

VERSCHILLEN TUSSEN REKENTOETS EN CENTRAAL EXAMEN WISKUNDE

Franziska van Dalen
Paul Drijvers
Hendrik Straat

De rekentoets staat sterk in de belangstelling. In 2014 adviseerde de commissie Bosker om de discrepantie te onderzoeken tussen de resultaten op wiskunde-examens en op de rekentoets 2F voor leerlingen in vmbo bb en vmbo kb. In het voorjaar van 2015 is dit onderzoek door Cito uitgevoerd; Franziska van Dalen, Paul Drijvers en Hendrik Straat vatten de belangrijkste resultaten samen.

Inleiding

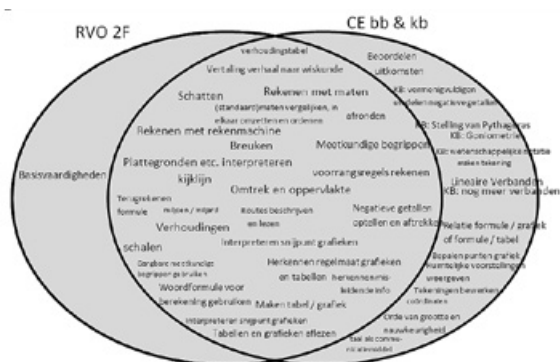
De resultaten van leerlingen op de rekentoets vo waren de afgelopen jaren lager dan die van de eindexamens wiskunde, terwijl met name de examens wiskunde vmbo ook veel rekenonderdelen bevatten. Dit was aanleiding voor de commissie Bosker^[1] om te adviseren de discrepantie tussen de resultaten op wiskunde-examens en op de rekentoets 2F voor leerlingen in vmbo bb en vmbo kb nader te onderzoeken. Omdat de resultaten van dit onderzoek, dat inmiddels is uitgevoerd en gepubliceerd^[2], interessant zijn voor wiskundeleraars, vatten we ze in dit artikel kort samen.

We beperken ons tot het vmbo bb en vmbo kb en onderzoeken de resultaten van deze leerlingen op de rekentoets 2F en het digitaal afgenomen centraal examen (CE) wiskunde. Immers, rekenen en wiskunde staan in deze schooltypen dicht bij elkaar en digitale toetsen zijn beter vergelijkbaar dan toetsen waarin verschillende media worden gebruikt. We bekijken de jaren 2013, 2014 en 2015, waardoor de resultaten van in totaal ruim 47.000 leerlingen in dit onderzoek zijn betrokken.

Verschillende toetsen?

De eerste vraag is in hoeverre de rekentoets en het CE wiskunde vmbo bb en vmbo kb van elkaar verschillen.

figuur 1 Overeenkomsten en verschillen tussen de rekentoetswijzer en de syllabi CE



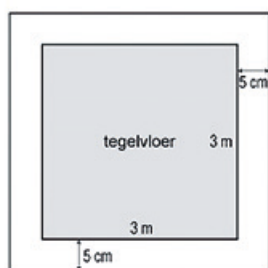
Inhoudelijk blijkt uit een analyse van de rekentoetswijzer en de syllabus van het examen dat er behoorlijk veel overlap bestaat tussen rekenen en wiskunde in het vmbo. In figuur 1 is deze overlap in kaart gebracht. De toetsen, dus de rekentoets en het centraal examen, zijn wel behoorlijk verschillend: de examens bevatten meer vragen bij eenzelfde context (*clustervragen*) terwijl de rekentoets in het algemeen één vraag per context stelt (al wordt sinds 2015 per toets één cluster met drie vragen bij dezelfde context opgenomen). Bij de rekentoets zijn items helemaal goed of helemaal fout, terwijl je bij het examen ook deelscores kunt halen. De rekentoets kent contextloze opgaven, die we in examens niet terugvinden en het aantal vragen in de rekentoets is veel groter dan het aantal vragen op het examen.

