

Actie nodig om vwo-niveaudaling wiskunde en natuurkunde te keren

Natuurkundigen Loek Zonnenberg en Paul Rutten hebben onderzoek gedaan naar de vwo-eindexamens wiskunde en natuurkunde in de afgelopen dertig jaar. Ze hebben daarover een rapport uitgebracht onder de vlag van McKinsey & Company. Hoewel voor de examens steeds hogere cijfers worden behaald, concluderen ze dat het niveau van de eindexamenleerlingen in de loop der jaren flink omlaag is gegaan.

Er is sprake van een paradox in het Nederlandse onderwijs. Enerzijds is het gemiddelde cijfer voor de centrale vwo-eindexamens wiskunde en natuurkunde en het aandeel leerlingen met een vwo-diploma fors gestegen in de afgelopen dertig jaar. Het gemiddelde vwo-cijfer voor het centraal examen (ce) in 1990-1995 ten opzichte van 2015-2021 steeg bij wiskunde B van 5,8 naar 7,0, bij wiskunde A van 6,3 naar 6,9 en bij natuurkunde van 6,2 naar 6,6. Reden om de vlag uit te hangen, zou je zeggen, maar anderzijds lopen de scores van Nederlandse leerlingen op andere nationale en internationale toetsen terug. Hoe valt dit te verklaren? Ondanks de inspanningen van vele gemotiveerde docenten, laat ons nieuwe onderzoek [1] zien dat deze hogere cijfers een aanzienlijke niveaudaling verbergen. Voor het onderzoek hebben we de ce's van de afgelopen dertig jaar bestudeerd en meer dan 1.500 opgaven geanalyseerd. Hierin hebben we in detail de periode 1990-1995 vergeleken met de periode 2015-2021; elk van beide perioden bevat zes jaar aan ce's (in 2020 was er geen ce vanwege de coronapandemie). Ons onderzoek laat zien dat het niveau van vwo-eindexamenleerlingen is gedaald en dat de ce's

aanzienlijk makkelijker zijn geworden. De geconstateerde effecten zijn bij wiskunde groter dan bij natuurkunde. Dit vraagt om actie.

De geconstateerde niveaudaling komt niet uit de lucht vallen. In Nederland waarschuwden de Onderwijsinspectie [2] en de Onderwijsraad [3] eerder al dat de reken- en taalvaardigheden van Nederlandse leerlingen teruglopen. Ook het zogeheten PISA-onderzoek, dat om de drie jaar het niveau van alle vijftienjarigen toetst in OESO-landen, signaleert deze trend al langer [4]: in Nederland liepen de prestaties op wiskunde en op natuurwetenschappen volgens dit onderzoek de laatste vijftien jaar al flink terug, zowel in absolute zin als in vergelijking met andere landen. Dit heeft al praktische gevolgen: Nederlandse studenten leggen het vaker af tegen internationale leeftijdsgenoten bij bijvoorbeeld het bemachtigen van een studieplek bij studies met een numerus fixus zoals Luchtvaart- en Ruimtevaarttechniek in Delft [5]. Maar ondertussen stijgen de cijfers voor de centrale vwo-eindexamens wiskunde en natuurkunde. Hoe kan dat?

Om de periode 1990-1995 te kunnen vergelijken met de periode 2015-2021 en de ce-cijferstijging te kunnen

verklaren hebben we alle factoren die van invloed zijn op de ce-cijferstijging benoemd en kwantitatief onderzocht. We hebben deze factoren geclusterd in drie categorieën:

1. Wat is de ontwikkeling in de moeilijkheidsgraad van het ce? Het gaat dan onder andere om de breedte/hoeveelheid en diepte/moeilijkheidsgraad van de stof, het vereiste tempo en uithoudingsvermogen, de moeilijkheidsgraad van de gestelde vragen, welke hulpmiddelen zijn toegestaan en hoe streng de correctie is.
2. Hoe streng is de normhandhaving (door middel van ijking) en gebeurt dit op meer absolute wijze (met behulp van ankervraagstukken door de jaren heen) of relatieve wijze (ten opzichte van de deelnemende populatie in dat jaar)?
3. Wat is de ontwikkeling van het niveau van de vwo-ce-kandidaten? Hierbij gaat het onder andere om het IQ van de gehele ce-populatie (alle onderwijsvormen), welk deel van de gehele ce-populatie deelneemt aan het specifieke vwo-ce voor een specifiek vak en de mate waarin de vwo-ce-kandidaten zijn voorbereid.

