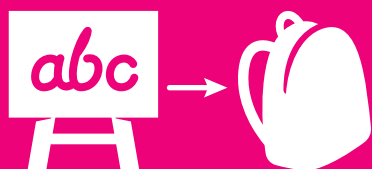




De waarde van eindtoetsen in het primair onderwijs



De eindtoets zou een grotere rol moeten spelen bij de overgang naar de middelbare school, maar daarvoor is een betere vergelijkbaarheid van de toetsadviezen noodzakelijk.

Leerlingen wisselen vaker van niveau op de middelbare school als de informatie uit de eindtoets niet wordt benut.

Toetsadviezen van de verschillende eindtoetsen zijn slechts beperkt vergelijkbaar en de toets is niet voor alle leerlingen even belangrijk.

Samenvatting

De eindtoets zou een grotere rol moeten spelen bij de overgang naar de middelbare school, maar daarvoor is een betere vergelijkbaarheid van de toetsadviezen noodzakelijk. Basisscholen gebruiken de eindtoets bij de overgang naar de middelbare school en de Inspectie voor het Onderwijs gebruikt de eindtoets bij de beoordeling van leerresultaten van scholen. Voor beide toepassingen functioneert de eindtoets momenteel niet optimaal om verschillende redenen. Allereerst zijn er meerdere eindtoetsen die toetsadviezen geven die niet vergelijkbaar zijn. Daarnaast beïnvloedt de eindtoets het schooladvies nu slechts bij een klein deel van de leerlingen, terwijl een combinatie van school- en toetsadvies minder wisselingen van niveau op de middelbare school lijkt op te leveren.

De rol die eindtoetsen innemen is sinds 2015 veranderd. In principe maken alle kinderen aan het eind van de basisschool een eindtoets. Deze eindtoets speelt echter een beperkte rol in de overgang van de basisschool naar de middelbare school, omdat deze wordt afgenomen nadat het schooladvies is gegeven. Alleen als het advies van de eindtoets hoger is dan het schooladvies, moet de school overwegen haar advies bij te stellen.

Leerlingen zitten in de derde klas van de middelbare school vaker op een ander niveau dan hun definitief schooladvies als de informatie uit de eindtoets niet wordt benut. Onze analyses laten zien dat er in Nederland systematische verschillen bestaan tussen het schooladvies en hoe kinderen scoren op de eindtoets. Zo krijgen kinderen van ouders met lagere inkomens vaker een lager schooladvies dan het toetsadvies en wordt hun schooladvies minder vaak bijgesteld. Ook wisselen leerlingen die recht hebben op een heroverweging, maar van wie het schooladvies niet wordt bijgesteld, vaker van niveau op het voortgezet onderwijs dan andere leerlingen. Dit suggereert dat de plaatsing van kinderen op de middelbare school beter is als informatie uit de toets wordt benut. Een objectieve eindtoets kan helpen om te corrigeren voor lagere verwachtingen die leraren, mogelijk onbewust, kunnen hebben van bepaalde groepen kinderen, terwijl het schooladvies helpt om de beperkte focus van een toets aan te vullen.

Toetsadviezen van de verschillende eindtoetsen zijn slechts beperkt vergelijkbaar over tijd en tussen leerlingen. Scholen kunnen kiezen uit verschillende eindtoetsen. Door de beperkte vergelijkbaarheid van deze toetsen hangt het advies dat leerlingen op basis van de eindtoets krijgen, af van welke toets ze gemaakt hebben. Daarnaast is de toets niet voor alle leerlingen even belangrijk: voor kinderen met een vwo-advies en voor kinderen die al tevreden zijn met hun schooladvies, is de eindtoets niet meer van belang, terwijl andere kinderen de toets juist willen gebruiken om hun advies te verhogen. Hierdoor doen kinderen niet allemaal even goed hun best. Deze variatie zorgt ervoor dat de eindtoets niet meer voor alle leerlingen een goede indicator van hun niveau is en daarmee ook minder geschikt is om de kwaliteit van scholen te vergelijken.

Door de eindtoets weer vóór het schooladvies af te nemen, kan deze direct worden benut. Hierdoor kan het schooladvies gebaseerd worden op zoveel mogelijk informatie en kunnen eventuele lagere verwachtingen die leerkrachten voor bepaalde groepen leerlingen worden gecorrigeerd, wat gelijke kansen verbetert. Bovendien hebben scholen dan het toetsadvies achter de hand als tegenwicht voor eventuele druk van ouders om het schooladvies te verhogen. Doordat de eindtoets belangrijker wordt, hebben wel weer meer ouders een prikkel om in te zetten op particuliere toetstraining, wat gelijke kansen kan tegengaan.

Voor het gebruik van de eindtoets als middel om scholen te vergelijken is het bovendien noodzakelijk dat de toetsadviezen goed vergelijkbaar zijn. De ingezette richting van het opnemen van ankers (een vaste set opgaven in de verschillende toetsen) lijkt hiervoor een geschikt middel. Een andere optie om tot vergelijkbare toetsadviezen te komen, is door alle kinderen dezelfde eindtoets te laten maken, zoals ook het geval is bij de centrale examinering op het voortgezet onderwijs.

1 Inleiding

De rol die eindtoetsen innemen is de afgelopen jaren veranderd. Tegenwoordig moeten vrijwel alle kinderen aan het eind van de basisschool een eindtoets maken. Scholen kunnen kiezen voor de Centrale Eindtoets (CET, voorheen de Citotoets) of een andere eindtoets die door de minister is toegelaten. In het schooljaar 2018/2019 kunnen scholen in totaal kiezen uit vijf eindtoetsen. Tegelijkertijd speelt deze eindtoets een minder grote rol in de overgang van de basisschool naar de middelbare school dan vroeger. Middelbare scholen mogen kinderen alleen op basis van het schooladvies toelaten. De eindtoets wordt pas afgenomen nadat het schooladvies is bepaald. Alleen als kinderen op de toets hoger scoren dan hun schooladvies, kan de eindtoets het schooladvies beïnvloeden. Het schooladvies moet dan worden heroverwogen, maar er is geen verplichting om het advies ook bij te stellen.

