

Gespreksnotitie Doorstroomtoets

Ten behoeve van rondetafelgesprek toetsing, examens en selectie op 24 april

Karen Heij (zelfstandig toetsexpert)

Geachte leden van de vaste Kamercommissie van onderwijs,

Aanstaande 24 april ben ik bij u uitgenodigd voor een rondetafelgesprek over toetsing, examens en selectie in het funderend onderwijs ter voorbereiding op het commissiedebat Toetsen en examens funderend onderwijs van 15 mei 2025.

Ik maak hierbij gebruik van de mogelijkheid om mijn standpunten voorafgaand aan de bespreking onder uw aandacht te brengen. Die zijn voor een belangrijk deel ontleend aan het promotieonderzoek dat ik deed naar de eindtoets basisonderwijs.¹

De doorstroomtoets² is in bijna zestig jaar³ uitgegroeid tot een nu onder regie van de overheid ontwikkeld en door de overheid verplicht gesteld instrument met twee belangrijke functies:

- het verdelen van toegangsbewijzen voor vervolgonderwijs (selectie), en
- toezicht houden op de kwaliteit van onderwijs.

De conclusie van dit promotie-onderzoek was dat de eindtoets (nu doorstroomtoets) niet voldoende valide is, niet voor selectie en niet als bepaler van kwaliteit van onderwijs. Er rust een veelheid van belangrijke, te vervullen taken, op de schouders van de doorstroomtoets die deze niet kan dragen.

Er is de afgelopen jaren het nodige gedaan om de eindtoetsen aan te passen om beide functies beter te vervullen waaronder een

¹ Proefschrift 'Van de kat en de bel, tellen en vertellen met de eindtoets basisonderwijs', 2021, <https://karenheij.bijzonderboeken.nl/>

² Er zijn weliswaar meerdere doorstroomtoetsen maar ze zijn in functie vergelijkbaar. Daarom kies ik ervoor te spreken van 'de' doorstroomtoets.

³ In eerste instantie was er de benaming van Amsterdamse Schooltoetsen (1966). Vanaf de oprichting van Cito was de toets onderdeel van Cito en kreeg de benaming 'Citotoets'. Vanaf 2014 is de toets verplicht voor alle scholen en is de overheid verantwoordelijk voor de 'Centrale Eindtoets'(CE) met daarnaast nog vier andere toetsen van andere toetsaanbieders. Vanaf nu geldt de benaming 'doorstroomtoets' met in totaal acht varianten geproduceerd door vijf verschillende toetsaanbieders, deels digitaal deels in papieren variant, onder centrale regie van het College voor Toetsen en Examens (CvTE).

naamsverandering. Die verandering van naam heeft niet geleid tot wezenlijke veranderingen in de opzet van de toets. Nog altijd wordt zowel leerlingen als scholen onvoldoende recht gedaan met deze toets. Ik ga in het vervolg van deze notitie nader in op de problemen van de doorstroomtoets in het licht van beide wettelijke functies.

Ad 1 Selectiefunctie (primair)

De wet is duidelijk wat betreft de functie van de doorstroomtoets: die is primair bedoeld voor advies over vervolgonderwijs. Het vervolgonderwijs kent in Nederland de volgende opties: vwo, havo, vmbo (tl, gemengde leerweg, kaderberoepsgerichte leerweg en basisberoepsgerichte leerweg) en praktijkonderwijs. Voor de doorstroomtoets geldt dat er alleen gecombineerde schooladviezen mogen worden gegeven, met uitzondering van vwo. Dit betekent dat de doorstroomtoets leerlingen moet verdelen over 6 verschillende adviezen.⁴

Vergelijken als uitgangspunt: discussie ‘hoger versus lager’ is gevolg van de manier van toetsen

