

Eindtoets basisschool: terug naar het doel

De eindtoets in groep 8 zorgt voor veel discussie. Logisch, want alle ouders, leerlingen en leraren hebben er op enig moment mee te maken. Machiel Karels, directeur van kennisplatform Wij-leren.nl, vindt discussie terecht, want volgens hem zijn er nogal wat problemen met de eindtoets. Aan de hand van een analyse signaleert Karels tien problemen en geeft hij een oplossingsrichting aan.

Tekst: Machiel Karels

In dit artikel wordt gesproken over 'de eindtoets' die in vijf verschillende varianten beschikbaar is. De genoemde problemen die er met de eindtoetsen zijn, gelden voor alle varianten. Op dit moment zijn dat de Centrale Eindtoets (van de overheid, gemaakt met expertise van Cito), de IEP, Route 8, DIA en AMN.

De eindtoets heeft twee functies. Hij dient als secundaire, onafhankelijke indicatie voor het schooladvies van de school. Daarnaast wordt hij gebruikt als element om scholen te beoordelen op hun kwaliteit. In dit artikel zal ik duidelijk maken waarom de eindtoets voor beide doelen nogal beperkt geschikt is of tenminste bijzonder veel nadelen heeft.

→ De eindtoets als onafhankelijke indicatie voor het schooladvies

Sinds schooljaar 2015-2016 is de afname van de eindtoets verplaatst naar een later tijdstip, zodat de uitslag pas bekend wordt ná het opstellen van het schooladvies door de school. Dat is een beweging ten goede, zodat scholen niet blindvaren op de uitslag van een nogal smalle toets.

Probleem 1: Lage dekkingsgraad

Het belangrijkste probleem van de eindtoets is de beperkte dekkingsgraad. Een eindtoets wekt de suggestie dat het een realistische weergave is van de leerstof die in de basisschoolperiode aangeboden is. Niets is echter minder waar. De eindtoets is een bijzonder smalle representatie van de houding, kennis en vaardigheden waar in de acht jaar durende

basisschoolperiode aan gewerkt wordt. De inhoud van de eindtoets dekt maximaal dertig procent van de kerndoelen voor het basisonderwijs (Karels, 2018).

Het wordt nog nijpender als we de persoonlijke vaardigheden van de leerlingen in deze berekeningen van de dekkingsgraad meenemen. Technisch gezien is dat natuurlijk niet te berekenen en die kant moet het ook niet op. Iedereen weet dat (toekomstig) schoolsucces maar ten dele afhangt van de harde meetbare getallen. Als we deze persoonlijke vaardigheden in de berekening van het dekkingspercentage meenemen, daalt die dertig procent drastisch. Ik wil hiermee niet aangeven dat we de persoonlijke vaardigheden ook moeten toetsen, als dat al mogelijk zou zijn. Het gaat erom dat we de eindtoets niet meer waarde toedichten dan deze verdient.

Vaardigheden die Cito niet toetst volgens wij-leren.nl

<i>Geduld</i>	<i>Verbazing</i>
<i>Discipline</i>	<i>Zelfsturing</i>
<i>Creativiteit</i>	<i>Motivatie</i>
<i>Accuratesse</i>	<i>Leiderschap</i>
<i>Mededogen</i>	<i>Moreel besef</i>
<i>Bescheidenheid</i>	<i>Behoedzaamheid</i>
<i>Betrouwbaarheid</i>	<i>Nieuwsgierigheid</i>
<i>Doorzettingsvermogen</i>	<i>Verantwoordelijkheid</i>
<i>Inlevingsvermogen</i>	<i>Verdraagzaamheid</i>
<i>Kritisch denken</i>	<i>Samenwerking</i>
<i>Daadkracht</i>	<i>Zelfstandigheid</i>
<i>Flexibiliteit</i>	<i>Veerkracht</i>
<i>Humor</i>	<i>Moed</i>

Bron: Wij-leren.nl (2016)

