



Maakt het voor een leerling uit op welke school hij of zij zit als het gaat om waardering van zijn of haar vaardigheid of kennis?

Schooleffecten op SE en CE

■ Marieke van Onna

Dr. M.J.H. van Onna is wetenschappelijk onderzoeker bij Cito. E-mail: marieke.vanonna@cito.nl

■ Anton Béguin

Dr. A.A. Béguin is directeur Centrale Toetsen en Examens bij Cito.

■ Remco Feskens

Dr. R.C.W. Feskens is hoofd Psychometrisch Onderzoek en Dienstverlening bij Cito

■ Paul van der Molen

Ir. P. van der Molen is manager bij Cito

In 2019 waren in het voortgezet onderwijs de vakgemiddelden op het schoolexamen (SE) gemiddeld gelijk aan de vakgemiddelden op het centraal examen (CE). Dit was in 2011 niet het geval, toen lagen de SE-cijfers gemiddeld hoger dan de CE-cijfers. Er bestaan verschillen tussen scholen in gemiddelde cijfers, zowel bij het SE als het CE. Echter, over het algemeen geven scholen leerlingen op het SE, gemiddeld over vakken heen, grotendeels dezelfde cijfers als op het CE. Aan de hand van een nieuwe berekeningswijze wordt dit aangetoond.

Inleiding

Met het afschaffen van de centrale eindexamens in het voortgezet onderwijs als gevolg van de corona-crisis, zijn de eindcijfers van examenkandidaten in 2020 gebaseerd op de SE-cijfers. Deze ontwikkeling wekt interesse op in de geschiktheid van SE-cijfers, en de vergelijking van SE- en CE-cijfers. In hun artikel 'Hoe statistiek het schoolexamen verdacht maakte en waarom schoolexamens dat niet verdienen' (Science Guide, 6 mei 2020) gaan Marion van Brederode en Martijn Meeter in op verschillen in cijfergeving tussen schoolexamen (SE) en centraal examen (CE), en het gebruik daarvan voor de evaluatie van schoolkwaliteit. Zij tonen aan dat verschillen niet goed bruikbaar zijn om evaluaties op schoolniveau te doen: de statistische verwachting komt niet altijd overeen met de intuïtieve verwachting dat de gemiddelde cijfers van SE en CE voor alle groepen leerlingen gelijk aan elkaar zouden

zijn. Statistiek op verschillen leidt inderdaad makkelijk tot interpretatiefouten. In dit artikel gaan we verder met statistiek, waar Van Brederode en Meeter ophouden. We verkennen wat we wél kunnen zeggen over SE- en CE-cijfers op schoolniveau, zonder naar verschillen te kijken. Welke uitspraken kunnen onderbouwd worden over SE- en CE-cijfers als de directe verschillen niet berekend worden? Daarvoor doen we enkele berekeningen op de examenresultaten van 2011 en 2019 uit het Basisregister Onderwijs (BRON, beheerd door Dienst Uitvoering Onderwijs). Leerlingen verschillen in de cijfers die zij krijgen bij het eindexamen. De vraag is of die verschillen alleen bepaald worden door vaardigheid of kennis, en eventueel wat toeval, of dat er structurele verschillen zijn tussen scholen in de hoogte van de cijfers die leerlingen krijgen. Maakt het voor een leerling uit op welke school hij of zij zit, als het gaat om

de waardering van zijn of haar vaardigheid of kennis? Structurele verschillen tussen scholen in de vaardigheid en kennis van leerlingen zouden naar verwachting min of meer in gelijke mate een effect moeten hebben op de hoogte van de cijfers op SE en CE. De school heeft vooral invloed op de waardering van vaardigheid en kennis bij het SE. Bij het CE wordt immers bij alle leerlingen de vaardigheid en kennis gemeten met dezelfde opgaven. En de correctievoorschriften zorgen voor een redelijk consistente waardering van het CE. De vraag is dus hoe groot de structurele verschillen tussen scholen zijn in waardering bij het SE.