Er zijn verschillende redenen voor beleid om te streven naar een zo goed mogelijk advies voor de middelbare school. Het is belangrijk dat kinderen op het niveau terecht komen waar ze het beste passen, omdat op deze manier het potentieel van een kind optimaal benut kan worden. Ook is het belangrijk dat kinderen met gelijke kwaliteiten gelijke kansen krijgen. Tot slot draagt een goede match van de kwaliteiten van een kind met het schoolniveau eraan bij dat het geld dat in onderwijs wordt geïnvesteerd, zo goed mogelijk wordt gebruikt.

In deze policy brief houden we, vanuit een economisch oogpunt en ondersteund door data, de rol van eindtoetsen in het primair onderwijs tegen het licht. We kijken hierbij binnen de kaders van het huidige stelsel. Deze policy brief geeft daarmee een antwoord op de volgende vragen: Waarvoor wordt de toets in de huidige situatie gebruikt? Hoe zit het met de vergelijkbaarheid van de prestaties van leerlingen en scholen? Hoe zit het met de objectieve rol van de eindtoets? Wat kunnen we zeggen over kansengelijkheid bij de overgang van het primair naar het voortgezet onderwijs? En tot slot, wat kunnen we doen om de huidige situatie te verbeteren?

2 Toepassingen eindtoets bijten elkaar

Een belangrijke functie van de eindtoets is om naast het schooladvies te dienen als tweede, objectief, gegeven bij de overgang van de basisschool naar de middelbare school. De drijfveer van de grondlegger van de Citotoets, Adriaan de Groot, was zijn ideaal dat ook het kind van de bakker naar het gymnasium kan. Een centrale toets waarbij iedereen langs dezelfde meetlat wordt gelegd, geeft een objectieve indicatie van het niveau van een kind. Zo'n objectieve indicatie kan nuttig zijn om te corrigeren voor eventuele bewuste of onbewuste vooroordelen bij de basisschool over welk niveau kinderen aankunnen in het voortgezet onderwijs.

Daarnaast wordt de eindtoets ook gebruikt als instrument voor externe verantwoording en de beoordeling van leerresultaten van scholen. Dit gebeurt bijvoorbeeld door de Inspectie van het Onderwijs. De gemiddelde score die leerlingen van een bepaalde school behalen op de eindtoets, is een van de manieren waarop de Inspectie kijkt of de onderwijsopbrengsten van scholen voldoen aan het vastgestelde minimum en of scholen hun leerlingen voldoende vaardigheden bijbrengen.

Deze twee toepassingen van de eindtoets kunnen elkaar soms bijten. Een duidelijk voorbeeld hiervan is dat scholen een prikkel hebben om leerlingen specifiek voor de eindtoets te laten oefenen, aangezien de toets door de Inspectie wordt gebruikt voor de beoordeling van basisscholen. Tegelijkertijd maakt dat het lastiger om de eindtoets als het objectieve gegeven te gebruiken bij de overgang naar het voortgezet onderwijs, omdat de resultaten op de eindtoets hierdoor een minder goed beeld geven van de vaardigheden van deze leerling. Een tweede voorbeeld waar de twee toepassingen van de eindtoets elkaar bijten als scholen kunnen kiezen uit meerdere eindtoetsen, is dat het bij de overgang naar de middelbare school niet zou moeten uitmaken welke eindtoets een leerling precies maakt voor het advies dat deze leerling op basis van de toets krijgt. Omdat de eindtoets echter gebruikt wordt door de Inspectie, hebben scholen een prikkel om over te stappen naar de eindtoets die de hoogste toetsadviezen geeft. In het tekstkader staat een uitgebreider overzicht van de prikkels die de eindtoets en het schooladvies opleveren.

Prikkels rondom de eindtoets en het schooladvies

Voor veel leerlingen is er geen prikkel om voor de eindtoets te oefenen, maar basisscholen hebben deze prikkel wel. Voor leerlingen die al tevreden zijn met hun schooladvies, is de eindtoets niet meer van belang, omdat het schooladvies alleen naar boven kan worden bijgesteld. Hetzelfde geldt voor leerlingen met een vwo-advies. Voor leerlingen die graag naar een hoger schoolniveau willen dan hun schooladvies, is de eindtoets nog wel belangrijk. Zulke leerlingen hebben een prikkel om zich specifiek op de eindtoets voor te bereiden. Voor basisscholen die gemiddeld niet zo goed presteren op de eindtoets, is er een prikkel om over te stappen naar een eindtoets die hogere toetsadviezen oplevert. De gemiddelde scores die kinderen op een bepaalde school halen zijn openbaar en beïnvloeden zodoende de toekomstige leerlingpopulatie op een school. Ook hebben basisscholen daarom een prikkel om leerlingen specifiek op de toets voor te bereiden. Voor middelbare scholen zijn er geen prikkels op het gebied van de eindtoetsen, omdat de eindtoets voor hen geen directe rol speelt in de toelating van leerlingen in het voortgezet onderwijs.