Een toets kan niet anders dan proberen een afspiegeling te geven van de werkelijkheid. Een toets modelleert de bestaande werkelijkheid. Zo is de doorstroomtoets van nu gemodelleerd naar de populatie van leerlingen die destijds, in 2018, in het derde jaar voortgezet onderwijs hebben gevolgd.⁵ De keuze voor een dergelijke modellering gaat er vanuit dat een populatie niet wezenlijk verandert en over het algemeen stabiel is. Zo is de aanname vanaf de eerste versie van de doorstroomtoets (1966) dat voor ongeveer 20% van de leerlingen een vwo-advies passend is en dat de helft van onze leerlingen via het beroepsgerichte voortgezet onderwijs zijn onderwijsloopbaan vervolgt. Deze aannames hangen samen met het idee dat de doorstroomtoets een vorm van intelligentie zou vaststellen die ‘normaal verdeeld is’.⁶ Daarmee is de basis van de doorstroomtoets een vergelijking van alle leerlingen met elkaar: Leerlingen maken dezelfde toets, hun scores worden vergeleken, de leerlingen geordend en verdeeld over de verschillende schooladviezen. De doorstroomtoets is géén geschiktheidsoordeel. Er is namelijk geen onderzoek waaruit blijkt wat aan kennis en vaardigheden, competenties of houding nodig is voor een bepaalde vorm van voortgezet onderwijs. De doorstroomtoets stelt dus ook geen ‘niveau’ vast. Er is niets anders dan de optelling van scores op (een klein deel van) taal en rekenen om te komen tot een verdeling.

⁴ Vwo, havo/vwo, vmbo gt/havo, vmbo kb/vmbo gt, vmbo bb/vmbo gt, praktijkonderwijs/vmbo bb

⁵ In de Terugblik Normering Doorstroomtoetsen van CvTE (2024) wordt gemeld dat het laatste cohortonderzoek dat gebruikt is voor de normering van 2018 is.

⁶ Zie de diverse verantwoordingen van Cito (1981,1992, 2010) en ook van CvTE (Verantwoording Centrale Eindtoet PO, 2015; Terugblik 2019)

Daarmee is de discussie over 'hoger' of 'lager' ook moeilijk te beteugelen want dit is precies wat er gebeurt: leerlingen worden langs dezelfde meetlat gelegd en kunnen daar hoger of lager op scoren. En bij een hoge score hoort een hoger advies gegeven.

Beperkingen cohortonderzoeken

De onderbouwing van deze manier van werken wordt vooral ontleend aan onderzoek naar waar leerlingen zitten na drie jaar (na het gegeven advies). Dit is een zichzelf bevestigende manier van onderzoeken omdat eerst de toets is ingezet voor het maken van de verdeling. Een leerling die eerder het advies kreeg voor vmbo wordt in die drie jaar geen vwo-leerling. Andersom wordt een leerling met een vwo-advies ook niet snel een vmbo-leerling. Hier werkt het principe van de 'selffulfilling prophecy': leerlingen, maar ook hun ouders en leraren geloven in het advies en gaan zich ernaar gedragen. Bovendien zijn de belangen in het voortgezet onderwijs groot om leerlingen te houden in de stromen waarin ze zijn geadviseerd. Alsnog blijkt een derde van de leerlingen na drie jaar meer dan een niveau onder het gegeven advies terecht te zijn gekomen. Het willen voorspellen van de schooltoekomst van leerlingen blijkt keer op keer een hachelijke zaak waarbij toetsen maar zeer beperkt van waarde kunnen zijn en vooral fungeren als 'waarmakers' van hun eigen voorspelling.