We proberen hierbij dus onder andere te bepalen of leerlingen van nu het beter of slechter doen dan leerlingen van dertig jaar geleden op de stof en toetsitems die in beide perioden in het centrale examen zitten. Het is niet onze bedoeling (of verwachting) dat leerlingen van nu het examen van dertig jaar geleden perfect zouden moeten kunnen maken. Ook evalueren we expliciet niet de beleidsbeslissingen die hebben geleid tot de geconstateerde feiten in ons rapport.

Langs elk van de bovenstaande factoren is een kwantitatieve data-analyse uitgevoerd.

1. De moeilijkheidsgraad van het ce

Allereerst toont ons onderzoek aan dat voor wiskunde en natuurkunde meer dan veertig procent(!) van de stof van het curriculum in 1990-1995 is verdwenen uit de centrale examens, vergeleken met 2015-2021. Voorbeelden van stof verdwenen uit het ce (en schoolexamen) zijn: onderdelen van het domein lading & veld (bijvoorbeeld condensator, halfgeleiderdiode, equipotentiaallijnen, een aantal specifieke baanberekeningen van geladen deeltjes in elektrische en magnetische velden), fysische informatica (regelsystemen en signaalverwerking), lichtbrekingen en lenzen (inclusief de tralie en laser uit de quantumoptica), statica en bijna de gehele wiskundige beschrijving van bewegingen (bijvoorbeeld de formules voor de eenparig rechtlijnig versnelde beweging en de wet van behoud van impuls inclusief deze behoudswet kunnen toepassen bij (in)elastische botsingen) en delen van trillingen/golven. Een deel van de stof geschrapt uit het ce komt nog wel voor op het schoolexamen (se): bijvoorbeeld de domeinen warmte & gasen (verplicht), biofysica oog en oor (keuzeonderwerp) en kern- en deeltjesprocessen (keuzeonderwerp) [6]. Uit de analyses blijkt ook dat er nauwelijks stof is toegevoegd. Wat wel is toegevoegd, staat niet in verhouding tot de vele thema's die verwijderd zijn. De moeilijkheidsgraad van de stof blijkt bij wiskunde A en B bovendien afgenomen. Dat volgt uit een analyse

McKinsey & Company

Toetsen getoetst

Het begrijpen van het langzaam verlaagde niveau van het vwo Centraal Examen (CE) Wiskunde en Natuurkunde.

Februari 2023

Auteurs
ir. Loek Zonnenberg, MBA
(docent wiskunde en Partner alumnus)
ir. Paul Rutten, MBA (Partner)

van p-waarden: het percentage punten dat landelijk voor een vraag wordt behaald. Hoe lager de p-waarde, hoe minder leerlingen de vraag goed hebben beantwoord en hoe moeilijker de vraag dus is. Daaruit blijkt dat de elementen die zijn weggehaald significant moeilijker zijn dan wat is overgebleven. Dit geldt niet voor natuurkunde, waar een klein deel moeilijker stof is geïntroduceerd (met name het beschrijven van het quantumtunneleffect inclusief het quantummodel van een deeltje opgesloten in een eendimensionale energieput met bijbehorende discrete energiespectra). Overigens bestaat het overgrote deel

van de examenstof behorende bij dit nieuwe subdomein Quantumwereld uit stof die voorheen al behoorde tot de centrale examenstof, voornamelijk uit het nu verdwenen domein Atoomfysica. Het gaat dan bijvoorbeeld om licht als golfverschijnsel kunnen benoemen en toelichten, de golf-deeltjesdualiteit kunnen toepassen bij het verklaren van interferentieverschijnselen bij elektromagnetische straling en materiedeeltjes, de onzekerheidsrelatie van Heisenberg kunnen toepassen, het quantummodel van een waterstofatoom beschrijven en mogelijke energieniveaus uitrekenen en het foto-elektrisch effect analyseren.

Om zeker te weten of het valide is om de puntenwaardes van examenopgaven te vergelijken over de jaren heen, is gekeken of scorepunten voor vergelijkbare vaak voorkomende wiskundige en natuurkundige handelingen ongeveer hetzelfde zijn. Dit is het geval. Als we deze puntenwaardes dan naast elkaar zetten voor vroeger en nu, blijkt dat de hoeveelheid inhoudelijk werk van de vragen op het ce afgenomen is. De maximaal te behalen punten in drie uur zijn met tien tot zeventien procent afgenomen.