Probleem 2: Relatieve meting

Een tweede belangrijk probleem van de eindtoets is dat het een relatieve meting is, in tegenstelling tot een absolute meting. Dit betekent dat de uiteindelijke verdeling van de eindscores gebaseerd is op het landelijk behaalde gemiddelde resultaat. Een voorbeeld: Stel dat alle leerlingen minder dan tien fouten maken op de eindtoets. Bij een absolute toets zouden ze dan allemaal een standaardscore van dicht tegen de 550 krijgen en een bijbehorend vwo-advies. Bij een relatieve toets worden de scores van alle leerlingen genomen en op volgorde gezet van hoog naar laag. De leerlingen met precies het gemiddelde krijgen een standaardscore van 535 en de anderen dus hoger of lager.

In dit voorbeeld krijgt dan de onderste helft van de leerlingen een vmbo-advies en alle leerlingen boven 536 krijgen een havo- of vwo-advies (Onderwijs in cijfers, 2017). Alhoewel de eindscores per soort eindtoets verschillen, is de onderliggende psychometrie bij elke toetsvariant hetzelfde. Doordat het bij de eindtoets om grote aantallen leerlingen gaat, is het niet zo extreem als in het voorbeeld, maar het is wel het onderliggende rekenmodel. Werken we met z'n allen heel hard om een zo hoog mogelijke eindtoetsscore te halen, dan wijzigt er uiteindelijk niets in de verhoudingen en de uitkomsten. Overigens werkt de onderliggende systematiek van de tussentoetsen precies zo. Hierover is al eerder gepubliceerd (Zwik, 2014).

Probleem 3: Bevestiging en versterking kansenongelijkheid

De eindtoets is bij uitstek talig, ook bij de rekenonderdelen. Dit verschilt wel iets per toets, maar het blijft een knelpunt. Leerlingen die om wat voor reden dan ook minder taalvaardig zijn, scoren dus lager dan op grond van hun intelligentie verwacht zou mogen worden (Berends & Van Lieshout, 2009). Dit zijn vaak kinderen uit achterstandssituaties, maar ook kinderen in hogere opleidingsniveaus kunnen problemen hebben met de taligheid van de eindtoets. Deze is deels bedoeld om als objectieve toets kansenongelijkheid tegen te gaan, maar het werkt door de taligheid dus precies omgekeerd.

Naast dat er op school voor de eindtoets wordt geoefend, zijn er allerlei oefenmaterialen verkrijgbaar waarmee de leerlingen thuis kunnen oefenen. Via internet kun je complete series historische eindtoetsen aanschaffen. Juist ouders met voldoende financiële middelen kunnen op deze manier schaduwonderwijs inkopen en zo de uitslag van de toets beïnvloeden. Hiermee komt de kansengelijkheid in het gedrang.

Daarnaast blijkt uit onderzoek dat de provincie Drenthe heeft laten doen, dat lagere adviezen en minder bijstellingen samenhangen met lagere verwachtingen van ouders en een minder ambitieuze leercultuur (Westerveld, 2018). Dit is een belangrijk probleem waarover veel is gepubliceerd.

Probleem 4: 'Teaching to the test'

Wanneer scholen deels 'afgerekend' worden op de resultaten van de eindtoets, is dit een perverse prikkel die hoe dan ook leidt tot *teaching to the test*. Het publiceren van eindtoetsgegevens op landelijke rankinglijsten werkt daar nog eens extra aan mee (Karels, 2014). Dergelijke prikkels zorgen voor een focus van de school op smallere, meetbare opbrengsten, waarvan de eindtoets een exponent is.