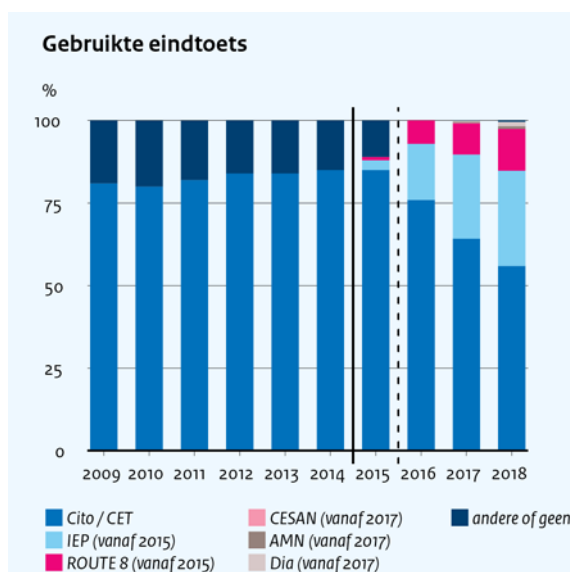
In het huidige systeem zitten vooral prikkels om het schooladvies hoger te laten uitvallen. Hoewel er voor leraren geen prikkels zijn om te hoge of te lage adviezen te geven, is het vanuit het perspectief van de basisschool prettig als er hoge adviezen worden gegeven. De adviezen die leerlingen op een bepaalde school krijgen, zijn namelijk net als de gemiddelde toetsscore openbaar en ouders kijken hiernaar bij het kiezen van een school voor hun kinderen. Voor ouders en leerlingen zijn er momenteel prikkels om de druk op de leerkracht uit te oefenen voor een hoger schooladvies. Tenslotte hebben middelbare scholen een prikkel om informatie over de kwaliteit van de schooladviezen terug te koppelen naar basisscholen. In de beoordeling van middelbare scholen is afstroom (leerlingen die terugzakken naar gemakkelijkere onderwijsniveaus) namelijk een belangrijke indicator. Hierdoor is het voor middelbare scholen prettig als het schooladvies een goede afspiegeling is van het niveau van de leerlingen.

Om de eindtoets goed te kunnen gebruiken bij de overgang naar de middelbare school en bij de beoordeling van scholen door de Inspectie, is het belangrijk dat toetsadviezen vergelijkbaar zijn en dat de eindtoets als objectief gegeven wordt gebruikt. In hoofdstuk 3 bekijken we daarom of de toetsadviezen, de adviezen die kinderen op basis van de eindtoets krijgen, vergelijkbaar zijn over de jaren heen en tussen leerlingen. Als dat niet zo is, worden leerlingen niet langs dezelfde meetlat gelegd en zijn de prestaties van scholen moeilijker met elkaar te vergelijken. In hoofdstuk 4 houden we de objectieve rol van de eindtoets tegen het licht. We onderzoeken of sommige groepen kinderen structureel van de toetsadviezen afwijkende schooladviezen krijgen. Vervolgens bekijken we van welke groep leerlingen het schooladvies vaker wordt bijgesteld en hoe zij het doen op het voortgezet onderwijs.

3 Vergelijkbaarheid van toetsadviezen is beperkt

Toetsadviezen van de verschillende eindtoetsen zijn slechts beperkt vergelijkbaar over tijd en tussen leerlingen. Dit heeft twee redenen. Enerzijds zijn de verschillende eindtoetsen onderling niet goed vergelijkbaar, waardoor ook de adviezen van deze verschillende eindtoetsen niet vergelijkbaar zijn. Dit betekent dat het advies dat leerlingen krijgen op basis van de eindtoets, deels afhangt van welke toets zij hebben gemaakt. Anderzijds is de toets niet voor alle leerlingen van even groot belang en hierdoor is het niet voor alle kinderen een goede indicatie van hun niveau. Het probleem van de beperkte vergelijkbaarheid wordt versterkt, omdat er steeds meer verschillende eindtoetsen worden gebruikt door basisscholen en omdat de scholen die overstappen van de CET (de Centrale Eindtoets, de opvolger van de Citotoets) naar andere eindtoetsen, aanvankelijk gemiddeld slechter scoorden op de CET. Onvergelijkbare toetsadviezen zorgen ervoor dat de eindtoets niet optimaal gebruikt kan worden bij de overgang naar de basisschool en de beoordeling van scholen door de Inspectie. In dit hoofdstuk lichten we deze factoren toe.

Figuur 3.1 Steeds minder scholen kiezen de laatste jaren voor de Centrale Eindtoets



Bron: Eigen berekening op basis van CBS Microdata. Noot: 2009 refereert naar het schooljaar 2008/2009. De streep geeft de invoering van de Wet Eindtoetsing PO aan in het schooljaar 2014/2015. Het eerste jaar was nog een overgangsjaar, dit is aangegeven door de stippellijn.

Scholen gebruiken steeds meer verschillende eindtoetsen (figuur 3.1). Terwijl in 2015 nog op 85% van de basisscholen de CET gedaan werd, gebeurde dit in 2018 nog maar op 56% van de scholen. Vooral het aantal scholen dat de IEP en ROUTE 8 afnemen, is de laatste jaren flink toegenomen, van ongeveer 4% van de scholen in 2015 tot 42% in 2018. Sinds 2017 zijn nog drie andere eindtoetsen beschikbaar: AMN, Dia en CESAN.¹ Het aantal scholen dat voor een van deze toetsen kiest, is echter beperkt (ongeveer 2% in 2018).

Scholen die lager scoorden op de CET, stappen vaker over naar andere eindtoetsen (figuur 3.2). Van de scholen die in 2015 behoren tot de 10% scholen die gemiddeld het laagst scoren op de CET, stapte 19% over naar een andere eindtoets, terwijl van de scholen die de hoogste gemiddelde toetsscore behaalden op de CET, slechts 6% overstapte.² De scholen die nu de CET nog afnemen, zijn dus scholen die voorheen gemiddeld hoger scoorden.³

Figuur 3.2 Scholen die gemiddeld lager scoorden op de CET, stappen vaker over naar andere eindtoetsen



Bron: Eigen berekening op basis van CBS Microdata. Alle scholen die in 2015 de CET afnamen, zijn verdeeld op basis van de gemiddelde score die er op een school behaald werd. De linkerbalk geeft de scholen weer die gemiddeld behoren tot de 10% laagst scorende scholen op de CET en de rechterbalk geeft de 10% van de scholen weer die de hoogste gemiddelde toetsscore behaalden op de CET.

