk of te soepel	Te moeilijk of te streng	% missing
1 Toetsing vaardigheden op niveau		100			55
2 Nieuwe examenprogramma in havo 4		100			66
3 Algemene Natuur Wetenschappen (ANW)	64	36			65

Kwaliteit toetsing met schriftelijke dan wel praktische opdrachten

De twee belangrijkste toetsvormen bij biologie zijn schriftelijke toetsen en praktische opdrachten. De uitkomsten van de beoordeling zijn weergegeven in Tabel 22 en 23. Te zien is dat de percentages ontbrekende waarnemingen zonder uitzondering zeer hoog zijn (mede als gevolg van het veelvuldig ontbreken van vooral de praktische opdrachten). Om deze reden zien we af van verdere bespreking van de resultaten.

Tabel 22 *Beoordeling van de kwaliteit van schriftelijke toetsing biologie (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen*

Indicator	Onvoldoende	Voldoende	Te gemakkelijk of te soepel	Te moeilijk of te streng	% missing
1 Verplichte onderdelen SE					100
2 Conform niveauspecificaties		100			75
3 Afnamemoment gerelateerd aan niveau		100			73
4 Moeilijkheid toetsen en opdrachten		95	5		72
5 Soepelheid strengheid beoordelingsmiddelen		89	11		73
6 Becijfering kandidaten		95	5		72
7 Conform producteisen		100			99
8 Vergelijkbaarheid opgaven		100			76
9 Verplichte toetsinhoud in overleg		100			80
10 Beoordelingsaspecten relevant	6	94			75
11 Beoordelingsaspecten volledig	11	89			75
12 Gebruik antwoord- c.q. beoordelingsmodel		100			76
13 Cijfers voor toetsen en opdrachten		100			73
14 Beoordeling in overleg					100
15 Scorepunten en/of cesuur vooraf vastgelegd	11	89			87
16 Scorepunten vroegtijdig meegedeeld	13	88			89
17 Bijstelling scorepunten cesuur beargumenteerd		100			92
18 Normering in overleg		100			93
19 Gewicht					100

Tabel 23 De kwaliteit van de toetsing met praktische opdrachten biologie (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen

Indicator	Onvoldoende	Voldoende	Te makkelijk of te soepel	Te moeilijk of te streng	% missing
1					100
2		100			94
3		100			94
4		75	25		94
5		80	20		93
6		100			92
7					100
8		100			94
9		100			96
10	20	80			93
11	20	80			93
12		100			93
13	13	88			89
14					100
15		100			92
16		100			92
17		100			96
18		100			96
19					100

3.3.4 Voorbeelden examenpraktijk

De vragenlijst *Inventarisatie Examenpraktijk* is ook voorgelegd aan de beoordelaars biologie. In deze vragenlijst noteerden de beoordelaars zowel voorbeelden van goede als minder goede examenpraktijken. Onderstaand overzicht bevat een bloemlezing uit hun antwoorden. Daarbij zijn de voorbeelden zoveel mogelijk in hun oorspronkelijke bewoording weergegeven. Opgemerkt wordt dat er niet voor alle indicatoren voorbeelden van goede of minder goede examenpraktijken gegeven zijn.

1 Verplichte onderdelen SE: Het SE voldoet aan de inhoudelijke specificaties uit het eindexamenprogramma

Voorbeelden minder goede examenpraktijk

- Er was geen, door de school geformuleerd, PTA aanwezig
- Er was wel een deel van het PTA, maar alleen voor H4 of alleen voor H5
- In veel gevallen is dit niet te controleren omdat er geen volledig PTA is bijgevoegd.
- De enquêtes zijn meestal wel ingevuld maar een duidelijke omschrijving van de onderdelen die in bepaalde toetsen aan bod komen is niet voorhanden.
- Doordat van veel scholen het materiaal in 2 porties binnen is gekomen, kan dit aspect moeilijk worden beoordeeld
- Bij een flink aantal scholen heb ik het aandeel van het praktisch werk niet terug kunnen vinden

Voorbeelden van voldoende of goede examenpraktijk

- Volledig PTA, en door de vakdocent een toelichting gegeven over hoe wij als beoordelaars dit dienden te interpreteren.
- Ik heb van goed beoordeelde scholen niet speciaal genoteerd waar dit aan lag. Het materiaal is al teruggestuurd, dus kan ik hier helaas weinig mee
- Voorbeelden van scholen die het PTA, met ingestuurd werk goed voor elkaar hebben zijn school A, B, C, D, E F en G (NB: geanonimiseerd).

2 Conform niveau specificaties: Het SE voldoet aan de niveau specificaties uit het eindexamenprogramma

Voorbeelden minder goede examenpraktijk

- Het is op veel scholen niet duidelijk hoe het praktische schoolexamen in het geheel wordt beoordeeld, omdat er geen leerlingenwerk is meegestuurd en ook niet duidelijk is welke eisen er aan gesteld worden. Soms is alleen een percentage van het eindcijfer bekend.
- Sommige scholen gebruiken alleen toetsmateriaal uit de toetsenbundel die bij de methode wordt geleverd. Soms is dit nogal beperkt en erg letterlijk uit de lesmethode.
- Soms komen in de schriftelijke toetsen heel oude meerkeuzeopgaven voor over onderwerpen die niet echt meer bij de stof horen

3 Afnamemoment gerelateerd aan niveau: Het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben

Voorbeelden minder goede examenpraktijk

- In het geval van praktische opdrachten maakte men gebruik van practica die je ook in de onderbouw kunt geven.
- Bij biologie is niet te beoordelen op welk moment een leerling een bepaald niveau heeft. Dit hangt af van de volgorde waarin de diverse onderdelen in de klas worden behandeld.
- In de meeste gevallen zien we een parallelle opbouw van delen van het boek en daaraan gekoppelde schriftelijke toetsen
- Op diverse scholen wordt in het examenjaar ook stof uit 4 havo als repetitie getoetst in 5 havo.

Voorbeelden van voldoende of goede examenpraktijk

- Opbouw van toetsing, zoals hierboven beschreven
- In examenjaar thematische toetsing van stof, dus niet volgens hoofdstukken uit het boek

4 Moeilijkheid toetsen en opdrachten: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Voorbeelden minder goede examenpraktijk

- Veel te simpel voor 5 HAVO
- Leerlingen kregen veel tussendoor overhoringen, die alleen een beroep deden op reproductie, maar wel in het PTA zijn opgenomen.
- De praktische opdrachten waren eigenlijk simpele éénuurs practica. Sommige kunnen ook in een tweede klas gegeven worden.
- Soms worden in 4 Havo kleine schriftelijke werkjes meegeteld voor het schoolexamen
- Soms worden alleen de meegeleverde toetsen bij de methode gebruikt voor toetsing
- In een enkel geval worden alleen oude meerkeuzevragen bij elkaar gescharreld (soms wel 30 jaar oud) die bij elkaar geplakt een schoolexamen vormen.
- Alleen gebruik van zeer recente eindexamenvragen in schoolexamens in 4 Havo. Deze zijn tamelijk eenvoudig op internet te vinden en voor 4 havo lastig te maken

Voorbeelden van voldoende of goede examenpraktijk

- Docenten hebben er naar gestreefd een voor de leerlingen zo aantrekkelijk mogelijk SE te maken en niet klakkeloos de toetsen uit de bekende methoden te kopiëren.
- Praktische opdrachten over diergedrag moesten door de leerlingen in de dierentuin worden uitgevoerd. Samen met de educatieve dienst van de dierentuin had de docent een voor de leerlingen uitgebalanceerde praktische opdracht gemaakt.

- Mooie toetsen van school X in Y (NB: geanonimiseerd). De schriftelijke toetsen bestaan uit 4 delen, met gevarieerde aspecten van toetsing. Eerst korte antwoordvragen, dan informatievaardigheden en twee blokken met complexere vraagstelling, deels meerkeuze en deels open vragen. Alles voorzien van aantal toe te kennen punten.
- Veel scholen gebruiken een mix van vragen uit de bij de methode behorende toetsenbundel, aangevuld met recente of wat oudere examenvragen. Bijvoorbeeld school A en B (vraag 1).

5 Soepelheid strengheid beoordelingsmiddelen: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Voorbeelden minder goede examenpraktijk

- Er worden antwoorden goed gerekend die fundamenteel onjuist waren.
- Bij sommige toetsen kreeg ik de indruk, dat de leerlingen over een antwoord doorzeurden, en de docent het dan alsnog goed rekende. Leerlingen schreven er dan bij, dat er 'na discussie' bijv. 2 punten bij moesten. Het cijfer werd inderdaad veranderd.
- Bij enkele scholen worden als praktische opdracht leerlingverslagen bijgevoegd die om onduidelijke criteria vrijwel allemaal een 8 krijgen voor minimaal werk, b.v. enkele tekeningen van preparaten. De status van deze 8 bij berekening eindcijfer SE is onduidelijk
- Soms wordt zelfs het invullen van een aantal vragen bij een videoband als praktisch werk opgevoerd.

Voorbeelden van voldoende of goede examenpraktijk

- Op school X krijgen leerlingen een heldere instructie bij opdracht over moderne medische technieken. Ook het beoordelingssjabloon wordt aan de leerling meegegeven.
- Soortgelijk op school Y in Z.
- Op diverse scholen doen leerlingen gedragsonderzoek met tevoren een beoordelingsblad met beoordelingsaspecten.

6 Becijfering kandidaten: De toegekende cijfers zijn noch te hoog noch te laag

Voorbeelden minder goede examenpraktijk

- Scholen volgen de normering die bij de proefwerken wordt geleverd.
- Men koos niet voor $((\text{score}/\text{totaal}) \times 9) + 1$, maar een variant waardoor de leerling veel hoger uitkomen dan ze op het examen zouden scoren.
- In enkele gevallen wordt een schriftelijke overhoring van een laag niveau meegeteld.
- Soms zijn de vragen van met name de eerste werken van 4 Havo van een te laag niveau

Voorbeelden van voldoende of goede examenpraktijk

- Zich houden aan de wijze van becijfering die ook op het examen gehanteerd wordt.
- Bij praktische opdrachten is het vaak moeilijk om tot een evenwichtige becijfering te komen. In sommige toelichtingen bij de praktische opdrachten gaven docenten voor de leerlingen zeer duidelijk aan hoe er beoordeeld zou worden. Hield men zich hier niet aan, dan waren de consequenties voor de leerlingen.
- Bij veel scholen wordt een berekening gegeven van de normering en de omzetting van score naar cijfer.

7 Conform producteisen: Het SE is zorgvuldig samengesteld en voldoet aan relevante producteisen

Voorbeelden minder goede examenpraktijk

- Men kopieert een toets uit de toetsbundel. Men wil niet alle vragen gebruiken. Men streept die vragen door en verandert de nummering. Vervolgens kopieert men weer, waardoor de kwaliteit van met name tekeningen zeer slecht wordt. Sommige leerlingen schreven dat dan ook op hun antwoordenblad.
- Als een antwoordenblad wordt meegeleverd dient dat zodanig samengesteld te zijn, dat 'elke' leerlingen ruimte heeft om het antwoord te formuleren. Dit was niet altijd het geval, zodat knoeiwerk onvermijdelijk werd.
- Bijna nooit volledig omdat in veel gevallen praktische opdrachten misten
- Soms bijna 100 % meerkeuze (>5 x)

- Soms voornamelijk weetjes, weinig inzicht (> 2 x)
- Onduidelijke vragen (2 x)
- De verzorging van het werk dat leerlingen onder ogen krijgen, varieert van een haast perfecte layout tot haastig plak-knip-kopieerwerk waarvan delen slecht leesbaar en kriskras door elkaar op het papier geplakt. Zo is school A in B wel erg slordig en gebruikt ook bijna alleen meerkeuzevragen

Voorbeelden van voldoende of goede examenpraktijk

- De toets was voorzien van een 'aantrekkelijk' voorblad, passend bij het onderwerp van de toets en op de eerste pagina duidelijke instructies voor de leerlingen over wat de docenten van de toets verwachtten.
- School A. Erg compleet met toetsen, tussentoetsen en Schoolexamenwerk Ook de instructie bij de praktische opdrachten.
- School B. Goede toetsen voor schoolexamen en praktisch werk met beoordelingen vooraf. Het werk werd ook digitaal per CD-rom aangeleverd.
- Er zijn nog diverse andere scholen met erg mooi verzorgd leerlingmateriaal.

8 Vergelijkbaarheid opgaven: Als met toetsvarianten wordt gewerkt, is de moeilijkheidsgraad ervan vergelijkbaar

Voorbeelden minder goede examenpraktijk

- Bij veel methoden zijn A- en B-versies. Veel docenten gebruiken de tweede versie voor de herkansingen. Vaak is die B-versie zo vergelijkbaar met de A-versie, dat de leerlingen in de herkansing duidelijk in het voordeel zijn.
- Als er toetsen in meerdere varianten werden aangeleverd, dan was dit werk doorgaans goed vergelijkbaar

9 Verplichte toetsinhoud in overleg: De vakcollega's stellen de inhoud van het SE in overleg vast

Voorbeelden minder goede examenpraktijk

- Op scholen (dependances) waar slechtst één docent aanwezig is, is dit natuurlijk niet mogelijk. In een geval had school slechts een docent
- Nogal wat scholen hebben hun klassen/groepen zo verdeeld dat er maar een docent op een klasselaag werkt. Door de vraagstelling is niet terug te vinden of er overleg is geweest. In de enquête geven de meermanssecties vrijwel allemaal aan in overleg te werken

10 Beoordelingsaspecten relevant: Alle beoordelingsaspecten zijn relevant

Voorbeelden minder goede examenpraktijk

- Bij diverse scholen heb ik geen beeld kunnen vormen over het onderdeel Praktisch Schoolexamen. Meestal kwam dit door het ontbreken van instructiemateriaal en/of leerlingwerk.
- In enkele gevallen is het niveau bedenkelijk met twee microscopische practicumlessen en een enkele video-kijkopdracht.

Voorbeelden van voldoende of goede examenpraktijk

- Voorbeelden te over!!

11 Beoordelingsaspecten volledig: Alle relevante beoordelingsaspecten worden in de beoordeling betrokken

Voorbeelden minder goede examenpraktijk

- Bij de meeste scholen is "volledigheid" niet te controleren omdat er delen van het materiaal niet zijn opgestuurd (of niet bij mij zijn aangekomen). Het praktische deel ontbreekt het meest.

12 Gebruik antwoord- c.q. beoordelingsmodel: De beoordeling vindt plaats met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan

Voorbeelden minder goede examenpraktijk

- Meestal wel, maar het is mij niet duidelijk op welk moment dit moment is gemaakt; vooraf aan de toets of pas bij het nakijken ervan

Voorbeelden van voldoende of goede examenpraktijk

- School A te B. Hier staan op het werk de beoordelingsnorm en de toe te kennen punten.
- School C te D idem
- Er zijn veel scholen waar dit goed voor elkaar is

13 Cijfers voor toetsen en opdrachten: Voor de toetsen en handelingsdelen worden cijfers respectievelijk de kwalificaties “naar behoren” of “onvoldoende” toegekend

14 Beoordeling in overleg: De beoordeling geschiedt door tenminste twee docenten

Voorbeelden minder goede examenpraktijk

- Op scholen (dependances) waar slechtst één docent aanwezig is, is dit natuurlijk niet mogelijk.
- Dit is echt uit geen toets of vraag op te maken. Dat kun je dus nooit weten: allemaal 9 ingevuld
- Niet controleerbaar

Voorbeelden van voldoende of goede examenpraktijk

- Vrijwel alle scholen beoordelen met cijfers
- Eén school met een eenmanssectie had overleg met een sectie van een andere school

15 Scorepunten en/of cesuur vooraf vastgelegd: De scorepunten en indien van toepassing ook de cesuur worden vóór de afname van het toetsmateriaal schriftelijk vastgesteld

Voorbeelden minder goede examenpraktijk

- Dit is nu achteraf al helemaal niet meer te controleren. Dit zou als vraag in de enquête moeten worden opgenomen voordat al het materiaal wordt opgevraagd

Voorbeelden van voldoende of goede examenpraktijk

- School A te B. Hier staan op het werk de beoordelingsnorm en de toe te kennen punten
- School C te D idem.
- Er zijn veel scholen waar dit goed voor elkaar is

16 Scorepunten vroegtijdig meegedeeld: De scorepunten worden vóór de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt

Voorbeelden minder goede examenpraktijk

- Veel afwijking van ingevulde formulier: Wordt ja gezegd, maar op toetsen niet te vinden
- zie 15
- Scholen werken met verschillende tabellen: Veelal schaal tussen 100 en 10 punten of tussen 100 en 0 punten. Er wordt ook nogal eens gewerkt met een omrekenformule. Deze is niet altijd vanzelfsprekend

Voorbeelden van voldoende of goede examenpraktijk

- Bij een aantal scholen wordt voor de vraag het aantal punten vermeld in de schoolexamen
- Op school D wordt bij de praktische opdracht per aspect tevoren het aantal punten vermeld. Ook bij andere scholen komt dit voor.

17 Bijstelling scorepunten cesuur beargumenteerd: **De scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd**

Voorbeelden van goede of minder goede examenpraktijk

- Dit is echt uit geen toets of vraag op te maken. Dat kun je dus nooit weten: allemaal 9 ingevuld.
- Regelmatig is er leerlingenwerk met 2 cijfers erboven. Het is dan niet duidelijk of er door een docent een nieuwe norm is gehanteerd of door leerlingen na inzage puntjes erbij “gesprokkeld”.
- Een enkele keer wordt bij het correctiemodel aangegeven dat er door een bepaalde omstandigheid (bv. te weinig tijd) een aanpassing van de norm heeft plaatsgevonden.

18 Normering in overleg: De vakcollega’s bepalen de normering in onderling overleg

Voorbeelden van goede of minder goede examenpraktijk

- Dit is echt uit geen toets of vraag op te maken. Dat kun je dus nooit weten: allemaal 9 ingevuld.
- Dit kan ik meestal met mijn data niet verifiëren.
- Hier zou expliciet in de enquête naar gevraagd moeten worden bij een vervolgonderzoek.

19 Gewicht: Het gewicht bij de bepaling van het SE cijfer is voor elke toets aangegeven

Voorbeelden minder goede examenpraktijk

- Vaak niet te scoren want vaak geen PTA aanwezig.
- Omdat er vrijwel nooit een volledig PTA is bijgevoegd is dit niet helder te controleren.

Voorbeelden van voldoende of goede examenpraktijk

- Mij lijkt dat veel scholen dit prima voor elkaar hebben. Er is dan in PTA of op het werk precies vermeld hoe zwaar een onderdeel meeweegt.
- Bijvoorbeeld school A te B heeft erg volledig en goed materiaal ontwikkeld
- Ook school C te D heeft de theoretische toetsing helder voor elkaar. Hier ontbreekt het praktische deel.
- Relatief veel scholen geven naar de leerlingen toe duidelijkheid over de beoordelingsnormen.

3.3.5 Conclusie en aanbevelingen

Conclusies

Een eerste conclusie is dat de beoordelaars de kwaliteit van het SE biologie gemiddeld bij 96% van de beoordelingen als voldoende hebben aangemerkt (waarbij negen indicatoren vanwege een te groot aantal ontbrekende oordelen buiten beschouwing zijn gelaten).

Een tweede conclusie is dat scholen volgens de beoordelaars van elkaar verschillen in de kwaliteit van het SE biologie. Zo scoorde de school met het “slechtste” SE biologie op vier van de tien bruikbare indicatoren een voldoende; van de 54 scholen (78%) met het “beste” SE biologie was de kwaliteit voor alle tien indicatoren in orde.

Een derde conclusie is dat de kwaliteit van het SE biologie naar de mening van de beoordelaars niet voor alle tien indicatoren even hoog is. De vijf “beste” indicatoren waren Cijfers voor toetsen en opdrachten (0% onvoldoende), Gebruik antwoord- c.q. beoordelingsmodel (0% onvoldoende), Scorepunten en/of cesuur vooraf vastgelegd (0% onvoldoende), Beoordelingsaspecten relevant (1% onvoldoende) en Afnamemoment gerelateerd aan niveau (1% onvoldoende). De vijf indicatoren die mogelijk het meest voor verbetering in aanmerking komen, zijn Beoordelingsaspecten volledig (2% onvoldoende), Vergelijkbaarheid opgaven (2% onvoldoende), Becijfering kandidaten (9% onvoldoende), Soepelheid strengheid beoordelingsmiddelen (9% onvoldoende) en Moeilijkheid toetsen en opdrachten (17% onvoldoende).

