

EXAMENCONSTRUCTIE: EEN LANGDURIG EN ZORGVULDIG PROCES

Over de rol van CEVO, Citogroep en docenten bij de totstandkoming van de eindexamens

[Ameling Algra en Ger Limpens]

Inleiding

Als het poldermodel niet bestond, dan zou het in examenland uitgevonden kunnen zijn. Ieder Nederlands eindexamen is namelijk het product van heel veel overleg van heel veel verschillende gremia, groepjes en belangenvertegenwoordigers. In onderstaand verhaal proberen we een beeld te geven van dit proces. In de ogen van veel leerlingen en ook van nogal wat docenten worden eindexamens vervaardigd door toetsenbakkers in een ivoren toren die slechts een keer per jaar de deur naar de echte wereld op een kier zetten. In werkelijkheid wordt een examen echter gemaakt door een groepje van docenten onder leiding van een toetsdeskundige van de Citogroep in Arnhem. Verder is er nog een vaksectie van de CEVO die de voorstellen van deze constructiegroep aandachtig bestudeert, van commentaar voorziet en uiteindelijk goedkeurt. Tot slot wordt er, na afname van het examen, een analyse gemaakt van de wijze waarop leerlingen dit examen volbracht hebben. Aan de hand van deze analyse, en van eventuele geconstateerde tekortkomingen, wordt door de CEVO de normerings-term vastgesteld.

CEVO

De Centrale Examencommissie Vaststelling Opgaven, kortweg CEVO, bestaat uit een Algemeen Bestuur, waarin het georganiseerde onderwijsveld is vertegenwoordigd, en uit een vaksectie voor ieder examenvak. De vaksectie bestaat uit een deskundig voorzitter, afkomstig uit het hoger onderwijs, en enkele leden die als leraar lesgeven aan examenkandidaten. De CEVO zelf produceert geen examenopgaven.

Zij geeft, onder verantwoordelijkheid van de Minister van Onderwijs, een gespecificeerde opdracht voor de productie van de opgaven en de toetsen voor het centraal examen aan de Citogroep. In hoofdlijnen wordt in de opdracht van CEVO aan Citogroep vastgelegd hoe de examens eruit moeten zien. De specificaties noemen de ook aan de scholen bekend-gemaakte exameneenheden en domeinen die in het centraal examen aan de orde moeten komen, en geven verder globale richtlijnen over het aantal scorepunten (met een marge) en waar van toepassing de verdeling over vraagvormen zoals open vragen en meerkeuzevragen (zoals bekend bij wiskunde steeds 100-0).

Citogroep en constructiegroep

Bij de Citogroep zijn toetsdeskundigen, vaak zelf afkomstig uit het onderwijs, betrokken bij de totstandkoming van deze opdracht. De Citogroep huurt per vak de expertise van een aantal docenten in. Deze docenten vormen, samen met een toetsdeskundige, de zogeheten constructiegroep. Voor een standaard-examen bestaat zo'n constructiegroep uit drie ervaren leerkrachten die elk voor een beperkt aantal uren (afhankelijk van het niveau van het examenvak, dit vanwege het feit dat een vmbo-examen nu eenmaal korter is dan een havo/vwo-examen) werkzaam zijn –meestal op detacheringsbasis– voor de Citogroep. Deze docenten staan voor de rest van hun werktijd dus gewoon voor de klas. Wat men verder ook van de kwaliteit van het examen kan vinden, kritiek in de vorm van opmerkingen waarmee geïnspireerd wordt dat examenmakers geen idee hebben van de realiteit

in de klas van tegenwoordig, is daarmee zeker onterecht. Een docent die constructiegroepslid wordt, wordt verondersteld voor een dag of dagdeel per week vrijgeroosterd te kunnen worden. Op dit lesvrije moment vindt diverse keren per jaar een constructiegroepsvergadering plaats. De andere weken kan deze dag of dit dagdeel gebruikt worden om thuis werkzaamheden voor de constructiegroep te verrichten.