Het advies dat leerlingen op basis van de eindtoets krijgen, hangt deels af van welke toets ze gemaakt hebben. Zo krijgen leerlingen op scholen die overstappen van eindtoets, in de jaren na de overstap substantieel andere toetsadviezen dan vóór de overstap (figuren 3.3 en 3.4). Dit wijst op verschillen in de toetsadviezen tussen de eindtoetsen.⁴ Zo komen op scholen die in 2016 overstapten van de CET naar de IEP, in de daaropvolgende jaren vaker vwo-toetsadviezen en minder vaak vmbo-toetsadviezen voor (figuur 3.3). Op deze scholen kreeg 14% van de leerlingen een vwo-toetsadvies op basis van de CET in 2015, terwijl in de daaropvolgende jaren 23% (in 2016), 20% (in 2017) en 18% (in 2018) van de leerlingen een vwo-toetsadvies kregen op basis van de IEP. Op scholen die overstappen van de CET naar ROUTE 8, zien we vooral in het eerste jaar na de overstap een sterk afwijkend patroon in de toetsadviezen (figuur 3.4). Zo bestond er in 2016 bij

¹ In 2019 zijn er vijf eindtoetsen, omdat de CESAN-eindtoets niet meer wordt aangeboden.

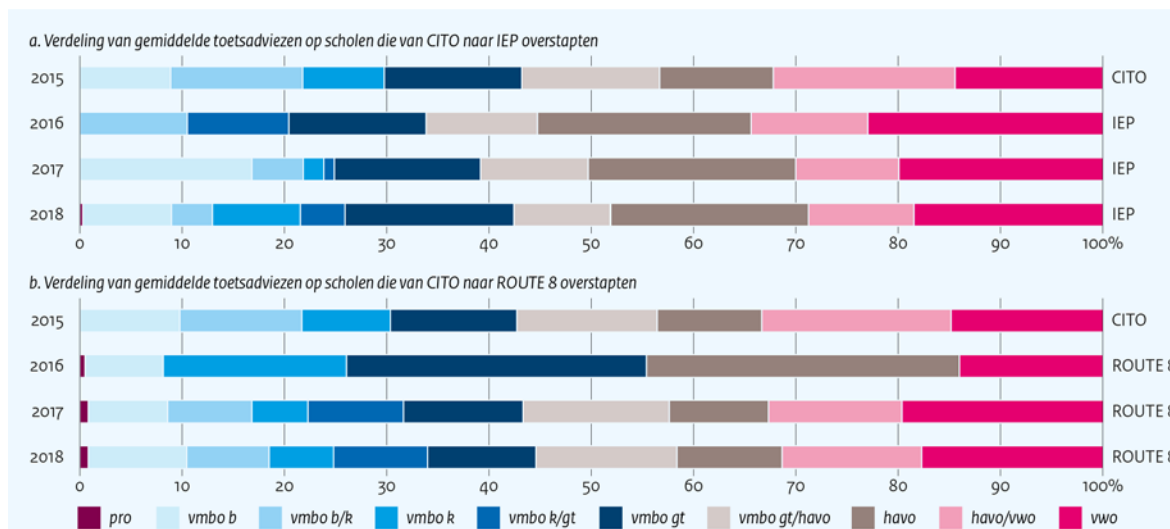
² Dit patroon is hetzelfde als we kijken naar scholen die in andere jaren overstappen, als we de decielen bepalen op basis van meerjarengemiddelden van de Cito/CET-score en als we rekening houden met achtergrondkenmerken van leerlingen. Overigens is het belangrijk om op te merken dat het niet duidelijk is welk van de eindtoetsen het 'beste' toetsadvies oplevert.

³ Dit heeft overigens geen invloed op de uitslagen van de leerlingen die wel de CET maken. De uitslag die een leerling behaalt op de eindtoetsen is niet afhankelijk van de uitslag van andere leerlingen.

⁴ Deze verschillen kunnen waarschijnlijk niet worden verklaard door veranderingen in het gemiddelde prestatieniveau van de leerlingen in deze periode, aangezien de trends van verschillende achtergrondkenmerken niet substantieel verschillen.

ROUTE 8 geen havo/vwo-advies en in 2016 is het aandeel leerlingen dat een havoadvies heeft gekregen, flink gestegen ten opzichte van 2015, terwijl het aandeel dat een vwo-advies kreeg, praktisch gelijk bleef. Dit illustreert dat de vergelijkbaarheid ook werd aangetast doordat de toetsen verschillende categorieën hanteerden voor het advies. Met name voor ROUTE 8 zien we dat na het eerste jaar de gemiddelde toetsadviezen weer meer lijken op die van voor de overstap. Wel is het aantal vwo-toetsadviezen in 2017 en 2018 op overgestapte scholen hoger dan toen op deze scholen nog de CET werd afgenomen.⁵

Figuur 3.3 Hogere toetsadviezen nadat scholen van de Centrale Eindtoets overstappen naar de IEP en aanvankelijk juist lagere toetsadviezen nadat scholen overstappen naar ROUTE 8



Noot: Deze figuur toont scholen die in 2015 de CET en in de jaren 2016 t/m 2018 de IEP (figuur 3.3a) of ROUTE 8 (figuur 3.3b) hebben afgenomen. In figuur 3.3a gaat het om ongeveer elfduizend leerlingen per jaar op 459 scholen die overstappen van de CET naar de IEP. In figuur 3.3b gaat het om ongeveer vijfduizend leerlingen per jaar op 192 scholen die overstappen van de CET naar ROUTE 8.