Over de mate waarin de aanvullende indicatoren voor de kwaliteit van het SE biologie gerealiseerd zijn, kan vanwege het grote percentage ontbrekende waarnemingen geen uitspraak worden gedaan.

Aanbevelingen

Op grond van deze studie kunnen twee aanbevelingen voor eventueel vervolgonderzoek worden gedaan:

- Controleer vóór het beoordelen of het materiaal volledig is. Als er toch onverhoopt onvolledig materiaal beoordeeld moet worden, laat de beoordelaar dan aangeven op welk materiaal het oordeel gebaseerd is (bijvoorbeeld PTA, vragenlijst, aantal toetsen Havo 4, aantal toetsen Havo 4, scorings- en beoordelingsmodel, leerlingenwerken, schriftelijke opdracht, praktische opdracht).
- Gezien de specificiteit van de praktische opdrachten bij de natuurwetenschappelijke vakken in het algemeen en die van biologie in het bijzonder verdient het aanbeveling voor de beoordeling van de kwaliteit van de praktische opdrachten een aanvullende set indicatoren te ontwikkelen. De leerlingen in de N&G en N&T profielen kiezen vaak een vervolgopleiding in deze richting en dan is een goede voorbereiding op dit natuurwetenschappelijk onderzoek onontbeerlijk.

3.4 De kwaliteit van het SE wiskunde

3.4.1 Verschillen tussen scholen

In hoeverre verschillen scholen in de kwaliteit van het SE wiskunde? Het antwoord op deze vraag wordt bemoeilijkt door het grote aantal indicatoren met zeer veel ontbrekende waarnemingen. Rapportage van de verschillen tussen scholen op basis van de somscore over alle negentien indicatoren is niet zinvol. Als we het criterium arbitrair op hooguit 20% ontbrekende waarnemingen stellen, resteren voor het berekenen van een somscore elf bruikbare indicatoren (d.w.z. indicator 3, 4, 5, 6, 7, 10, 11, 12, 13, 15 en 16). Van een school ontbraken de oordelen voor alle elf indicatoren waardoor we deze school niet in de berekening van de somscore hebben meegenomen. Van de resterende 68 scholen waren er voor ten minste acht indicatoren oordelen beschikbaar. De verdeling van de oordelen is weergegeven in Tabel 24. Van de scholen is de kwaliteit van het SE wiskunde op gemiddeld 10.34 (94%) van de 11 bruikbare indicatoren als voldoende beoordeeld (met een standaard deviatie van 1.16). De school met het "slechtste" SE wiskunde kreeg voor vier van de elf indicatoren een voldoende beoordeling; de 39 (57%) scholen met het "beste" SE kregen voor alle elf indicatoren het oordeel "voldoende".

Tabel 24 Verdeling van de scholen naar de algemene kwaliteit van het SE wiskunde

Aantal indicatoren voldoende	Aantal scholen	Percentage	Valide percentage
4	1	1	1
7	2	3	3
9	2	3	3
10	24	35	35
11	39	57	57
Valide	68	99	100
Missing	1	1	
Totaal	69	100	

3.4.2 Algemene kwaliteit per indicator

De uitkomsten van de beoordeling van de algemene kwaliteit van het SE wiskunde zijn weergegeven in Tabel 25. De vijf indicatoren waarvoor de kwaliteit van het SE wiskunde het vaakst als voldoende beoordeeld werd en die derhalve het minst voor verbetering in aanmerking komen, zijn (waarbij indicatoren met meer dan 20% ontbrekende waarnemingen buiten beschouwing zijn gelaten):

- Afnamemoment gerelateerd aan niveau (0% onvoldoende)
- Beoordelingsaspecten relevant (1% onvoldoende)
- Beoordelingsaspecten volledig (1% onvoldoende)
- Soepelheid strengheid beoordelingsmiddelen (3% onvoldoende)
- Bcijfering kandidaten (3% onvoldoende)

De vijf indicatoren waarvoor de kwaliteit van het SE wiskunde het vaakst als onvoldoende beoordeeld werd en die mogelijk het meest voor verbetering in aanmerking komen, zijn (waarbij indicatoren met meer dan 20% ontbrekende waarnemingen buiten beschouwing zijn gelaten):

- Scorepunten en/of cesuur vooraf vastgelegd (3% onvoldoende)
- Moeilijkheid toetsen en opdrachten (4% onvoldoende)
- Gebruik antwoord- c.q. beoordelingsmodel (4% onvoldoende)
- Cijfers voor toetsen en opdrachten (4% onvoldoende)
- Scorepunten vroegtijdig meegedeeld (36% onvoldoende)

Tabel 25 Percentage onvoldoende en voldoende oordelen per indicator (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen voor wiskunde

Indicator	Onvoldoende	Voldoende	Te makkelijk of te soepel	Te moeilijk of te streng	% missing
1 Verplichte onderdelen SE		100			62
2 Conform niveauspecificaties	2	98			23
3 Afnamemoment gerelateerd aan niveau		100			3
4 Moeilijkheid toetsen en opdrachten		96	4		1
5 Soepelheid strengheid beoordelingsmiddelen		97	3		1
6 Becijfering kandidaten		97	3		6
7 Conform producteisen	3	97			9
8 Vergelijkbaarheid opgaven		100			35
9 Verplichte toetsinhoud in overleg	50	50			57
10 Beoordelingsaspecten relevant	1	99			3
11 Beoordelingsaspecten volledig	1	99			3
12 Gebruik antwoord- c.q. beoordelingsmodel	4	96			1
13 Cijfers voor toetsen en opdrachten	4	96			1
14 Beoordeling in overleg	64	36			64
15 Scorepunten en/of cesuur vooraf vastgelegd	3	97			12
16 Scorepunten vroegtijdig meegedeeld	36	64			4
17 Bijstelling scorepunten cesuur beargumenteerd	40	60			86
18 Normering in overleg	62	38			62
19 Gewicht	15	85			33

3.4.3 Kwaliteit per vakonderdeel

De beoordelaars hebben naast de negentien algemene kwaliteitskenmerken ook de kwaliteit per vakonderdeel beoordeeld, te weten a) Algebraïsche verbanden, b) Rekenen, meten en schatten, c) Meetkunde en d) Statistiek. De uitkomsten zijn weergegeven in Tabel 26. Het percentage ontbrekende waarnemingen is met 6% relatief laag. De lage percentages doet vermoeden dat de beoordelaars de kwaliteit per vakonderdeel altijd beoordeeld te hebben (voorzover het benodigde materiaal aanwezig en de indicator beoordeelbaar was), dus ook als de *quick scan* geen aanleiding gaf tot twijfel aan de algemene kwaliteit. Hoe het ook zij, de beoordelaars hebben de kwaliteit van het SE voor de drie vakonderdelen Algebraïsche verbanden, Rekenen, meten en schatten en Meetkunde vrijwel zonder uitzondering als voldoende beoordeeld. De kwaliteit van het SE-onderdeel Statistiek is met 22% onvoldoendes minder goed beoordeeld.

Tabel 26 Percentage onvoldoende en voldoende oordelen per indicator (rijpercentages optellend tot 100%) en het percentage ontbrekende waarnemingen voor wiskunde

Indicator	Onvoldoende	Voldoende	% missing
1 Algebraïsche verbanden	2	98	6
2 Rekenen, meten en schatten	3	97	6
3 Meetkunde	3	97	6
4 Statistiek	22	78	6

3.4.4 Voorbeelden examenpraktijk

De vragenlijst *Inventarisatie Examenpraktijk* is ook voorgelegd aan de beoordelaars wiskunde. In deze vragenlijst noteerden de beoordelaars zowel voorbeelden van goede als minder goede examenpraktijken. Onderstaand overzicht bevat een bloemlezing uit hun antwoorden. Daarbij zijn de voorbeelden zoveel mogelijk in hun oorspronkelijke bewoording weergegeven. Opgemerkt wordt dat er niet voor alle indicatoren voorbeelden gegeven zijn en dat de beoordelaars wiskunde geen duidelijk onderscheid hebben gemaakt tussen goede en minder goede examenpraktijk.

1 Verplichte onderdelen SE: Het SE voldoet aan de inhoudelijke specificaties uit het eindexamenprogramma

In een enkel geval ontbrak verrijking.
Niet voor alle vakonderdelen is materiaal aangeleverd. Vooral materiaal over statistiek ontbreekt.

2 Conform niveauspecificaties: Het SE voldoet aan de niveauspecificaties uit het eindexamenprogramma

Bijna alle scholen gebruiken materiaal uit de bundel. De meesten kopiëren eenvoudigweg de aangeleverde proefwerken en gebruiken die zeer waarschijnlijk elk jaar weer. Slechts weinigen nemen de moeite zelf te knippen en te plakken. Eigen werk is een zeldzaamheid. Deze indicator is niet te beoordelen omdat het PTA ontbreekt.

3 Afnamemoment gerelateerd aan niveau: Het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben

De kwaliteit van het SE is wat dit betreft overal in orde. Deze indicator is ook zonder PTA wel te beoordelen.

4 Moeilijkheid toetsen en opdrachten: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

De meeste scholen maken gebruik van materiaal van de uitgever. Dat is in orde. Een ROC uit het Westen van het land toetst op een zeer laag niveau. In een enkel geval is te veel gebruik gemaakt van toetsopgaven en/of diagnostische toetsen uit het boek. Schrift meetellen als onderdeel van het SE is te eenvoudig.

Toelichting van de toetsdeskundige WS: Om de leerlingen te motiveren tot het maken van huiswerk, zijn er docenten die aan het eind van een periode het schrift laten inleveren en daar een cijfer voor geven, dat dan voor een deel het schoolexamencijfer bepaalt.

5 Soepelheid strengheid beoordelingsmiddelen: De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

De meeste scholen (85%) gebruiken de correctiemodellen en de normering van de uitgever en hanteren deze als zodanig. Het niveau is dan wel goed, maar op den duur fraudegevoelig. Bij zelf samengestelde SE's komt meestal een eigen normering die over het algemeen acceptabel is.

6 Becijfering kandidaten: De toegekende cijfers zijn noch te hoog noch te laag

Scholen volgen de normering die bij de proefwerken wordt geleverd.

7 Conform producteisen: Het SE is zorgvuldig samengesteld en voldoet aan relevante producteisen

De proefwerkbundels zijn goed en de kwaliteit van de zelf uit verschillende proefwerken samengestelde toetsen zijn ook van voldoende kwaliteit. De toetstijd wordt vaak niet op de toets vermeld. Het werken met uit het boek gekopieerde opgaven is als onvoldoende beoordeeld.

8 Vergelijkbaarheid opgaven: Als met toetsvarianten wordt gewerkt, is de moeilijkheidsgraad ervan vergelijkbaar

Alle leerlingen krijgen meestal hetzelfde proefwerk. Als er varianten zijn, zijn deze vergelijkbaar (inhaalwerk en/of herkansing).

9 Verplichte toetsinhoud in overleg: De vakcollega's stellen de inhoud van het SE in overleg vast

Is niet te controleren en nauwelijks te beoordelen. De beoordelaars moeten afgaan op de vragenlijst. Wat de school hierover vermeld heeft, klopt vaak niet. De school zegt dat proefwerken worden samengesteld door docent(en), maar men maakt gewoon gebruik van een kopie van de bundel.

10 Beoordelingsaspecten relevant: Alle beoordelingsaspecten zijn relevant

Is bijna altijd zo.

11 Beoordelingsaspecten volledig: Alle relevante beoordelingsaspecten worden in de beoordeling betrokken

Is bijna altijd zo.

12 Gebruik antwoord- c.q. beoordelingsmodel: De beoordeling vindt plaats met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan

Als de school het correctiemodel van de bundel gebruikt, is de kwaliteit in orde.

13 Cijfers voor toetsen en opdrachten: Voor de toetsen en handelingsdelen worden cijfers respectievelijk de kwalificaties "naar behoren" of "onvoldoende" toegekend

Is altijd het geval. Cijfers geven is "standaard" bij wiskunde.

14 Beoordeling in overleg: De beoordeling geschiedt door tenminste twee docenten

Is niet uit het werk op te maken. Het zal in de praktijk niet voorkomen dat meer dan één docent het werk nakijkt. In een enkel geval zal er incidenteel onderling overleg zijn geweest, maar dit is door ons niet te beoordelen.

15 Scorepunten en/of cesuur vooraf vastgelegd: De scorepunten en indien van toepassing ook de cesuur worden vóór de afname van het toetsmateriaal schriftelijk vastgesteld

Bij het correctiemodel zijn ook de te behalen punten opgenomen. Scholen brengen daar geen veranderingen in aan en gebruiken het zoals het is.

Als dit niet op het werk staat, is niet te beoordelen of dit *van tevoren* gebeurd is. Een cesuur zal in de praktijk ook achteraf gebeuren.

16 Scorepunten vroegtijdig meegedeeld: De scorepunten worden vóór de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt

Dit is heel wisselend. Bij kopieën uit de bundel staan de te behalen punten vaak niet vermeld. Bij zelf samengesteld werk wordt het meestal wel gedaan.

Vaak niet, ongeveer vijftig procent.

17 Bijstelling scorepunten cesuur beargumenteerd: De scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd

Geen enkele keer gezien. Niet te beoordelen, wordt in ieder geval geen melding van gemaakt.

18 Normering in overleg: De vakcollega's bepalen de normering in onderling overleg

Bij zelf samengesteld werk is dat wel eens het geval. Meestal niet, men volgt de norm van de bundel. Is niet te beoordelen of dit in overleg gebeurd is.

19 Gewicht: Het gewicht bij de bepaling van het SE cijfer is voor elke toets aangegeven

Staat in het PTA en is alleen te controleren als het PTA is meegezonden. Is alleen te beoordelen als PTA aanwezig is of als het gewicht op de toets staat.

3.4.5 Conclusie en aanbevelingen

Conclusies

Een eerste conclusie is dat de beoordelaars de kwaliteit van het SE wiskunde gemiddeld bij 94% van de beoordelingen als voldoende hebben aangemerkt (waarbij acht indicatoren vanwege een te groot aantal ontbrekende oordelen buiten beschouwing zijn gelaten).

Een tweede conclusie is dat scholen naar het oordeel van de beoordelaars van elkaar verschillen in de kwaliteit van het SE wiskunde. Zo scoorde de school met het "slechtste" SE wiskunde op vier van de elf bruikbare indicatoren een voldoende; van de 39 scholen (57%) met het "beste" SE wiskunde was de kwaliteit voor alle elf indicatoren in orde.

Een derde conclusie is dat de kwaliteit van het SE wiskunde naar de mening van de beoordelaars niet voor alle elf indicatoren even hoog is. De vijf "beste" indicatoren waren Afnamemoment gerelateerd aan niveau (0% onvoldoende), Beoordelingsaspecten relevant (1% onvoldoende), Beoordelingsaspecten volledig (1% onvoldoende), Soepelheid strengheid beoordelingsmiddelen (3% onvoldoende) en Becijfering kandidaten (3% onvoldoende). De vijf indicatoren die mogelijk het meest voor verbetering in aanmerking komen, zijn Scorepunten en/of cesuur vooraf vastgelegd (3% onvoldoende), Moeilijkheid toetsen en opdrachten (4% onvoldoende), Gebruik antwoord- c.q. beoordelingsmodel (4% onvoldoende), Cijfers voor toetsen en opdrachten (4% onvoldoende) en Scorepunten vroegtijdig meegedeeld (36% onvoldoende).

Een vierde conclusie is dat de kwaliteit van het SE wiskunde niet voor elk vakonderdeel gelijk is. De beoordelaars hebben de kwaliteit van de vakonderdelen Algebraïsche verbanden, Rekenen, meten en schatten en Meetkunde vrijwel zonder uitzondering als voldoende aangemerkt (respectievelijk 98%, 97% en 97%). Op bijna ruim een vijfde (22%) van de scholen is het vakonderdeel Statistiek volgens de beoordelaars van onvoldoende kwaliteit.

Aanbevelingen

Op grond van deze studie kunnen de volgende aanbevelingen worden gedaan ter verbetering van de kwaliteit van het SE wiskunde:

- Laat docenten het werk onafhankelijk van elkaar met hetzelfde nakijkmodel nakijken.
- Verplicht docenten de lengte, het toetsmoment, het niveau (schooltype) en de normering van de toets op het opgavenblad aan te geven.
- Voorkom dat thuis gemaakt werk een onderdeel van het schoolexamen vormt (met uitzondering van het handelingsdeel).
- Voorkom dat opgaven uit het leerboek het schoolexamen te sterk bepalen. Stel een bovengrens op voor de hoeveelheid schoolexamenopgaven uit een bundel, zowel in totaal als per schoolexamen. De reden is de fraudegevoeligheid. Elke methode heeft een eigen invalshoek en manier van bevragen. Huiswerkinstituten beschikken over de bundels. Een andere reden is dat opgaven die direct uit het boek komen vooral toetsen of een leerling zijn huiswerk maakt, niet zozeer of hij het vak wiskunde beheerst. Juist als de leerling in staat is een niet methodegebonden vraag te beantwoorden weet je of hij dat onderdeel beheerst.

Voor eventueel vervolgonderzoek kunnen de volgende aanbevelingen worden gedaan:

- Onderzoek een kleinere steekproef dan nu is gedaan, maar zorg er wel voor dat het materiaal compleet is voordat de beoordelaars met de beoordeling beginnen.
- Betrek ook de cijferlijsten van alle schoolexamencijfers, het berekende eindcijfer en de aan de leerling meegegeven uiteindelijke cijferlijst in de beoordeling (waarin ook duidelijk de rekenwijze staat).
- Zorg ervoor dat het PTA van alle scholen beschikbaar is voor de beoordelaars.

4 Conclusies, discussie en aanbevelingen

4.1 Conclusies

De belangrijkste conclusie uit het onderzoek is dat de beoordelaars van oordeel zijn dat de vakinhoudelijke en toetstechnische kwaliteit van het SE op de overgrote meerderheid van de onderzochte scholen voor voortgezet onderwijs van voldoende kwaliteit is. Naar het oordeel van de beoordelaars voldoen de schoolexamens in het voortgezet onderwijs in voldoende mate aan redelijke eisen ten aanzien van de inhoudelijke dekking van het eindexamenprogramma, het moment van toetsing, de toetstechnische kwaliteit van de opgaven, de moeilijkheid van de toetsen, de strengheid/soepelheid van de beoordelingsmiddelen en de becijfering, de volledigheid en relevantie van de beoordelingsaspecten, het vooraf vastleggen en vroegtijdig meedelen van de scorepunten, het aangeven van het gewicht bij de bepaling van het cijfer en de samenwerking tussen docenten bij de toetsontwikkeling, beoordeling en normering.

Het vervolg van deze paragraaf geeft de belangrijkste conclusies per indicator van de kwaliteit van het SE. Zonder informatie over wat er beoordeeld is, is het geven van zuiver cijfermatige gegevens over de kwaliteit van het SE weinig zinvol. Om een indruk te geven van waar de beoordelaars bij het beoordelen op gelet hebben, zijn de conclusies gelardeerd met voorbeelden van examenpraktijken die de beoordelaars als voldoende of onvoldoende hebben aangemerkt. Deze voorbeelden zijn ontleend aan de toelichtingen die de beoordelaars gaven in de vragenlijst Inventarisatie Examenpraktijk.

De beoordelaars hebben niet voor alle vakken en indicatoren voorbeelden gegeven die zich lenen voor rapportage. Zoals eerder vermeld is de belangrijkste oorzaak het ontbreken van bronnenmateriaal. Vaak was niet voldaan aan een of meer van de volgende voorwaarden: a) het PTA moet beschikbaar zijn, b) er moet voldoende informatie over de indicator in het PTA te vinden zijn, c) de door de vakdocent ingevulde vragenlijst moet beschikbaar zijn, d) de vragenlijst moet voldoende informatie bevatten, e) de benodigde toetsen, opdrachten, scorings- en beoordelingsmiddelen en leerlingenwerken moeten beschikbaar zijn (voor alle leerjaren, leerwegen en vakonderdelen) en f) de docent moet de benodigde informatie over bijvoorbeeld de scorepunten, de cesuur en de weging op de toetsen en opdrachten vermeld hebben. De conclusies uit het onderzoek moeten dan ook met de grootst mogelijke voorzichtigheid en terughoudendheid geïnterpreteerd worden.