Als we ons beperken tot vragen die zowel op de rekentoets als op het examen gesteld hadden kunnen worden (en dus contextloze opgaven uitsluiten omdat die op het CE niet worden gesteld) dan is de eerste opvallende bevinding dat de rekenvragen van de rekentoets inhoudelijk complexer zijn dan die van het examen, met name voor bb. De inhoudelijke complexiteit is geoperationaliseerd door een systematiek te ontwikkelen om de complexiteit van een vraag uit te drukken in het aantal benodigde denk- en rekenstappen. In figuur 2 staan twee voorbeelden van opgaven met de bijbehorende complexiteit. Het algemene beeld na het coderen (en controleren en hercoderen...) van 745 vragen van beide toetsen is dat de gemiddelde complexiteit van de vragen van de rekentoets gelijk is aan 3,2 rekenstappen en die van de CE's wiskunde bb en kb aan 2,7 rekenstappen. Uit significantietoetsen (*t*-toetsen) blijkt dat het aantal benodigde rekenstappen van de contextvragen van de rekentoets in 2013, 2014 en 2015 significant hoger ligt dan het aantal benodigde rekenstappen van de rekenvragen van het CE wiskunde vmbo bb en kb met als uitzondering vmbo kb in 2014 en 2015. De rekentoetsvragen zijn dus inhoudelijk complexer dan de rekenonderdelen van de CE's, met name bij vmbo bb. De tweede opvallende indruk betreft de taligheid van de items, een punt waarop de rekentoets vaak wordt bekritiseerd. Een analyse van bescheiden

Blokhut



Onder de tegelvloer wil Brenda plastic leggen. Dit plastic moet aan alle kanten 5 cm onder de tegelvloer uitsteken.



(3p) Bereken hoeveel m^2 plastic Brenda minstens nodig heeft.
Typ je berekening in.

figuur 2 De complexiteit van twee items van reken-toets en centraal examen.



Kranenbrug

Woensdag, 28 november 2012

In 2013 veel deelnemers aan

Er worden in 2013 veel hardlopers verwacht bij de marathon van Kranenburg. Het aantal lopers zal minimaal anderhalf keer zo groot zijn als dit jaar.

In 2012 deden 1520 hardlopers mee aan deze marathon.

Hoeveel hardlopers worden in 2013 minimaal verwacht?

hardlopers



figuur 2a Dit item uit het CE wiskunde vmbo bb van 2013 kan in vier stappen opgelost worden. Stap 1: Omzetten, 5 cm omzetten naar 0,05 m. Stap 2: Vermenigvuldigen, 0,05 m met 2 vermenigvuldigen omdat het er aan beide kanten bij moet. Stap 3: Optellen, 0,10 m bij 3 m optellen om de afmetingen van het vierkante stuk plastic te bepalen. Stap 4: Vermenigvuldigen, 3,1 m met 3,1 m vermenigvuldigen om de oppervlakte te berekenen.

figuur 2b Dit item uit de rekentoets 2013 kan in drie stappen opgelost worden. Stap 1: Bron, er staan overbodige getallen zoals de datum van het krantenartikel in het plaatje. Stap 2: Omzetten term, anderhalf moet omgezet worden naar 1,5. Stap 3: Vermenigvuldigen, 1,5 vermenigvuldigen met 1520.

omvang door een taaldeskundige suggereert dat het CE wiskunde vmbo bb 2013 qua taligheid complexer is dan de rekentoets tijdvak 1 2013. Dat de rekentoets te talig zou zijn, wordt hiermee dus ontkracht, al zou een grootschaliger onderzoek op dit punt (zie bijvoorbeeld [5]) nodig zijn om uitspraken met meer zekerheid te kunnen doen.

Verschillende resultaten?

De volgende vraag is of er significante verschillen zijn tussen de resultaten van kandidaten op de rekentoets en het CE wiskunde vmbo bb en vmbo kb. Allereerst is er gekeken naar de samenhang tussen de scores op de CE's wiskunde vmbo en de scores op de rekentoets. De correlatiecoëfficiënten tussen het CE wiskunde bb 2013 en 2014 en de rekentoets zijn respectievelijk 0,75 en 0,68. Voor kb 2013 en 2014 zijn deze waarden 0,65 en 0,53.

Er is dus een behoorlijke samenhang tussen deze resultaten, die voor bb sterker is dan voor kb. Om de vraag naar de verschillen te beantwoorden zijn scores geanalyseerd van leerlingen vmbo bb en kb die in een van de afnamejaren 2013, 2014 of 2015 aan zowel de rekentoets als aan het CE wiskunde hebben meegedaan. In totaal gaat het om ruim 47.000 leerlingen. In het onderzoek zijn enkel de scores meegenomen op vragen die in beide toetsen hadden kunnen voorkomen; er is dus op item- of vraagniveau gekeken en niet naar de scores op de toetsen als geheel. Om de scores te kunnen vergelijken zijn de vmbo opgaven, waarvoor je meestal meer punten kunt halen, omgeschaald naar 0 en 1: 1 punt als de opgave helemaal goed is en anders 0, net zoals bij de rekentoets het geval is. Met t-toetsen is vervolgens bepaald of de behaalde scores van kandidaten op de rekentoets en het

figuur 3 Scores rekenopgaven CE wiskunde vergeleken met scores contextitems rekentoets