Huidige wetgeving en functies eindtoets in het kort

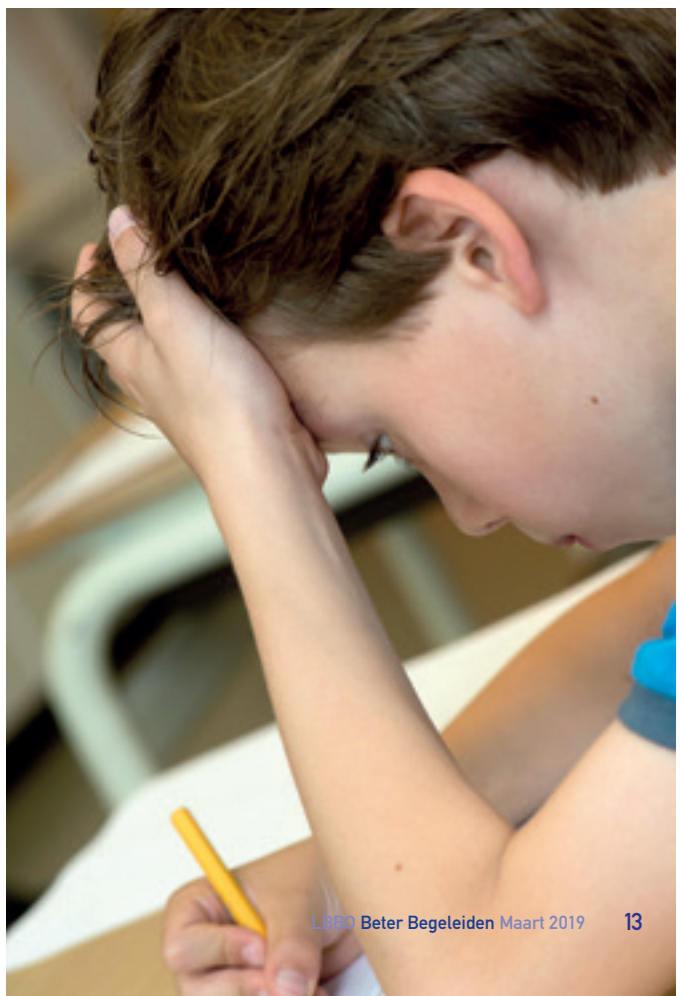
Huidige wetgeving

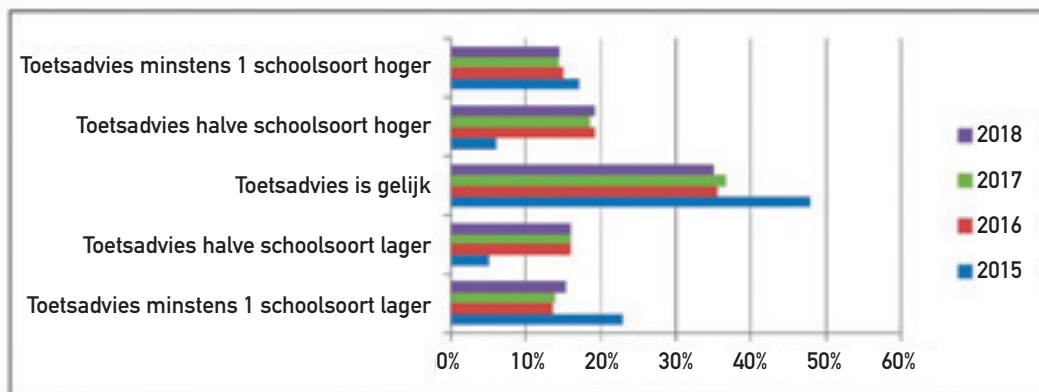
- Schooladvies uiterlijk 1 maart gegeven, leidend voor toelating in het voortgezet onderwijs.
- Verplichte afname eindtoets voor alle leerlingen in het achtste leerjaar, behalve leerlingen die onder de ontheffingsgronden vallen.
- Keuzevrijheid in door de minister toegelaten eindtoetsen.
- Afname tussen 15 april en 15 mei.
- Hoger toetsadvies dan schooladvies leidt tot een verplichte heroverweging en eventueel bijstelling. Wanneer niet wordt bijgesteld, moet dit beargumenteerd worden.
- Toets doet uitspraak over meest passend type voortgezet onderwijs en beheersing van de referentieniveaus lezen, taalverzorging en rekenen.