Verplichte onderdelen SE

Op 66% van de scholen voldoet het SE Engels naar het oordeel van de beoordelaars in voldoende mate aan de inhoudelijke specificaties uit het examenprogramma en van het SE Nederlands werd de kwaliteit voor deze indicator op alle onderzochte scholen als voldoende beoordeeld. Voor biologie en wiskunde waren er te veel ontbrekende waarnemingen om hierover een conclusie te kunnen trekken.

Voorbeelden van examenpraktijken die beoordelaars als onvoldoende hebben aangemerkt betreffen het te grote accent op woordenschat, spellen of idioom bij het toetsing van schrijfvaardigheid Engels, een andere invulling van gedocumenteerd schrijven Nederlands dan bedoeld door de opstellers van de eindtermen, het ontbreken van schrijven op basis van documentatie in het SE Nederlands of het ontbreken van verrijking in het SE wiskunde.

Voorbeelden van examenpraktijken die de beoordelaars als voldoende hebben aangemerkt zijn het gebruik van de luistertoetsen Engels van het Cito, het opnemen van schrijf- en luistervaardigheid in het SE Nederlands, het gebruik van toetsen die zich richten op één onderdeel van het vak Nederlands (zodat de docent over de beheersing van deze vaardigheden zuivere uitspraken kan doen) en de aanwezigheid van een volledig PTA biologie met een toelichting van de vakdocent over hoe de beoordelaars dit dienden te interpreteren.

Conform niveauspecificaties

Op 68% van de scholen voldoet het SE-Engels aan de niveauspecificaties uit het examenprogramma en bij Nederlands is dit percentage 100%. Bij biologie en wiskunde laat het hoge percentage ontbrekende waarnemingen geen cijfermatige conclusie toe.

Voorbeelden van “onvoldoende” examenpraktijken zijn volgens de beoordelaars een te sterk accent op grammatica, idioom en/of spelling bij de toetsing van schrijfvaardigheid Engels, het uitsluitend gebruiken van methodegebonden toetsmateriaal bij biologie en het gebruik van oude meerkeuzeopgaven biologie over onderwerpen die niet meer bij de stof horen. Een voorbeeld bij wiskunde is dat bijna alle scholen materiaal uit de bundel gebruiken. De meeste docenten wiskunde kopiëren eenvoudigweg de aangeleverde

proefwerken en gebruiken die waarschijnlijk elk jaar weer. Slechts weinigen nemen de moeite zelf te knippen en te plakken. Eigen werk is bij wiskunde een zeldzaamheid.

Een voorbeeld van een “voldoende” examenpraktijk is het gebruik van Cito-toetsen Engels (waarbij een beoordelaar opmerkte dat hij of zij de varianten die docenten op basis van deze Cito-toetsen zelf ontwikkelen ook als voldoende beoordeelde).

Afnamemoment gerelateerd aan niveau

Scholen blijken het toetsmateriaal vrijwel zonder uitzondering af te nemen op het moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben. De percentages voldoende voor de vakken Engels, Nederlands, biologie en wiskunde zijn respectievelijk 96%, 100%, 99% en 100%.

Een voorbeeld van een “onvoldoende” examenpraktijk is het alleen toetsen van schrijf- of luistervaardigheid Engels of fictie Nederlands in het vóórexamenjaar en niet ook in het laatste examenjaar. Voorbeelden bij biologie zijn het gebruik van te gemakkelijke practica die men ook in onderbouw zou kunnen geven en het als repetitie toetsen van stof uit 4-havo in 5-havo.

Moelijkheid toetsen en opdrachten

De moelijkheid van de toetsen en opdrachten van het SE is volgens de beoordelaars doorgaans in orde: de percentages voldoende voor de vakken Engels, Nederlands, biologie en wiskunde bedragen respectievelijk 89%, 100%, 83% en 96%.

Voorbeelden van examenpraktijken die de beoordelaars als onvoldoende aanmerkten, zijn een te ingewikkelde en te abstracte schrijfopdracht Engels, een te hoge moeilijkheidsgraad van een methodegebonden toets Nederlands, het toetsen van verschillende moeilijke bijzinnen in een SE Nederlands, te moeilijke dictees en grammaticatoetsen Nederlands en onbegrijpelijke vragen bij zelf gemaakte lees- en schrijfopdrachten Nederlands (die daardoor soms moeilijker werden dan nodig en bedoeld was). Voorbeelden uit het vakgebied biologie zijn het geven van te veel op reproductie gerichte tussentijdse overhoringen, het uitsluitend gebruiken van methodegebonden toetsen biologie, het gebruik van wel dertig jaar oude meerkeuze opgaven biologie en het gebruik van zeer recente eindexamenvragen biologie in 4-havo die tamelijk eenvoudig op het Internet te vinden zijn (en dus aanleiding kunnen geven tot examenfraude). Voorbeelden bij wiskunde zijn het gebruik van toetsmateriaal van de uitgever (waarvan de beoordelaar de kwaliteit dan automatisch als “in orde” beschouwde), een te sterk gebruik van toetsopgaven of diagnostische toetsen uit het boek en het meetellen van het schrift waardoor het examen te eenvoudig wordt.

Voorbeelden van qua moelijkheid passende toetsen of opdrachten zijn niet of nauwelijks gegeven. Veel lof oogstte een uitgebalanceerde praktische opdracht biologie over diergedrag die de docent samen met de educatieve dienst van een diertuin ontwikkeld had; de kandidaten voerden deze opdracht in de diertuin uit. Een andere prijzenswaardige opdracht biologie betrof een set van veertien schriftelijke toetsen, bestaande uit vier delen en gevarieerde aspecten van toetsing (eerst kort-antwoordvragen, dan informatievaardigheden en vervolgens twee blokken met een complexere vraagstelling, deels meerkeuze en deels open vragen).

Een voorbeeld van een qua moelijkheid passende opdracht Engels was een schrijfopdracht waarbij de kandidaat gevraagd werd een brief te schrijven om zichzelf te presenteren en te informeren naar vrijetijdsbesteding, stageplek en opleiding.

Soepelheid/strengheid beoordelingsmiddelen

De soepelheid of strengheid van de scorings- en beoordelingsmiddelen is naar het oordeel van de beoordelaars over het algemeen passend bij het niveau dat van de kandidaten verwacht mag worden. De percentages voldoende voor de vakken Engels, Nederlands, biologie en wiskunde bedragen bij deze indicator respectievelijk 86%, 100%, 91% en 97%.

Een voorbeeld van een te strenge scorings- of beoordelingsmiddel is de te strenge en eenzijdige beoordeling van grammaticale correctheid of spelling bij schrijfvaardigheid Engels (met veronachtzaming van opbouw en effectieve communicatie). Een voorbeeld uit het vakgebied biologie is het goed rekenen van antwoorden die fundamenteel onjuist waren, leerlingenverslagen die om onduidelijke redenen vrijwel alle een acht kregen voor minimaal werk (bijvoorbeeld enkele tekeningen van preparaten) en het als praktisch werk opvoeren van het invullen van een aantal vragen bij een videoband. Beoordelaars wiskunde gaven aan dat de meeste scholen (85%) de correctiemodellen en de normering van de uitgever hanteren. Het niveau was dan wel passend, maar omdat dit materiaal niet geheim was, kon het aanleiding geven tot examenfraude (bijvoorbeeld door huiswerkinstituten).

Voorbeelden van passende scorings- of beoordelingsmiddelen zijn onder meer het gebruik van Cito-toetsen Engels (die naar de mening van een beoordelaar objectief waren en bij schrijfvaardigheid een evenwichtige beoordeling toelieten). Een van de beoordelaars biologie noemden een voorbeeld waarbij leerlingen een heldere instructie bij een opdracht over moderne medische technieken kregen en waarbij het beoordelingssjabloon aan de kandidaten werd meegegeven. Op een andere school deden leerlingen gedragsonderzoek waarbij zij een beoordelingsblad met beoordelingsaspecten van tevoren meekregen.

Becijfering kandidaten

Over het algemeen zijn de toegekende cijfers volgens de beoordelaars noch te hoog en noch te laag. Het percentage scholen waarop de kwaliteit van het SE voor deze indicator in orde bevonden werd, bedroeg voor de vakken Engels, Nederlands, biologie en wiskunde respectievelijk 87%, 100%, 91% en 97%.

Gevraagd naar voorbeelden van een onvoldoende examenpraktijk merkte een van de beoordelaars Engels op dat er verbazend genoeg overal redelijke cijfers uit de bus kwamen en dat de beoordeling samen met de normering uiteindelijk op een voor de meesten acceptabel gemiddelde bleek uit te komen. Een beoordelaar Nederlands bracht naar voren dat scholen de normering volgen die bij de proefwerken wordt geleverd en dat vrijwel alle toetsen een normaal beeld laten zien, met een duidelijke spreiding van de cijfers, en een te verwachten verhouding tussen voldoende en onvoldoende. Bij biologie werd geconstateerd dat enkele scholen een schriftelijke overhoring van een te laag niveau meetelden voor het cijfer en dat de vragen van met name de eerste werken van 4-Havo soms van een te laag niveau waren.

Als voorbeelden van een voldoende of acceptabele examenpraktijk noemden de beoordelaars biologie een becijfering van het SE als bij het centraal examen, het voor leerlingen inzichtelijk maken van de beoordeling en de consequenties voor leerlingen en het geven van de berekening van de normering en de omzetting van score naar cijfer. Een van de beoordelaars wiskunde merkte op dat scholen de normering volgen die de uitgevers bij de proefwerken leveren.

Conform producteisen

Op de overgrote meerderheid van de scholen zijn de schoolexamens volgens de beoordelaars voldoende zorgvuldig samengesteld en voldoet het toetsmateriaal in voldoende mate aan relevante producteisen. De percentages scholen waarvan de kwaliteit als voldoende beoordeeld is, bedragen voor Engels, Nederlands en wiskunde respectievelijk 70%, 100% en 97%. Vanwege het grote aantal ontbrekende waarnemingen kan over de kwaliteit van het SE biologie geen uitspraak worden gedaan.

Als een voorbeeld van een onvoldoende examenpraktijk noemden de beoordelaars Engels dat de toetsen soms zonder overleg door één persoon geconstrueerd werden (terwijl de sectie uit meerdere docenten bestond). Bij biologie en wiskunde noemden de beoordelaars te vaak gekopieerde en dus slecht leesbare toetsen uit de methodegebonden bundels, onvoldoende ruimte op het antwoordblad, onduidelijke vragen en het werken met uit het boek gekopieerde opgaven.

Voorbeelden van als voldoende aangemerkt examenpraktijken waren het gebruik van Cito-toetsen voor luisteren en schrijven Engels of van toetsen en/of beoordelingsmodellen die daarvan afgeleid waren (waarbij de desbetreffende beoordelaar dan maar aannam dat deze goed geconstrueerd waren), toetsen biologie die waren voorzien van een aantrekkelijk voorblad dat paste bij het onderwerp van de toets met duidelijke instructies voor de leerlingen. Een voorbeeld uit het vakgebied wiskunde was het gebruik van proefwerkbundels waarvan de kwaliteit volgens de beoordelaar wiskunde goed was (waarbij de beoordelaar opmerkte dat dit ook gold voor de toetsen die docenten wiskunde zelf uit de proefwerkbundels samenstelden).

Een beoordelaar Engels merkte overigens op dat een van de best doordachte examens het werk was van een docente zonder vakcollega's binnen haar school.

Vergelijkbaarheid opgaven

Als er met toetsvarianten gewerkt wordt, is de moeilijkheidsgraad ervan naar het oordeel van de beoordelaars op nagenoeg alle scholen vergelijkbaar. De percentages scholen waarvan de kwaliteit van het SE voor deze indicator in orde is, bedragen voor de vakken Engels, Nederlands en biologie respectievelijk 98%, 100% en 98% (voor wiskunde waren er te veel ontbrekende oordelen).

Van een als onvoldoende aangemerkt examenpraktijk is volgens de beoordelaars sprake als dezelfde toetsen Engels op verschillende momenten aan het gehele leerjaar worden gegeven (en kandidaten elkaar dus kunnen "helpen"), als leerlingen zelf een keuze mogen maken uit drie mogelijk onvergelijkbare opdrachten, als toetsvarianten schrijfvaardigheid Nederlands ontbreken waar de beoordelaar deze wel verwacht had. Een beoordelaar biologie noemt als voorbeeld dat docenten de voor de herkansing gebruikte

B-versie van een methodegebonden toets zo vergelijkbaar is met de A-versie dat herkansers duidelijk in het voordeel zijn.

Verplichte toetsinhoud in overleg

Bij de vakken Engels en Nederlands stellen de vakcollega's de inhoud van het SE naar het oordeel van de beoordelaars op vrijwel alle scholen in voldoende mate in overleg vast (respectievelijk 93% en 100% voldoende). Bij biologie en wiskunde verhindert het grote aantal ontbrekende waarnemingen het trekken van een zinvolle conclusie.

Als voorbeelden van een als onvoldoende aangemerkte examenpraktijk noemden de beoordelaars onder meer docenten Engels die geen overleg met collega's plegen (ook al waren er wel degelijk vakcollega's aan de school verbonden) en dat nogal wat scholen hun klassen/groepen bij het vak biologie zo verdelen dat er maar één docent op een klasselaag werkt. Een beoordelaar wiskunde noemt als voorbeeld dat wat de vakdocent hierover in de vakvragenlijst heeft vermeld vaak niet klopt: de school zegt dat de docenten zelf proefwerken samenstellen terwijl zij in de praktijk gewoon een kopie uit de bundel gebruiken.

Beoordelingsaspecten relevant

Alleen bij het vak Engels voldoet een substantieel deel van de scholen volgens de beoordelaars in onvoldoende mate aan de eis dat de beoordelingsaspecten relevant moeten zijn voor de geëxamineerde vaardigheid. De percentages voldoende bedragen voor Engels, Nederlands, biologie en wiskunde respectievelijk 70%, 100%, 99% en 99%.

Als voorbeelden van een "onvoldoende" examenpraktijk noemden de beoordelaars onder meer de (eenzijdige) beoordeling van schrijfvaardigheid Engels op aspecten van grammatica, idioom, spelling en tekstkennis, de vermenging van gespreksvaardigheid Engels met literatuur en bij biologie het bedenkelijke niveau van twee microscopische practicumlessen.

Als een voorbeeld van voldoende examenpraktijk noemt een beoordelaar dat niet alle scholen de beoordelingsmodellen Engels van het Cito onverkort toepassen, maar dat ze dan in ieder geval wel blijken te hebben nagedacht over de systematiek die ze wel willen toepassen.

Beoordelingsaspecten volledig

Met uitzondering van het vak Engels voldoen nagenoeg alle scholen naar het oordeel van de beoordelaars in voldoende mate aan de eis dat alle relevante beoordelingsaspecten in de beoordeling van de kandidaten betrokken worden. Voor de vakken Engels, Nederlands, biologie en wiskunde bedragen de percentages voldoende respectievelijk 66%, 100%, 98% en 99%.

Voorbeelden van een "onvoldoende" examenpraktijk zijn dat bij schrijfvaardigheid Engels niet alle vereiste aspecten in de beoordeling betrokken werden (bijvoorbeeld opbouw, conventies of effectiviteit) en dat er van schrijfvaardigheid in de schrijfvaardigheidstoets Engels vaak geen sprake was.

Een voorbeeld van een als voldoende aangemerkte examenpraktijk was het gebruik van de Cito-luistertoetsen Engels.

Gebruik antwoord- c.q. beoordelingsmodel

Op nagenoeg alle scholen vindt de beoordeling naar het oordeel van de beoordelaars in voldoende mate plaats met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan. De percentages scholen die voor deze indicatoren een voldoende kregen, bedragen voor de vakken Engels, Nederlands, biologie en wiskunde respectievelijk 99%, 100%, 100% en 96%.

Voorbeelden van een "voldoende" examenpraktijk zijn de afwezigheid of onvolledigheid van het antwoord-c.q. beoordelingsmodel, een te ingewikkelde en onduidelijke methodiek voor schrijfvaardigheid Engels, het ontbreken van de gewichten bij de beoordeling van spreek- en gespreksvaardigheid Nederlands en een te globale beoordeling van Fictie Nederlands zonder informatie over de beoordelingscriteria en het gewicht ervan,

Voorbeelden van als voldoende beoordeelde examenpraktijken waren onder meer het vermelden van de beoordelingscriteria, het gewicht en de punten op de opgaven van de kandidaten en het gebruik van Cito-toetsen Engels of toetsen en beoordelingsmodellen die daarop geïnspireerd zijn. Tenminste één beoordelaar wiskunde heeft de kwaliteit van deze indicator als voldoende beoordeeld als de school het correctiemodel van de bundel gebruikte.

Cijfers voor toetsen en opdrachten

Met het vak Engels als uitzondering voldoen nagenoeg alle scholen volgens de beoordelaars aan de eis dat voor de toetsen en handelingsdelen cijfers respectievelijk de kwalificaties “naar behoren” of “onvoldoende” worden toegekend. De percentages voldoende bedragen voor de vakken Engels, Nederlands, biologie en wiskunde respectievelijk 84%, 100%, 100% en 96%.

Als voorbeelden van een “onvoldoende” examenpraktijk noemden de beoordelaars dat literatuur werd meegenomen in gespreksvaardigheid Engels en dus een cijfer kreeg, dat voortgangstoetsen en idioomtoetsen mede het cijfer Engels bepaalden en dat handelingsdelen met een cijfer beoordeeld werden dat meetelde voor het totaal gemiddelde Engels.

Een voorbeeld van een “voldoende” examenpraktijk was dat toetsen Engels en wiskunde altijd een cijfer kregen.

Beoordeling in overleg

Volgens de beoordelaars Engels voldoet 56% van de scholen aan de eis dat de beoordeling door ten minste twee docenten moet plaatsvinden. Het overeenkomstige percentage voor Nederlands is 100%. Over de kwaliteit van het SE op deze indicator kon bij de vakken biologie en wiskunde vanwege het grote aantal ontbrekende waarnemingen geen uitspraak worden gedaan.

Voorbeelden van een “onvoldoende” examenpraktijk zijn dat gespreksvaardigheid Engels vaak niet door twee docenten beoordeeld wordt en het ontbreken van een “bijzitter” bij een mondeling literatuurexamen Nederlands (omdat dit te veel lesuren kost).

Voorbeelden van een “voldoende” examenpraktijk zijn dat een behoorlijk aantal scholen probeert om meerdere docenten bij de beoordeling van schrijfvaardigheid Engels te betrekken (al gebeurt dit soms heel summier en alleen in twijfelgevallen) en dat een school met een eenmanssectie biologie overleg had met een sectie van een andere school.

Scorepunten en/of cesuur vooraf vastgelegd

Op nagenoeg alle scholen worden de scorepunten en indien van toepassing ook de cesuur volgens de beoordelaars in voldoende mate vóór de afname van het toetsmateriaal schriftelijk vastgesteld (de percentages voldoende voor Engels, Nederlands, wiskunde en biologie bedragen respectievelijk 97%, 100%, 100% en 97%).

Een voorbeeld van een “onvoldoende” examenpraktijk is het gebruik van de Cito-luistertoetsen Engels (waarbij de cesuur immers pas achteraf verstrekt wordt).

Voorbeelden van als voldoende aangemerkte examenpraktijk waren dat de scorepunten en de cesuur op de aan de kandidaten uitgereikte toetsen Engels vermeld stonden en de inzet van twee docenten bij de toetsing van gespreksvaardigheid Engels.

Scorepunten vroegtijdig meegedeeld

Alleen voor de vakken Nederlands en wiskunde konden bij deze indicator bruikbare gegevens worden verzameld. Bij het vak Nederlands worden de scorepunten volgens de beoordelaars in voldoende mate voorafgaand aan de afname van het toetsmateriaal aan de kandidaten bekend gemaakt, maar voor het vak wiskunde bedraagt het overeenkomstige percentage ‘slechts’ 64%.

Voorbeelden van een “onvoldoende” examenpraktijk zijn het bestaan van discrepanties tussen de antwoorden van vakdocenten biologie op de ingevulde vragenlijst en hetgeen er op het toetsmateriaal voor de kandidaten vermeld was. Volgens de vragenlijst was de kwaliteit in orde, terwijl de scorepunten op het werk van de kandidaten ontbraken.

Voorbeelden van als voldoende aangemerkte examenpraktijk waren onder meer dat de beoordelingsmethode (d.w.z. beoordelingsaspecten, weging en maximumscore) en de scoring helder op de opgaven Engels vermeld stonden, dat er Cito-luistertoetsen Engels gebruikt werden en dat het aantal punten per beoordelingsaspect bij een praktische opdracht biologie vermeld was.