Vroege selectie en kansengelijkheid

Uit internationaal onderzoek blijkt dat vroege selectie nadelig is in het licht van kansengelijkheid. De selectie in Nederland is, zeker in vergelijking met andere westerse landen, vroeg te noemen. De OESO (o.a. verantwoordelijk voor de PISA-onderzoeken) wijst Nederland ook op de gevolgen van vroege selectie in het licht van het bieden van zo gelijk mogelijke onderwijskansen. Te vroeg selecteren is nadelig voor kinderen die met achterstand het onderwijs binnenkomen. Dit geldt zowel voor leerlingen met een beperkt sociaal cultureel kapitaal (laag SES) als voor leerlingen met een andere thuistaal dan het standaard Nederlands. Door het verplichte gebruik van een Leerlingvolgsysteem (LVS) worden leerlingen in Nederland al vanaf groep 3 met elkaar vergeleken en geordend in vijf groepen die net als de eindtoets geordend worden van 'hoog' naar 'laag' scorend in percentielgroepen.⁷ Daarmee start de selectie feitelijk al ver voor het moment van de doorstroomtoets.

Een probleem van de huidige manier van selecteren is dat hier de suggestie vanuit gaat dat de inschatting van het schooladvies gebaseerd is op een stabiel gegeven, een kenmerk van een persoon dat uit te drukken is in een getal, vergelijkbaar met een intelligentiescore. De afgelopen zestig jaar is er veel onderzoek gedaan waaruit blijkt dat

⁷ Dit geldt voor het meest gebruikte LVS-systeem Leerling in Beeld van Cito.

intelligentie voor een deel wellicht stabiel is maar zeker ook voor een belangrijk deel ontwikkelbaar is mits er goed en uitdagend onderwijs wordt gegeven. Op een leeftijd van 12 jaar is zeker nog geen stabiele inschatting te maken. Hoe jonger de inschatting, hoe bepalender de rol van de herkomst of thuissituatie van een leerling. Vroege selectie is dan ook af te raden als het willen bieden van meer kansengelijkheid serieus genomen wordt. Dit is ook het standpunt dat de Onderwijsraad onlangs in een brief aan de TK onder de aandacht bracht.⁸

LVS-toetsen en de effecten van 'ondergemiddeld' scores

Om van het selectiemoment in groep 8 geen momentopname te maken is bij de verplichting tot deelname aan de doorstroomtoets ook een verplichting tot het gebruik van een LVS⁹ opgenomen in de wet. Hiermee zou een stabiel beeld van de leerling kunnen worden verkregen wat scholen zouden moeten gebruiken bij het opstellen van een schooladvies. Hier kleven een aantal belangrijke bezwaren aan.

Het eerste is dat scholen daarmee ook nog maar heel smal naar een leerling kijken: wat getoetst wordt, bepaalt de blik op de leerling. Daarmee is de kijk op de leerling in zijn totaliteit, alle kwaliteiten meewegend, naar de achtergrond geschoven. In plaats daarvan wordt nu vooral gekeken naar de getoetste scores op begrijpend lezen, taalverzorging (grammatica en spelling) en rekenen.

Het tweede bezwaar is dat leerlingen nu steeds vroeger geordend worden ten opzichte van elkaar en dus steeds vroeger beginnen aan het selectieproces. Omdat ook LVS-toetsen kunnen leiden tot selffulfilling prophecy werken LVS-toetsen niet uit als objectieve inschatters die bijdragen aan kansengelijkheid maar precies in het tegendeel: leerlingen die met achterstand het onderwijs binnenkomen, worden steeds jonger 'bevestigd' in hun achterstand en hebben steeds minder tijd het been bij te trekken. Dit pakt met name nadelig uit voor leerlingen met een lage SES-positie of leerlingen met een andere thuistaal dan het Nederlands. Niet omdat meertaligheid een probleem is maar omdat de onderwijstijd simpelweg te kort is om zich het Nederlands voldoende eigen te maken zodat een toetsing 'eerlijk' of 'gelijk' voor hen kan uitpakken. Want ook rekentoetsen zijn in hoge mate talig (denk aan de zogenaamde 'verhaaltjessommen').

