

Het rapport *Toetsen* getoetst onder de loep

Het rapport *Toetsen getoetst* door Loek Zonnenberg en Paul Rutten [1] heeft veel discussie opgeroepen [2]. Het rapport constateert dat de eindexamencijfers van natuurkunde en wiskunde zijn gestegen, maar dat de PISA-scores zijn gedaald en verkent mogelijke oorzaken. Vervolgens worden er stevige conclusies getrokken: de examens zijn de afgelopen dertig jaar volgens de auteurs van het rapport makkelijker geworden en het niveau van de leerlingen zou zijn gedaald. Echter, op basis van de gehanteerde onderzoeksmethode kunnen deze conclusies niet getrokken worden.

De analyse van het rapport geeft onvoldoende zicht op onderliggende oorzaken. Dat maakt het moeilijk, zo niet onmogelijk, om eventuele gerichte actie te ondernemen. Het gevaar bestaat namelijk dat we aan knoppen gaan draaien die niets opleveren en mogelijk ongewenste effecten hebben. In deze reactie willen we de lezer op weg helpen om het rapport kritisch te lezen, zodat deze zelf kan nagaan of er goed onderzoek is gedaan en of de getrokken conclusies wel kloppen. We richten ons op een selectie van de passages uit het rapport die relevant zijn voor het vak natuurkunde. Voor wie niet bekend is met de structuur van het examen en de daarbij gebruikte begrippen is figuur 1 opgenomen.

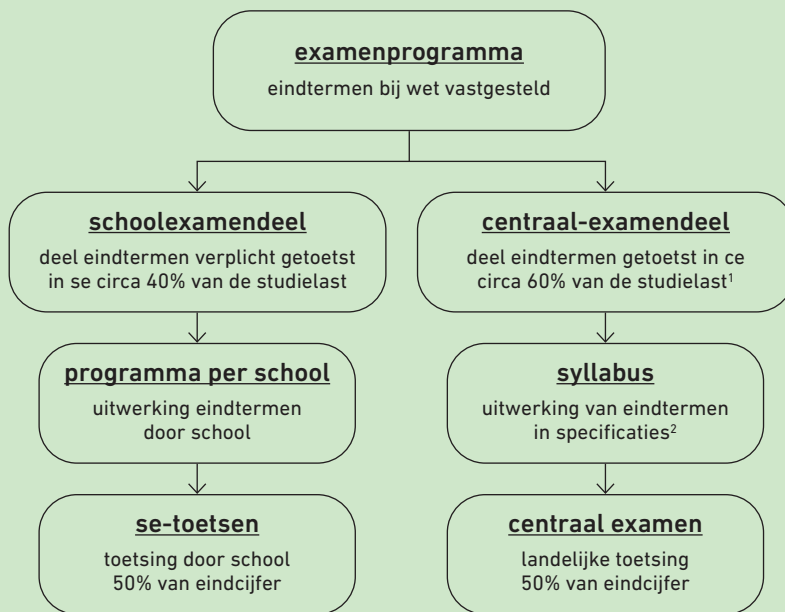
Argumentatiestructuur

Het rapport licht de gevolgde argu-

mentatie toe middels een schematisch beeld (figuur 2). De grafiek is onwetenschappelijk en uiterst suggestief: er ontbreken waarden langs de assen en het suggereert een perfecte anticorrelatie tussen twee verbanden die niet wordt onderbouwd. Er wordt in het rapport niet duidelijk een onderzoeksvraag met bijbehorende deelvragen geformuleerd. De discrepantie tussen PISA- en de resultaten voor het centraal examen (ce) is de aanleiding voor het onderzoek. De hoofdvraag lijkt te zijn: 'Wat zou de oorzaak [van die discrepantie] kunnen zijn?' Er worden drie oorzaken onderzocht. Daarbij horen weer drie vragen: twee voor PISA, één voor het ce. We lopen die vragen en de gevolgde methode langs. Daarnaast spelen signalen in vervolgopleidingen ook een rol in het rapport.