Bijstelling scorepunten cesuur beargumenteerd

Op de vraag of de scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd kan geen antwoord worden gegeven. Daarvoor waren er bij alle vier vakken te veel ontbrekende waarnemingen. Een voorbeeld van de moeilijke beoordeelbaarheid is dat er leerlingewerk met twee cijfers voorkomt. Het is dan niet duidelijk of de docent een nieuwe norm heeft gehanteerd of dat leerlingen er na inzage puntjes “bijgesprokkeld” hebben.

Normering in overleg

Alleen bij Nederlands kon worden vastgesteld dat de vakcollega's de normering op alle onderzochte scholen in onderling overleg vaststelden. De andere vakken hadden te zeer te leiden onder incomplete en onbeoordeelbare gegevens om hierover een uitspraak te kunnen doen. Volgens een van de beoordelaars Nederlands mag uit de vragenlijst die door de vakdocent was ingevuld worden afgeleid dat er bij de schriftelijke toetsen - d.w.z. de algemene taalvaardigheidstoetsen en de luister-, lees- en schrijftoetsen - doorgaans aan deze indicator wordt voldaan.

Gewicht

Alleen bij Engels en Nederlands zijn er voldoende waarnemingen beschikbaar om de vraag naar het gewicht bij de bepaling van het SE cijfer voor elke toets is aangegeven te kunnen beantwoorden. De beoordelaars Engels waren van mening dat aan deze indicator op 67% van de scholen in voldoende mate voldaan was en bij Nederlands voldeden alle scholen aan deze eis.

Voorbeelden van als onvoldoende aangemerkte examenpraktijk zijn dat docenten Engels zich niet hielden aan de verdeling 1/3, 1/3, 1/3 voor respectievelijk schrijf-, spreek- en luistervaardigheid en dat de weging erg ingewikkeld in elkaar zat of onduidelijk was.

Voorbeelden van een 'voldoende' examenpraktijk zijn dat docenten Engels zich aan de verdeling 1/3, 1/3, 1/3 hielden, dat de weging op de toets Engels vermeld stond, dat veel scholen het gewicht van de onderdelen biologie in het PTA en/of het werk van de kandidaten precies vermeldden.

Een van de beoordelaars Nederlands merkte overigens op dat het gewicht van een toets vaak moeilijk uit de vragenlijst en het PTA te herleiden viel. Volgens een beoordelaar biologie lijken veel scholen het gewicht van de onderdelen in het PTA en/of het werk van de kandidaten precies te vermelden.

4.2 Discussie

In dit onderzoek hebben ervaren docenten vastgesteld dat de kwaliteit van het SE op het overgrote deel van de scholen en indicatoren voldoende is. Dit doet vermoeden dat de kwaliteit van de schoolexamens het laatste decennium sterk verbeterd is. De algemene conclusie sluit aan bij ander onderzoek dat eveneens wijst op een voldoende kwaliteit van het schoolexamen. Zo blijkt uit onderzoek naar de Kwaliteitsmonitor Schoolexamens dat docenten over het algemeen zeer tevreden zijn over de kwaliteit van het schoolexamen voor hun vak (Kuhlemeier & Van Weerden, 2006; Erkens, Kuhlemeier & Van Weerden, 2007).

Verstralen (2007) presenteert gegevens die erop wijzen dat de betrouwbaarheid van het schoolexamen even hoog of zelfs hoger is dan die van het centraal examen. Van den Bergh, Rohde en Zwarts (2003) constateren dat de stabiliteit van het schoolexamens over een reeks van jaren groter is dan die van het centraal examen.

De algemene uitkomsten van het onderzoek wijken echter af van eerdere rapportages van de Inspectie (Inspectie van het onderwijs, 1995, 1996, 1998, 1999) en van een niet gepubliceerd onderzoek dat het Cito eind jaren negentig in opdracht van de Inspectie uitvoerde (drs. Tom Erkens, persoonlijke communicatie d.d. 25 november 2008). In dit laatste onderzoek is de inhoudelijke en toetstechnische kwaliteit van de toenmalige schoolexamens beoordeeld voor de vakken geschiedenis en economie in het havo en twee beroepsgerichte vakken in het vmbo. Een van de conclusies was dat de representativiteit van de schoolexamens ten opzichte van het examenprogramma sterk te wensen overliet, terwijl er ook relatief vaak onderwerpen werden geëxamineerd die docenten volgens het eindexamenprogramma niet zouden mogen toetsen. Tegelijkertijd was er veel aan te merken op de toetstechnische kwaliteit van de toetsen (zoals op de leesbaarheid en lay-out van de examenopgaven). De onderzoekers lieten een steekproef van examenwerken opnieuw beoordelen onder gecontroleerde omstandigheden. Daarbij vergeleken zij de door de docenten toegekende cijfers met die van de onafhankelijke beoordelaars. De conclusie was dat docenten de examenwerken van de kandidaten doorgaans te soepel beoordeelden (drs. Tom Erkens, persoonlijke communicatie d.d. 27 november 2008). Al met al lijken de overwegend positieve bevindingen van ons onderzoek moeilijk verenigbaar met die van voorafgaand onderzoek naar de kwaliteit van de schoolexamens (al dient gezegd te worden dat de eindexamenprogramma's en regelgeving destijds explicieter waren dan tegenwoordig).

De gunstige bevindingen bij de "niveau-indicatoren" 4, 5 en 6 lijken ook op gespannen voet te staan met de prestatieverschillen tussen scholen voor leesvaardigheid Nederlands op de centrale examens en met uitkomsten van landelijk peilingsonderzoek naar het onderwijsniveau. Hieruit komt naar voren dat de populatie van scholen een aantal zeer hoog presterende scholen bevat die naar verwachting in het SE vooral moeilijke toetsen en opdrachten afnemen en/of strenge beoordelingscriteria hanteren. Daarnaast

kent de populatie een aantal zeer laag presterende scholen die vooral gemakkelijk toetsmateriaal en/of soepele beoordelingscriteria zullen hanteren. Dat de beoordelaars bij Nederlands geen enkele school aantreffen die over de hele linie te moeilijke of te gemakkelijke (lees)toetsen gebruikten of die de meerderheid van de (lees)toetsen te streng of te soepel beoordeelden, is verbazingwekkend. Daarnaast zou men zich kunnen afvragen hoe de positieve onderzoeksresultaten bij Nederlands, biologie en wiskunde zich verhouden tot de uitkomsten van onderzoek naar discrepanties tussen het cijfer op het SE en het CE (De Lange & Dronkers, 2006; Rekers-Mombargs & Harms, 2008; Inspectie van het onderwijs, 2007a,b). Uit dit onderzoek komt naar voren dat de populatie een aantal scholen kent waar het SE in vergelijking met het CE over een reeks van vakken en jaren uitzonderlijk gemakkelijk of moeilijk is (zowel in vmbo-bb, vmbo-kb, vmbo-gl, vmbo-tl, havo als vwo) en waarbij de grote discrepantie tussen het SE en CE gepaard gaat met lage opbrengsten. Het wekt dan ook verbazing dat de beoordelaars Nederlands geen enkele school aantreffen waarvan “het niveau” van het examenmateriaal van het SE en/of de becijfering duidelijk hoger of lager was dan men op grond van het door de eindtermen geïmpliceerde niveau en de moeilijkheid van de toetsen en opdrachten uit het centraal examen zou mogen verwachten. Tot slot lijken de positieve bevindingen bij de vakken Nederlands, biologie en wiskunde zo op het eerste gezicht in tegenspraak met de constatering van de Inspectie dat ongeveer een kwart van de scholen het SE niet op orde heeft (Inspectie van het Onderwijs, 2007c). Uiteraard zij hierbij opgemerkt dat de Inspectie tijdens de schoolbezoeken (ook) op andere aspecten van de kwaliteit van het SE let en haar oordeel op meer vakken baseert dan alleen Nederlands. Een uitgebreide toelichting op de onderzoeksresultaten is opgenomen in Bijlage 6. In de volgende paragrafen 4.2.1 en 4.2.2 worden vakinhoudelijke, organisatorische en beoordelingstechnische kanttekeningen bij de onderzoeksresultaten geplaatst.

4.2.1 Vakinhoudelijke kanttekeningen

De vigerende eindexamenprogramma's kennen een heldere domeinindeling, maar de afzonderlijke eindtermen zijn doorgaans globaal geformuleerd en specifieke niveauaanduidingen ontbreken. De marges van de regelgeving voor het schoolexamen zijn zeer breed met veel mogelijkheden voor schoolspecifieke bepaling van bijvoorbeeld het aantal schoolexamens, de te toetsen onderwerpen en vaardigheden en het niveau waarop de stof behandeld wordt. Naarmate de kaders voor examinering vager zijn er in het schoolexamen meer geoorloofd is, wordt het voor beoordelaars moeilijker om afwijkingen van de wettelijke regels te constateren (zie ook Bijlage 6).

De leerlingen blijken in de tweejarige examenperiode grote hoeveelheden toetsen en opdrachten te krijgen. Dat de beoordelaars de kwaliteit van het SE niet vaker als onvoldoende hebben aangemerkt, heeft ook met deze grote hoeveelheid toetsen te maken. Het aantal als onvoldoende aangemerkte toetsen was in verhouding tot het totale aantal afgenomen toetsen meestal te klein om het oordeel “onvoldoende” te rechtvaardigen (zie ook Bijlage 6).

Docenten maken veel gebruik van toetsen en opdrachten uit de gebruikte methode. Daarnaast gebruiken zij voor bepaalde vakonderdelen “oude” opgaven uit de centrale examens of toetsmateriaal dat het Cito speciaal voor het SE ontwikkeld heeft. Voor zover docenten “eigen” opgaven en toetsen maken, betreft het meestal materiaal dat uit deze bronnen is samengesteld of er sterk door geïnspireerd is. Naar het oordeel van de beoordelaars is er met het (her)gebruikte toetsmateriaal “weinig mis” (zie ook Bijlage 6). Tegelijkertijd speelde ook het ontbreken van materiaal een rol. We lichten dit toe aan de hand van Engels havo. Bij dit vak bleek de kwaliteit van het SE gespreksvaardigheid vrijwel niet te beoordelen (zie ook paragraaf 3.1.3). De oordelen over de algemene kwaliteit van het SE zijn daarom voor een belangrijk deel gebaseerd op de beoordelingen van de schoolexamens luistervaardigheid en schrijfvaardigheid. Voor het schoolexamen luistervaardigheid Engels gebruiken vrijwel alle scholen de kijk-luistertoetsen van het Cito. Voldoende beoordelingen liggen dan voor de hand. Ook voor de toetsing van schrijfvaardigheid Engels gebruiken veel scholen toetsen en beoordelingsmiddelen van het Cito of volgens het model van het Cito. Dit zal eveneens in veel gevallen hebben geleid tot een voldoende beoordeling. Speciaal bij de toetsing en beoordeling van gespreksvaardigheid echter zijn de risico's op niet-valide toetsing en onbevredigende beoordelingspraktijken groot. Er zouden ongetwijfeld minder scholen voor Engels voldoende zijn beoordeeld als genoeg informatie over de beoordeling van het SE gespreksvaardigheid voorhanden was geweest (zie ook paragraaf 3.1.5).

Docenten blijken maar weinig toetsmateriaal volledig zelf te ontwikkelen. De beoordelaars constateren dat de kwaliteit van de geheel zelf ontwikkelde vragen en toetsen vrijwel zonder uitzondering zeer veel lager is dan die van het externe toetsmateriaal (en het materiaal dat daarop gebaseerd is). Vanwege het kleine aandeel van het zelf ontwikkelde toetsmateriaal in het totale schoolexamen was er voor beoordelaars vrijwel nooit aanleiding tot het toekennen van het oordeel “onvoldoende”.

4.2.2 Organisatorische en beoordelingstechnische kanttekeningen

De Inspectie heeft de scholen uit de steekproef ruim voor aanvang van het schooljaar 2008-2009 over het onderzoek geïnformeerd. Volgens sommige beoordelaars is het niet ondenkbaar dat scholen hun schoolexamen in de wetenschap van Inspectietoezicht in positieve zin hebben aangepast (bijvoorbeeld door examenmateriaal van mindere kwaliteit of bijzonder gemakkelijke of moeilijke toetsen en opdrachten achter te houden).

De toetsingskaders Engels en biologie zijn grotendeels gebaseerd op documenten die in een vooronderzoek ontwikkeld zijn. De toetsingskaders voor Nederlands en wiskunde zijn een afgeleide van de documenten voor Engels en biologie. De toetsingskaders bieden weliswaar een omschrijving van de standaarden en indicatoren, maar geen uitgewerkt beoordelingskader met eenduidige beslisregels voor het toekennen van oordelen en karakteristieke voorbeelden van examenpraktijken op een schaal van lage naar hoge kwaliteit. Een bijkomend probleem is de openheid van de wettelijke voorschriften voor de vorm en inhoud van het SE die scholen veel ruimte voor eigen invullingen biedt. Hoe het ook zij, het toetsingskader, het eindexamenprogramma en de zeer grote vrijheid van de scholen iets wel of niet te toetsen, gaf de beoordelaars relatief weinig houvast bij de beoordeling. Daarnaast speelde het probleem van het ontbreken van belangrijk materiaal. In het geval van ontbrekend materiaal hadden de beoordelaars de opdracht op het beoordelingsformulier de code voor "Materiaal niet aanwezig en moet nog worden opgevraagd" te noteren en hun oordeel op te schorten tot het materiaal compleet was. Vaak ontbrak echter slechts een deel van het materiaal van de school, zonder dat er duidelijke regels waren over hoe zij in dat geval moesten handelen. Deze onzekerheid zien we terug in het gebruik van de beoordelingscategorieën. De ene beoordelaar maakte veel meer gebruik van de categorie "Materiaal niet aanwezig" dan de andere, zelfs al beoordeelden zij dezelfde school (zoals in het onderzoek naar de overeenstemming tussen de beoordelaars het geval was). Verder bleken de vragen in de vragenlijst die de scholen moesten invullen vaak niet precies genoeg om de vraag naar de kwaliteit van een indicator eenduidig als voldoende of onvoldoende te kunnen beoordelen. Het wekt dan ook geen verbazing dat sommige beoordelaars in een staat van voortdurende onzekerheid over de juistheid van hun oordelen verkeerden en de kwaliteit van het SE regelmatig als "Niet beoordeelbaar" moesten typeren. De onzekerheid bij het beoordelen wordt weerspiegeld in de resultaten van het onderzoek naar de beoordelaarsovereenstemming. Die was niet altijd duidelijk hoger dan men op basis van toeval zou mogen verwachten. Het valt zeker niet uit te sluiten dat de twijfel aan de juistheid van hun oordelen de beoordelaars ertoe gebracht heeft de scholen wat vaker "het voordeel van de twijfel" te geven dan de werkelijke kwaliteit van het SE rechtvaardigt.⁴

De beoordelaars hadden de opdracht te werk te gaan volgens het principe van "beoordeling bij gebrek aan contra-indicatie" zoals dat bij de Inspectie voor bepaalde indicatoren gebruikelijk is. Was er tijdens de *quick scan* geen reden tot twijfel aan de algemene kwaliteit van het SE, dan moesten de beoordelaars de algemene kwaliteit voor de desbetreffende indicator bij gebrek aan contra-indicatie als voldoende beoordelen. Leek de kwaliteit bij de *quick scan* twijfelachtig of onvoldoende, dan moesten zij het oordeel over de algemene kwaliteit opschorten tot na de *deep scan*. Het is denkbaar dat beoordeling bij gebrek aan contra-indicatie tijdens de *quick scan* een toegeeflijke beoordeling in de hand heeft gewerkt (zeker als beoordelaars het ontbreken van materiaal of de moeilijke beoordeelbaarheid ervan zouden verwarren met een gebrek aan contra-indicatie). Een *quick scan* leidt niet per definitie tot een toegeeflijke beoordeling en de beoordelaars hebben zeker niet bewust toegeeflijk beoordeeld, maar het ontbreken van materiaal kan er wel voor gezorgd hebben dat zij scholen wat eerder "het voordeel van de twijfel" hebben gegeven. Hoewel dit bij Engels onwaarschijnlijk is omdat daar vrijwel steeds elke vaardigheid apart beoordeeld is, geeft een van de beoordelaars aan zich in deze redenering te kunnen vinden.

De beoordelaars moesten de kwaliteit van het SE dichotoom scoren. Afgezien van de restcategorieën "Niet beoordeelbaar" en "Materiaal niet aanwezig" konden zij bij het beoordelen van de kwaliteit alleen kiezen uit de scores 0 (onvoldoende) en 1 (voldoende). Uit onderzoek is bekend dat dichotome scoring aanleiding kan geven tot toegeeflijke beoordelingen en een gebrek aan spreiding (range restriction). Tijdens de voorbereidingsfase is in overleg met de Inspectie van polytome scoring afgezien omdat beoordelaars dan naar verwachting vooral de "veilige" middencategorieën zouden kiezen.

Bij een van de vakken heeft het team niet volledig onafhankelijk beoordeeld. De beoordeling vond plaats in één ruimte (zij het dat niet alle beoordelaars altijd tegelijkertijd in deze ruimte aanwezig waren) waarbij er geen maatregelen genomen werden om overleg tussen beoordelaars te voorkomen. Vanuit de intentie tot een zo goed mogelijk oordeel te komen is ook na de training in twijfelgevallen overleg gevoerd.

⁴ Een mogelijke tegenwerping is dat er bij Engels wel een flink percentage onvoldoendes werden gegeven. Een mogelijke verklaring is dat de beoordelaars Engels ook aan het vooronderzoek hadden meegewerkt en derhalve reeds over ruime beoordelingservaring beschikten.

Per vak waren er slechts drie beoordelaars. De belangrijkste reden was dat het erg moeilijk bleek om geschikte beoordelaars te werven. De beoordelaars moesten immers beschikken over ruime onderwijservaring in het desbetreffende vak en onderwijssegment, ervaring met het construeren van centrale examens (voldoende zicht op het niveau dat van kandidaten verwacht mag worden) en ruimte in hun takenpakket. Door deze eisen was de spoeling zeer dun en beperkt tot CG-leden die werkzaam waren (geweest) voor het betreffende examenniveau en het enthousiasme voor deze klus konden opbrengen. We konden met andere woorden niet meer beoordelaars vinden. Drie beoordelaars is een wankel basis voor generalisatie. Toevallige verschillen tussen de beoordelaars in strengheid of soepelheid of in spreidend vermogen kunnen het gemiddeld oordeel en de spreiding van de oordelen hebben beïnvloed.

Zonder nader onderzoek kan over de plausibiliteit van de hierboven vermelde verklaringen niets met zekerheid worden gezegd. Wel geven de alternatieve verklaringen aanleiding om de onderzoeksresultaten met de grootst mogelijke voorzichtigheid en terughoudendheid te interpreteren.

4.3 Aanbevelingen

Op grond van deze studie kunnen aanbevelingen worden gedaan voor de verbetering van de wettelijke voorschriften, het eindexamenprogramma, de praktijk van het schoolexamen en mogelijk vervolgonderzoek naar de kwaliteit van het SE.

4.3.1 Regelgeving

Ten aanzien van de wettelijke voorschriften kunnen de volgende aanbevelingen worden gedaan:

- Verplicht of stimuleer scholen tot transparantie over hun examenpraktijken.
- Limiteer het aantal toetsingen in het schoolexamen (zowel voor het laatste als het voorlaatste examenjaar) en schrijf een minimum aantal toetsen per vakonderdeel voor.
- Beperk het aantal facultatieve onderdelen in het SE.
- Verstrek heldere voorschriften voor de tweede correctie/beoordeling van de vakonderdelen van het SE (bijvoorbeeld voor de mondelinge vaardigheden en grotere schrijfoopdrachten bij Nederlands).
- Verstrek nadere regels over het gewicht van de vakonderdelen in de bepaling van het SE-cijfer.

4.3.2 Het eindexamenprogramma

Ten aanzien van het eindexamenprogramma kunnen de volgende aanbevelingen worden gedaan:

- Expliciteer de eindtermen (bijvoorbeeld ten aanzien van wat kandidaten bij Nederlands op het vlak van spelling en grammatica moeten beheersen).
- Neem weer niveaueisen op.
- Neem het gewicht van de onderdelen in het programma op.