Omdat de eindtoets niet voor alle leerlingen van even groot belang is, is de toets niet voor iedereen een goede indicatie van hun niveau. Voor kinderen die al tevreden zijn met hun schooladvies en voor alle kinderen met een vwo-schooladvies, is de eindtoets van minder groot belang, terwijl andere kinderen de toets juist willen gebruiken om hun advies te verhogen. Leerlingen tonen niet altijd dezelfde inzet bij belangrijke toetsen als bij toetsen die minder belangrijk zijn (Finn, 2015). Bovendien is het verschil in resultaat tussen toetsen die van groot en minder groot belang zijn niet voor iedereen gelijk. Dit verschil is bijvoorbeeld groter voor mannen dan voor vrouwen (Attali *et al.*, 2011). Bovendien geeft de eindtoets geen goede indicatie meer van het verschil in niveau tussen leerlingen met lage en met hoge vaardigheden, omdat in ieder geval voor alle leerlingen met een vwo-schooladvies de eindtoets niet meer belangrijk is.

Onvergelijkbare toetsadviezen zorgen ervoor dat de eindtoets niet optimaal gebruikt kan worden en ze kunnen leiden tot kansenongelijkheid. De informatie uit de eindtoets is pas objectief als leerlingen met dezelfde capaciteiten op elke eindtoets hetzelfde toetsadvies halen. Daarnaast is het ook voor het toezicht door de Inspectie noodzakelijk dat de resultaten van verschillende eindtoetsen vergelijkbaar zijn, ook over tijd. De Inspectie signaleert dat dat vooralsnog niet zo is (Inspectie, 2017, 2018). In de huidige situatie lijken leerlingen met dezelfde capaciteiten niet op elke eindtoets hetzelfde toetsadvies te krijgen. Dit kan leiden tot kansenongelijkheid als de kans op heroverweging van het schooladvies afhangt van welke eindtoets gemaakt wordt. Er is wel een aantal ontwikkelingen in gang gezet die de vergelijkbaarheid van de verschillende

⁵ Op scholen die in al deze jaren de CET afnamen, is geen trend te zien in het aantal vwo-toetsadviezen.

eindtoetsen in de nabije toekomst moeten verbeteren, zie het tekstkader. Het is nog een openstaande vraag of deze maatregelen afdoende zijn om de toetsadviezen onderling en over tijd vergelijkbaar te maken.

Maatregelen ter verbetering van de vergelijkbaarheid

Sinds de invoering van de Wet Eindtoetsing PO, is een aantal maatregelen genomen om de vergelijkbaarheid van de resultaten van de verschillende eindtoetsen te verbeteren (Kamerstukken, 2018).

1. Sinds 2018 is een set van opgaven Nederlandse taal, rekenen en wiskunde aan alle eindtoetsen toegevoegd. Het doel van dit 'anker' is om tot vergelijkbare toetsadviezen te komen. Dit betekent dat de kans dat een leerling bijvoorbeeld een havoadvies krijgt op basis van de ene toets, gelijk is aan de kans op een havoadvies bij een andere toets. Omdat een eindtoets door veel leerlingen wordt gemaakt, hoeven niet alle leerlingen elke vraag uit het anker te beantwoorden. Elke leerling krijgt slechts een paar vragen uit het anker voorgelegd. Een vergelijking van de prestaties van leerlingen op basis van alleen de ankervragen is dus niet mogelijk.
2. Tot 2018 hanteerden de eindtoetsen verschillende categorieën voor het toetsadvies (zie ook figuur 3). Vanaf 2019 hanteren alle eindtoetsen dezelfde zes categorieën voor het toetsadvies: pro/vmbo-bb, vmbo-bb/vmbo-kb, vmbo-kb/vmbo-gl-tl, vmbo-gl-tl/havo, havo/vwo en vwo.

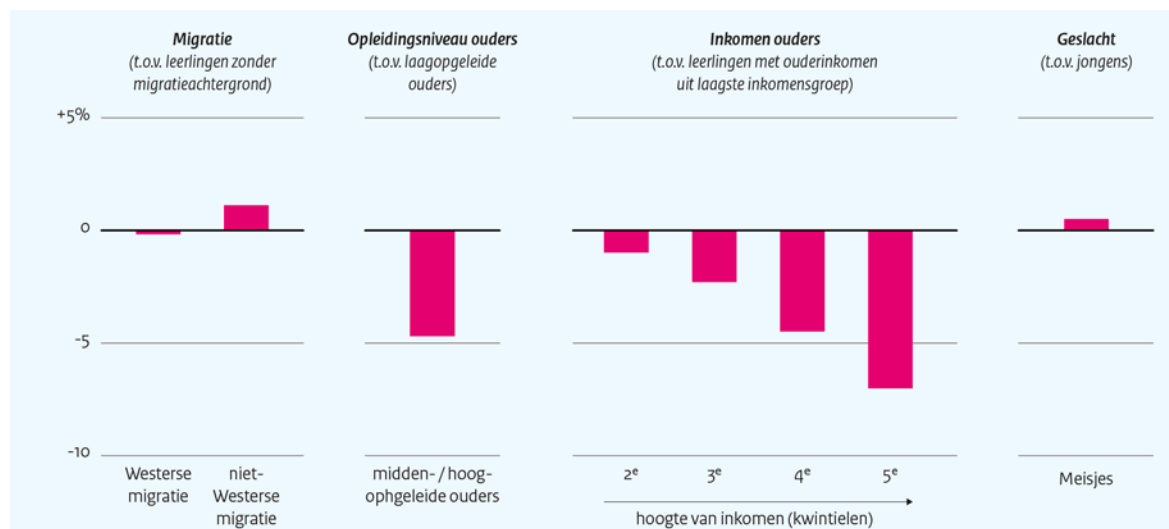
4 Eindtoets verbetert plaatsing

Leerlingen wisselen meer van niveau op de middelbare school als de informatie uit de eindtoets niet wordt benut. De internationale literatuur laat zien dat leraren sommige groepen kinderen structureel anders beoordelen dan uit de resultaten van een toets blijkt. Een objectieve toets kan helpen om te corrigeren voor (onbewuste) vooroordelen bij leraren, terwijl het schooladvies helpt om de beperkte focus van een toets aan te vullen. Onze analyses laten zien dat er ook in Nederland systematische verschillen zijn tussen hoe sommige groepen kinderen beoordeeld worden in het schooladvies en hoe deze kinderen scoren op de toets. Leerlingen die recht hebben op een heroverweging, maar van wie het schooladvies niet wordt bijgesteld, blijken vaker van niveau te wisselen op het voortgezet onderwijs dan andere kinderen. Dit suggereert dat de plaatsing van kinderen op de middelbare school beter is als ook informatie uit de toets wordt benut. Dit hoofdstuk gaat nader in op deze aspecten.