PISA

Om de discrepantie tussen PISA en het ce te kunnen verklaren kijken de auteurs naar de kwaliteit van de PISA-toets. Deze kwaliteit wordt in dit artikel niet ter discussie gesteld. Wat problematisch is, is het gesuggereerde verband tussen de resultaten van PISA en het ce. De PISA-toets bevat vragen in drie categorieën: leesvaardigheid, wiskunde en science. De sciencevragen bij PISA betreffen hoofdzakelijk het toetsen van de *nature of science* van leerlingen en niet de natuurkundige inhouden en onderwerpen van het ce. In tegenstelling tot het vwo-eindexamen natuurkunde zijn de meeste PISA-vragen meerkeuzevragen, waarbij nauwelijks een toelichting en/of een berekening hoeft te worden gegeven. Voor de beantwoording van de sciencevragen is bovendien geen enkele (natuurkunde)formule nodig. Slechts



Figuur 1. Structuur examenprogramma en gebruikte begrippen.

¹ Dit percentage geldt voor natuurkunde sinds 2013, daarvoor was het meer.
² Syllabus onder verantwoordelijkheid van het College voor Toetsen en Examen, zie www.examenblad.nl.

een klein deel van de sciencevragen betreft niet-levende natuur (natuurkunde en scheikunde): in 2006 ging het om 19 van de 104 vragen, dat is 18%. De PISA-populatie betreft een representatieve steekproef van alle 15-jarigen, over alle niveaus. De ce-resultaten gaan over een beperktere groep, namelijk vwo-leerlingen met natuurkunde als eindexamenvak. De gemiddelde leeftijd van deze populatie ligt tussen de zeven- en achttien jaar.

De auteurs van het rapport zijn van mening dat de PISA-toets en het ce vergelijkbaar zijn. De onderbouwing van die bewering (schema 23, bladzijde 52, zie ook noot [3]) is onvoldoende: acht regels met argumenten als “Als iemand goed is in bijvoorbeeld wiskunde, dan verwacht je dat deze leerling zowel op PISA als op het ce daar goed in is” en “De vorm van de door PISA gestelde vragen (bijvoorbeeld contextrijk) sluit goed aan bij het onderwijs in Nederland.”

Een verband tussen PISA en het ce, laat staan hoe sterk dat verband is, is hiermee derhalve niet aangetoond. Voor het vervolg is het belangrijk dat PISA niet relevant is voor de vraag of het niveau van de examens natuurkunde is gedaald. Die vraag moet los van de PISA-scores beantwoord kun-

nen worden. Een aanwijzing voor een dalend niveau denken de auteurs van het rapport in de vervolgopleidingen gevonden te hebben.

Vervolgopleidingen

Het rapport concludeert dat een vwo-diploma met natuurkunde en wiskunde B onvoldoende garantie geeft voor studiesucces en dat het ce *niet meer dezelfde waarde heeft* (bladzijde 28). Dit wordt onderbouwd met twee statistieken: 1) voor de periode 2000-2004 voor de studie technische natuurkunde het percentage studenten dat de prope-deuse in één studiejaar heeft behaald in relatie tot het gemiddelde cijfer voor wiskunde B en natuurkunde (schema 2, bladzijde 41) en 2) voor het cohort gestart aan de TU Delft in 2014 het percentage studenten dat het tweede jaar haalt in relatie tot het cijfer wiskunde B (schema 10, bladzijde 45).

Het is al jaren bekend dat bij technische universitaire opleidingen de vwo-cijfers de kans op studiesucces goed voorspellen. Dat wordt door beide schema's ook weer bevestigd. Er ontbreken echter data uit de periode 1990-1995 zodat niet geconcludeerd kan worden dat hierin iets is veranderd in de laatste dertig jaar. Bovendien, uit bovengenoemde schema's blijkt dat in de periode 2000-

2004 het studiesucces voor leerlingen met gemiddeld een 8 of hoger voor wiskunde B en natuurkunde gelijk is aan 67%. In 2014 is het studiesucces opgelopen tot zeker 85%. Er zou dus geconcludeerd kunnen worden dat de voorspellende waarde van het ce juist is *toegenomen*.