Constructiegroepsleden worden via advertenties in de grote landelijke dagbladen geworven en worden per jaar benoemd met een maximale aanstellingsduur van negen jaar. Dat een constructiegroepslid voor een maximale termijn van 'slechts' negen jaar aan de slag mag binnen deze constructiegroep heeft als achtergrond dat men op deze manier hoopt te voorkomen dat een constructiegroep te zeer een eigen gezicht krijgt, en te zeer zijn eigen stempel gaat drukken op het betreffende examen.

Examenconstructie

Voor ieder van de verschillende wiskundevakken is in de loop der jaren een zogenoemde bank ontstaan. Dit is een verzameling van losse opgaven die in meer of mindere mate door de constructiegroep geschikt geacht worden voor opname in een examen. De constructiegroep heeft dan ook als kerntaak het aanleveren van geschikte examenopgaven, meestal vergezeld van een rudimentair correctievoorschrift. Daar blijft het echter niet bij.

Als de constructie van het betreffende examen een aanvang neemt, wordt een greep gedaan uit deze opgavenbank. In onderling overleg wordt door de constructiegroep een examenconcept samengesteld. Bij deze greep wordt onder meer rekening gehouden met zaken als diversiteit van contexten en lay-out (denk daarbij aan illustraties e.d.). Uiteraard is deze greep nog niet voldoende om er een volledig examen van te maken. Zo zou een op het eerste gezicht geschikte greep voor een vwo A12-examen bijvoorbeeld een verzameling opgaven kunnen opleveren waarbij zoets als een hypothesetoets ontbreekt. Als een dergelijk aspect toch in het betreffende examen opgenomen dient te zijn, betekent dit dat de greep heroverwogen dient te worden dan wel de bewuste verzameling opgaven uitgebreid dient te worden met een vraag rond het verschijnsel hypothesetoets. Na een proces van slijpen en vijlen wordt op zeker moment een eerste concept inclusief correctievoorschriften aan de CEVO-vaksectie voorgelegd. Daar wordt een eerste oordeel gegeven over het concept. Dit kan er soms toe leiden dat bepaalde opgaven uit het concept gegooid worden en vervangen door andere. Vaker gebeurt het echter dat niet zozeer hele opgaven als wel specifieke vragen minder goed vallen, en in een dergelijk geval krijgt de constructiegroep de opdracht mee deze activiteiten te vervangen door andere die beter passen bij de visie die de vaksectie over het betreffende examen heeft.

Bij de beoordeling van de kwaliteit van examenconcepten maakt de vaksectie ook wel gebruik van

deskundigen die niet direct betrokken zijn bij de examenconstructie. Dat kunnen leden zijn van andere constructiegroepen, maar dat kunnen ook deskundigen zijn afkomstig uit bijvoorbeeld vervolgonderwijs. Deze zogeheten screeners voorzien de vaksectie van onafhankelijk commentaar en geven tevens vaak suggesties hoe sommige contexten c.q. vragen verbeterd zouden kunnen worden. Ook komt het voor dat opgaven uitgetest worden in een try-out bij een populatie die op de een of andere wijze vergelijkbaar is met de uiteindelijke doelgroep. Zo kan een constructiegroepslid opgaven waarover men informatie wil hebben, enkele jaren voor afname uittesten in zijn eigen examenklas van het betreffende niveau. Op deze wijze is te controleren of een opgave niet allerlei onbedoelde misvattingen bij leerlingen oproept, of toch veel te eenvoudig dan wel te moeilijk is. Na evaluatie van een dergelijke proeftest wordt de opgave waar nodig bijgesteld.

Op zeker moment wordt het examen als nagenoeg afgerond beschouwd en op grond van de vergaarde informatie wordt ook vastgesteld hoe moeilijk het examen naar alle waarschijnlijkheid in zijn geheel is. Vaak blijkt pas op dat moment, dat het betreffende examen als totaliteit toch moeilijker dan wenselijk is met als gevolg dat er toch weer het een en ander vertimmerd dient te worden.