Kinderen van laagopgeleide ouders of van ouders met een laag inkomen hebben vaker recht op een heroverweging van hun schooladvies (figuur 4.1). Deze kinderen worden bij het schooladvies dus lager ingeschat dan verwacht mag worden op basis van hun toetsresultaat, waardoor ze in aanmerking komen voor een heroverweging. Kinderen met een migratieachtergrond hebben gemiddeld ongeveer net zo vaak als kinderen zonder migratieachtergrond recht op een heroverweging, als in de analyse ook gecontroleerd wordt

voor het inkomen en het opleidingsniveau van ouders, omdat ze de toets beter maken dan het advies van de school.

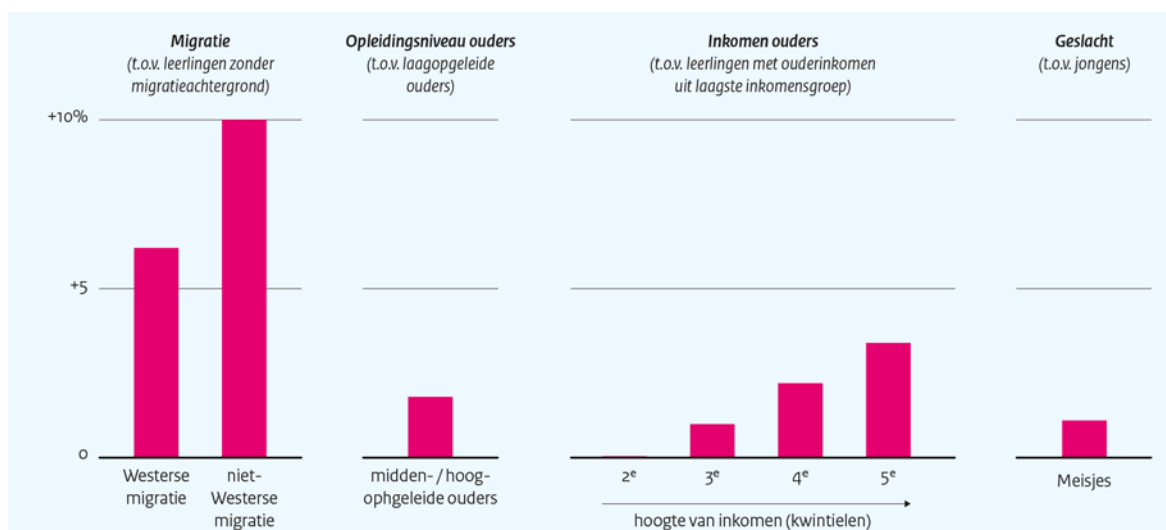
Figuur 4.1 De kans dat kinderen recht hebben op een heroverweging van het schooladvies, verschilt tussen groepen leerlingen



Noot: In deze figuur wordt getoond welke kinderen grotere kans hebben op een heroverweging, oftewel een schooladvies hebben dat lager is dan het toetsadvies. De figuur laat de ondergrens van de kans op heroverweging zien, omdat alleen verschillen van minimaal een heel niveau worden meegenomen. In deze analyses is rekening gehouden met andere achtergrondkenmerken en verschillen tussen scholen die niet veranderen over de tijd, zoals de locatie van de school, de onderwijsachtergrond en eventuele niet-geobserveerde verschillen.

Bovendien verschilt de kans dat het schooladvies wordt bijgesteld tussen groepen leerlingen (figuur 4.2). Van de leerlingen die in aanmerking komen voor een heroverweging, wordt van ongeveer een derde het schooladvies ook daadwerkelijk bijgesteld. Kinderen met een migratieachtergrond hebben een grotere kans op een bijstelling, gegeven dat ze recht hebben op een heroverweging. Hetzelfde geldt voor kinderen van ouders met een hoger inkomen. Kinderen van ouders met een laag opleidingsniveau hebben daarentegen een lagere kans op een bijstelling.

Figuur 4.2 De kans op bijstelling van het schooladvies (gegeven dat kinderen recht hebben op een heroverweging) verschilt ook tussen verschillende groepen kinderen.



Noot: In deze figuur worden alleen kinderen van wie het toetsadvies ten minste één niveau hoger is dan het schooladvies, meegenomen. Al deze kinderen hebben dus recht op een heroverweging van het schooladvies. Het gaat om 14% van de leerlingen en van hen wordt bij één derde van de kinderen het schooladvies bijgesteld. In de analyse gebruiken we data van de CET, IEP en ROUTE 8. Het schooljaar 2014/2015 laten we buiten beschouwing, omdat dit een overgangsjaar is. In de onderliggende regressies is rekening gehouden met dezelfde achtergrondkenmerken als bij figuur 4.1.