Niveau examenpopulatie

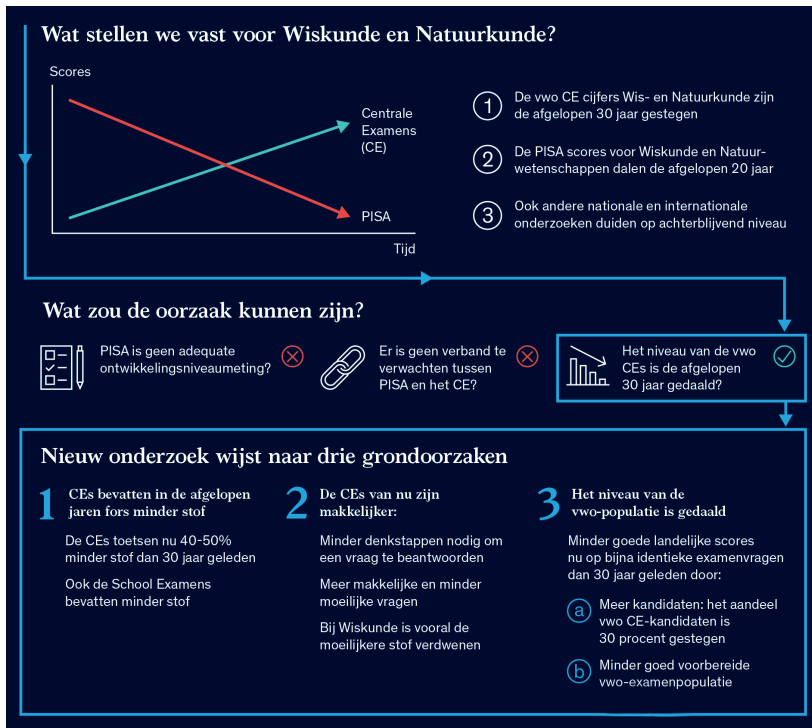
Het rapport neemt aan dat het niveau van de vwo-populatie is gedaald aangezien het aandeel leerlingen dat eindexamen vwo doet de afgelopen jaren is *toegenomen*. Deze discutabele uitspraak wordt nergens onderbouwd met data. Bovendien wordt deze veronderstelde niveaudaling gepresenteerd als één van de drie oorzaken van de niveaudaling van het examen, dus zonder dat de niveaudaling van het examen zelf daarmee is aangetoond.

Niveau van het examen

Zijn de examens wis- en natuurkunde nu *echt* makkelijker geworden? Het rapport doet uitspraken over zowel het ce als het se.

Het examenprogramma: inhoud, niveau en toetsing

Volgens het rapport zijn er drie oorzaken voor het gestegen ce-cijfer.



Figuur 2. Argumentatieschema uit het rapport.

Eén hebben we kort besproken (het niveau van de examenpopulatie). De andere twee zijn: het ce bevat 40-50% minder stof en de ce's zijn makkelijker geworden. De auteurs trekken hier een conclusie zonder een vooraf gevalideerd meetinstrument te gebruiken. Daarnaast hebben de auteurs de analyse zelf uitgevoerd in plaats van dit door meerdere onafhankelijke beoordelaars te laten doen. Dit maakt het genoemde percentage en het oordeel over niveau van het examen onbetrouwbaar.

Hoeveelheid en niveau lesstof

Het is een beleidsmatige keuze geweest om de hoeveelheid stof in het ce te verminderen en die in het se te vergroten. Veel examenstof is dus niet verdwenen uit het examenprogramma, maar wordt niet meer getoetst in het ce. Een voorbeeld hiervan is het keuzedomein kern- en deeltjesprocessen. Dit maakt het se moeilijker.

Het rapport houdt onvoldoende rekening met de moeilijkheid dat specificaties uit het examenprogramma niet zo maar kwantificeerbaar zijn en dat het niveau ervan zich niet zomaar laat

vergelijken en bepalen. Enkele voorbeelden hiervan:

1. *Verschillende specificaties, vergelijkbare toetsing is mogelijk.*
 - a) De specificatie dat de arbeid de integraal is van de kracht over de verplaatsing is verdwenen ($W = \int F ds$). Maar in het programma staat nog wel: "De arbeid bepalen uit een kracht-verplaatsingsdiagram". Dus wiskundig klopt het dat er iets uit is, maar conceptueel wordt dezelfde natuurkunde getoetst. b) De formules voor snelheid en versnelling van een harmonisch trillend voorwerp zijn ook verdwenen, maar leerlingen moeten de snelheid en versnelling nog wel grafisch uit een plaats-tijd-diagram kunnen bepalen. Bovendien moeten zij door middel van het differentiëren van een gegeven goniometrische uitwijkingsfunctie de snelheids- en versnellingsfunctie nog wel kunnen bepalen.
2. *Niet geëxpliciteerde lesstof wordt getoetst.* Enkele specificaties zijn uit het programma verdwenen, zoals de horizontale worp, halfgeleiders met bandenstructuur en het hall-effect,

maar daar worden op het ce nog wel regelmatig vragen over gesteld (zie bijvoorbeeld het ce van 2021-2 en 2022-2). De examenopgave Vliegen (2016-1), over het getal van Strouhal, dat niet in het examenprogramma voorkomt, is een complexe opgave waarbij diverse wiskundige verbanden geanalyseerd en gecombineerd moeten worden. Dit soort opgaven maken een ce juist moeilijker, omdat leerlingen hun natuurkundige kennis op een voor hen relatief nieuwe situatie moeten toepassen.