4.3.3 De praktijk van het schoolexamen

Voor de praktijk van het schoolexamen kunnen de volgende aanbevelingen worden gedaan:

- Limiteer het aantal toetsen in het schoolexamen.
- Verplicht scholen om voor alle betrokkenen (dus ook de Inspectie) nauwkeurig te omschrijven welke vakonderdelen meetellen voor het SE en met welk gewicht.
- Verplicht scholen om voor alle betrokkenen te verantwoorden waarom zij de bovenstaande keuze hebben gemaakt.
- Verplicht scholen om voor alle betrokkenen duidelijk te maken welk(e) eindniveau(s) zij nastreven voor de verschillende onderdelen van het SE. Voor Engels kan daarbij bijvoorbeeld de *Handreiking schoolexamen moderne vreemde talen* (Meijer & Fasoglio, 2006) dienen.
- Controleer jaarlijks een steekproef van de bovengenoemde verantwoordingsdocumenten en wijs scholen zo nodig op hun verantwoordelijkheden.
- Stimuleer deskundigheidsbevordering bij scholen en docenten ten aanzien van de valide, betrouwbare en efficiënte toetsing en beoordeling van de verschillende onderdelen van het SE.
- Werk aan bewustwording bij scholen en docenten ten aanzien van de inrichting en opzet van de schoolexamens.
- Stel een bovengrens op voor het aandeel van methodegebonden toetsmateriaal (om fraude te voorkomen).

4.3.4 Vervolgonderzoek

Voor vervolgonderzoek naar de kwaliteit van het schoolexamen kunnen de volgende organisatorische en methodologische aanbevelingen worden gedaan.

Organisatorische aanbevelingen

- Check te beoordelen materiaal op volledigheid en consistentie alvorens het aan beoordelaars voor te leggen.
- Laat beoordelaars pas met het beoordelen van een school beginnen als al het materiaal van die school aantoonbaar volledig is.

Methodologische aanbevelingen

- Ga na of scholen al het examenmateriaal hebben opgestuurd (en geen examenmateriaal van mindere kwaliteit of bijzonder gemakkelijke of moeilijke toetsen en opdrachten hebben achtergehouden).
- Ontwerp een betrouwbaar en hanteerbaar categorieënsysteem voor de beoordeling van de kwaliteit van het SE met eenduidige beslissingsregels voor het toekennen van oordelen aan scholen en karakteristieke voorbeelden van examenpraktijken die de beoordelaars bijvoorbeeld als inacceptabel, slecht, matig, redelijk, voldoende, goed of uitstekend moeten beoordelen.
- Ontwikkel een trainingsprogramma met ondersteunend materiaal voor het instrueren van beoordelaars. School beoordelaars in het herkennen en bestrijden van storende beoordelaarfouten, zoals een te toegeeflijke beoordeling en het aanbrengen van minder spreiding dan de feitelijke verdeling rechtvaardigt.
- Laat beoordelaars pas met hun beoordelingstaak beginnen als hun geschiktheid als beoordelaar is aangetoond.
- Ga tussentijds regelmatig na in hoeverre de beoordelaars betrouwbaar en zoals bedoeld beoordelen. Neem zo nodig bijtijds maatregelen ter verhoging van de betrouwbaarheid.
- Tref maatregelen die afhankelijke beoordeling voorkomen.
- Zie af van “beoordeling bij gebrek aan contra-indicatie”.
- Vervang dichotome door polytome beoordeling. Hanteer bijvoorbeeld bij het beoordelen van de moeilijkheid van de toetsen en opdrachten (indicator 4) in plaats van alleen voldoende of onvoldoende de beoordelingscategorieën veel makkelijker, makkelijker, even gemakkelijk/moeilijk, moeilijker en veel moeilijker (in vergelijking met het door de eindtermen geïmpliceerde niveau en de moeilijkheid van de toetsen en opdrachten uit het centraal examen).
- Trek een representatieve steekproef uit de vele toetsen, opdrachten, scorings- en beoordelingsmiddelen en examenwerken (in plaats van telkens alles in één keer te willen beoordelen).
- Betrek een groter aantal beoordelaars in de beoordeling, bij voorkeur zoveel als voor een betrouwbare beoordeling noodzakelijk is. Dit biedt tevens de mogelijkheid om “dissidente” beoordelaars te verwijderen.

Aanbevelingen voor detailstudies

Het onderzoek heeft niet voor alle vakonderdelen voldoende informatie opgeleverd over toetsing en beoordeling (denk bijvoorbeeld aan het onderdeel gespreksvaardigheid bij het vak Engels). Dit is deels te wijten aan het feit dat weinig scholen alle door de Inspectie opgevraagde materialen hebben opgestuurd. Daarnaast is echter geconstateerd dat zulke materialen alleen niet genoeg inzicht opleveren in de toets- en beoordelingspraktijk. Het verdient daarom aanbeveling om bij een steekproef van scholen alle relevante materialen op te vragen en op basis van deze materialen een leidraad op te stellen ten behoeve van een interview met een verantwoordelijke examinerator dat de gewenste extra duidelijkheid moet verschaffen. Hierbij zou men kunnen beginnen bij de scholen die in dit onderzoek de algemene beoordeling “onvoldoende” hebben gekregen.

Literatuur

- Bergh, H. van den, Rohde, E., & Zwarts, M. (2003). Is het ene examen het andere? Over de stabiliteit van schoolonderzoek en centraal examen. *Pedagogische Studiën*, 80, 176-191.
- Inspectie van het Onderwijs (1995). *Schoolonderzoek onderzocht I*. Zwolle: Inspectie van het Onderwijs.
- Erkens, T., Kuhlemeier, H., & Weerden, J. van (2007). De kwaliteitsmonitor schoolexamens van het Cito: Op weg naar betere schoolexamens. *Examens*, 3, 5-9.
- Inspectie van het Onderwijs (1996). *Schoolonderzoek onderzocht II*. Zwolle: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (1997). *Schoolonderzoek onderzocht III*. Zwolle: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (1999). *Schoolonderzoek onderzocht IV*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2006). *De staat van het onderwijs: Onderwijsverslag 2004-2005*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2007a). *Discrepanties tussen de cijfers op het schoolexamen en het centraal examen*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2007b). *Discrepanties tussen de cijfers op het schoolexamen en het centraal examen in de periode 2004-2006*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2007c). *De staat van het onderwijs: Onderwijsverslag 2005-2006*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2008). *De staat van het onderwijs. Onderwijsverslag 2006/2007*. Utrecht: Inspectie van het Onderwijs.
- Kuhlemeier, H., & Weerden, J. van (2006). *De meetkwaliteit van de Kwaliteitsmonitor Schoolexamens: Rapportage van de uitkomsten van een onderzoek met de conceptversie*. Arnhem: Cito.
- Kuhlemeier, H., Weeren, J. van, & Werf, M.P.C. van der (2006). *Scheiding van opleiden en examineren in Nederland en omringende landen. Studie verricht in opdracht van de Onderwijsraad*. Cito: Arnhem.
- Lange, M. De, & Dronkers, J. (2006). *Hoe gelijkwaardig blijft het eindexamen tussen scholen? Discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen in het voortgezet onderwijs tussen 1998 en 2005*. Nijmegen/Fiesiole: Radboud Universiteit Nijmegen/Europees Universitair Instituut, San Domenico di Fiesole, Italië.
- Lange, M. De, & Dronkers, J. (2007). *Groeide de ongelijkwaardigheid van het eindexamen tussen scholen verder in 2005? Discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen in het voortgezet onderwijs: 2005 versus 1997-2004*. Paper voor de Onderwijsresearchdagen 2007 in Groningen, 6-8 juni 2007. Nijmegen/Fiesiole: Radboud Universiteit Nijmegen/Europees Universitair Instituut, San Domenico di Fiesole, Italië.
- Meijer, D., & Fasoglio, D. (2006). *Handreiking schoolexamen moderne vreemde talen; Duits, Engels, Frans*. Enschede: SLO
- Meijer, D., & Noijons, J. (red.) (2008). *Gemeenschappelijk Europees Referentiekader voor Moderne Vreemde Talen: Leren, onderwijzen, beoordelen*. Den Haag: Nederlandse Taalunie
- Nichols, D. (1997). MKAPPASC.SPS/MKAPPASC.TXT. Online beschikbaar via <ftp://ftp.spss.com/pub/spss/statistics/nichols/macros/> en <http://www.spsstools.net/Syntax/Matrix/CohensKappa.txt>
- Rekers-Mombarg, L.T.M., & Harms, G.J. (2008). *Metten met twee maten: De discrepantie tussen de cijfers op het schoolexamen en het centraal examen VO van allochtone leerlingen*. Groningen: Gion.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press

Schouten-van Parreren, C. (1985). *Woorden leren in het vreemde-talenonderwijs*. Apeldoorn: Van Walraven

Verstralen, H. (2007). *Het schatten van de betrouwbaarheid met behulp van een covariaat met bekende betrouwbaarheid en equivalering van cijfers*. Arnhem: Cito.

Bijlagen

Bijlage 1

Toetsingskader Engelse taal havo

In de linker kantlijn wordt steeds aangegeven of een tekstgedeelte van toepassing is op HAVO 5 (H5) of op HAVO 4 (H4) of op beide (H5/H4). Raadpleeg steeds de toelichting op bladzijde 5 en volgende.

INHOUD

1 Verplichte onderdelen SE

H5/H4 *De inhoud van het schoolexamen voldoet aan de specificaties van het betreffende examenprogramma voor Engels HAVO.*

H5 Het schoolexamen bestaat uit een examendossier dat bevat:

1.1 toetsen:

- luistervaardigheid
- gespreksvaardigheid
- schrijfvaardigheid

1.2 handelingsdeel dat de volgende onderdelen omvat:

- leesvaardigheid: extensief lezen, samenvatten als strategie en informatie verwerven met behulp van ICT
- luistervaardigheid: extensief luisteren en het maken van aantekeningen
- gespreksvaardigheid: reële communicatieve situaties en presentaties
- schrijfvaardigheid: correspondentie (mede met behulp van ICT)
- letterkunde: samenstelling leesdossier en verwerkingsopdrachten

H4 Het schoolexamen bestaat uit door het bevoegd gezag gespecificeerde onderdelen.

NIVEAU

De inhoud van de toetsen en (eventuele) overige onderdelen van het schoolexamen voldoet aan de eisen voor de moeilijkheidsgraad die verwacht mogen worden bij voltooiing van een HAVO-opleiding Engels.

2

H5 De toetsen voldoen aan de specificaties van niveau 3 in bijlage 19 van het Examenprogramma.

3

H5/H4 Alle toetsen die meetellen voor het schoolexamen worden afgenomen op een zodanig tijdstip dat redelijkerwijs verwacht mag worden dat de kandidaten het gewenste eindniveau bereikt kunnen hebben.

4

De toetsen zijn van de juiste moeilijkheidsgraad, dat wil zeggen, niet te moeilijk en niet te gemakkelijk en zijn passend voor het niveau.

5

De scoring en de beoordeling van alle leerlingen is billijk, dat wil zeggen niet te streng en niet te soepel.

6

De kandidaten hebben een cijfer gekregen passend bij het niveau van hun kennis en vaardigheden.

TOETSCONSTRUCTIE

De toetsen/onderdelen van het schoolexamen zijn zorgvuldig samengesteld en voldoen aan relevante producteisen.

7

- H5/H4 Bij de samenstelling van de toetsen is rekening gehouden met de volgende eisen:
- aansluiting bij de eindtermen en bijbehorende specificaties van het Examenprogramma;
 - eenduidigheid van vragen en opdrachten;
 - vakinhoudelijke correctheid;
 - voldoende afnametijd.

8

- H5/H4 De kandidaten van een bepaald examencohort krijgen bij de toetsen gelijke of vergelijkbare opgaven voorgelegd.

9

- H5/H4 De inhoud van de toetsen wordt, waar mogelijk, in overleg met vakcollega's bepaald.

BEOORDELING

De beoordeling van de toetsen en (eventuele) overige onderdelen van het schoolexamen gebeurt zorgvuldig, is relevant en transparant.

10

- H5/H4 In de beoordeling van de toetsen worden uitsluitend aspecten betrokken die relevant zijn voor de toetsing van de betreffende vaardigheid.

11

- H5/H4 In de beoordeling van de toetsen worden alle relevante beoordelingsaspecten betrokken.

12

- H5/H4 De beoordeling geschiedt waar mogelijk aan de hand van een antwoord- c.q. beoordelingsmodel, met daarin vermelding van de beoordelingscriteria en het gewicht ervan.

13

- H5 Voor de toetsen worden cijfers toegekend; voor handelingsdelen de kwalificaties "naar behoren" of "onvoldoende".

14

- H5/H4 De beoordeling geschiedt door twee docenten.

NORMERING

De normering van de toetsen is zo veel mogelijk van tevoren vastgelegd en wordt consequent toegepast.

15

- H5/H4 Vóór de afname van de toetsen worden de normeringssystematiek en zo mogelijk de puntenwaardering per opgave schriftelijk vastgelegd.

16

- H5/H4 De normeringssystematiek wordt voorafgaand aan of bij de afname van de toets kenbaar gemaakt aan de kandidaten.

17

- H5/H4 Bijstelling van de puntentoekenning en/of cesuur achteraf gebeurt op grond van vooraf daarvoor geformuleerde argumenten.

18

H5/H4 De normering wordt voor zover mogelijk in onderling overleg door de vakcollega's bepaald.

WEGING

De weging van de verschillende onderdelen van het schoolexamen is helder aangegeven in het pta.

19

H5/H4 Per toets is aangegeven wat het gewicht ervan is bij de bepaling van het schoolexamencijfer.

TOELICHTING BIJ HET TOETSINGSKADER ENGELS HAVO

ALGEMEEN

Verschillende examenprogramma's

Leerlingen die in het schooljaar 2007-2008 in HAVO 5 zitten, moeten worden onderwezen volgens het *Examenprogramma profielen havo/vwo* van mei 1998. In dit toetsingskader wordt ervan uitgegaan dat leerlingen die zijn blijven zitten in HAVO 4, worden opgeleid volgens het *Examenprogramma moderne vreemde talen en literatuur havo/vwo*, dat sinds 1 augustus 2007 van kracht is.

Centrale vraag

De centrale vraag is:

Voldoet het schoolexamen Engels (voortaan: "schoolexamen") aan redelijke eisen ten aanzien van:

- inhoud, niveau en toetsconstructie;
- beoordeling, normering en weging van de verschillende onderdelen.

INHOUD

De beoordeling van de inhoud van het schoolexamen vindt plaats aan de hand van de beschrijving van de inhoud door docenten en van concrete opgaven, beoordelings- c.q. antwoordmodellen en eventueel (nagekeken) werken.

Examenstof

H5/H4 Het onderwijsprogramma moet de stof uit het Examenprogramma volledig dekken. Dat wil niet per se zeggen dat alle kandidaten op alle examenstof getoetst moeten worden; per leerling kan de toetsing betrekking hebben op een steekproef van de examenstof. Evenmin hoeft elk schoolexamen volledig het betreffende (sub)domein van het Examenprogramma te dekken. Het mag echter niet zo zijn dat bepaalde onderdelen van het Examenprogramma systematisch in de toetsing worden genegeerd.

H5 In toetsen en handelingsdeel mogen geen eindtermen aan de orde komen die niet voorkomen in het Examenprogramma HAVO (en wel bijvoorbeeld in het examenprogramma vwo).

Handelingsdeel

H4 In het nieuwe Examenprogramma wordt niet gesproken over een handelingsdeel. Het bevoegd gezag kan echter ook aan leerlingen in HAVO 4 andere schoolexamenonderdelen dan toetsen opleggen; denk bijvoorbeeld aan een leesverslag voor de literatuur.

Profielwerkstuk

H5/H4 Het examen kan ook een profielwerkstuk omvatten. Regelgeving die het profielwerkstuk betreft, is eerder te verwachten in het algemene gedeelte van het schoolexamenreglement dan in het vakspecifieke programma van toetsing en afsluiting (voortaan: "pta").

NIVEAU

2

H5 De niveauspecificaties voor Engels taal en letterkunde HAVO staan in bijlage 19 van *Examenprogramma's profielen vwo/havo Nederlands/moderne vreemde talen* (1998). Voor alle taalvaardigheden geldt niveau 3.

3

Als bijvoorbeeld al in de vierde klas HAVO bij alle leerlingen toetsen worden afgenomen die meetellen voor het schoolexamen, is het de vraag of deze kunnen voldoen aan het vereiste niveau. Individuele leerlingen

kunnen echter wel het vereiste niveau hebben bereikt, bijvoorbeeld omdat ze opgegroeid zijn in een land waar Engels de voertaal is.

4-6

Het niveau wordt om te beginnen bepaald door de moeilijkheidsgraad van de opgaven. Maar ook de scoring en beoordeling hebben invloed. Een moeilijke opgave kan bijvoorbeeld soepel worden beoordeeld. In het ideale geval is de moeilijkheidsgraad van de opgave gepast en zijn de scoring en beoordeling billijk. Dit is te beoordelen aan de hand van opgaven en scorings- en beoordelingsmodellen. Maar dan nog is het mogelijk dat de kandidaten feitelijk strenger of soepeler beoordeeld worden dan je op grond van opgaven en score- en beoordelingsmodellen zou verwachten. Dit is alleen na te gaan aan de hand van de scores en cijfers die op leerlingwerken zijn geschreven.

TOETSCONSTRUCTIE

De vorm van (elk onderdeel van) het schoolexamen moet adequaat zijn vanuit het oogpunt van validiteit: nadere toelichting zie onder. De omvang van (elk onderdeel van) het schoolexamen moet adequaat zijn vanuit het oogpunt van toetsbetrouwbaarheid en efficiëntie.

Validiteit

Het gaat hierbij bijvoorbeeld om de volgende punten:

- De toetsing van de luistervaardigheid vindt plaats aan de hand van de luistertoetsen van Cito of een gelijkwaardig alternatief;
- Een schrijfpodracht mag bijvoorbeeld niet zo geformuleerd zijn dat het maken van de opdracht eerder vertaalwerk dan productief schrijven vereist;
- De gespreksvaardigheid moet daadwerkelijk worden getoetst in een mondelinge toets en niet bijvoorbeeld gedeeltelijk in schriftelijke vorm.

Omvang van de toetsen

De toetsen moeten minimaal een zodanige lengte hebben dat zij de betreffende vaardigheid betrouwbaar kunnen toetsen. Zij moeten echter niet langer zijn dan strikt noodzakelijk is.

Handelingsdeel

Aan de taken die de school formuleert ten aanzien van (eventuele) overige schoolexamenonderdelen, moeten de kandidaten in redelijkheid kunnen voldoen. Hiervoor liggen geen criteria vast. De beoordeling van de eisen die elke school stelt, moet worden gebaseerd op ervaringsgegevens.

9

Als bijvoorbeeld de sectie Engels op een school uit slechts één docent bestaat, is overleg met een of meer collega's Engels van andere scholen of met een of meer docenten van andere moderne vreemde talen van de eigen school wenselijk.

BEOORDELING

10

Bijvoorbeeld: bij een luistervaardigheidstoets met open vragen mag de stel- en spelvaardigheid van de kandidaten geen rol spelen in de beoordeling.

11

Vooraf de correctiemodellen van de schrijfvaardigheidstoetsen kunnen een indicatie geven van de kwaliteit van de beoordeling op de school.

10-11

Kernvraag: Welke beoordelingsaspecten worden meegewogen?

H5 Volgens het Examenprogramma moeten de volgende aspecten in de beoordeling worden betrokken:

- bij gespreksvaardigheid: inhoud (volledigheid, uitvoerigheid, begrijpelijkheid), correctheid en gepastheid van het taalgebruik, vlotheid en uitspraak;
- bij schrijfvaardigheid: inhoud (volledigheid, originaliteit, duidelijkheid), opbouw en correctheid en gepastheid van het taalgebruik

H4 Het bevoegd gezag bepaalt welke beoordelingsaspecten worden meegewogen.

NORMERING

15-16

De mate van gedetailleerdheid van de informatie hangt samen met de aard van de opgave(n) en de gekozen beoordelingswijze. Bij een sterk gestructureerde opgave is bijvoorbeeld aan te geven hoeveel punten kandidaten met elke deelopgave kunnen verdienen. Maar opgaven voor de schrijfvaardigheid en zeker de gespreksvaardigheid zijn niet altijd (tot in detail) gestructureerd. De beoordelingswijze kan uiteenlopen van analytisch/gedetailleerd tot globaal/impressionistisch.

Hoe dan ook moeten de kandidaten op de hoogte zijn van de volgende informatie:

- de te behalen maximum totaalscore
- de onderlinge weging van de verschillende beoordelingsaspecten
- de cesuur

Deze informatie is op de school waarschijnlijk slechts globaal vastgelegd. En als de school bijvoorbeeld luistertoetsen van Cito gebruikt, is de kans heel klein dat deze informatie op school is vastgelegd.