Kinderen die recht hebben op een heroverweging, maar voor wie het schooladvies niet wordt bijgesteld, wisselen vaker van niveau op het voortgezet onderwijs. 22% van de kinderen die recht hebben op een heroverweging en waarvan het uiteindelijke advies ook wordt bijgesteld, zitten in de derde klas van de middelbare school op een ander niveau dan hun definitieve schooladvies aangaf. Van de kinderen van wie het schooladvies wel wordt heroverwogen, maar niet wordt bijgesteld, wisselt 32% van niveau.⁶ Veruit het grootste deel van hen (86%) stroomt op naar een hoger niveau. Dit suggereert dat de toets een belangrijke rol kan spelen bij de juiste plaatsing van kinderen op het voortgezet onderwijs. Eerder onderzoek voor Limburg liet al zien dat een combinatie van het schooladvies en de eindtoets beter voorspelt wat de onderwijsprestaties van een leerling zijn, dan het schooladvies alleen (Feron *et al.*, 2016).

Deze systematische verschillen tussen het schooladvies en het toetsadvies zijn niet uniek voor Nederland. Uit de internationale literatuur blijkt dat ook in andere landen leraren sommige groepen kinderen systematisch anders beoordelen dan een toets. Zo laten Burgess en Greaves (2013) voor het Verenigd Koninkrijk zien dat 11-jarige kinderen met een Caribische of Afrikaanse migratieachtergrond systematisch lager worden ingeschat door leraren dan op basis van hun toetsresultaat verwacht mag worden. Het omgekeerde geldt voor kinderen met een Chinese of Indiase migratieachtergrond. Cornwell *et al.* (2013) laten voor de Verenigde Staten zien dat jongens vaker hoger scoren dan meisjes op toetsen voor lezen, rekenen en wetenschap. Leraren schatten meisjes juist vaker hoger in bij lezen. Lavy (2008) laat bij middelbareschoolexamens in Israël zien dat leraren jongens systematisch lager beoordelen in alle vakken dan op basis van hun toets mag worden verwacht.

⁶ Het gaat ook hier weer alleen om kinderen die minimaal één niveau hoger scoren op de toets dan het schooladvies. Wisselen van niveau betekent hier dat de plaatsing van het kind in het derde jaar van het voortgezet onderwijs minimaal één niveau afwijkt van het definitieve schooladvies. Van de leerlingen die geen recht hebben op een heroverweging, wisselt 19% van niveau. Deze verschillen blijven bestaan als wordt gecorrigeerd voor kenmerken van de leerlingen, ouders, het definitief advies en verschillen tussen po- en vo-scholen die niet veranderen over tijd.

Er zijn verschillende mogelijke verklaringen voor systematische verschillen tussen school- en toetsadviezen. Zo is een toets een momentopname⁷ en meet die niet alle vaardigheden die van belang zijn voor succes in het voortgezet onderwijs. Een leraar kan daar mogelijk beter zicht op hebben. Daarnaast zouden ouders met een hoger inkomen bijvoorbeeld vaker kunnen zorgen voor hulp bij het huiswerk van hun kinderen, waardoor de leraar denkt dat het kind een hoger niveau op de middelbare school aankan. Mogelijk zijn deze ouders ook beter in het overtuigen van de leraar om een hoger advies te geven voor hun kind.⁸

Een andere verklaring voor de verschillen zijn (onbewuste) vooroordelen bij leraren. Leraren kunnen systematisch lagere verwachtingen hebben van kinderen met een lagere sociaaleconomische achtergrond, of van jongens tegenover meisjes. Zij hoeven zich hier helemaal niet bewust van te zijn. Uit de wetenschappelijke literatuur weten we dat mensen dit soort, vaak onbewuste, vooroordelen hebben.⁹ Dit kan bijvoorbeeld gaan over geslacht (Reuben *et al.*, 2014) of etniciteit (Glover *et al.*, 2017). Deze onbewuste vooroordelen zijn te vinden bij alle groepen in de samenleving. Bewustwording kan helpen deze vooroordelen te voorkomen. Een recent onderzoek laat zien dat wiskundeleraren in Italië lagere cijfers geven aan migrantenkinderen, terwijl ze hetzelfde zouden moeten scoren gegeven hun toets (Alesina *et al.*, 2018). De onderzoekers laten zien dat dit samenhangt met de impliciete stereotypen die deze leraren hebben. Zodra ze alle leraren vertellen over hun onbewuste vooroordelen, geven zij de migrantenkinderen hogere cijfers. Dit geldt echter ook voor de leraren die al geen onderscheid maakten in hun cijfers, waardoor zij juist te hoge cijfers geven.

5 Beleidsopties

Door de eindtoets een grotere rol te laten spelen bij de overgang naar de middelbare school, kan betere plaatsing bewerkstelligd worden, maar daarbij is een betere vergelijkbaarheid van de toetsadviezen belangrijk. Basisscholen gebruiken de eindtoets bij de overgang naar de middelbare school en de Inspectie gebruikt de eindtoets bij de beoordeling van leerresultaten van scholen. Door de beperkte vergelijkbaarheid van de eindtoetsen kan de Inspectie de toets niet goed inzetten voor de beoordeling van leerresultaten en worden leerlingen niet langs dezelfde meetlat gelegd. Daarnaast is de informatie uit de eindtoets nu slechts voor een klein deel van de leerlingen relevant bij de overgang naar de middelbare school, terwijl een combinatie van school- en toetsadvies minder op- en afstroom oplevert.

Beleidsoptie 1: de eindtoets moet kunnen worden meegenomen in het schooladvies om het schooladvies zo goed mogelijk te maken.¹⁰ Een schooladvies is zo goed mogelijk als het op zoveel mogelijk informatie gebaseerd kan worden. Hiervoor is het nodig dat de eindtoets wordt afgenomen voordat het schooladvies wordt gegeven. Het voordeel hiervan is dat scholen dan het toetsadvies achter de hand hebben als tegenwicht voor druk van ouders om het schooladvies te verhogen. In de huidige situatie gebruiken basisscholen nu ook informatie uit toetsen bij het vaststellen van het schooladvies (Oberon en Universiteit Twente, 2017). Deze toetsen uit het leerlingvolgsysteem (LVS) zijn echter vaak toetsen die bedoeld zijn om de leerbehoeften van leerlingen vast te stellen, zogeheten formatieve toetsen. Deze toetsen zijn dus niet bedoeld om te gebruiken als externe evaluatie (Onderwijsraad, 2018). Wel zorgt de belangrijkere rol van de eindtoets ervoor dat alle kinderen weer een prikkel hebben om particuliere toetstraining te volgen om zo hun toetsadvies te verhogen. Hierdoor is het effect van een dergelijke maatregel op kansgelijkheid onduidelijk:

⁷ Dit betekent dat de toets de vaardigheden van een kind niet altijd goed meet. Het kan bijvoorbeeld dat een kind een slechte dag heeft en daardoor slechter scoort dan we zouden verwachten op basis van zijn normale resultaten (Diris en Borghans, 2017).