3. Hoe zwaar telt nieuwe stof?

Er zijn ook specificaties bijgekomen: met name Numeriek Modelleren, het domein Natuurwetten en Modellen en het domein Quantumwereld. Numeriek modelleren neemt in de regel een belangrijk deel van de examens in en vergt inzicht van de leerlingen, met name op het natuurkundig computationele vlak. Hoewel enkele quantumconcepten in het oude examenprogramma voorkwamen, werden hierover op het ce niet altijd vragen gesteld in de periode 1990-1995. Sinds het domein Quantumwereld in 2013 in het programma is opgenomen zijn hier nagenoeg jaarlijks opgaven in het ce terechtgekomen: in de regel zijn dat uitdagende opgaven die een groot beroep doen op het natuurkundig conceptuele denkvermogen van leerlingen.

Verder wordt in het rapport melding gemaakt van "verdwenen eindtermen", die helemaal niet verdwenen zijn. Voorbeelden hiervan zijn het dopplereffect (als rood- en blauwverschuiving in het domein Elektromagnetische Straling en Materie) en de formule voor de inductiespanning, als (negatieve) tijdsafgeleide van de magnetische flux. Gesteld dat we tot een kwantificering en niveau van de examenstof kunnen komen, hoeft een eventuele vermindering hierin nog geen makkelijker eindexamen op te leveren. Dat zal los daarvan vastgesteld moeten worden.

Niveau van de examenvragen

Het niveau van de vragen wordt op een aantal manieren beoordeeld. Over de statistische analyse heeft Paul

Drijvers in *Euclides* het een en ander beschreven [2].

Kijken we naar de inhoudelijke analyse, dan is deze beperkt en kent een aantal mankementen. Er is gekeken naar negen paren van (bijna) identieke vragen uit de periode 1990-1995 en de periode 2015-2021. Als extra controle zijn drie van deze vragenparen aan zeven leerlingen van dezelfde school voorgelegd. Die leerlingen scoorden slechter op de vragen uit de periode 1990-1995. Er wordt vervolgens geconcludeerd dat het kennisniveau is gedaald.

De problemen met deze analyse zijn:

1. Als de drie vragenparen gelijk zijn, dan moeten de leerlingen ook gelijk scoren op beide vragen binnen een vragenpaar. Zo niet, dan zouden de vragen niet gelijk kunnen zijn en kan niet geconcludeerd worden dat het niveau is gedaald.
2. De steekproef bevat een relatief klein aantal leerlingen ($N=7$) van één middelbare school. De auteurs beargumenteren dat de steekproef representatief is op basis van de “score op exact dezelfde ce-vragen van nu” (bladzijde 73 van het rapport). Het is niet duidelijk wat dat precies betekent. Het gebruik van één respondentkenmerk is een onvoldoende basis om de conclusie te trekken dat het kennisniveau van de leerlingen is gedaald.
3. Er wordt niet onderbouwd of de specificaties die met deze vragen getoetst worden wel voldoende vergelijkbaar zijn. Dat is belangrijk om te beoordelen of leerlingen van toen op dezelfde wijze op het examen zijn voorbereid als de leerlingen van nu en daarmee of de scores op deze vragen van toen vergeleken kunnen worden met de scores van nu.
4. Inhoudelijke analyse van de vragenparen laat zien dat ze niet altijd even makkelijk zijn: vaak is er in de recente vragen sprake van een voor leerlingen nieuwe context, zoals gespecificeerd in het examenprogramma. Dat maakt de vragen juist moeilijker.

Op basis van deze inhoudelijke analyse kan niet de conclusie getrokken worden dat het examen makkelijker is geworden.

Het verschil se-ce

Het rapport raket een oude discussie op over de vraag of scholen het se gebruiken om hun examenresultaten ‘kunstmatig’ op te hogen. De auteurs concluderen: “Dat een vwo-diploma onvoldoende garantie geeft voor studiesucces kan deels worden verklaard doordat scholen 10-15 jaar geleden mindere ce-prestaties compenseerden met te makkelijke schoolexamens (se)” (schema 12, bladzijde 46).