Bijlage 2

Toetsingskader Nederlandse taal BB, KB en GT

INHOUD VAN HET SCHOOLEXAMEN

Standaard 1 Exameneenheden

In het schoolexamen zijn alle verplichte exameneenheden opgenomen, conform het Examenprogramma Nederlandse taal BB, KB en GT.

Indicatoren

1.3 Het **schoolexamen BB** bestaat uit de volgende exameneenheden:

EXAMENEENHEDEN		VERPLICHT IN SE	MAG IN SE
NE/K/1	Oriëntatie op leren en werken	X	
NE/K/2	Basisvaardigheden	X	
NE/K/3	Leervaardigheden voor het vak Nederlands	X	
NE/K/4	Luister- en kijkvaardigheid		X
NE/K/5	Spreek- en gespreksvaardigheid	X	
NE/K/6	Leesvaardigheid		X
NE/K/7	Schrijfvaardigheid		X
NE/K/8	Fictie	X	

Norm

In het SE worden de verplichte exameneenheden getoetst, al dan niet geïntegreerd, al dan niet in de vorm van handelingsdelen.

1.4 Het **schoolexamen KB** bestaat uit de volgende exameneenheden:

EXAMENEENHEDEN		VERPLICHT IN SE	MAG IN SE
NE/K/1	Oriëntatie op leren en werken	X	
NE/K/2	Basisvaardigheden	X	
NE/K/3	Leervaardigheden voor het vak Nederlands	X	
NE/K/4	Luister- en kijkvaardigheid	X	
NE/K/5	Spreek- en gespreksvaardigheid	X	
NE/K/6	Leesvaardigheid		X
NE/K/7	Schrijfvaardigheid		X
NE/K/8	Fictie	X	

Norm

In het SE worden de verplichte exameneenheden getoetst, al dan niet geïntegreerd, al dan niet in de vorm van handelingsdelen.

1.5 Het **schoolexamen GT** bestaat uit de volgende exameneenheden:

EXAMENEENHEDEN		VERPLICHT IN SE	MAG IN SE
NE/K/1	Oriëntatie op leren en werken	X	
NE/K/2	Basisvaardigheden	X	
NE/K/3	Leervaardigheden voor het vak Nederlands	X	
NE/K/4	Luister- en kijkvaardigheid	X	
NE/K/5	Spreek- en gespreksvaardigheid	X	
NE/K/6	Leesvaardigheid		X
NE/K/7	Schrijfvaardigheid		X
NE/K/8	Fictie	X	
NE/V/1	Verwerven, verwerken en verstrekken van informatie	X	
NE/V/2	Schrijven op basis van documentatie	X	
NE/V/3	Vaardigheden in samenhang	X	

Norm

In het SE worden de verplichte exameneenheden getoetst, al dan niet geïntegreerd, al dan niet in de vorm van handelingsdelen.

Standaard 2 Niveau

Niveau, aard en inhoud van de toetsen van het schoolexamen komen overeen met de specificaties gegeven bij de onderscheiden eindtermen, conform het overzicht in de bijlage.

Indicatoren

- 2.1 De toetsen voldoen aan de specificaties beschreven in het programma.
- 2.2 Alle toetsen die meetellen voor het schoolexamen worden afgenomen op een zodanig tijdstip dat redelijkerwijs verwacht mag worden dat de kandidaten het gewenste eindniveau bereikt kunnen hebben.
- 2.3a De toetsen zijn van de juiste moeilijkheidsgraad, dat wil zeggen, niet te moeilijk en niet te makkelijk en zijn passend voor het niveau.
- 2.3b De scoring en de beoordeling van alle leerlingen is billijk, dat wil zeggen niet te streng en niet te soepel
- 2.3c De kandidaten hebben een cijfer gekregen passen bij het niveau van hun kennis en vaardigheden.

Standaard 3 Toetsconstructie

De toetsen/onderdelen van het schoolexamen zijn zorgvuldig samengesteld en voldoen aan relevante producteisen.

Indicatoren bij producteisen

- 3.1 Bij de samenstelling van de toetsen is rekening gehouden met de volgende eisen:
 - aansluiting bij de eindtermen en bijbehorende specificaties van het examenprogramma;
 - eenduidigheid van vragen en opdrachten;
 - (vakinhoudelijke correctheid);
 - voldoende afnametijd.

- 3.2 De kandidaten van een bepaald examencohort krijgen bij de toetsen gelijke of vergelijkbare opgaven en/of opdrachten voorgelegd.
- 3.3 De inhoud van de toetsen wordt, waar mogelijk, in overleg met vakcollega's bepaald.

BEOORDELING, NORMERING EN WEGING VAN DE TOETSEN VOOR HET SCHOOLEXAMEN

Standaard 4 Beoordeling

De beoordeling van de toetsen voor het schoolexamen gebeurt zorgvuldig, is relevant en transparant.

Indicatoren

- 4.1 In de beoordeling van de toetsen en eventuele handelingsdelen worden uitsluitend aspecten betrokken die relevant zijn voor de toetsing van de betreffende vaardigheid.
- 4.2 In de beoordeling van de toetsen worden alle relevante beoordelingsaspecten betrokken.
- 4.3 De beoordeling geschiedt waar mogelijk aan de hand van een antwoord- c.q. beoordelingsmodel, met daarin vermelding van de beoordelingscriteria en het gewicht ervan.
- 4.4 Voor de toetsen worden cijfers toegekend; voor handelingsdelen de kwalificaties "naar behoren" of "onvoldoende"
- 4.5 De beoordeling geschiedt door twee docenten.

Standaard 5 Normering

De normering van de toetsen is zo veel mogelijk van tevoren vastgelegd en wordt consequent toegepast.

Indicatoren

- 5.1 Vóór de afname van de toetsen wordt de systematiek van normeren en zo mogelijk de puntenwaardering per opgave of opdracht schriftelijk vastgelegd.
- 5.2 De systematiek van normeren wordt voorafgaand aan of tijdens de afname van de toets kenbaar gemaakt aan de kandidaten.
- 5.3 Bijstelling van de puntentoekenning en/of cesuur achteraf gebeurt op grond van vooraf daarvoor geformuleerde argumenten.
- 5.4 Voor alle kandidaten geldt dezelfde normering.
- 5.5 De normering wordt in onderling overleg met vakcollega's bepaald.

Standaard 6 Weging

De weging van de verschillende onderdelen van het schoolexamen is helder aangegeven in het PTA

Indicator

Per toets is aangegeven wat het gewicht ervan is bij de bepaling van het schoolexamencijfer.

Norm

Aan de indicator moet worden voldaan.

Toelichting bij het toetsingskader Nederlandse taal BB, KB, GT

ALGEMEEN

De centrale vragen zijn:

Voldoet het schoolexamen Nederlands vmbo aan redelijke eisen ten aanzien van:

- inhoud, vorm, omvang, niveau en toetsconstructie;
- beoordeling, normering en weging van de verschillende onderdelen;
- de organisatie;
- de communicatie naar alle betrokkenen.

INHOUD VAN HET SCHOOLEXAMEN

De beoordeling van de inhoud van het schoolexamen vindt plaats aan de hand van de beschrijving van de inhoud door docenten en van concrete opgaven, beoordelings- c.q. antwoordmodellen en eventueel (nagekeken) werken.

Standaard 1 inhoud

Eindtermen

De toetsen en het handelingsdeel hebben tezamen betrekking op de betreffende eindtermen uit het examenprogramma. Het onderwijsprogramma moet alle eindtermen uit het examenprogramma dekken. Dat wil niet per se zeggen dat alle kandidaten op alle eindtermen getoetst moeten worden; per leerling kan de toetsing betrekking hebben op een steekproef van de eindtermen. Evenmin hoeft elk schoolexamen volledig het betreffende (sub)domein van het examenprogramma te dekken. Het mag echter niet zo zijn dat bepaalde onderdelen systematisch in de toetsing worden genegeerd. In toetsen en handelingsdeel mogen geen eindtermen aan de orde komen die niet voorkomen in het examenprogramma.

Standaard 2 Niveau

De vereiste beheersingsniveaus moeten per exameneenheid passend zijn bij de onderscheiden leerwegen.

Standaard 3 Toetsconstructie

Validiteit, betrouwbaarheid en efficiency

De toetsen meten wat, gelet op de vakinhoud, relevant is. Zij zijn valide (zowel naar inhoud als naar het gewenste niveau) en betrouwbaar. Zij meten op een efficiënte en doelmatige wijze. De beoordeling van de prestaties van de leerlingen geschiedt aan de hand van een transparant beoordelingsmodel.

Omvang van de toetsen

De toetsen moeten minimaal een zodanige lengte hebben dat zij de betreffende vaardigheid betrouwbaar kunnen toetsen. Zij moeten echter niet langer zijn dan strikt noodzakelijk is.

Handelingsdeel

Aan de taken die de school formuleert ten aanzien van een handelingsdeel, moeten de kandidaten in redelijkheid kunnen voldoen. Hiervoor liggen geen criteria vast. De beoordeling van de eisen die elke school stelt, moet worden gebaseerd op ervaringsgegevens.

BEOORDELING, NORMERING EN WEGING

Standaard 4 Beoordeling

Centraal bij dit aspect staat de vraag of in de beoordeling van de onderscheiden vaardigheden de juiste aspecten betrokken worden, mede bepaald door de specificaties bij de eindtermen.

Standaard 5 Normering

De mate van gedetailleerdheid van de informatie m.b.t. de normering hangt samen met de aard van de opgave(n) en de gekozen beoordelingswijze. Bij een sterk gestructureerde opgave is bijvoorbeeld aan te geven hoeveel punten kandidaten met elke deelopgave kunnen verdienen. Maar opgaven voor de schrijfvaardigheid en zeker de gespreksvaardigheid zijn niet altijd (tot in detail) gestructureerd. De beoordelingswijze kan uiteenlopen van analytisch/gedetailleerd tot globaal/impressionistisch. Hoe dan ook moeten de kandidaten op de hoogte zijn van de volgende informatie:

- de te behalen maximum totaalscore
- de onderlinge weging van de verschillende beoordelingsaspecten
- de grens tussen voldoende en onvoldoende

Standaard 6 Weging

Het PTA dient helder aan te geven hoe de onderscheiden toetsen meetellen voor het schoolexamencijfer. Als niet wordt voldaan aan de eisen van een handelingsdeel van het PTA, dient helder aangegeven te zijn wat daarvan de consequenties zijn.

De weging der onderscheiden onderdelen is in principe vrij, maar de verschillende examenonderdelen dienen qua gewicht in een redelijke verhouding tot elkaar te staan.

Bijlage bij Toetsingskader Nederlands

Specificatie eindtermen Nederlandse taal vmbo BB / KB/ GL / TL

NE/K/1 Oriëntatie op leren en werken

1. De kandidaat kan zich oriënteren op de eigen loopbaan en het belang van Nederlands in de maatschappij.

NE/K/2 Basisvaardigheden

2. De kandidaat kan basisvaardigheden toepassen die betrekking hebben op communiceren, samenwerken, en informatie verwerven, verwerken en presenteren.

NE/K/3 Leervaardigheden in het vak Nederlands

3. De kandidaat kan strategische vaardigheden toepassen die bijdragen tot:

- het bereiken van verschillende lees-, schrijf-, luister- en kijk-, en spreek- en gespreksdoelen
- de bevordering van het eigen taalleerproces
- het compenseren van eigen tekortschietende taalkennis of communicatieve kennis.

NE/K/4 Luister- en kijkvaardigheid

4. De kandidaat kan:

- luister- en kijkstrategieën hanteren
- compenserende strategieën kiezen en hanteren
- het doel van de makers van een programma aangeven
- de belangrijkste elementen van een programma weergeven
- een oordeel geven over een programma en dit toelichten
- een instructie uitvoeren.

5. De kandidaat kan:

- luister- en kijkstrategieën hanteren
- compenserende strategieën kiezen en hanteren
- het doel van de makers van een programma aangeven
- de belangrijkste elementen van een programma weergeven
- een oordeel geven over een programma en dit toelichten
- een instructie uitvoeren
- de waarde en betrouwbaarheid aangeven van de informatie die door de massamedia verspreid wordt.

NE/K/5 Spreek- en gespreksvaardigheid

6. De kandidaat kan:

- relevante informatie verzamelen en verwerken ten behoeve van de spreek- en gesprekssituatie
- strategieën hanteren ten behoeve van de spreek- en gesprekssituatie
- compenserende strategieën kiezen en hanteren
- het spreek-/luisterdoel in de situatie tot uitdrukking brengen
- het spreek-/luisterdoel en taalgebruik richten op verschillende soorten publiek
- het spreekdoel van anderen herkennen en de reacties van anderen inschatten
- in spreek- en gesprekssituaties taalvarianten herkennen en daar adequaat op inspelen.

NE/K/6 Leesvaardigheid

7. De kandidaat kan:

- leesstrategieën hanteren
- compenserende strategieën kiezen en hanteren
- functie van beeld en opmaak in een tekst herkennen
- het schrijfdoel van de auteur aangeven
- een tekst indelen in betekenisvolle eenheden en de relaties tussen die eenheden benoemen
- het hoofdonderwerp en de hoofdgedachte van een tekst aangeven
- een oordeel geven over de tekst en dit oordeel toelichten.

8. De kandidaat kan:

- leesstrategieën hanteren
- compenserende strategieën kiezen en hanteren
- functie van beeld en opmaak in een tekst herkennen
- het schrijfdoel van de auteur aangeven en de talige middelen die hij hanteert om dit doel te bereiken
- een tekst indelen in betekenisvolle eenheden en de relaties tussen die eenheden benoemen
- het hoofdonderwerp en de hoofdgedachte van een tekst aangeven en een samenvatting geven
- een oordeel geven over de tekst en dit oordeel toelichten.

NE/K/7 Schrijfvaardigheid

9. De kandidaat kan:

- relevante informatie verzamelen en verwerken ten behoeve van het schrijven
- schrijfstrategieën hanteren
- compenserende strategieën kiezen en hanteren
- het schrijfdoel in teksten tot uitdrukking brengen
- het schrijfdoel en taalgebruik richten op verschillende soorten lezerspubliek
- conventies hanteren met betrekking tot schriftelijk taalgebruik
- elektronische hulpmiddelen gebruiken bij het schrijven
- concepten van de tekst herschrijven op basis van geleverd commentaar.

NE/K/8 Fictie

10. De kandidaat kan:

- verschillende soorten fictiewerken herkennen
- de situatie en het denken en handelen van de personages in het fictiewerk beschrijven
- de relatie tussen het fictiewerk en de werkelijkheid toelichten
- een persoonlijke reactie geven op een fictiewerk en deze toelichten met voorbeelden uit het werk.

11. De kandidaat kan:

- verschillende soorten fictiewerken herkennen
- de situatie en het denken en handelen van de personages in het fictiewerk beschrijven
- de relatie tussen het fictiewerk en de werkelijkheid toelichten
- kenmerken van fictie in het fictiewerk aanwijzen
- relevante achtergrondinformatie verzamelen en selecteren
- een persoonlijke reactie geven op een fictiewerk en deze toelichten met voorbeelden uit het werk.

NE/V/1 Verwerven, verwerken en verstrekken van informatie

12. De kandidaat kan zelfstandig informatie verwerven, verwerken en verstrekken in het kader van het sectorwerkstuk.

NE/V/2 Schrijven op basis van documentatie

13. De kandidaat kan een doel- en publiekgerichte tekst schrijven:

- overeenkomstig de voor de tekstsoort geldende conventies
- onder gebruikmaking van documentatie.

NE/V/3 Vaardigheden in samenhang

14. De kandidaat kan de vaardigheden uit het kerndeel in samenhang toepassen

Bijlage 3

Toetsingskader biologie

TOETSINGSKADER VOOR BIOLOGIE IN HET HAVO

De centrale vraag hierbij is:

Voldoet het schoolexamen biologie HAVO (voortaan biologie genoemd) aan redelijke eisen ten aanzien van de inhoud, de vorm, de omvang, de constructie, de normering en het niveau?

Om deze vraag te kunnen beantwoorden is onderstaand model ontwikkeld.

Standaard 1: Inhoud

havo 4 (vernieuwde programma)

De beschrijving van de inhoud van het schoolexamen is dekkend voor dat deel van het examenprogramma biologie dat niet in het centraal examen getoetst wordt.

Indicatoren:

- 1.1 Het schoolexamen bestaat uit een examendossier dat bevat:
 - toetsen met gesloten en/of open vragen
 - praktische opdrachten
- 1.2 De toetsen en de praktische opdrachten hebben tezamen betrekking op de algemene vaardigheden uit de domeinen A1 en de eindtermen in de subdomeinen B2, C2, C3, D2 en E5.
- 1.3 De toetsen en de praktische opdrachten hebben betrekking op de algemene vaardigheden uit domein A2 zoals ze door school in het PTA zijn opgenomen.
- 1.4 De toetsen en de praktische opdrachten hebben betrekking op de door de school in het PTA aangegeven eindtermen in de subdomeinen B1, C1, D1, D3 en E1 t/m E4.
- 1.5 De toetsen en de praktische opdrachten hebben betrekking op de door de school in het PTA aangegeven onderdelen die buiten het examenprogramma vallen.

Norm:

Aan alle indicatoren moet worden voldaan.

Toelichting:

De indicatoren zijn afgeleid van het examenprogramma biologie dat geldt voor leerlingen die in 2007 in havo 4 gestart zijn.

havo 5 (oude programma)

De beschrijving van de inhoud van het schoolexamen is dekkend voor het examenprogramma biologie.

Indicatoren:

- 1.6 Het schoolexamen bestaat uit een examendossier dat bevat:
- toetsen met gesloten en/of open vragen
 - praktische opdrachten
- 1.7 De toetsen en de praktische opdrachten hebben tezamen betrekking op de 49 algemene vaardigheden uit domein A en de 167 eindtermen in de domeinen B, C, D en E, met uitzondering voor het betreffende cohort van de uitsluitingen, aangegeven door de CEVO en de uitsluitingen die door de school in het PTA zijn aangegeven.

Norm:

Aan alle indicatoren moet worden voldaan.

Toelichting:

De indicatoren zijn afgeleid van het examenprogramma biologie dat geldt voor leerlingen die in 2008 examen doen.

Standaard 2: Vaardigheden

De toetsing van voor biologie belangrijke vaardigheden in het schoolexamen is adequaat

Indicatoren:

- 2.1 In het schoolexamen komen vragen en opdrachten aan de orde waarbij met name de volgende vaardigheden van belang zijn:
Subdomein: Onderzoeksvaardigheden

De kandidaat kan:

- een natuurwetenschappelijk probleem herkennen en specificeren.
- verbanden leggen tussen probleemstellingen, hypothesen, gegevens en aanwezige natuurwetenschappelijke voorkennis.
- een natuurwetenschappelijk probleem herleiden tot een onderzoeksvraag.
- hypothesen opstellen en verwachtingen formuleren.
- prioriteiten, mogelijkheden en randvoorwaarden vaststellen om een natuurwetenschappelijk onderzoek uit te voeren.
- een werkplan maken voor het uitvoeren van een natuurwetenschappelijk onderzoek ter beantwoording van een onderzoeksvraag.
- relevante waarnemingen verrichten en (meet)gegevens verzamelen.
- conclusies trekken op grond van verzamelde gegevens van uitgevoerd onderzoek.
- oplossingen, onderzoeksgegevens, resultaat en conclusies evalueren.

- 2.2 In het schoolexamen komen vragen en opdrachten aan de orde waarbij met name de volgende vaardigheden van belang zijn:
Subdomein: Vaardigheden, specifiek voor biologie

De kandidaat kan

- aangeven dat in de levende natuur relaties complex van aard zijn en dat verschijnselen vaak niet monocausaal kunnen worden verklaard, terwijl in onderzoek meestal één factor wordt onderzocht:
 - het geheel is meer dan de delen: systeemdenken.
- biologische verschijnselen op verschillende organisatieniveaus – cel, organisme, ecosysteem – met elkaar in verband brengen.