⁸ Het is ook mogelijk dat de toets juist bepaalde kinderen bevoordeelt, maar gegeven dat we hetzelfde patroon bij alle eindtoetsen zien, lijkt dit niet waarschijnlijk.

⁹ In de literatuur wordt dit 'implicit bias' genoemd.

¹⁰ Deze beleidsoptie is in lijn met het voorstel van de PO- en de VO-raad.

de toetsscore kan mogelijk onterecht lagere verwachtingen van leerkrachten corrigeren, wat gelijke kansen verbetert, terwijl particuliere toetstraining juist zorgt voor minder gelijke kansen. Er is geen evidentie dat het eventueel eerder afnemen van de eindtoets ervoor zorgt dat leerlingen minder effectieve lestijd zouden overhouden in groep 8.¹¹

Beleids optie 2: maak eindtoetsen vergelijkbaarder om de toetsadviezen vergelijkbaarder te krijgen. De ingezette richting van ankers opnemen in de verschillende toetsen lijkt hiervoor een geschikt middel, maar het is belangrijk om komend schooljaar te toetsen of dit inderdaad het geval is. Een andere optie om tot vergelijkbare toetsadviezen te komen is door alle kinderen dezelfde eindtoets te laten maken, zoals ook het geval is bij de centrale examinering op het voortgezet onderwijs. Een belangrijk voordeel hiervan is dat alle leerlingen dan langs dezelfde meetlat worden gelegd. Bovendien is de eindtoets dan weer beter te gebruiken voor beoordeling van schoolprestaties door de Inspectie, omdat de resultaten dan weer onderling en over tijd vergelijkbaar zullen zijn. Tegelijkertijd heeft de toetsaanbieder die in een dergelijke situatie de enige eindtoets aanbiedt, minder prikkels om de toets te innoveren en gaat het ten koste van de keuzevrijheid van basisscholen om een eindtoets te kiezen die aansluit bij hun onderwijsconcept.

6 Referenties

Alesina, A., M. Carlana, E. La Ferrara, P. Pinotti, 2018, Revealing stereotypes: Evidence from immigrants in schools. *National Bureau of Economic Research*, 25333.

Attali, Y., Z. Neeman, A. Schlosser, 2011, Rise to the challenge or not give a damn: differential performance in high vs. low stakes tests. IZA Discussion Paper No. 5693.

Burgess, S., E. Greaves, 2013, Test scores, subjective assessment, and stereotyping of ethnic minorities, *Journal of Labor Economics*, 31(3), pp.535-576.

Cornwell, C., D.B. Mustard, J. Van Parys. 2013, Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school, *Journal of Human resources*, 48(1), pp. 236-264.

CPB, 2019, Verschillen in leerresultaten tussen basisscholen, CPB Notitie.

Diris, R., L. Borghans, 2017, Meetfout in de relatie tussen opleidingsniveau ouders, citoscore en schooladvies, *Educatieve Agenda Limburg*, <https://www.educatieveagendalimburg.nl/limburg-in-cijfers/meetfout-de-relatie-tussen-opleidingsniveau-ouders-cito-en-advies>

Feron, E., T. Schils, B. ter Weel, 2016. Does the teacher beat the test? The value of the teacher's assessment in predicting student ability. *De Economist*, 164(4), pp.391-418.

Finn, B., 2015, Measuring motivation in low-stakes assessments, *ETS Research Report Series*, 2015(2), pp.1-17.

¹¹ Uit de tussenrapportage van de evaluatie van de Wet Eindtoetsing PO blijkt dat leerkrachten niet het idee hebben dat de onderwijstijd in groep acht sinds de invoering van de wet effectiever wordt besteed (Oberon en Universiteit Twente, 2017). Daarnaast laat een pilot uit 2011 waar een groep scholen de eindtoets twee maanden later afnam, ook zien dat dit niet leidde tot meer aandacht voor taal en rekenen in groep acht (Roeleveld et al. , 2011).

Glover, D., A. Pallais, W. Pariente, 2017, Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores, *The Quarterly Journal of Economics*, 132(3), pp. 1219–1260.

Inspectie van het Onderwijs, 2017, De staat van het onderwijs 2015/2016, Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.

Inspectie van het Onderwijs, 2018, De staat van het onderwijs 2016/2017, Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.

Kamerstukken, 2018, Schooladviezen en resultaten op de eindtoets in schooljaar 2017-2018, 31293, nr. 422.

Lavy, V., 2008, Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment, *Journal of Public Economics*, 92(10-11), pp.2083-2105.

Oberon en Universiteit Twente, 2017, Evaluatie Wet Eindtoetsing PO: Tussenrapportage.

Onderwijsraad, 2018, Toets wijzer: Naar een eigen(tijdse) wijze van toetsen en examineren, Den Haag: Onderwijsraad.

Roeleveld, J., L. Mulder, T. Paas, 2011, Gevolgen van een latere afname van de Cito Eindtoets basisonderwijs, Nijmegen/Amsterdam: ITS/Kohnstamm Instituut.

Rueben, E., S. Paola, L. Zingales, 2014, How stereotypes impair women's careers in science, *Proceedings of the National Academy of Sciences*, 111(12), pp. 4403-4408.