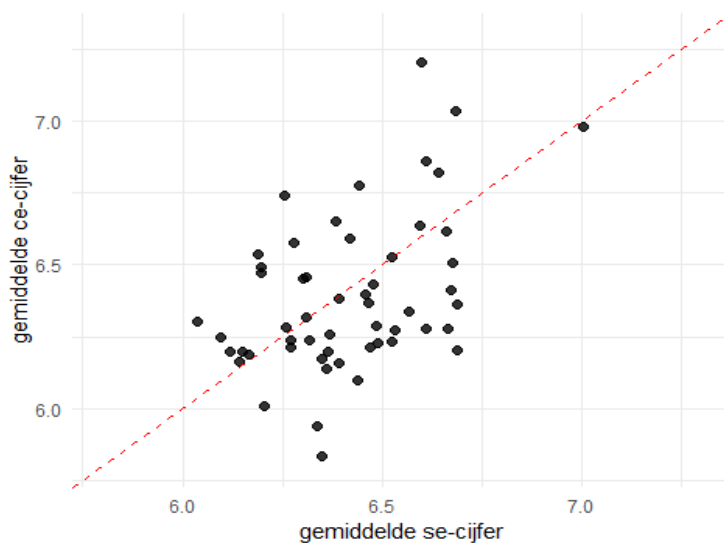
In dit artikel maken we onderscheid tussen verschillen tussen leerlingen, en verschillen tussen schoolgemiddelden. Stap voor stap gaan we de gelaagdheid van de gegevens benaderen en de resultaten bespreken. Daarbij kiezen we ervoor om geen gebruik te maken van multilevel-modellen. Deze zijn weliswaar ontworpen om de gegevens op twee niveaus (leerlingen en scholen) in één keer te analyseren. Maar dat zijn complexe modellen, die

makkelijk foutief te interpreteren zijn. Zeker als het gaat om de relatie van twee variabelen (SE- en CE-cijfer), en naast de twee niveaus leerling en school, ook nog een effect van vak, of vakdocent.

Om de invloed van toevalsfluctuaties te beperken, laten we in de analyses scholen met minder dan 15 leerlingen voor een vak buiten beschouwing. Ook vakken waarbij er minder dan 15 scholen over bleven zijn buiten beschouwing gelaten. Een school die alle onderwijsniveaus aanbiedt, wordt hier als vijf verschillende scholen geteld. Er zijn immers examens op vijf niveaus: drie leerwegen binnen vmbo (basisberoepsgericht - bb, kaderberoepsgericht - kb en gemengd of theoretisch - gl/tl), havo en vwo.

Gemiddelde cijfers van leerlingen

We beginnen met de cijfers van alle eindexamenleerlingen uit 2019 op het schoolexamen (SE) en het centraal examen (CE) van hetzelfde vak. Voor 54 (grote) algemene vakken van de vijf onderwijsniveaus berekenden we de



Figuur 1. Vakgemiddelden op SE en CE (2019)