2.3 In het schoolexamen worden één of meer kortdurende practica uitgevoerd, waarbij ook de technisch-instrumentele vaardigheden getoetst worden
Subdomein: Technisch-instrumentele vaardigheden

De kandidaat kan

- gebruik maken van stoffen, instrumenten en apparaten:
 - voor het in de praktijk uitvoeren van experimenten en technische ontwerpen met betrekking tot de in de domeinen B t/m E genoemde vakinhoud, voor zover veiligheid, milieueisen, kosten en beschikbaar instrumentarium dit toelaten;
 - onder meer: gebruik van loupe en microscoop.
- bij het raadplegen, verwerken en presenteren van informatie en bij het inzichtelijk maken van processen gebruik maken van toepassingen van ICT.
- gebruik maken van micro-elektronica systemen voor het meten en regelen van grootheden.
- aangeven met welke technieken en apparaten de belangrijkste grootheden uit de natuurwetenschappen worden gemeten.
- verantwoord omgaan met stoffen, instrumenten en organismen, zonder daarbij schade te berokkenen aan mensen, dieren en milieu.

Norm:

Aan beide indicatoren moet worden voldaan.

Standaard 3: Omvang en spreiding

Het schoolexamen is evenwichtig gespreid in de tijd over een aantal onderdelen die redelijk van omvang zijn.

Indicatoren:

- 3.1 de omvang van de totale stof voor het schoolexamen;
- 3.2 het aantal toetsen;
- 3.3 de verdeling van de examenstof over de afzonderlijke toetsen;
- 3.4 de spreiding van de toetsmomenten in de tijd.

Norm:

Het intersubjectieve oordeel van deskundigen over bovenstaande indicatoren.

Standaard 4: Kwaliteit

De toetsen/onderdelen van het schoolexamen zijn zorgvuldig samengesteld en voldoen aan relevante producteisen.

Indicatoren:

- 4.1 bij de samenstelling van de toetsen wordt rekening gehouden met de volgende producteisen:
 - eenduidigheid van formulering;
 - betrouwbaarheid;
 - validiteit;
 - transparantie;
 - - adequate vormgeving (lay-out);
 - voldoende afnametijd;
 - geen vakinhoudelijke onjuistheden;
- 4.2 de kandidaten krijgen bij schriftelijke herkansingstoetsen dezelfde/in moeilijkheidsgraad overeenkomstige opgaven en bij praktische opdrachten dezelfde/in vaardigheden en bewerkelijkheid overeenkomstige opdrachten voorgelegd.
- 4.3 de toetsen worden in overleg met vakcollega's samengesteld.

Norm:

Voor de indicatoren 4.1 geldt het intersubjectieve oordeel van deskundigen.
Aan 4.2 moet worden voldaan.

Toelichting:

Naar de mening van de inspectie dienen de toetsen aan relevante producteisen te voldoen. Voor het opstellen van goede toetsen is het niet noodzakelijk dat dit gebeurt in overleg met andere docenten. Docenten kunnen zonder steun van anderen goede toetsen opstellen, als zij evaluatiegegevens voldoende laten meewegen. Het constructieproces zal echter in veel gevallen aan kwaliteit winnen, als dit in overleg met anderen gebeurt.

Standaard 5: Normering

De normering van de toetsen en de praktische opdrachten is op verantwoorde wijze in correctiemodellen vastgelegd en consequent toegepast.

Indicatoren:

- 5.1 vóór de afname van de toetsen en de praktische opdrachten wordt een correctiemodel opgesteld, waarin de puntenwaardering per vraag en de cesuur zijn aangegeven;
- 5.2 de puntenwaardering voor de verschillende vragen wordt aan de kandidaten bij het afnemen van de toets schriftelijk meegedeeld;
- 5.3 de kandidaten zouden vooraf een indicatie moeten hebben van de puntenwaardering voor de verschillende onderdelen van de praktische opdracht
- 5.4 als de puntenwaardering en/of cesuur achteraf wordt bijgesteld, gebeurt dit op grond van argumenten;
- 5.5 bij dezelfde toets en opdracht geldt voor alle leerlingen dezelfde normering.

Norm:

Aan de indicatoren 5.1, 5.2, 5.3 en 5.5 moet worden voldaan.
Voor indicator 5.4 geldt het intersubjectieve oordeel van deskundigen.

Standaard 6: Niveau

De opgaven in het schoolexamen hebben een niveau (moeilijkheidsgraad) dat in vakkringen geldt als HAVO-niveau.

Indicator: Het intersubjectieve oordeel van deskundigen hierover.

Bijlage 4

Toetsingskader wiskunde BB, KB en GT

INHOUD VAN HET SCHOOLEXAMEN

Standaard 1 Inhoud

De inhoud van het schoolexamen voldoet aan de specificaties van het betreffende examenprogramma voor Wiskunde BB, KB of GT, overeenkomstig het schooltype waarvoor het schoolexamen wordt afgenomen.

Indicatoren

1.6 (A.1) Het **schoolexamen BB** bestaat uit de volgende exameneenheden:

EXAMENEENHEDEN		VERPLICHT IN SE	MAG IN SE
WI/K/1	Oriëntatie op leren en werken	X	
WI/K/2	Basisvaardigheden	X	
WI/K/3	Leervaardigheden voor het vak Wiskunde	X	
WI/K/4	Algebraïsche verbanden		X
WI/K/5	Rekenen, meten en schatten		X
WI/K/6	Meetkunde		X
WI/K/7	Informatieverwerking, statistiek	X	
WI/K/8	Geïntegreerde wiskundige activiteiten	X	

Het **schoolexamen KB** bestaat uit de volgende exameneenheden:

EXAMENEENHEDEN		VERPLICHT IN SE	MAG IN SE
WI/K/1	Oriëntatie op leren en werken	X	
WI/K/2	Basisvaardigheden	X	
WI/K/3	Leervaardigheden voor het vak Wiskunde	X	
WI/K/4	Algebraïsche verbanden		X
WI/K/5	Rekenen, meten en schatten		X
WI/K/6	Meetkunde		X
WI/K/7	Informatieverwerking, statistiek	X	
WI/K/8	Geïntegreerde wiskundige activiteiten	X	

Het schoolexamen GT bestaat uit de volgende exameneenheden:

EXAMENEENHEDEN		VERPLICHT IN SE	MAG IN SE
WI/K/1	Oriëntatie op leren en werken	X	
WI/K/2	Basisvaardigheden	X	
WI/K/3	Leervaardigheden voor het vak Wiskunde	X	
WI/K/4	Algebraïsche verbanden		X
WI/K/5	Rekenen, meten en schatten		X
WI/K/6	Meetkunde		X
WI/K/7	Informatieverwerking, statistiek	X	
WI/K/8	Geïntegreerde wiskundige activiteiten	X	
WI/V/2	Verrijkingsoopdrachten	X	
WI/V/3	Verwerven, verwerken en verstrekken van informatie	X	

Norm

In het SE worden de verplichte exameneenheden getoetst, al dan niet geïntegreerd, al dan niet in de vorm van handelingsdelen.

Standaard 2 Niveau

Niveau, aard en inhoud van de toetsen van het schoolexamen komen overeen met de specificaties gegeven bij de onderscheiden eindtermen, conform het overzicht in de bijlage.

Indicatoren

- 2.1 (A.2) De toetsen voldoen aan de specificaties van het betreffende examenprogramma voor Wiskunde BB, KB of GT. (Conform het overzicht in de bijlage.)
- 2.2 (A.3) Alle toetsen die meetellen voor het schoolexamen worden afgenomen op een zodanig tijdstip dat redelijkerwijs verwacht mag worden dat de kandidaten het gewenste eindniveau bereikt kunnen hebben.
- 2.3a (A.4) De toetsen zijn van de juiste moeilijkheidsgraad, dat wil zeggen, niet te moeilijk en niet te makkelijk en zijn passend voor het niveau.
- 2.3b (A.5) De scoring en de beoordeling van alle leerlingen is billijk, dat wil zeggen niet te streng en niet te soepel.
- 2.3c (A.6) De kandidaten hebben een cijfer gekregen passend bij het niveau van hun kennis en vaardigheden.

Standaard 3 Toetsconstructie

De toetsen/onderdelen van het schoolexamen zijn zorgvuldig samengesteld en voldoen aan relevante producteisen.

Indicatoren bij producteisen

- 3.4 (A.7) Bij de samenstelling van de toetsen is rekening gehouden met de volgende eisen:
- aansluiting bij de eindtermen en bijbehorende specificaties van het examenprogramma;
 - eenduidigheid van vragen en opdrachten;
 - (vakinhoudelijke correctheid);
 - voldoende afnametijd.
- 3.5 (A.8) De kandidaten van een bepaald examencohort krijgen bij de toetsen gelijke of vergelijkbare opgaven en/of opdrachten voorgelegd.
- 3.6 (A.9) De inhoud van de toetsen wordt, waar mogelijk, in overleg met vakcollega's bepaald.

BEOORDELING, NORMERING EN WEGING VAN DE TOETSEN VOOR HET SCHOOLEXAMEN

Standaard 4 Beoordeling

De beoordeling van de toetsen voor het schoolexamen gebeurt zorgvuldig, is relevant en transparant.

Indicatoren

- 4.6 (A.10) In de beoordeling van de toetsen en eventuele handelingsdelen worden uitsluitend aspecten betrokken die relevant zijn voor de toetsing van de betreffende vaardigheid.
- 4.7 (A.11) In de beoordeling van de toetsen worden alle relevante beoordelingsaspecten betrokken.
- 4.8 (A.12) De beoordeling geschiedt waar mogelijk aan de hand van een antwoord- c.q. beoordelingsmodel, met daarin vermelding van de beoordelingscriteria en het gewicht ervan.
- 4.9 (A.13) Voor de toetsen worden cijfers toegekend; voor handelingsdelen de kwalificaties "naar behoren" of "onvoldoende"
- 4.10(A.14) De prestaties van de kandidaten bij de toetsen worden zo mogelijk door twee docenten beoordeeld.

Standaard 5 Normering

De normering van de toetsen is zo veel mogelijk van tevoren vastgelegd en wordt consequent toegepast.

Indicatoren

- 5.6 (A.15) Vóór de afname van de toetsen wordt de systematiek van normeren en zo mogelijk de puntenwaardering per opgave of opdracht schriftelijk vastgelegd.
- 5.7 (A.16) De systematiek van normeren wordt voorafgaand aan of tijdens de afname van de toets kenbaar gemaakt aan de kandidaten.
- 5.8 (A.17) Bijstelling van de puntentoekenning en/of cesuur achteraf gebeurt op grond van vooraf daarvoor geformuleerde argumenten.
- 5.9 (A.18) De normering wordt in onderling overleg met vakcollega's bepaald.

Standaard 6 Weging

De weging van de verschillende onderdelen van het schoolexamen is helder aangegeven in het PTA

Indicator

- 6.1 (A.19) Per toets is aangegeven wat het gewicht ervan is bij de bepaling van het schoolexamencijfer.

Toelichting bij het toetsingskader Wiskunde BB, KB, GT

ALGEMEEN

De centrale vragen zijn:

Voldoet het schoolexamen Wiskunde vmbo aan redelijke eisen ten aanzien van:

- inhoud, vorm, omvang, niveau en toetsconstructie;
- beoordeling, normering en weging van de verschillende onderdelen;
- de organisatie;
- de communicatie naar alle betrokkenen.

INHOUD VAN HET SCHOOLEXAMEN

De beoordeling van de inhoud van het schoolexamen vindt plaats aan de hand van de beschrijving van de inhoud door docenten en van concrete opgaven, beoordelings- c.q. antwoordmodellen en eventueel (nagekeken) werken.

Standaard 1 inhoud

Eindtermen

De toetsen en het handelingsdeel hebben tezamen betrekking op de betreffende eindtermen uit het examenprogramma. Het onderwijsprogramma moet alle eindtermen uit het examenprogramma dekken. Dat wil niet per se zeggen dat alle kandidaten op alle eindtermen getoetst moeten worden; per leerling kan de toetsing betrekking hebben op een steekproef van de eindtermen. Evenmin hoeft elk schoolexamen volledig het betreffende (sub)domein van het examenprogramma te dekken. Het mag echter niet zo zijn dat bepaalde onderdelen systematisch in de toetsing worden genegeerd.

In toetsen en handelingsdeel mogen geen eindtermen aan de orde komen die niet voorkomen in het examenprogramma.

Standaard 2 Niveau

De vereiste beheersingsniveaus moeten per exameneenheid passend zijn bij de onderscheiden leerwegen.

Standaard 3 Toetsconstructie

De vorm van (elk onderdeel van) het schoolexamen moet adequaat zijn vanuit het oogpunt van validiteit. De omvang van (elk onderdeel van) het schoolexamen moet adequaat zijn vanuit het oogpunt van toetsbetrouwbaarheid en efficiëntie.

Omvang van de toetsen

De toetsen moeten minimaal een zodanige lengte hebben dat zij de betreffende vaardigheid betrouwbaar kunnen toetsen. Zij moeten echter niet langer zijn dan strikt noodzakelijk is.

Handelingsdeel

Aan de taken die de school formuleert ten aanzien van een handelingsdeel, moeten de kandidaten in redelijkheid kunnen voldoen. Hiervoor liggen geen criteria vast. De beoordeling van de eisen die elke school stelt, moet worden gebaseerd op ervaringsgegevens.

BEOORDELING, NORMERING EN WEGING

Standaard 4 Beoordeling

Centraal bij dit aspect staat de vraag of in de beoordeling van de onderscheiden vaardigheden de juiste aspecten betrokken worden, mede bepaald door de specificaties bij de eindtermen.

Standaard 5 Normering

De mate van gedetailleerdheid van de informatie m.b.t. de normering hangt samen met de aard van de opgave(n) en de gekozen beoordelingswijze. Bij een sterk gestructureerde opgave is bijvoorbeeld aan te geven hoeveel punten kandidaten met elke deelopgave kunnen verdienen. Maar praktische opdrachten zijn niet altijd (tot in detail) gestructureerd. De beoordelingswijze kan uiteenlopen van analytisch/gedetailleerd tot globaal/impressionistisch.

Hoe dan ook moeten de kandidaten op de hoogte zijn van de volgende informatie:

- de te behalen maximum totaalscore
- de onderlinge weging van de verschillende beoordelingsaspecten
- de grens tussen voldoende en onvoldoende

Standaard 6 Weging

Het PTA dient helder aan te geven hoe de onderscheiden toetsen meetellen voor het schoolexamencijfer. Als niet wordt voldaan aan de eisen van een handelingsdeel van het PTA, dient helder aangegeven te zijn wat daarvan de consequenties zijn.

De weging der onderscheiden onderdelen is in principe vrij, maar de verschillende examenonderdelen dienen qua gewicht in een redelijke verhouding tot elkaar te staan.

BIJLAGE

Specificatie eindtermen Wiskunde vmbo BB / KB/ GL / TL

WI/K/1 Oriëntatie op leren en werken

- 1 De kandidaat kan: **BB/KB/GL/TL**
- zich oriënteren op het belang van wiskunde voor de eigen loopbaan en voor zijn functioneren in de maatschappij
 - een relatie leggen tussen wiskundige kennis en vaardigheden en de beroepspraktijk.

WI/K/2 Basisvaardigheden

- 2 De kandidaat kan basisvaardigheden toepassen die betrekking hebben op communiceren, samenwerken, en informatie verwerven, verwerken en presenteren. **BB/KB/GL/TL**

WI/K/3 Leervaardigheden in het vak Wiskunde

- 3 De kandidaat kan structuren en verbanden opsporen in voor hem herkenbare situaties en verbanden leggen met wiskundige begrippen, en daarbij: **BB/KB/GL/TL**
- wiskundige technieken kiezen en gebruiken om problemen op te lossen, waaronder basisalgoritmen en standaardmethodes
 - communiceren door middel van adequaat (wiskundig) taalgebruik
 - adequate onderzoeks- en redeneerstrategieën toepassen.

WI/K/4 Algebraïsche verbanden

- 4 De kandidaat kan problemen oplossen waarin verbanden tussen variabelen een rol spelen, en daarbij: **BB**
- tabellen, grafieken en woordformules hanteren, in het bijzonder bij lineaire verbanden
 - geschikte wiskundige modellen gebruiken.
- 5 De kandidaat kan problemen oplossen waarin verbanden tussen variabelen een rol spelen, en daarbij: **KB**
- tabellen, grafieken en (woord)formules hanteren bij verschillende typen verbanden
 - geschikte wiskundige modellen gebruiken.
- 6 De kandidaat kan problemen oplossen waarin verbanden tussen variabelen een rol spelen, en daarbij: **GL/TL**
- tabellen, grafieken en formules hanteren bij verschillende typen verbanden
 - geschikte wiskundige modellen gebruiken.

WI/K/5 Rekenen, meten en schatten

- 7 De kandidaat kan efficiënt rekenen en cijfermatige gegevens kritisch beoordelen, en daarbij: **BB/KB/GL/TL**
- schatten en rekenen met gangbare maten en grootheden
 - op een verstandige manier de rekenmachine gebruiken.

WI/K/6 Meetkunde

- 8 De kandidaat kan voorstellingen maken, onderzoeken en interpreteren van objecten en hun plaats in de ruimte, en daarbij: **BB**
- redeneren over meetkundige figuren en deze tekenen
 - afmetingen meten, schatten en berekenen
 - meetkundige begrippen, instrumenten en apparaten hanteren.

- 9 De kandidaat kan voorstellingen maken, onderzoeken en interpreteren van objecten en hun plaats in de ruimte, en daarbij:
- redeneren over meetkundige figuren en deze tekenen
 - afmetingen meten, schatten en berekenen
 - meetkundige begrippen en formules, instrumenten en apparaten hanteren.
- KB/GL/TL**
- WI/K/7 Informatieverwerking, statistiek**
- 10 De kandidaat kan informatie verzamelen, weergeven en analyseren met behulp van grafische voorstellingen, en daarbij:
- statistische representatievormen en een graaf hanteren
 - op basis van de verwerkte informatie verwachtingen uitspreken en conclusies trekken.
- BB/KB/GL/TL**
- WI/K/8 Geïntegreerde wiskundige activiteiten**
- 11 De kandidaat kan problemen in alledaagse situaties vertalen naar wiskundige problemen, en daarbij:
- de hierboven genoemde vaardigheden geïntegreerd gebruiken
 - conclusies trekken die relevant zijn voor de bewuste probleemsituatie.
- BB/KB/GL/TL**
- WI/V/1 Aanvullende eisen**
- 12 De kandidaat kan:
- op de verschillende verbanden toegespitste technieken toepassen
 - formules en verbanden op een meer formele manier hanteren
 - complexe rekentechnieken verrichten met behulp van de rekenmachine
 - complexe meetkundige technieken gebruiken.
- GL/TL**
- WI/V/2 Verrijksopdrachten**
13. De leerling verricht complexe opdrachten, waarbij het proces van het probleemgebied kiezen, de probleemsituatie identificeren en mathematiseren, het probleem oplossen, de oplossing terugplaatsen in de oorspronkelijke situatie en reflecteren op het proces wordt doorlopen.
- GL/TL**
- WI/V/3 Verwerven, verwerken en verstrekken van informatie**
14. De kandidaat kan zelfstandig informatie verwerven, verwerken en verstrekken in het kader van het sectorwerkstuk.
- GL/TL**
- WI/V/4 Vaardigheden in samenhang**
15. De kandidaat kan de vaardigheden uit het kerndeel in samenhang toepassen
- GL/TL**

Bijlage 5

Vragenlijst Inventarisatie Examenpraktijk

VRAGENLIJST INVENTARISATIE EXAMENPRAKTIJK

ONDERZOEK KWALITEIT SE

IDENTIFICATIE

Vak: Nederlands Engels Biologie Wiskunde

Nummer beoordelaar:

Naam beoordelaar:

INSTRUCTIE VOOR HET BEANTWOORDEN VAN DE VRAGEN

Met behulp van het scoreformulier heb je beoordeeld in hoeverre het schoolexamen voldoet aan de wettelijke voorschriften en een aantal niet-wettelijk verankerde kwaliteitseisen. Het formulier bevat geen ruimte voor het noteren van de aard van de geconstateerde afwijkingen. Deze informatie is echter nodig om de cijfermatige gegevens in het eindrapport te kunnen interpreteren. Vandaar dat wij je nu vragen de meest frequente afwijkingen te beschrijven. In de vragenlijst is hiervoor onder het kopje "Welke afwijkingen heb je geconstateerd?" ruimte gereserveerd.

Daarnaast vragen we je per indicator enkele voorbeelden van *good practice* te geven. In het rapport willen we niet alleen aandacht schenken aan de tekortkomingen die voor verbetering vatbaar zijn, maar ook een beeld schetsen van hoe een kwalitatief hoogwaardig SE eruit ziet (en de scholen complimenteren met de zaken die wel goed gaan). Je wordt daarom verzocht om naast de afwijkingen ook voorbeelden/kenmerken te geven van een voldoende, goed of uitstekend verzorgd SE. In de vragenlijst noteer je deze voorbeelden/kenmerken onder het kopje "Voorbeelden van good practice".