Productie- en controlefase

Nadat de vaksectie er haar fiat aan gegeven heeft, wordt het examen klaargemaakt voor productie. Tijdens het constructieproces zijn de noodzakelijke illustraties al vervaardigd en is ook een beslissing genomen over de volgorde waarin de opgaven in het examen gepresenteerd worden. Toch vraagt die laatste fase van het productieproces nog een behoorlijke tijd. Alle opgaven en vragen worden nogmaals door iemand die niet bij de constructie betrokken is, gecontroleerd in samenhang met het gedachte antwoordmodel. Ook de meegeleverde illustraties worden op dat moment weer opnieuw tegen het licht gehouden. Op zeker moment wordt dan het licht op groen gezet voor het drukproces dat als eerste resultaat een drukproef oplevert die ook weer gecontroleerd wordt. Dit is echter niet zozeer een inhoudelijke controle als wel een controle waarbij geverifieerd wordt of er geen veranderingen plaats hebben gevonden tussen gedachte en gedrukte versie. Toch wordt er nog wel eens bij deze drukproefcontrole een inhoudelijke fout rechtgezet. En tot slot vindt er circa twee maanden voor de afnamedatum nog een controle onder de naam 'erratumprocedure' plaats waarbij Citogroepvoorzitter deskundige en vaksectievoorzitter onafhankelijk van elkaar het betreffende werk doornemen op zoek naar fouten van welke aard dan ook. Mocht er op dat moment nog iets worden aangetroffen dat rechtgezet dient te worden, dan treft men dit naderhand tijdens de examencampagne aan in de vorm van een erratum dat voorgelezen dan wel uitgereikt wordt tijdens de examensessie of bij de verstrekking van het correctievoorschrift aan docenten wordt meegeleverd.

Toch nog een fout?

Ondanks al deze controles blijkt toch af en toe pas bij of na de afname, dat in een examen of in het correctievoorschrift een storende fout is blijven staan. Als een docent zo'n fout constateert, moet hij dat melden bij de Examenlijn van de CEVO. Blijkt het inderdaad een fout te zijn, dan hangt het vooral van het moment van melding af, wat er gaat gebeuren.

Als de fout tijdig gemeld is, wordt een aanvulling op het correctievoorschrift gepubliceerd, die bijvoorbeeld meldt dat elk leerlingantwoord moet worden goed gerekend. Als er al wat tijd is verstreken, wordt zo'n aanvulling niet meer gepubliceerd. Als de CEVO dat wel zou doen, zou de correctieprocedure erg vertraagd en gefrustreerd worden. Voor deze vraag moet de hele stapel weer worden doorgenomen, en misschien gaan docenten dan ook wachten op aanvullingen die niet komen. Zo'n latere melding wordt betrokken bij de vaststelling van de normeringsterm.

Een enkele keer komt het voor dat er sprake is van een kennelijke fout, maar wordt er toch geen aanvulling uitgebracht. Een voorbeeld (van een ander vak; wiskunde kent geen meerkeuzevragen): bij een vierkeuzevraag bleken alternatief 3 en 4 identiek. Omdat alternatief 1 het juiste antwoord was, leverde deze fout geen problemen op, noch voor de corrector, noch voor de kandidaat. Voor de laatste hoogstens lichte verbazing, maar omdat het de facto een driekeuzevraag is geworden, werd hij alleen maar iets eenvoudiger.