Als tegenwerping wordt vaak gesteld dat de ce's wiskunde en natuurkunde nu moeilijker zijn, omdat ze steeds taliger geworden zijn. Dit blijkt het geval te zijn, als je dit afmeet aan het aantal gebruikte woorden in de vragen bij het ce: dit is met ongeveer 300-700 woorden toegenomen (voor natuurkunde van ~1.900 woorden in 1990-1995 naar ~2.200 woorden in 2015-2021). De vragen zijn dus contextrijker geworden. Dit representeert echter slechts vijf à tien minuten van de examentijd en heeft daarmee geen grote invloed op het vereiste tempo voor een gemiddelde leerling (we hebben dit niet bekeken voor leerlingen die moeite hebben met lezen). Samenvattend is het tempo dat voor het ce van leerlingen geëist wordt nu zo'n vijf à tien procent lager dan dertig jaar geleden. Als de stof verminderd is, is het nog steeds mogelijk dat leerlingen meer uitgedaagd worden door moeilijkere vragen gesteld te krijgen (althoewel er natuurlijk minder stofcombinatiemogelijkheden voor een vraag zijn). Het tegenovergestelde is echter het geval. Deze conclusie volgt wederom uit de p-waarde van de toetsitems. Uit de data blijkt dat er een verband is tussen het aantal denkstappen per vraag en de p-waarde: hoe meer denkstappen per vraag des te lager de p-waarde, dus hoe meer denkstappen per vraag hoe moeilijker de vraag. Kijkend naar de ce's worden er nu minder complexe of samengestelde vragen gesteld (minder denkstappen per vraag) dan dertig jaar geleden (bij natuurkunde is dit ongeveer 0,4 scorepunt minder per toetsitem).

We hebben ook gekeken welke vragen

relatief makkelijk zijn voor de deelnemers aan het ce en welke vragen relatief moeilijk zijn. Makkelijke vragen definiëren we als een p-waarde van 85 procent of hoger en moeilijke vragen als een p-waarde van 35 procent of minder. Daaruit blijkt dat er nu meer makkelijke en aanzienlijk minder moeilijke vragen op het ce zijn dan dertig jaar geleden. Bij natuurkunde is het aandeel moeilijke vragen afgenomen van negentien procent naar tien procent.

Een mogelijke verklaring voor dit effect zou kunnen zijn dat de ce-kandidaten over de gehele linie van de ce-stof nu veel sterker zijn of beter voorbereid zijn dan dertig jaar geleden, waardoor zij de moeilijke vragen beter maken. Dan zou voor een kleiner deel van het ce de p-waarde 35 procent of lager zijn, en zou een groter deel van het examen een p-waarde van 85 procent of hoger hebben. Echter, vooruitlopend op onderdeel drie van dit artikel zijn er data die duiden op een niveaudaling van de vwo-ce-kandidaten en dus in ieder geval niet op een significante niveaustijging.

Daarnaast zijn er nu ook meer hulpmiddelen toegestaan in het ce, waaronder de grafische rekenmachine, staat een deel van de wiskundeformules op papier en worden er hints gegeven, zoals aangegeven in vaktijdschriften en in het correctievoorschrift (bijvoorbeeld hulpgrafieken en het opknippen van vragen in deelstukken waarbij in de eerste vraag het antwoord gegeven wordt: "toon aan dat..."). Ook is opgevallen dat bij natuurkunde minder vragen gesteld worden waarbij een beroep op wiskundige vaardigheden wordt gedaan en leerlingen bij grotere vraagstukken meer 'aan de hand worden meegenomen' door de benodigde deelstappen om het probleem op te lossen als recept erbij te geven in plaats van de leerling deze zelf te laten bedenken.

Ten slotte is ook het correctievoorschrift iets versoepeld, door bijvoorbeeld minder precieze notaties te vereisen bij wiskunde of meer en eerdere deelscores bij natuurkunde. Er wordt verwacht dat dit leidt tot hogere scores voor soortgelijke ce's. De inschatting is

dat dit gemiddeld één à twee extra scorepunten per kandidaat oplevert, dus een relatief klein effect op het totale ce-cijfer (ruwweg 0,1 tot 0,2 punt op het hele ce).

2. De normhandhaving

Kort samengevat blijkt uit de data-analyses dat de normhandhaving geen verklaring voor de gestegen cijfers biedt. Dus zullen we er in dit artikel niet nader op ingaan.