Het geven van positieve voorbeelden is ook van belang bij de indicatoren waarbij de kwaliteit van het SE op alle of bijna scholen als voldoende werd beoordeeld (zoals bij het vak Nederlands het geval was). Beschrijf in dat geval ook afwijkingen waar je naar gezocht hebt, maar die je (vrijwel) niet hebt kunnen constateren omdat ze zich (bijna) niet voordeden.

Je wordt verzocht je antwoorden zodanig te formuleren dat de beschrijving volledig begrijpelijk is voor een leek die niets van het desbetreffende schoolexamen afweet (zodat we de voorbeelden "panklaar" in het eindrapport kunnen opnemen).

Er is geen onder- of bovengrens voor de lengte van de antwoorden zolang je de afwijkingen of voorbeelden van good practice maar beschrijft in taal die voor derden begrijpelijk is.

Er kunnen natuurlijk belangrijke afwijkingen of voorbeelden van *good practice* zijn die maar heel weinig voorkomen. Wil je deze belangrijke maar weinig voorkomende voorbeelden dan toch beschrijven?

Een aantal indicatoren was niet of moeilijk te beoordelen. In dat geval kun je volstaan met de opmerking dat je deze indicator niet of onvoldoende hebt kunnen beoordelen. Graag met opgave van de reden of oorzaak.

Een vraag van een van de beoordelaars die de beoordeling al had afgerond was waarom we de vragenlijst niet eerder verzonden hebben (zodat hij nog in de gelegenheid zou zijn geweest om zijn antwoorden aan de hand van het ingezonden materiaal te controleren). Het antwoord is dat we jullie hier tijdens de vakantieperiode niet mee lastig wilden vallen. Daarnaast is het niet de bedoeling dat jullie het materiaal nogmaals bestuderen. Je wordt verzocht de vragen "geheel vanuit het geheugen" te beantwoorden.

Het doen van aanbevelingen voor verbetering van de kwaliteit van het SE is een taak van de toetsdeskundigen. Mochten jullie als beoordelaars echter suggesties hebben, zou je die dan op het formulier willen noteren (bij de indicator in kwestie)?

De tijd die je nodig hebt om de vragen te beantwoorden kun je uiteraard volgens de bekende procedures declareren.

Toelichting bij het geven van voorbeelden van afwijkingen

Het geven van voorbeelden van afwijkingen behoeft wellicht enige toelichting. Onder het kopje “Welke afwijkingen heb je geconstateerd” vragen we je de vijf meest voorkomende afwijkingen te beschrijven (zo mogelijk op volgorde van afnemende frequentie). Uiteraard mag je meer afwijkingen beschrijven (dus een F, G, H etc. toevoegen). Heb je tijdens het beoordelen minder dan vijf afwijkingen geobserveerd, dan hoef je er uiteraard geen vijf te beschrijven.

Enkele voorbeelden die laten zien wat de bedoeling is

- Overeenkomstig indicator 1 moet het SE voldoen aan de inhoudelijke specificaties uit het eindexamenprogramma. Noteer bij vraag 1 s.v.p. de vijf meest frequente afwijkingen van de inhoudelijke specificaties die je geconstateerd hebt, zo mogelijk op volgorde van afnemende frequentie.
- Overeenkomstig indicator 2 moet het SE voldoen aan de niveauspecificaties uit het eindexamenprogramma. Noteer bij vraag 2 s.v.p. de vijf meest frequente afwijkingen van de niveauspecificaties die je geconstateerd hebt, zo mogelijk op volgorde van afnemende frequentie.
- Overeenkomstig indicator 4 mogen de toetsen en opdrachten niet te moeilijk en niet te makkelijk zijn (gegeven het door de eindtermen geïmpliceerde niveau en in vergelijking met de moeilijkheid van het centraal examen). Geef bij vraag 4 vijf voorbeelden van te moeilijke of te makkelijke toetsen en opdrachten, zo mogelijk op volgorde van afnemende frequentie.
- Overeenkomstig indicator 7 moet het SE zorgvuldig zijn samengesteld en moeten de toetsen en opdrachten voldoen aan relevante producteisen. Geef bij vraag 7 vijf voorbeelden van een onzorgvuldige samenstelling en geef voorbeelden van toetsen op opdrachten die NIET aan de toetstechnische kwaliteitseisen voldoen (zo mogelijk op volgorde van afnemende frequentie). Bij Engels is bijvoorbeeld een schrijftoets met een afnameduur van een kwartier als onvoldoende beoordeeld omdat schrijfvaardigheid volgens de beoordelaars in een zo korte afnameduur niet voldoende betrouwbaar getoetst kan worden.

Kortom:

- Noteer onder het kopje “Welke regelafwijkingen heb je geconstateerd” vijf afwijkingen die je tijdens het beoordelen geconstateerd hebt (zo mogelijk op volgorde van afnemende frequentie).
- Noteer onder het kopje “Voorbeelden van good practice” vijf voorbeelden of kenmerken van een kwalitatief voldoende, goed of uitstekend SE.

Graag hadden we de ingevulde vragenlijst **vrijdag 12 september a.s.** terugontvangen (graag digitaal per e-mail aan wijnanda.winkel@cito.nl).

1 Verplichte onderdelen SE

Het SE voldoet aan de inhoudelijke specificaties uit het eindexamenprogramma

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

2 Conform specificaties

Het SE voldoet aan de niveauspecificaties uit het eindexamenprogramma

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

- 3 Afnamemoment gerelateerd aan niveau**
Het toetsmateriaal wordt afgenomen op een moment dat de kandidaten het vereiste niveau redelijkerwijs bereikt kunnen hebben

Welke afwijkingen heb je geconstateerd?

- A
- B
- C
- D
- E

Voorbeelden van good practice

- A
- B
- C
- D
- E

- 4 Moeilijkheid toetsen en opdrachten**
De toetsen en opdrachten zijn noch te moeilijk noch te gemakkelijk (gegeven het door de eindtermen geïmpliceerde niveau)

Welke afwijkingen heb je geconstateerd?

- A
- B
- C
- D
- E

Voorbeelden van good practice

- A
- B
- C
- D
- E

5 Soepelheid/strengheid beoordelingsmiddelen
De beoordelingsmiddelen zijn noch te streng noch te soepel

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

6 Becijfering kandidaten
De toegekende cijfers zijn noch te hoog noch te laag

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

7 Conform producteisen

Het SE is zorgvuldig samengesteld en voldoet aan relevante producteisen

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

8 Vergelijkbaarheid opgaven

Als met toetsvarianten wordt gewerkt, is de moeilijkheidsgraad ervan vergelijkbaar

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

- 9 Verplichte toetsinhoud in overleg**
De vakcollega's stellen de inhoud van het SE in overleg vast

Welke afwijkingen heb je geconstateerd?

- A
- B
- C
- D
- E

Voorbeelden van good practice

- A
- B
- C
- D
- E

- 10 Beoordelingsaspecten relevant**
Alle beoordelingsaspecten zijn relevant

Welke afwijkingen heb je geconstateerd?

- A
- B
- C
- D
- E

Voorbeelden van good practice

- A
- B
- C
- D
- E

11 Beoordelingsaspecten volledig

Alle relevante beoordelingsaspecten worden in de beoordeling betrokken

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

12 Gebruik antwoord- c.q. beoordelingsmodel

De beoordeling vindt plaats met een antwoord- c.q. beoordelingsmodel met vermelding van de beoordelingscriteria en het gewicht ervan

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

13 Cijfers voor toetsen en opdrachten

Voor de toetsen en handelingsdelen worden cijfers respectievelijk de kwalificaties “naar behoren” of “onvoldoende” toegekend

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

14 Beoordeling in overleg

De beoordeling geschiedt door tenminste twee docenten

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

15 Scorepunten en/of cesuur vooraf vastgelegd

De scorepunten en indien van toepassing ook de cesuur worden vóór de afname van het toetsmateriaal schriftelijk vastgesteld

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

16 Scorepunten vroegtijdig meegedeeld

De scorepunten worden vóór de afname van het toetsmateriaal aan de kandidaten kenbaar gemaakt

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

17 Bijstelling scorepunten/cesuur beargumenteerd

De scorepunten en indien van toepassing de cesuur worden bijgesteld op basis van argumenten die vóór of tijdens het SE zijn/worden vastgelegd

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

18 Normering in overleg

De vakcollega's bepalen de normering in onderling overleg

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

19 Gewicht

Het gewicht bij de bepaling van het SE cijfer is voor elke toets aangegeven

Welke afwijkingen heb je geconstateerd?

A

B

C

D

E

Voorbeelden van good practice

A

B

C

D

E

VAN HARTE BEDANKT VOOR HET BEANTWOORDEN VAN DEZE VRAGEN!

Bijlage 6

Vakinhoudelijke en andere kanttekeningen bij de resultaten Nederlands

Inleidende opmerking

Bij de beoordeling van de kwaliteit van de huidige schoolexamens Nederlands vmbo kunnen de volgende kanttekeningen worden gemaakt:

- Het vigerende examenprogramma Nederlands vmbo kent weliswaar een heldere domeinindeling, de afzonderlijke eindtermen/kerndoelen zijn doorgaans globaal geformuleerd en specifieke niveauaanduidingen, zoals we die bijvoorbeeld sinds kort kennen bij de moderne vreemde talen (zie het Europese Referentiekader), ontbreken. Ook ontbeert het huidige schoolexamenprogramma een beschrijving van taalcontexten of specifieke taalgebruikssituaties waaraan indicaties van het gewenste niveau ontleend zouden kunnen worden. Te verwachten valt dat in het komende decennium deze niveauaanduidingen en overige specificaties wel in de examenprogramma's voor de moedertaal zullen worden opgenomen (zie bijvoorbeeld de voorstellen van de commissie Doorlopende Leerlijnen en het Raamwerk Nederlands van Cinop). In de onderwijspraktijk lijken docenten Nederlands zich weinig te laten hinderen door wellicht te globaal geformuleerde eindtermen: zij voegen zich naar lang bestaande tradities binnen het vak, dikwijls met behulp van een schoolboekmethode inclusief bijbehorende toetsen.
- Een van de redenen waarom er met het SE weinig mis gaat, is dat de docenten zich veelal bedienen van oud examenmateriaal (of afgeleiden daarvan), bij lezen en schrijven, de luistertoetsen bij Cito inkopen en zich voor het overige bedienen van door de schoolboekmethoden aangeleverde toetsen voor het meten van zaken als spelling en woordenschat (onderdelen die overigens niet verplicht in deze vormen getoetst hoeven worden). Verder geven docenten in de enquêtes keurig aan wanneer wordt samengewerkt (en dat is op vrijwel altijd op alle relevante onderdelen) en geven zij aan met relevante beoordelingscriteria te werken (in hoeverre hier een en ander sociaal wenselijk is ingevuld, is niet na te gaan). Enkele onderdelen onttrekken zich goeddeels aan de waarneming (mondelijke vaardigheden). Waarin scholen zich waarneembaar onderscheiden, is fictieonderwijs, maar daarin worden de scholen door het programma bewust vrij gelaten. Wat bijzonder is, is dat, anders dan op havo en vwo, op de allerlaagste niveaus veel combinatietoetsen (toetsen met een hybride samenstelling van subvaardigheden) voorkomen. Dit type toets verdient uit zuiver meetperspectief geen schoonheidsprijs, maar daar staat tegenover dat een dergelijke toets in hoge mate aansluit bij het gegeven onderwijs, de gebruikte methoden, en het verschijnsel van de korte spanningsboog bij de leerlingen (zeker op BB en KB-niveau), die zich maar met moeite langdurig op één en dezelfde taak (subvaardigheid) kunnen concentreren. Wie binnen BB en KB, zo lijkt de filosofie, niet heel veel (kort) toetst, demotiveert de leerlingen.
- De marges van de regelgeving voor de schoolexamens Nederlands zijn zeer breed. Het programma zelf staat veel vrijheden toe en leent zich derhalve goed voor een eigen schoolspecifieke invulling. Een en ander komt tot uitdrukking in:
 - een grote variëteit in het aantal schoolexamens;
 - een grote variëteit in (al dan niet gecombineerde) toetsonderwerpen;
 - een grote variëteit in het niveau van de stofbehandeling bij enkele vakonderdelen (hieronder verder uitgewerkt).

Volgens de beoordelaars zijn er in het programma te weinig aanwijzingen die tot een uniforme aanpak bij schoolexamens leiden. Daarin is niets geregeld over de frequentie, over de omvang, over de normering, over het gewicht, over beoordelingscriteria, over de volgorde. Ook niveauaanduidingen ontbreken. Dat wil niet zeggen dat de toetsen niet beoordeelbaar zijn. De redenen waarom vrijwel alles voldoende is, zijn hierboven uiteengezet (men speelt op safe en bedient zich van erkende toetsen en beoordelingscriteria).

Algemene opmerkingen per vakonderdeel

Fictie

- Dit vakonderdeel wordt veelal mondeling getoetst (aan de hand van leeslijsten en/of fictiedossiers, waarin boekbesprekingen zijn opgenomen of analyses aan de hand van veelal voorgestructureerde vragenlijsten); ook worden onderdelen van het fictieonderwijs wel als handelingsdeel ingericht; soms wordt het vakonderdeel Fictie gecombineerd getoetst met het vakonderdeel Spreek- en gespreksvaardigheid;
- Aangehouden zijn leeslijsten (fictie) van in kwantitatieve en kwalitatieve zin uiteenlopende samenstelling, waarin soms duidelijk de signatuur van de school spreekt. Doorgaans sluiten deze lijsten goed aan bij de belevingswereld van de leerlingen. Een landelijke erkende en gehanteerde canon voor (jeugd)literatuur ontbreekt. "Aansluiten bij de belevingswereld" is overigens een buitengewoon subjectief begrip. Enkele maanden terug is Theo Witte (vakdidacticus aan de universiteit van Groningen) gepromoveerd op een proefschrift waarin voor het eerst iemand criteria aanreikt voor het onderscheiden van zes bruikbare niveaus, bestemd voor het VO. Een verplichte canon is er nog steeds niet (en die komt er vermoedelijk ook niet, omdat het de vrijheid van scholen raakt). Het staat (gereformeerde) scholen vrij vierdeklassers uitsluitend stichtelijke literatuur te laten lezen. De Inspectie en het Ministerie van OC&W heeft die vrijheid al lang onderkend. Niettemin kunnen de beoordelaars op grond van hun eigen ervaringen heel goed vaststellen of de lijsten niet te zwaar of niet te licht zijn.
- Aangehouden bij enkele GT-klassen zijn zuivere kennistoetsen aangaande de literaire theorieën inzake verhaal- en in een enkel geval ook wel poëzieanalyse; strikt genomen is dit geen verboden en zelfs bruikbare kennis, maar het is wenselijk deze kennis te koppelen aan door de leerling gelezen werken en vooral ook de leeservaring in de toetsing te betrekken.

Spelling en grammatica (Basisvaardigheden)

- aangehouden is een grote hoeveelheid (kleinere) toetsjes spelling, grammatica (zinsontleding) en woordkennis (veelal door schoolboekmethoden zelf aangeleverd);
- met name op het vlak van spelling zien we een diversiteit aan moeilijkheidsgraden en onderwerpen; de meeste scholen besteden veel aandacht aan de werkwoordspelling; een duidelijke 'taalkundige' canon ontbreekt evenwel, zodat de verschillen in aandacht voor de andere onderdelen (zoals hoofdletters, interpunctie en samenstellingen e.d.) groot zijn.

Schrijfvaardigheid

- Aangehouden is een grote hoeveelheid aan veelal ('functionele') schrijfoopdrachten, veelal een afspiegeling van het type opdracht, inclusief beoordelingsmodel, dat men ook in het centrale examen kan aantreffen (e-mailbericht, zakelijke brief, waaronder de sollicitatiebrief, formulieren, vragenlijsten en stageverslagen);
- In een enkele toets werden te veel aansturingelementen aangehouden, waarbij bijvoorbeeld op één school een brief die geschreven moest worden, een zeer hybride karakter kreeg. De reden dat de beoordelaars hier niet het oordeel "onvoldoende" gaven, is dat het hier slechts één opdracht betrof in een reeks van opdrachten die verder wel in orde waren.

'Gedocumenteerd schrijven'

Waargenomen is dat lang niet alle scholen grip hebben op dit onderdeel (vermoedelijk veroorzaakt door de onduidelijke eindtermen ter zake); ook de notie 'gedocumenteerd schrijven' is in het programma onvoldoende van een vakinhoudelijke duiding voorzien, zodat scholen dit onderdeel laten rusten dan wel bij een (een combinatie van) andere vakken onderbrengen (of koppelen aan andere onderdelen, zoals het fictiedossier of sectorwerkstuk).

Mondelinge vaardigheden

Lang niet alle scholen werken met een protocol of beoordelingsformulier (althans ze werden zelden aangehouden). Dergelijke formulieren zijn natuurlijk niet verplicht, maar wel uiterst bruikbaar en ze komen een vergelijkende beoordeling zeker ten goede.

Leesvaardigheid

Een enkele keer werd een zuivere theorietoets aangehouden (met vragen als 'welke vier functies kan de inleiding van een tekst hebben'). Feitelijk gaat het dan om een toets die niet afsluitend zou mogen zijn. De reden dat de beoordelaars hier geen onvoldoende hebben gegeven is dat docenten geen regels overtreden als zij extra activiteiten in de vrije ruimte ontplooiën. Wel moet het aangeboden natuurlijk niet te moeilijk zijn voor de leerlingen en relevantie voor het vak hebben.

Slotbeschouwing

- De beoordelaars zijn van mening dat de scholen hun schoolonderzoekprogramma vrijwel zonder uitzondering goed hebben ingericht en verzorgd aanbieden. De meeste toetsen lijken passend bij het niveau van de leerlingen van de onderscheiden leerwegen, de becijfering is billijk, de inhoud is doorgaans valide. Docenten beoordelen het werk van de leerlingen doorgaans op relevante beoordelingsaspecten. Waar nodig, werken docenten samen. Eén enkele school werkt met meer onbevoegde dan bevoegde docenten, wat zorgwekkend is. Dat de beoordelaars de kwaliteit van het SE niet vaker onvoldoende hebben aangemerkt, heeft vooral te maken met de hoeveelheid toetsen. Stel dat een school vijf schrijfp opdrachten aanbiedt, waarvan er één niet helemaal aan alle kwaliteitscriteria voldoet, is dat voor de beoordelaars geen reden geweest om de inspectie te adviseren deze school een reprimande te geven.
- Daar waar docenten zelf toetsen of onderdelen van toetsen maken, laat de vraagtechniek (in het bijzonder de vraagformulering) wel eens te wensen over (in het bijzonder geldt dat voor meerkeuzevragen).
- Opvallend is de ruime en herhaalde aandacht voor deelvaardigheden, zoals de spelling van de werkwoorden, die in de toetsing in klas drie en vier aan bod komt.
- De schrijfp opdrachten zijn doorgaans zinvol en functioneel en zijn een afspiegeling van het normale taalverkeer.
- Diffuser is de vormgeving van het onderdeel Fictie en de bijbehorende toetsing en op dit punt lijken de verschillen tussen de scholen het grootst.
- Met het verrijksdeel (V1, V2, V3), ruwweg te vertalen in 'Gedocumenteerd schrijven', lijken veel docenten slecht uit de voeten te kunnen en lopen de interpretaties over wat het programma voorschrijft sterk uiteen (NB dit is bij havo en vwo op dit onderdeel niet anders). Bij een toekomstige revisie van het programma zouden de onduidelijkheden met betrekking tot deze eindtermen moeten worden weggenomen.
- Voorts zou het gewenst zijn, uit het oogpunt van landelijke uniformiteit, die bovendien een volgende kwaliteitsbeoordeling zou kunnen vergemakkelijken, in de programma's meer indicaties op te nemen over het gewenste taalvaardigheidsniveau (bijvoorbeeld door ook contexten aan te geven) en de afzonderlijke onderdelen van de taalbeheersing nader te specificeren.
- Voor wat betreft de weging van de verschillende onderdelen zou, uit het oogpunt van gelijkwaardige behandeling van de leerlingen, een landelijk voorschrift gewenst zijn. Uniformiteit in het programma kan voorts bevorderd worden door onderdelen die in het centraal examen zitten, maar die door veel scholen toch ook in het schoolonderzoekprogramma zijn opgenomen, voor alle scholen ook verplicht te stellen voor de schoolexamens.