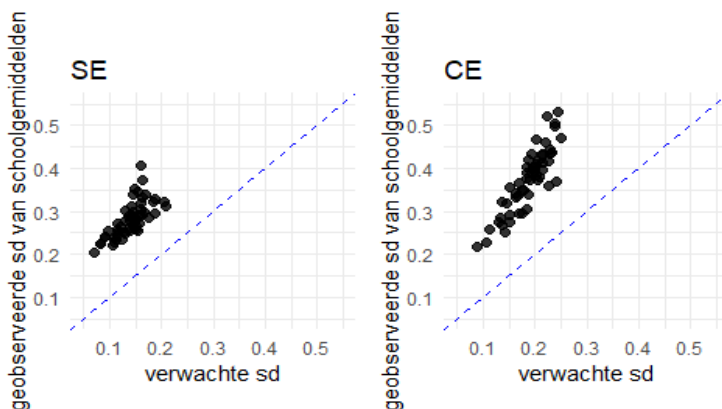
gemiddelde cijfers voor SE en CE. Deze gemiddelde cijfers kunnen verschillen tussen SE en CE, zoals te zien in Figuur 1. De rode lijn geeft aan wanneer de twee gemiddelden gelijk zijn aan elkaar. Dat komt niet vaak precies voor. Bij sommige vakken is het gemiddelde SE-cijfer hoger, bij andere het gemiddelde CE-cijfer, maar ze verschillen nooit meer dan 0,6 cijferpunt. Gemiddeld is het SE-cijfer 0,02 cijferpunt hoger dan het CE-cijfer. Dit verschil is verwaarloosbaar klein. Overigens was dit gemiddelde verschil in 2011 nog 0,27 cijferpunt en lag de puntenwolk grotendeels onder de rode lijn. Dus bij de meeste vakken waren in 2011 de SE cijfers gemiddeld hoger dan de CE-cijfers. De verschillen tussen SE- en CE-cijfers op landelijk niveau zijn dus in de loop van de jaren vermindert. In een nader onderzoek zou een verdere opsplitsing van de vakken in clusters als 'exact' of 'talen' gemaakt kunnen worden, of een analyse per onderwijsniveau.

Schoolgemiddelden per vak

Ook per school berekenden we per vak wat het gemiddelde was op het SE en op het CE. De gemiddelden van deze schoolgemiddelden weken niet veel af van de gemiddelden van alle afzonderlijke leerling-cijfers. Dit toont aan

dat er geen systematisch effect was van de schoolgrootte op de cijfers van de leerlingen. De standaarddeviatie van de schoolgemiddelden is ook interessant. Als scholen een vergelijkbare instroom zouden hebben, en de cijfers van de leerlingen niet beïnvloed worden door de school waarop zij zitten, dan kunnen we de verwachte standaarddeviatie van de schoolgemiddelden afleiden van de standaarddeviatie van afzonderlijke leerling-cijfers en het harmonisch gemiddelde van het aantal leerlingen per school (Sniijders & Bosker, 1994). In Figuur 2 worden deze verwachtingen, als er dus geen structurele verschillen tussen scholen zouden zijn, uitgezet tegen de geobserveerde standaarddeviaties van de schoolgemiddelden.

Bij alle vakken zien we dat de daadwerkelijke spreiding in schoolgemiddelden altijd groter is dan de toevallige spreiding die je verwacht als er geen structurele verschillen tussen scholen zijn. Zowel bij het SE als bij het CE. De punten liggen namelijk boven de blauwe stippellijn die gelijke waarden op beide assen weergeeft. Er zijn dus verschillen tussen scholen in de gemiddelde cijfers die hun leerlingen per vak hebben, die niet door toeval, noch door schoolgrootte, verklaard kunnen worden.



Figuur 2. Spreiding van schoolgemiddelden per vak (2019)

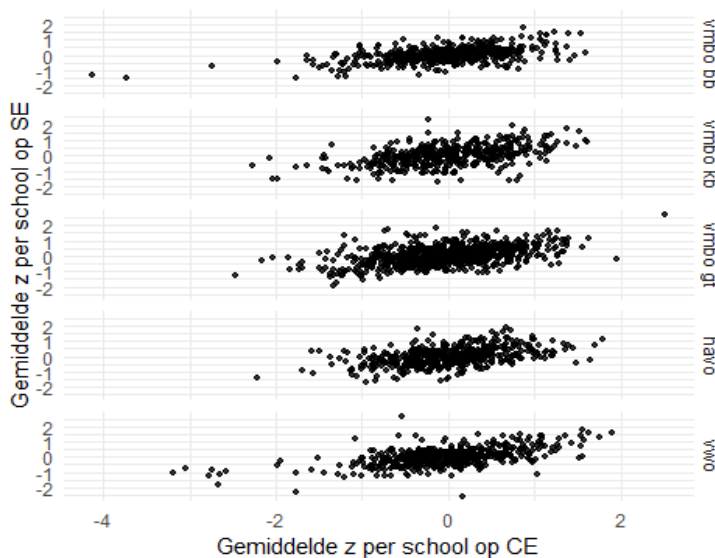
Zowel bij het SE als bij het CE. Dit kan aan de leerling-instroom liggen, of aan keuzes van scholen of hun docenten. Met de gegevens die we gebruiken, kunnen we daar per vak geen verklaring voor geven; daarvoor zijn aanvullende achtergrondgegevens nodig. Wel kunnen we meer inzicht krijgen via de vergelijkingen tussen SE en CE over vakken heen.