Leerwegsector en afnamejaar	M_{RVO} (SD_{RVO})	M_{CE} (SD_{CE})	N	Significantie	Cohen's d
Vmbo bb 2013	.294 (.147)	.533 (.171)	5426	.000	1.50
Vmbo kb 2013	.445 (.169)	.576 (.195)	6981	.000	0.72
Vmbo bb 2014	.354 (.151)	.504 (.178)	9764	.000	1.19
Vmbo kb 2014	.467 (.158)	.488 (.210)	13038	.000	0.11
Vmbo bb 2015	.419 (.163)	.567 (.173)	5534	.000	0.88
Vmbo kb 2015	.575 (.158)	.672 (.178)	6453	.000	0.58

figuur 4 Percentage voldoende per toets/examen en schooltype in tijdvak 1

Leerwegsector en afnamejaar	Rekentoets	CE wiskunde
Vmbo bb 2013	25,8%	80,1%
Vmbo kb 2013	25,3%	65,9%
Vmbo bb 2014	29,9%	83,6%
Vmbo kb 2014	27,9%	68,9%
Vmbo bb 2015	40,7%	82,0%
Vmbo kb 2015	50,7%	76,6%

CE wiskunde vmbo bb of kb significant verschillen. De resultaten van deze exercitie vindt u in de tabel van figuur 3. In de eerste kolom staan leerweg en afnamejaar. In de tweede kolom staat de gemiddelde score. Het getal .294 betekent bijvoorbeeld dat de vmbo bb leerlingen in 2013 bij de rekentoets ruim 29% van de contextopgaven correct hebben beantwoord. Tussen haakjes staat daarachter de bijbehorende standaarddeviatie. In de derde kolom staan vergelijkbare gegevens, maar dan voor de rekenvragen van het centraal examen. Kolom 4 bevat de aantallen leerlingen. In kolom 5 staat of het verschil tussen beide gemiddeldes significant is. De laatste kolom geeft de effectgrootte van het verschil. Dit is de grootte van het verschil uitgedrukt in de gezamenlijke standaardafwijking. De effectgrootte is dus een maat voor het effect die niet afhangt van de spreiding van de gegevens. Dus 1.50 geeft aan dat het gemiddelde verschil anderhalve standaardafwijking is. Ter vergelijking, in de sociale wetenschappen wordt doorgaans vanaf een effectgrootte van 0.8 gesproken van een groot effect. Uit de tabel blijkt dat voor alle zes groepen leerlingen geldt dat het resultaat op de rekentoets significant lager is dan dat van het centraal examen. In deze zin kunnen we dus inderdaad van een discrepantie spreken. De effectgroottes zijn redelijk tot groot te noemen, met uitzondering van vmbo kb 2014. In het algemeen zijn de effectgroottes voor bb hoger dan die voor kb. Wel nemen deze effectgroottes voor bb in de loop van de tijd af, wat zou kunnen betekenen dat de scores op de rekentoets en het CE naar elkaar toe groeien. Samengevat kunnen we dus concluderen dat de scores voor de rekentoets significant lager zijn dan voor het CE wiskunde en dat dit met name voor vmbo bb het geval is, al lijken de verschillen kleiner te worden.

'DUIDELIJK DAT HET MOEILIJKER IS OM EEN VOLDOENDE VOOR DE REKENTOETS TE HALEN.'

Verklarende factoren?

De derde en laatste vraag is hoe we de geconstateerde verschillen tussen de resultaten kunnen verklaren. Factoren die hierbij mogelijk een rol spelen zijn het feit dat de leerlingen op de rekentoets geen en op het CE wel deelscores kunnen behalen en het feit dat het CE wel en de rekentoets geen clustervragen bevatten. Uit ander onderzoek blijkt dat deze factoren vermoedelijk geen grote rol spelen.^[3, 4] Evenmin lijkt de taligheid een verklaring te bieden, want de indruk is dat de rekentoets juist minder talig is dan het CE. Wat natuurlijk wel een verklaring kan zijn, is de hogere rekencomplexiteit die we hebben geconstateerd bij de rekentoets in vergelijking met het CE: voor de vragen van de rekentoets waren immers gemiddeld meer rekenstappen nodig dan voor de CE-rekenvragen. Maar er is nog iets meer aan de hand. Een aantal paren van rekenvragen uit de rekentoets en CE is op twee

verschillende manieren met elkaar vergeleken. Ten eerste is een aantal paren onderzocht dat dezelfde rekencomplexiteit had. Daarbij bleek dat leerlingen de betreffende items van het CE significant beter maakten dan de qua