3. Het niveau van de vwo-ce-kandidaten

De ontwikkeling van het niveau van de vwo-ce-kandidaten in de afgelopen dertig jaar is moeilijk direct vast te stellen. Daarom hebben we dit op verschillende manieren benaderd. We kijken hiervoor onder andere naar het IQ van de populatie, de voorbereiding van leerlingen en waar mogelijk de scores op identieke vragen.

Het gemiddelde IQ van de hele populatie middelbare scholieren is afgelopen veertig jaar nagenoeg constant gebleven (of zeer licht afgenomen) volgens onderzoek [7]. Aan het ce wiskunde A en B en natuurkunde nemen nu relatief significant meer kandidaten deel dan dertig jaar geleden, dus is deze ce-populatie gemiddeld zwakker geworden, omdat een significant grotere groep uit de hele populatie (iedereen in Nederland) nu deelneemt aan het vwo-ce voor wiskunde A, wiskunde B en natuurkunde. Al met al is het aandeel vwo-ce-kandidaten met ongeveer dertig procent gestegen (van zestien naar eenentwintig procent). Binnen het vwo-ce is het aandeel wiskunde B met vier procentpunt gestegen en natuurkunde met negen procentpunt. Wel is het aandeel wiskunde A met dertien procentpunt gedaald, maar omdat het aantal vwo-ce-kandidaten in absolute zin sterk gestegen is, leidt dit netto toch tot een groter aandeel wiskunde A van de totale populatie ce-kandidaten in Nederland.

Ook is het niet aannemelijk dat de huidige ce-populatie beter is voorbereid, om drie redenen. Ten eerste is de voorbereidingstijd niet toegenomen: de lestijd voor wiskunde en natuurkunde is afgenomen [8] en de huidige

generatie heeft meer afleidingen tot haar beschikking. Ten tweede is de kwaliteit en hoeveelheid van het oefenmateriaal vergelijkbaar: dertig jaar geleden was ook vijftien à twintig jaar aan oefenexamens beschikbaar. Ten derde is de hoeveelheid en de kwaliteit van de begeleiding ook niet toegenomen: het aandeel academische vwo-bètadocenten is afgenomen en bijlessen waren dertig jaar geleden ook beschikbaar (weliswaar minder formeel en geïnstitutionaliseerd dan nu). Het totale effect (niet voor een individuele leerling) van de bijlesindustrie wordt beperkt geschat, omdat de omzet van de bijlesindustrie ten opzichte van regulier vo-lerarensalarissen minder dan twee procent is [9]. Harde data over het gebruik van bijles bij informele aanbieders ontbreken over de afgelopen dertig jaar.

Er wordt soms aangedragen dat het belang van het ce is toegenomen. Dat is niet eenduidig vast te stellen in de data: het ce was ook vroeger al maatgevend voor vervolgcopleidingen. In de afgelopen decennia zijn de zakslaagregelingen eerst afgezwakt, en in de afgelopen tien jaar weer aangescherpt. In 2013 is de kernvakkenregeling ingevoerd, waarbij in het vwo voor Nederlands, Engels en wiskunde maximaal één vijf mocht worden gehaald. Dit is vergelijkbaar met hoe streng het examen was in de periode rond 1990-1995, toen er maximaal één vier mocht zijn of twee vijven, maar waarbij het gemiddelde 6,0 of hoger moest zijn, en waarbij het ce voor vijftig procent meetelde.

Om de ontwikkeling in niveau beter in te kunnen schatten, hebben we gezocht naar andere manieren om het niveau te kunnen vergelijken. Hoewel het Cito geen identieke vragen stelt, zijn er wel vragen te vinden in de ce's van dertig jaar geleden en nu die zeer sterk op elkaar lijken. We hebben deze opgaves opgezocht (veertien vragen in totaal) en zien dat leerlingen nu minder goede scores (p-waarden) hebben op vrijwel identieke vragen: zestien procentpunt minder goed bij wiskunde B, veertien procentpunt bij wiskunde A en zes procentpunt bij natuurkunde.

We kunnen met deze analyse geen onderscheid maken tussen de extra instroom van vwo-leerlingen of de mate van voorbereiding van de ce-kandidaten. Gegeven de grootte van de daling is het aannemelijk dat beide een rol spelen.