Huidige functies eindtoets

- Tweede inschatting van het best passende schooltype voortgezet onderwijs voor de leerling.
- Bepaling van de beheersing van de referentieniveaus lezen, taalverzorging en rekenen van de leerling.
- Reflectie van scholen op hun onderwijs en verantwoording daarover.
- Extern toezicht op kwaliteit van afzonderlijke scholen en het stelsel als geheel.

Bron: PO-Raad (2018)





Figuur 1: Schematische weergave van het verschil tussen schooladvies en toetsadvies in de afgelopen vier jaar (bron: brief van de onderwijsminister aan de voorzitter van de Tweede Kamer, november 2018).

Veel scholen oefenen in de weken voor het afnamemoment klassikaal voor de eindtoets met behulp van toetsen van eerdere schooljaren. Daarnaast is het zo dat leraren die ervaring met de eindtoets hebben, bewust of onbewust extra aandacht geven aan elementen van het onderwijs die op eindtoetsen aan de orde komen.

Probleem 5: Schooladvies en toetsadvies lopen vaak uit elkaar

Landelijk onderzoek laat zien dat het door de school opgestelde schooladvies en het toetsadvies dat uit de eindtoets voortvloeit, maar in 35 procent van de gevallen gelijk zijn (Onderwijs in Cijfers, 2018). We zeggen meestal dat het oordeel van de leraar en het schooladvies het belangrijkste zijn. Maar hoe valide en betrouwbaar is dan een toets die maar in 35 procent van de gevallen hetzelfde aangeeft als het schooladvies? En kun je van een toets verwachten dat die de toekomst van een leerling kan voorspellen?

→ De eindtoets als element om de kwaliteit van het onderwijs te evalueren

De resultaten van de eindtoets worden door de Onderwijsinspectie gebruikt om scholen te beoordelen op hun kwaliteit. Op zich is er niets mis met het beoordelen van scholen op hun kwaliteit met betrekking tot de referentieniveaus taal en rekenen. Hiervoor is immers een opbrengstverplichting afgesproken (Overheid.nl, 2014). Voor de beoordeling van de onderwijskwaliteit is recent het onderwijsresultatenmodel ontwikkeld (Onderwijsinspectie.nl, 2018). In dit model gaat het om de percentages leerlingen die aan het einde van het basisonderwijs het fundamentele niveau 1F en 1S halen voor rekenen, taalverzorging en lezen. In het 'oude model' werd er voor de beoordeling gebruikgemaakt van de gemiddelde score op de eindtoets. Binnen het onderwijsresultatenmodel wordt op een nieuwe en gedetailleerdere manier naar de schoolweging gekeken. Dit is een stap in de goede richting, maar het lost niet alle problemen met deze functie van de eindtoets op.

Probleem 6: De eindtoets doet geen recht aan brede ontwikkeling

De beperkte dekkingsgraad is ook bij het beoordelen van de kwaliteit van scholen een knelpunt. Op zich is de gemiddelde score op de eindtoets maar een onderdeel van het uiteinde-

lijke oordeel van de Onderwijsinspectie. Toch wringt de smalle kijk op alleen taal en rekenen bij veel scholen, en dan met name de scholen die de brede ontwikkeling van kinderen op het oog hebben. De eindtoets doet namelijk weinig recht aan scholen die naast de cognitieve vakken veel aandacht geven aan de persoonsvorming en talentontwikkeling van kinderen. Scholen hebben de neiging om hun lesprogramma op de eindtoets af te stemmen. De eindtoets zorgt zo voor een perverse prikkel die verschraling van het onderwijsaanbod tot gevolg heeft (Onderwijsraad, 2018).