Tot slot is er een belangrijk derde bezwaar. Namelijk dat leerlingen het basisonderwijs verlaten met een 'fixed mindset' wat betreft hun kennis en kunde. De helft van de leerlingen scoort op de LVS-toetsen 'onder

⁸ <https://www.onderwijsraad.nl/publicaties/brieven/2024/11/19/brief-aan-tweede-kamer>

⁹ Met een leerlingvolgsysteem worden systemen van gestandaardiseerde, door de overheid goedgekeurde toetsen bedoeld.

gemiddeld'. Dit is inherent aan hoe de toetsen zijn genormeerd. Maar deze oordelen zetten zich vast in het hoofd van leerlingen als dat zij niet goed genoeg zouden zijn, of zelfs 'dom'. Dit is zeer nadelig voor de ambities die Nederland heeft als het gaat om bijvoorbeeld de verdere lees- en rekenvaardigheidsontwikkeling. Als de helft van de leerlingen bij het verlaten van de basisschool het gevoel heeft niet goed genoeg te zijn in lezen en rekenen, dan nemen zij dat mee naar de rest van hun onderwijscarrière.

Doorstroomtoets een second opinion?

Omdat scholen hun eigen schooladvies moeten verantwoorden in relatie tot het advies dat uit de doorstroomtoets komt, is er zeker geen sprake van een toets als 'second opinion'. In tegendeel. De toets is de norm. Scoort een leerling hoger op de toets dan het eerder gegeven schooladvies, dan is de school verplicht het schooladvies bij te stellen. Dit maakt dat scholen meer en meer hun advies gaan baseren op de uitkomsten van LVS-toetsen omdat ze er daarmee voor zorgen dat schooladvies en toetsadvies in lijn zijn met elkaar.

Het verplicht bijstellen van het schooladvies als de doorstroomtoets 'hoger' adviseert en de druk die er is door de waarde van LVS-toetsuitkomsten (met name vanaf groep 6) maken dat bijles en toetstraining op grote belangstelling kunnen rekenen, zeker van ouders die dat kunnen betalen.

Van het beeld van de toets als 'gelijkmaker' of 'eerlijke selecteur' blijft al met al nauwelijks iets over.

Ad 2 Kwaliteit van onderwijs

Er zijn 58 kerndoelen die sinds 2006 gelden voor het basisonderwijs en ze bestrijken de leergebieden: Nederlands, Engelse taal, (Friese taal), rekenen/wiskunde, oriëntatie op jezelf en de wereld, kunstzinnige oriëntatie en bewegingsonderwijs. Aanvullend daarop zijn de referentieniveaus voor taal en rekenen verplicht gesteld.

Beperkte afspiegeling leerdoelen

Zowel ten opzichte van de kerndoelen als ten opzichte van het referentiekader geldt dat de inhoud van de doorstroomtoets een beperkte afspiegeling vormt van wat het basisonderwijs beoogt. Zeker zijn taal en rekenen belangrijke vaardigheden maar wat betreft taal is de doorstroomtoets inhoudelijk maar een zeer beperkte afspiegeling van wat taalontwikkeling behelst. Alleen begrijpende lezen en taalverzorging komen in de toetsing aan bod. Spreken, luisteren en schrijven ontbreken.

En door het gebruik van meerkeuzevragen worden de mogelijkheden van het toetsen van tekstbegrip en taalverzorging nog verder beperkt. Die beperkte afspiegeling van taal en van de leerdoelen in de volle breedte maakt een oordeel over de kwaliteit van onderwijs op basis van de doorstroomtoets lastig: de toets is maar een zeer beperkte weergave van het curriculum en dus van wat scholen aan onderwijs bieden als zij volgens de leerdoelen hun aanbod vormgeven. Het is begrijpelijk dat het onderwijs zich tekort gedaan voelt als uitsluitend wordt gekeken naar scores op de doorstroomtoets en dus een zeer beperkt deel van het verplichte curriculum.