Beschrijvende statistieken op schoolniveau

Van Brederode en Meeter kijken niet altijd alleen per vak, maar ook naar gemiddelde eindlijst-cijfers van scholen op SE en op CE. Dus per school één gemiddelde over alle vakken heen. Hier berekenen we ook gemiddelden per school. Om betere vergelijkingen mogelijk te maken, doen we dat iets anders dan zij deden. Vakken verschillen in gemiddelde en standaarddeviatie, ook tussen SE en CE binnen een vak. In Figuur 1 zijn de verschillen tussen gemiddelden per vak te zien. Ook niet alle scholen bieden alle vakken aan. Daarom hebben we de schoolgemiddelden per vak (op ofwel SE, ofwel CE) eerst gestandaardiseerd.

Dit levert per school een rijtje z-scores op, die aangeven in hoeverre de school gemiddeld bij ieder vak boven of onder het landelijke gemiddelde scoorde op SE of CE.

Een z-score of z-waarde geeft het aantal standaarddeviaties aan dat de geobserveerde waarde boven of onder het gemiddelde ligt. Bij een gemiddelde van 50, en een standaarddeviatie (sd) van 10, krijgt de score 60 een z-score van 1.00. Een score van 25 krijgt een z-score van -2.50. Per vak kijken we nu naar de verdeling van de schoolgemiddelden. Stel dat een vak op het SE een gemiddeld cijfer van 6,5 heeft, en de sd van de schoolgemiddelden is 0,3. Een school met voor dit vak een gemiddeld SE-cijfer van 6,5 heeft dan voor dit vak op het SE een z-score van 0,00. Een andere school met een gemiddeld SE-cijfer voor dit vak van 6,2 heeft dan een z-score van -1,00. Negatieve z-scores geven aan dat de school onder het gemiddeld



Figuur 3. Samenhang tussen schoolafwijkingen op SE en CE (2019)

scoorde, positieve z-scores dat de school boven het gemiddelde scoorde. Omdat de gemiddelde landelijke cijfers per vak verschillen, bieden z-scores een beter beeld van de relatieve prestatie van een school, dan de geobserveerde gemiddelden per vak. Bij het ene vak is een 6,5 namelijk heel gemiddeld, bij een ander vak kan dat juist boven of onder gemiddeld zijn. Z-scores hebben per vak een gemiddelde van 0 en een standaarddeviatie van 1, en zijn dus tussen vakken te vergelijken.

Nu nemen we per school het gemiddelde van de z-scores per vak. Deze gemiddelde z-scores per school zijn minder extreem dan per vak, omdat binnen een school afwijkingen naar boven bij het ene vak worden weggemiddeld met afwijkingen naar beneden. Alleen de systematische afwijking van scholen van het landelijke gemiddelde wordt met de gemiddelde z-score weergegeven. En dat zowel voor SE als voor CE.