Het eerste deel van dit citaat (vwo-diploma geeft onvoldoende garantie voor studiesucces) is eerder in dit artikel weerlegd. Het tweede deel van het citaat is recentelijk weerlegd door Marion van Brederode en Martijn Meeter [4]. Zij hebben aangetoond dat het verband dat in schema 12 wordt getoond een artefact is. Er is derhalve geen bewijs dat scholen se-cijfers zouden manipuleren. (Het zou te ver voeren die analyse hier te herhalen; het wordt aangeraden om [4] te raadplegen.)

Dit voorbeeld laat zien dat de auteurs ‘data’ gebruiken die goed aansluiten bij hun hypothese en onvoldoende verder zoeken om hun hypothese te ontcrachten. Dit toont nog eens hoe belangrijk het is om tot een gevalideerde methode en meetinstrument te komen en die onafhankelijk, dus niet door de auteurs zelf, te laten uitvoeren.

Er is dus juist bewijs dat het se een eerlijke meting is [5]. Als er stof uit het ce naar het se gaat, dan is het examen als geheel niet fundamenteel wezenlijk anders.

Conclusie

Het rapport *Toetsen getoetst* stelt de vraag wat de oorzaak van de discrepantie tussen PISA- en ce-cijfers zou kunnen zijn en komt tot de conclusie dat het niveau van de examens wiskunde en natuurkunde is gedaald. Die conclusie wordt niet eenduidig door het geleverde bewijs ondersteund. Er is niet goed gedefinieerd wat er met ‘niveau’ wordt bedoeld. Het is goed te bedenken dat de examenprogramma's vele doelen kennen: meer dan alleen kwalificatie voor een ‘harde’ bèta-opleiding. De examenprogramma's zijn er juist voor een breed scala aan leerlingen en vervolopleidingen.

Het rapport roept wel een belangrijke en interessante vraag op: bereidt het examenprogramma wiskunde B en natuurkunde nog wel voor op een bètastudie? Dit zou nader onderzocht moeten worden. Echter, zelfs als uit een gedegen wetenschappelijk onderzoek zou blijken dat het voorbereidingsniveau voor een bètaopleiding niet voldoet, dan is het de vraag of alle leerlingen een wiskundig zoaarder programma opgelegd moeten krijgen. We zien mogelijkheden om voor beide groepen leerlingen een geschikt programma vorm te geven.

Roeland Boot geeft ruim dertig jaar natuurkundelessen aan examenklassen havo en vwo. Hij heeft meegewerkt aan de ontwikkeling en implementatie van het in 2013 ingevoerde eindexamenprogramma natuurkunde (havo en vwo) en is vier jaar verbonden geweest aan het College voor Toetsen en Examens als lid van de vaststellingscommissie centrale eindexamens natuurkunde havo en vwo. Ook doet hij promotieonderzoek op het gebied van natuurkundig modelleren door vwo-bovenbouwleerlingen.
r.j.boot@uu.nl

Lodewijk Koopman studeerde theoretische natuurkunde aan de UvA. Hij is gepromoveerd op de vakdidactiek van de quantummechanica aan eerstejaarsnatuur- en scheikundestudenten. Momenteel is hij werkzaam als eerstegraadsdocent natuurkunde op het vwo en als lerarenopleider aan de TU Delft. Daarnaast is hij auteur bij een bovenbouwmethode voor vwo natuurkunde.

BRONNEN EN NOTEN

- 1 www.mckinsey.com/nl/our-insights/toetsen-getoetst.
- 2 CvTE: www.cvte.nl/actueel/nieuws/2023/02/27/kun-je-examens-vergelijken, Paul Drijvers, *Euclides* 98(6), 12-13 (2023), Hans van Bommel, NVOX, mei, 16-18 (2023).
- 3 Schema 23 bespreekt nog vijf andere tegenargumenten. Die zijn echter ondergeschikt aan het tegenargument dat PISA iets anders toetst dan het ce. Bijvoorbeeld: als PISA niet hetzelfde toetst, dan is het niet relevant of de scores van PISA op 15-jarige leeftijd voorspellend zijn voor de vaardigheden van 18-jarigen.
- 4 www.scienceguide.nl/2020/05/hoer-statistiek-het-schoolexamen-verdacht-maakte.
- 5 In het huidige beoordelingskader van de Onderwijsinspectie wordt gekeken naar het verschil se-ce. Het streven is dit verschil kleiner dan 0,5 te houden. Meer hierover is te vinden op www.onderwijsinspectie.nl.