Probleem 7: Eindtoetsscore zegt weinig over kwaliteit

Als de score op de eindtoets veel zegt over de onderwijskwaliteit, zou daar een direct verband tussen moeten zijn. Maar dat is niet het geval. Goede scholen halen niet per definitie hoge eindtoetsscores en zwakkere scholen niet per definitie lagere scores. De eindtoets doet overigens ook geen recht aan scholen die qua gedrag en leren een zware leerlingpopulatie hebben (PO-Raad, 2018). Bij dit aspect speelt ook al sterk de vraag wat er wordt verstaan onder onderwijskwaliteit. Gaat het dan alleen om de scores op taal en rekenen? Dat is een nogal smalle definitie van onderwijskwaliteit. Dit is echter wel precies wat de huidige eindtoets in beeld brengt.

Een eindtoets wekt de suggestie dat het een realistische weergave is van de leerstof

Daarbij komen nog de fenomenen 'momentopname' en 'leerlingpopulatie'. Hoe complex je de weging van de leerlingpopulatie ook maakt, de werkelijkheid is altijd weerbarstiger en zorgt ervoor dat je scholen niet op deze manier rechtvaardig met elkaar kunt vergelijken.

Overigens is het nieuwe onderwijsresultatenmodel hierin wel veel beter dan de werkwijze die voorheen gehanteerd werd. Maar dan nog blijft het op deze manier beoordelen van

de kwaliteit een optelsom van stevig van elkaar verschillende individuen en onderwijsgebieden in een gecombineerd getal. Dit geeft de school niet de feedback waar zij wat mee kan voor het verbeterbeleid.

Probleem 8: Oneigenlijk gebruik door externe partijen

Wanneer de gemiddelde uitslag van de eindtoets gebruikt wordt voor de beoordeling van de onderwijskwaliteit, wekt dat de suggestie dat kwaliteit zich op die manier laat meten en in één getal laat vangen. Aan die suggestie doet de overheid mee middels Vensters voor Verantwoording en door rapportages naar de Tweede Kamer hierop te baseren. Maar het werkt daarnaast ook het produceren van rankinglijsten in de hand. De rankinglijst van RTL is daar een bekend voorbeeld van. Een stuitend voorbeeld, als je bedenkt dat RTL jarenlang een financieel belang heeft gehad in de toets-oefensoftware *Squla* (Karels, 2017).

Probleem 9: Tegenstrijdige belangen

De twee functies van de huidige eindtoets – schooladvisering en kwaliteitsmeting – staan op gespannen voet met elkaar. Dit komt door de tegenstrijdige prikkels die bij elk doel spelen. Als scholen de eindtoets willen gebruiken als tweede, objectief criterium voor het schooladvies, is het prima als de eindtoets een realistisch beeld van de leerling geeft. Een te hoog eindtoetsadvies is niet handig, want dan krijgt de leerling het zwaar in het voortgezet onderwijs. Te laag ook niet, want dan zit de leerling ook niet op zijn plek. De prikkel voor de school is

hier dus niet: hoe hoger hoe beter, maar: hoe realistischer, hoe beter. Overigens kan er bij de ouders een andere prikkel leven, die het alleen nog maar complexer maakt.

Bij de functie 'inzicht geven in de onderwijskwaliteit' ligt het anders. Om een goede beoordeling te krijgen, willen scholen zo hoog mogelijk scoren. Dat effect wordt versterkt als er landelijke vergelijkingslijsten gemaakt worden van de uitslag. Hier is de prikkel dus: hoe hoger, hoe beter (College voor Toetsen en Examens, 2018). De prikkel van de kwaliteitsbeoordeling komt hier dus in conflict met de prikkel van de schoolkeuze.