Hoe belangrijker de toets, hoe bepalender de inhoud

Scholen en ouders/leerlingen vinden elkaar in het belang om veel tijd te besteden aan dat wat in de toetsing aan bod komt. Voor scholen zijn hoge scores belangrijk om aan te tonen dat zij kwaliteit van onderwijs leveren. Voor leerlingen zijn hoge scores belangrijk voor hun kansen op succes bij plaatsing in het voortgezet onderwijs. Op deze manier is toetsing geen middel meer om de ontwikkeling of prestaties van leerlingen te verbeteren maar is het andersom: leerlingen moeten met hun prestaties laten zien dat de school op orde is. De beperkte inhoudelijke afspiegeling in combinatie met het belang dat de doorstroomtoets heeft voor zowel de leerling als de school, zorgt ervoor dat wat in de doorstroomtoets getoetst wordt onevenredig veel aandacht krijgt in het onderwijs. Ook methodemakers laten zich in hoge mate leiden door de inhoud van de doorstroomtoets. Voor het curriculum als geheel betekent dit een enorme verschraling kijkend naar de 58 leerdoelen in totaal.

Schoolgewicht pakt niet uit zoals bedacht

Voor sommige scholen is het gezien hun doelgroep lastiger om hoog te scoren op de doorstroomtoets dan voor andere. Dat maakt dat de overheid de schoolpopulatie 'meeweegt' in het oordeel over de resultaten: hoe 'moeilijker' de populatie hoe lager de te verwachten gemiddelde score. Scholen moeten dus zorgen dat hun scores (minimaal) in lijn zijn met wat er gezien hun populatie van hen verwacht wordt. Hier kan een verkeerde prikkel vanuit gaan. Maar het kan ook zijn dat een school enorme prestaties verricht met een moeilijke doelgroep van leerlingen maar dat niet zichtbaar weet te maken in de scores op de doorstroomtoets omdat lang niet alles wat een populatie 'lastig' maakt zijn weg vindt naar het schoolgewicht. Ook kan het ertoe leiden dat scholen selectief worden in de aanname van leerlingen, zeker in de bovenbouw: leerlingen die de gemiddelde score van een school mogelijkwijs kunnen 'drukken' bij de doorstroomtoets vinden lastig een plek terwijl potentieel 'hoog' scorende leerlingen bij verhuizing graag gezien worden.

Referentieniveaus: moeilijk vast te stellen, wel bepalend

De referentieniveaus taal en rekenen zijn ingevoerd om scholen te kunnen vergelijken op een 'onafhankelijk criterium'. Deze referentieniveaus zijn ingevoegd in de doorstroomtoets die óók, en primair, de functie heeft het verschil tussen leerlingen zo goed mogelijk in beeld te brengen. Dat vraagt om vergelijking van leerlingen met elkaar (en/of met eerdere cohorten leerlingen). Voor een oordeel over beheersing van de referentieniveaus is het niet nodig leerlingen te vergelijken met elkaar maar juist met een criterium, in dit geval het al dan niet beheersen van een niveau (1F, 2F of 1S). Daarmee heeft de doorstroomtoets een dubbele opdracht: een deel van de items is bedoeld voor de ene opdracht en een deel voor de andere opdracht. Daarmee is de set opgaven voor de bepaling van de referentieniveaus relatief klein gegeven de totale omvang van de toets, maar wel bedoeld om meerdere niveaus mee vast te kunnen stellen én met grote belangen wat betreft de gevolgen van de scores in het kader van toezicht. De beperkte itemset voor het vaststellen van meerdere referentieniveaus beperkt de zekerheid van een oordeel over het al dan niet beheersen van een bepaald referentieniveau. Bovendien blijkt het in de praktijk lastig om op basis van expertoordelen items in te delen in referentieniveaus. De onzekerheden die samenhangen met het vaststellen van een referentieniveau hebben niet geleid tot voorzichtigheid bij het gebruik van die scores. Want de scores op de referentieniveaus zijn ook voor leerlingen niet zonder gevolgen. De scores gaan mee in het onderwijskundig rapport van de leerling naar het voortgezet onderwijs. Zo 'dragen' leerlingen de scores, met alle daarmee samenhangende onzekerheden en 'foute' inschattingen, met zich mee van basis- naar voortgezet onderwijs. Die scores bepalen wel hoe leerlingen naar zichzelf kijken wat betreft lezen en rekenen, hoe leraren naar hen kijken en zelfs het lesaanbod dat leerlingen verplicht zijn te volgen in het voortgezet onderwijs.