Vervolgens zijn we gaan kijken naar de samenhang tussen de gemiddelde z-score van een school bij het SE, en bij het CE (Figuur 3). Als de cijfergeving uitsluitend afhangt van de leerling-instroom, dan zouden we verwachten dat de gemiddelde z-score bij SE en CE dezelfde kant op wijst. Dit zou vervolgens resulteren in een hoge positieve correlatie. De correlatie tussen de gemiddelde z-scores bij het SE en bij het CE is bij vmbo-kb het laagst met 0,45 en bij vwo het hoogst met 0,56. Dit betekent dat de gemiddelde cijfergeving van scholen bij SE en CE wel positief samenhangt, maar niet exact gelijk opgaat: de correlatie is niet in de buurt van 1. Verschillen in gemiddelde cijfers tussen scholen kunnen slechts deels verklaard worden door een gemeenschappelijke factor als instroom. Merk op dat fluctuaties op basis van toeval hier nog maar een beperkte rol speelt. Bij scholen zijn vakken met minder dan 15 leerlingen niet meegenomen, en er wordt over



Als op het CE vergelijkbaar gepresteerd wordt, zijn de verschillen in gemiddelde cijfers op het SE meestal rond 0,1 à 0,2 cijferpunt

meerdere vakken gemiddeld: bij vmbo-bb ligt dat aantal het laagst met meestal vier vakken, bij havo en vwo betreft het meestal meer dan elf vakken. Dat de correlatie niet perfect is, kan gezien worden als een indicatie voor verschillen tussen scholen in omgang met SE en CE. De verdeling van aandacht voor SE en CE is bijvoorbeeld niet op iedere school hetzelfde, of de traditie van strengheid en soepelheid bij het SE kan verschillen tussen scholen.

Wat betekent dit nu voor leerlingen? Stel, je zit op een school waar vrij gemiddeld gepresteerd wordt op het CE, hoeveel punten op het SE had je dan naar verwachting meer of minder gehad op een andere school? Daartoe doen we regressies van de gemiddelde z-score op SE op de gemiddelde z-scores op CE, per onderwijsniveau. De residuele variantie zegt dan

iets over de spreiding tussen scholen, gegeven eenzelfde gemiddelde z-score op CE. In de Tabel 1 is in de tweede kolom de residuele standaarddeviatie van de gemiddelde z-scores op het SE weergegeven. Bij vmbo-bb is deze waarde 0,58. Dit betekent dat twee willekeurig gekozen scholen, met eenzelfde gemiddelde afwijking op het CE, naar verwachting 0,58 in gemiddelde z-score op het SE van elkaar verschillen. Dan weten we nog niet hoeveel cijferpunten op het SE dit nu betekent. Daarvoor is een extra stap nodig in de berekening. We vertellen deze residuele verschillen in gemiddelde z-waarden naar residuele verschillen in gemiddelde cijfers. Daarvoor grijpen we terug op de standaarddeviaties van de schoolgemiddelden binnen één vak (Figuur 2). In kolom 3 van Tabel 1 staat de gemiddelde waarde van deze standaarddeviaties op schoolniveau per vak. Dit is een getal dat aan de reguliere cijferschaal gerelateerd kan worden. We vermenigvuldigen kolom 2 met kolom 3 en het resultaat is de residuele variant op het SE in cijferpunten (kolom 4). In cijferpunten lag de residuele variantie op het SE in 2019 tussen 0,11 en 0,18, afhankelijk van het onderwijsniveau. Dit betekent dat als we willekeurig twee scholen zouden trekken, bij eenzelfde gemiddelde prestatie op het CE, het gemiddelde SE cijfer 0,11 tot 0,18 had kunnen verschillen. Als je naar extremen kijkt, verschillen scholen ongeveer 4 standaarddeviaties van elkaar. Daarmee kan het verschil tussen de scholen met de laagste cijfergeving

op SE en de hoogste cijfergeving oplopen tot $(4 \times 0,18 =) 0,7$ cijferpunt bij vmbo bb. Bij havo is de residuele variantie in cijferpunten het kleinst en zou een verschil tussen twee scholen tot maximaal zo'n 0,4 cijferpunt kunnen oplopen. Over het algemeen zullen de verschillen tussen scholen dus veel kleiner zijn, zie de vierde kolom. Ofwel, de gemiddelde cijfers bij het SE verschillen tussen scholen, die vergelijkbaar scoren bij het CE, maar in cijferpunten is dit verschil meestal rond 0,1 à 0,2.