Probleem 10: Onzuiverheid door verhuizingen enzovoort

De gemiddelde score van de eindtoets meewegen in de beoordeling van de kwaliteit van een school is een manier die aan vrij grote onzuiverheden onderhevig is. Enkele leerlingen die vrij hoog of juist laag scoren, hebben een relatief grote impact op het gemiddelde. Daarnaast zijn er de verhuizingen van leerlingen. Instroom in groep 7 en 8 hoeft niet meegerekend te worden, maar heeft wel impact op de inspanningen van de

Er zijn allerlei oefenmaterialen te koop waarmee de leerlingen thuis kunnen oefenen





school op het eindresultaat. Dit effect is sterker naarmate de betreffende school kleiner is. Er zijn daarnaast overigens wel uitzonderingsbepalingen voor leerlingen die nog maar kort in Nederland zijn (College voor Toetsen en Examens, 2018). Onlangs was ik op een school met minder dan honderd leerlingen. De directeur vertelde mij dat er vorig jaar een meerbegeefde leerling enkele weken voor de afname van de eindtoets verhuisde naar een naburig dorp. Voor beide scholen had dit een grote impact op de gemiddelde score, zonder dat zij daar invloed op hadden gehad in relatie tot de onderwijskwaliteit. Met name op zeer kleine scholen speelt het probleem van de onzuiverheid sterk, aangezien daar vaak minder dan tien leerlingen in groep 8 zitten.

→ Waar moet het heen met de eindtoets?

Er zijn dus nogal wat problemen met de eindtoets. Moet een instrument dat zoveel nadelen kent, dan maar afgeschaft worden? Op zich zou dat een optie zijn, maar dat is wellicht wat kort door de bocht. Recent onderzoek door de Algemene Vereniging Schoolleiders (AVS) maakt duidelijk dat zestig procent van de schoolleiders de eindtoets als *second opinion* wil handhaven. Maar er moet in ieder geval wat gebeuren.

Wees bewust van de beperkte waarde

Allereerst is het belangrijk dat er rondom de eindtoets duidelijkheid is over de smalle focus en de daarmee samenhangende beperkte waarde. Helderheid over de dekkingsgraad in

Zestig procent van de schoolleiders wil de eindtoets handhaven als *second opinion*

relatie tot de kerndoelen is daarin een eerste stap. Daarnaast is goede voorlichting van leerlingen, ouders en scholen belangrijk. De eindtoetsscore gebruiken in een plaatsingswijzer voor het voortgezet onderwijs is in dit verband ook uiterst ongewenst. Overigens kan alles qua inhoud en dekkingsgraad op termijn weer op de schop gaan, gezien de ontwikkelingen rond Curriculum.nu.

Breid mogelijkheden van late selectie uit

Scholen voor voortgezet onderwijs zouden veel meer mogelijkheden moeten bieden voor een latere keuze voor een specifiek schooltype. Dat zou de druk verlagen om al aan het eind van groep 8 tot een precieze keuze voor het juiste onderwijsniveau te komen. Een eindtoets veroorzaakt in veel gevallen te vroegtijdige labeling en selectie op basis van een momentopname. Na drie jaar zit 25 procent van de leerlingen op een ander niveau dan bij het schooladvies was geadviseerd

(Onderwijs in Cijfers, 2013). Dat is een behoorlijk percentage. Van dakpanklassen en brede brugklassen zijn prima voorbeelden bekend (De Adelhart Toorop & Werner, 2018). Dergelijke onderwijstypen maken het voor leerlingen mogelijk om rustiger een onderwijssoort in te groeien.

Gebruik de eindtoets niet voor kwaliteitsbeoordeling

De gemiddelde score op de eindtoets zou niet meegenomen moeten worden in de beoordeling tijdens het vierjaarlijks onderzoek van de Onderwijsinspectie. Zelfs als er een beter gewogen score beschikbaar is, middels het onderwijsresultatenmodel, zou dit moeten leiden tot gesprek en niet tot beoordeling. Feitelijk is er sinds het nieuwe toezicht al meer gesprek over de eigen kwaliteitsambities van scholen en besturen. Met de eindtoetsen zou dezelfde aanpak gehanteerd kunnen worden, zoals nu bij de tussentoetsen gangbaar is: wel over de score in gesprek, maar er geen normeringen op loslaten. Daarmee zou de aanpak van de Onderwijsinspectie onder het nieuwe toezichtkader meteen ook meer congruent zijn. De functie van het beoordelen van de opbrengstverplichting moet niet uitgevoerd worden door een instrument dat gemaakt is voor een ander doel.