Omdat de scoredaling op (bijna) identieke vragen bij natuurkunde kleiner is dan bij wiskunde A en B, is een klein deelonderzoek uitgevoerd om te testen of deze scoredaling toe te schrijven is ofwel aan een niveau-daling van leerlingen ofwel doordat de vragen van nu (2015-2021) uit de vraagparen moeilijker zijn dan die van dertig jaar geleden (1990-1995). Hierbij zijn dezelfde leerlingen vraagparen van bijna identieke vragen uit beide perioden voorgelegd, waarmee de opzet van een statistische tekentoets ontstaat. Het voordeel van dergelijke opzet is dat niet een groot aantal deelnemers nodig is om tot enigszins statistisch betrouwbare uitkomsten te komen [10]. Uit dit kleine deelonderzoek blijkt dat het niet verklaard kan worden doordat de vragen van nu moeilijker zouden zijn.

Conclusie

We concluderen dat er over de laatste dertig jaar sluipenderwijs een substantiële niveaudaling heeft plaatsgevonden. Omdat we resultaten over de laatste drie decennia vergeleken, werd dit zichtbaar en kunnen we nu wel goed onderbouwde uitspraken doen. De geconstateerde effecten zijn bij wiskunde groter dan bij natuurkunde.

Hoe kunnen we het tij keren? Allereerst moeten we het eens zijn dat de lat omhoog moet. Onderwijs heeft een bredere verantwoordelijkheid dan enkel kwalificaties: het moet helpen leerlingen breed te vormen, maar kennis en vaardigheden blijven essentieel voor onze welvaart en welzijn. Wiskunde en natuurkunde zijn twee vakken die aan de basis liggen van de oplossingen voor uitdagingen van de toekomst. Een (haalbaar) hogere lat is ook in het belang van de leerlingen zelf: enkel zo worden zij uitgedaagd om zich ten volle te ontwikkelen. Ten tweede moet het probleem com-

pleter in kaart worden gebracht. Zien we op havo en vmbo-niveau hetzelfde patroon? Hoe zit het met andere belangrijke vakken, zoals Nederlands? Door dit in kaart te brengen, kunnen we het niveau van ons onderwijs gericht aanpakken.

Ten slotte moeten we systematisch werken aan het verstevigen van het onderwijssysteem. Hierbij hoort het invoeren van een verbetercultuur op alle scholen, waarin leraren elkaar elke dag ondersteunen om leerlingen zo goed mogelijk op te leiden. Wij denken aanwijzingen gevonden te hebben dat we ons wiskunde- en natuurkundeonderwijs de afgelopen dertig jaar onvoldoende aandacht hebben gegeven, en onbedoeld en ongemerkt hebben laten wegglijden. Leerkrachten die al kampen met een enorme werkdruk hebben dit niet kunnen tegenhouden. Het tij keren zal een inspanning van de lange adem zijn en het is nu aan alle betrokken partijen om in gezamenlijkheid structureel te werken aan het verhogen van het niveau.

Loek Zonnenberg is docent wiskunde en voormalig partner McKinsey & Company. Paul Rutten is partner McKinsey & Company. De auteurs hebben beiden technische natuurkunde in Delft gestudeerd.

REFERENTIES

- 1 Toetsen Getoetst: Het begrijpen van het langzaam verlaagde niveau van het vwo centraal examen (ce) Wiskunde en Natuurkunde, www.mckinsey.com/nl/our-insights/toetsen-getoetst.
- 2 Verschillende publicaties van De Staat van het Onderwijs door de Inspectie van het Onderwijs 2015, 2017, 2018, 2021 en 2022 (zie ook het rapport Toetsen Getoetst schema 5).
- 3 Advies van de Onderwijsraad Taal en rekenen in het vizier (2022).
- 4 PISA – PISA (oecd.org) (zie ook het rapport Toetsen Getoetst schema's 7 en 8).
- 5 Zie de schema's 13 t/m 15 uit het rapport Toetsen Getoetst voor de analyse van de studie luchtvaart- en ruimtevaarttechniek aan de TU Delft.
- 6 Zie schema's 40 tot en met 46 uit het rapport Toetsen Getoetst voor het gehele overzicht.
- 7 Woodley en Meisenberg, *In the Netherlands the anti-Flynn effect is a Jensen effect* (2013).
- 8 Op basis van 9 lessentabellen van scholen uit 1990-95 en 30 lessentabellen van scholen uit 2015-21.
- 9 Zie CBS, onder andere *Onderwijs; uitgaven aan onderwijs en CBS/OESO indicatoren*.
- 10 Zie schema's 59 tot en met 66 uit het rapport Toetsen Getoetst.
- 11 Zie McKinsey rapport *Een verstevigd fundament voor iedereen*.