We hebben bovenstaande analyses ook uitgevoerd op de examen-data van 2011. Als we kijken naar de spreiding in gemiddelde afwijkingen, gegeven eenzelfde resultaat op het CE, dan vonden we resultaten die ongeveer gelijk waren aan die in 2019. Met uitzondering van vmbo bb, waar de residuele standaarddeviatie in cijferpunt in 2019 lager is dan in 2011. Zoals we al eerder aangaven, verschilden de landelijke gemiddelde cijfers van SE en CE in 2011 wel behudidend meer van elkaar dan in 2019. Over het algemeen lijkt de cijfergeving van het SE in 2019 dus meer op de cijfergeving van het CE, dan in 2011. Maar de verschillen tussen scholen lijken, behalve bij vmbo bb, niet anders geworden.

Discussie

Van Brederode en Meeter toonden aan dat statistiek op verschillen tussen SE- en CE-cijfers tot verkeerde conclusies kan leiden. In dit artikel hebben we andere analyses gedaan op de SE- en CE-cijfers. Wat weten we na deze analyses?

Tabel 1. Omvang van de residuele verschillen op het SE tussen scholen

niveau	residuele sd	gem. sd van SE-schoolgemiddelden	res.sd in cijferpunt
vmbo bb	0.58	0.31	0.18
vmbo kb	0.53	0.32	0.17
vmbo gt	0.54	0.27	0.15
havo	0.42	0.26	0.11
vwo	0.48	0.28	0.13

- Het landelijke gemiddelde op SE was in 2019 ongeveer gelijk aan het landelijke gemiddelde op CE. In 2011 was het gemiddelde op het SE nog hoger.
- Scholen verschillen in gemiddelde cijfers, zowel bij SE als CE. Dat kan slechts deels door instroom verklaard worden.
- Gegeven eenzelfde gemiddelde afwijking in cijfers van landelijke gemiddelden op het CE, zijn de verschillen tussen scholen in gemiddelde cijfers op het SE meestal rond 0,1 à 0,2 cijferpunt.

Dit betekent voor leerlingen dat er verschillen in gemiddelde examencijfers tussen scholen zijn, zowel op het SE als op het CE. Waardoor dat komt, kunnen we hier niet zeggen omdat we geen aanvullende gegevens hebben. Mogelijke verklaringen zijn verschillen in voorbereidingstijd tussen SE en CE, verschillen in inhoud van het SE, of verschillen in prestatie-eisen op het SE. Dat zou nader onderzocht moeten worden. We zien echter ook dat als op het CE vergelijkbaar gepresteerd wordt, de verschil-

len in cijfergeving op het SE tussen scholen niet groot zijn. Merk op dat we hier kijken naar het gemiddelde verschil tussen scholen. Voor een individuele leerling zou het door toeval wel kunnen uitmaken op welke school hij of zij een SE-cijfer krijgt. In de praktijk is dat niet te constateren, omdat leerlingen nu eenmaal maar op één school eindexamen doen. Hier constateren we dat, over alle vakken heen, scholen de opgedane vaardigheid en kennis van leerlingen over het algemeen op min of meer dezelfde wijze waarderen. ■

Literatuur

- Snijders, T.A.B. & Bosker, R.J. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, 22, 342-363.
- Van Brederode, M. & Meeter, M. (2020). Hoe statistiek het schoolexamen verdacht maakte en waarom schoolexamens dat niet verdienen. *ScienceGuide*, 6 mei 2020.