Bevorder de diversiteit bij eindtoetsen

De huidige instructies en criteria van de overheid om eindtoetsen toe te laten, zijn zodanig op uniformiteit gericht, dat er voor scholen nauwelijks iets te kiezen valt. Dit is nodig, omdat de overheid de uitslag van de eindtoets wil gebruiken voor onderlinge vergelijking, management en beheersbaarheid (Overheid.nl, 2018). De overheid moet echter weer terug naar het oorspronkelijke doel van de eindtoets: de leerlingen een optimaal advies geven. En dan moet er juist diversiteit zijn en moet het spectrum aan eindtoetsen aansluiten bij de diversiteit die er is aan pedagogische en didactische werkwijzen en culturen in de scholen.

In het onderzoekskader stimuleert de inspectie het ontwikkelen van een eigen visie en lerende cultuur. Daar moet het instrumentarium aan eindtoetsen bij aansluiten (Ministerie van OCW, 2018). Op het moment dat een eindtoets goed in lijn is met de pedagogische en didactische visie van de school, zullen schooladvies en toetsadvies waarschijnlijk ook dichter bij elkaar komen.

Als er veel diversiteit in eindtoetsen is, en deze ook elk jaar doorontwikkeld worden, wordt het voor alle betrokkenen ook veel lastiger om te oefenen voor de eindtoets.

Teaching to the test en de daarmee samenhangende kansenongelijkheid doven daarmee uit. Ruimte voor diversiteit in eindtoetsen maakt het voor toetsleveranciers ook mogelijk om wat te doen aan de taligheid van toetsen en de nadelen die dat met zich meebrengt.

Ga wel in gesprek over de score, maar laat er geen normeringen op los

Laat scholen zelf beslissen over publicatie van resultaten

Als de Onderwijsinspectie in gesprek gaat over de scores, maar er niet op beoordeelt, is het niet nodig om de resultaten van de eindtoets openbaar te maken, ook niet middels Vensters voor Verantwoording. Met het loslaten van de openbaarmaking steek je een spaak in het wiel van de externe partijen die rankinglijsten maken. Scholen kunnen dan zelf beslissen met wie ze de resultaten van de eindtoets delen. En ze kunnen zelf bepalen hóé ze die delen en welke toelichting zij daarbij geven.

Samenvattend

We moeten als onderwijssector weer terug naar het oorspronkelijke doel van de eindtoets: de leerlingen een optimaal schooladvies geven. Dat vraagt om diversiteit qua eindtoetsen die aansluiten bij de pedagogische en didactische werkwijzen en culturen in de scholen. Oneigenlijk gebruik van de eindtoets voor kwaliteitsmeting en ranking van scholen moet zo snel mogelijk stoppen.

Literatuur

Bij dit artikel hoort een uitgebreide literatuurlijst. Deze vind je in Beter Begeleiden Digitaal van april 2019 en in de materialenbank bij het artikel op www.lbbo.nl.

In de materialenbank op www.lbbo.nl vind je onder meer de volgende artikelen over toetsen:

- *Een andere kijk op toetsen (november 2008)*
- *Vaardigheidsscores: Wat kan een leerling? (oktober 2011)*
- *'Inspectie, stop met beoordelen onderwijskwaliteit op basis van leerresultaten' (mei 2014)*

TIP



Machiel Karels is directeur van het onderwijskundig kennisplatform *Wij-Leren.nl*. Daarnaast begeleidt hij verandertrajecten op weg naar gepersonaliseerd onderwijs. E-mail: machiel@wij-leren.nl.

AUTEUR