Ook in een andere situatie verwachten docenten soms een aanvulling maar wordt die niet uitgebracht. Kandidaten komen wel eens met een vakinhoudelijk volledig correcte oplossing die echter in het correctievoorschrift niet is voorzien. Dan is er geen sprake van een fout in het

correctievoorschrift, maar is artikel 3.3 van de algemene correctieregels van toepassing ('indien een antwoord op een open vraag niet in het beoordelingsmodel voorkomt en dit antwoord op grond van aantoonbare, vakinhoudelijke argumenten als juist of gedeeltelijk juist aangemerkt kan worden, moeten scorepunten worden toegekend naar analogie of in de geest van het beoordelingsmodel') en kan de docent gewoon de punten geven. Hij moet het dan wel met de tweede corrector eens worden over de vakinhoudelijke juistheid.

Analyse

Na afloop van het examen worden de resultaten van de versnelde correctieprocedure met spanning afgewacht bij de Citogroep. Iedere school stuurt de resultaten van de eerste vijf kandidaten van de examenkandidatenlijst tijdig naar Arnhem. Aan de hand van deze representatieve steekproef wordt een analyse gemaakt van de wijze waarop de kandidaten het examen gemaakt hebben. Daarbij wordt uiteraard niet alleen gekeken naar het gemiddelde resultaat over het gehele examen, maar ook naar de gemiddelde score per vraag en de mate waarin het antwoordgedrag van de gemiddelde leerling bij een specifieke vraag overeenstemt met het antwoordgedrag over het hele examen. Dit in de veronderstelling dat een goede vraag eigenlijk een afspiegeling dient te zijn van het examen in zijn geheel. Een goede vraag is daarmee een vraag die door de betere leerling (hij die goed scoort voor het hele examen) beter gemaakt wordt dan door een minder goede leerling (iemand die minder goed scoort voor het hele examen). Ook de deelscores per vraag worden geanalyseerd zodat we een beeld krijgen van de wijze waarop de populatie het totaal aantal punten bij iedere vraag verdiend heeft.

FIGUUR 1

Voorbeeld van een schema van activiteiten rond examenconstructie betreffende het examen van jaar n

jaar	periode	activiteit
$n - 3$	voorjaar	Eerste werkzaamheden aan examen: selectie opgaven voor try-outversie en geschikt maken van deze geselecteerde opgaven voor examenconcept
	mei	Presentatie try-outversie aan vaksectie
	juni/september	Verwerken commentaar vaksectie
	september/oktober	Presentatie try-outversie screeners
	oktober/november	Evalueren commentaar screeners
	november/december	Verwerken commentaar screeners en visie vaksectie op commentaar en try-outversie
$n - 2$	voorjaar	Voorleggen try-outversie aan een pretestpopulatie
	oktober/december	Reparaties aan testversie op basis van commentaren/ervaringen/resultaten try-out
$n - 1$	januari	Presentatie concept aan vaksectie
	februari/augustus	Reparaties aan conceptexamen
	september/oktober	Vaststelling en inlevering definitieve examen
	december	Drukproefcontrole
n	januari	Drukproefcontrole
	maart	Erratumprocedure
	mei/juni	Afname examen
	juni	Analyse en vaststelling N-term

(De constructie van nieuwe opgaven vindt voortdurend tussen de in het schema opgenomen activiteiten door plaats.)

Vaststelling normeringsterm

Aan de hand van de analyse wordt door de CEVO de normeringsterm (N-term) vastgesteld; zie voor een uitgebreidere toelichting van het begrip normeringsterm bijvoorbeeld www.kennisnet.nl/vo/examen/docenten/normering/ (doorklikken via examenblad). Zowel de gebleken totale moeilijkheidsgraad van het examen als vragen die om de een of andere reden mislukt zijn, spelen een rol bij de vaststelling. Ook reacties van het veld in de vorm van bijvoorbeeld de verslagen van de regionale examenbesprekingen worden meegenomen in deze discussie.

Tijdpad

Bovenstaande activiteiten staan schematisch en voorzien van een tijdpad in [figuur 1](#).