complexiteit dus vergelijkbare vragen van de rekentoets. Ten tweede is een aantal paren van vragen afkomstig uit de rekentoets en het CE onderzocht die in praktijk even goed zijn gemaakt door de leerlingen. Vervolgens zijn deze paren aan experts voorgelegd met de vraag welke zij het meest complex vonden. De uitkomst hiervan was dat men significant vaker de CE-opgave als de meest complexe aanwees. Informeel geformuleerd: vragen die inhoudelijk even moeilijk lijken, worden op de rekentoets slechter gemaakt dan op het CE. Andersom, van opgaven die even goed gemaakt worden, worden die van het CE als inhoudelijk moeilijker ingeschat. Dit suggereert dat leerlingen beter presteren op het CE dan op de reken-

toets. Mogelijke verklaringen hiervoor zouden kunnen zijn dat men voor het CE beter is gemotiveerd dan voor de rekentoets, of dat de leerlingen er beter op zijn voorbereid. Een derde mogelijke verklaring voor het verschil in resultaten is de manier van cesuurbepaling. Die van de rekentoets is gebaseerd op de referentieniveaus rekenen, terwijl de normering van het CE plaatsvindt op basis van een referentie-examen. In figuur 4 staan de percentages voldoende per toets/examen en schooltype. We zien dat deze percentages sterk verschillen. Het verschil in complexiteit gecombineerd met het verschil in percentages voldoende maakt duidelijk dat het moeilijker is om een voldoende voor de rekentoets te halen dan voor het CE wiskunde. De cesuur ligt voor de rekentoets hoger dan voor het CE wiskunde.

Conclusie

In dit artikel zijn we nagegaan in hoeverre de resultaten op de rekentoets verschillen van die van het centraal examen wiskunde voor vmbo en waarom dat zo zou kunnen zijn. De eerste conclusie is dat er grote inhoudelijke overlap bestaat maar dat er ook grote verschillen bestaan in de wijze van toetsen. De tweede conclusie is dat de leerlingen vmbo bb en kb inderdaad significant hoger scoren voor het CE wiskunde dan voor de rekentoets en dat met name bij vmbo-bb er sprake is van behoorlijke effectgroottes, al lijken die in de loop van de jaren af te nemen. De belangrijkste mogelijke verklaringen voor deze verschillen zijn (1) de hogere rekencomplexiteit van de items van de rekentoets in vergelijking met die van de rekenvragen in het CE, (2) de lagere prestaties van leerlingen op de rekentoets door lagere motivatie of slechtere voorbereiding, en (3) de cesuur die vanuit een ander referentiekader is opgesteld. Andere factoren zoals taligheid van rekenvragen, het al dan niet toekennen van deelscores en het gebruik van clustervragen lijken niet van doorslaggevend belang te zijn.

Noten

- [1] Bosker, R., & Vorle, R. Van de (Red.) (2014). *Advies over de uitwerking van de referentieniveaus 2F en 3F voor rekenen in toetsen en examens*. Enschede: SLO.
- [2] Cito (2015). *Onderzoek Discrepantie Rekentoets vo 2F en CE wiskunde vmbo bb en kb*. Eindrapportage juni 2015. Arnhem: Stichting Cito. www.hetcvte.nl/item/overige_publicaties
- [3] Cito (2015). *De mogelijkheden en consequenties van het werken met deelscores* Eindrapport, mei 2015. Arnhem: Stichting Cito. www.hetcvte.nl/item/overige_publicaties
- [4] Cito (2015). *Clustervragen in rekentoetsen en -examens* Eindrapport, september 2015. Arnhem: Stichting Cito. www.hetcvte.nl/item/overige_publicaties
- [5] Evers-Vermeul, J. (2015). *Toetsing in context vraagt door-denkend van de rol van taal*. *De Cascade*, 12, 17-19.

Over de auteurs

Franziska van Dalen schrijft dit artikel als resultaat van het master onderzoek dat zij bij Cito uitvoerde en is inmiddels docent wiskunde aan het Oosterlicht College te Nieuwegein. Paul Drijvers schrijft dit artikel als wetenschappelijk onderzoeker bij Cito en is tevens hoogleraar in de didactiek van de wiskunde bij het Freudenthal Instituut van de Universiteit Utrecht. Hendrik Straat is wetenschappelijk onderzoeker bij de afdeling Psychometrie en Onderzoek van Cito.

E-mailadressen: vandalenfranziska@gmail.com;
Paul.drijvers@cito.nl; Hendrik.straat@cito.nl

ADV
MATH 4 ALL