Tweede en derde tijdvak

Bovenstaande werkzaamheden hebben betrekking op de constructie van het eerste tijdvak. Voor tweede en derde tijdvak vinden min of meer dezelfde werkzaamheden plaats, uiteraard enigszins verschoven in de tijd. Ook bij het tweede tijdvak wordt tegenwoordig een analyse gemaakt en bestudeerd om de N-term te kunnen bepalen. Natuurlijk is het daarbij wel van belang te bedenken dat de populatie die deelneemt aan dat tweede tijdvak niet vergelijkbaar is met de totale populatie die deelneemt aan het eerste tijdvak. Betere leerlingen zullen bij dit tweede tijdvak ongetwijfeld in veel mindere mate aanwezig zijn. In deze analyse wordt ook gekeken naar het resultaat dat de deelnemende kandidaat behaald heeft voor zijn eerste tijdvakexamen.

Bij het derde tijdvak, dat tijdens de zomervakantie wordt afgenomen, worden geen analyses gemaakt. Kwantitatieve evaluatie van deze examens is dan ook niet mogelijk.

Slotwoord

Ondanks de grote mate van zorgvuldigheid waarmee examens gemaakt worden, blijkt toch ieder jaar weer dat fouten hierbij niet vermeden kunnen worden. Alle bij de examens betrokken medewerkers nemen dan ook steevast ergens na de eerste tijdvakken de tijd om 'hun' examens te evalueren. Procedures worden tegen het licht gehouden en waar nodig bijgesteld. En serieuze reacties uit het veld worden gebruikt om eventuele blinde vlekken bij de constructeurs weg te werken. Gelukkig is het merendeel van de respons uit het veld zodanig, dat examenconstructeurs ook iets met de kritiek kunnen. De enkeling die in zijn reactie blijkt geeft geen inzicht te hebben in de huidige examens en examenprogramma's, daarmee in wezen uitdragend geen enkel geloof te hebben in de kwaliteit van het huidige voortgezet onderwijs, wordt ook door de examenmakers niet serieus genomen.

Over de auteurs

Ameling Algra is projectmanager bij de CEVO (website: www.cevo.nl).

Ger Limpens is wiskundemedewerker en toetsdeskundige bij de Citogroep (website: www.citogroep.nl).

Hun e-mailadressen zijn A.Algra@cevo.nl en

Ger.Limpens@citogroep.nl.

Aankondiging / Studiedag 'UITWISKELING LIVE'



Al twintig jaar confronteert het tijdschrift UITWISKELING de wiskundeleerkracht in Vlaanderen en Nederland met nieuwe ideeën uit de wiskunde-didactiek. Om de afsluiting van dit tweede en de aanhef van het derde decennium te vieren

organiseert UITWISKELING op zaterdag 20 november 2004 een jubileumviering in de gebouwen van de Ehsal, op wandelafstand van het centraal station in Brussel.

Feesten en werken gaan hand in hand. Daarom wordt aan de meevierende leerkrachten gevraagd tekenmateriaal en een grafische zakrekenmachine mee te brengen en zich op voorhand in te schrijven voor drie

workshops. Op de website (www.uitwiskeling.be) kunnen belangstellenden een elektronische inschrijfstrook vinden met tien keuze-onderwerpen voor deze workshops, bv. 'rekenen met tandwielen', 'is de Belg een bedreigde diersoort?', 'wiskunde en aardrijkskunde', 'meetkundig denken rond een cirkel'.

Bij de selectie van de onderwerpen voor deze seminars is rekening gehouden met de onmiddellijke bruikbaarheid in de klas, met de nieuwe leerplannen en met de evenwichtige spreiding over de drie graden.

Als afsluiter van deze (inter)nationale feestdag werd de Schotse vuurspuwer-éénwieler-jongleur-en-doctor-in-de-zuivere-wiskunde (Cambridge University) Colin Wright uitgenodigd. Of zijn activiteiten direct toepasbaar zullen zijn in onze klassen, kan niet waterdicht en vuurbestendig door de organisatoren verzekerd worden.