Het is goed om ook nog eens te memoreren dat de referentieniveaus nooit bedoeld zijn geweest voor de ontwikkeling van 'harde' normen maar bedoeld waren als een denkraam voor de beschrijving van de gemiddelde ontwikkeling van een gemiddelde leerling wat betreft een beperkte selectie van wat er toe doet in het licht van goede taal- en rekenvaardigheid.

Toezicht en toetsing

Voor toezicht door de Inspectie van Onderwijs zijn de gemiddelde scores (in relatie tot het schoolgewicht) van groot belang. In het toezichtkader dat de Inspectie hanteert bij het beoordelen van de kwaliteit van scholen, is het te zien als het meest zwaarwegende criterium: bij een onvoldoende op

de resultaten kan een school geen voldoende oordeel meer krijgen over de kwaliteit als geheel. Hoe goed de kwaliteit van een school ook is. Daarbij hanteert de Inspectie een drie-jaarlijks gemiddelde en de eis dat dit gemiddelde moet stijgen door de jaren heen. Deze eis/wens tot stijgen van de resultaten klinkt aantrekkelijk in het licht van het willen stimuleren van de taal- en rekenvaardigheidsontwikkeling van onze leerlingen maar de referentieniveaus zijn zo nooit bedoeld en uitgewerkt. Niveau 1F is uitgewerkt als het niveau dat het gros van de basisschoolleerlingen zou moeten kunnen behalen aan het eind van het basisonderwijs. Niveau 2F is uitgewerkt voor de populatie leerlingen aan het eind van het vmbo of pas te voldoen aan het eind van mbo-1, mbo-2 of mbo-3. De eis/wens om aan het eind van het basisonderwijs steeds meer leerlingen te laten eindigen op 2F (taal) of 1S (rekenen) is niet realistisch en niet in lijn met de gedachte achter het Referentiekader.

Op deze manier is van een criteriumgerichte beoordeling voor de kwaliteit van onderwijs alsnog een normgerelateerde beoordeling gemaakt: de prestaties moeten steeds verder omhoog om tot een voldoende te kunnen komen in het toezicht. Dit is voor leerlingen en scholen een niet te vervullen opdracht gegeven het huidige referentiekader.

Zou, met de bedoeling van het Referentiekader, naar de huidige resultaten worden gekeken, dan zou er tevredenheid kunnen zijn over de onderwijsopbrengsten. Het overgrote deel van de leerlingen behaalt niveau 1F voor begrijpend lezen (98,4%), taalverzorging (94,1%) en rekenen (88,7%). Meer dan de gedachte was bij het opstellen van het referentiekader, toen ging men er vanuit dat driekwart van de leerlingen 1F zou kunnen behalen. De slagingspercentages van het vmbo voor het vak Nederlands en de toetsing in het mbo geven eveneens geen reden tot ontevredenheid in relatie tot de referentieniveaus.

Conclusie

De wetsverandering waar de huidige doorstroomtoets uit voortkomt, was in eerste instantie bedoeld om de druk op toetsing te verminderen. Ruim tien jaar na invoering van de wet kunnen we constateren dat daar geen sprake van is. De druk op leerlingen, scholen en de samenleving die van de toetsing uitgaat, is groter dan ooit en de selectie start vroeger dan ooit. De doorstroomtoets, zeker in combinatie met het verplichte gebruik van LVS, leidt daarbij niet tot meer kansengelijkheid. Het vroege selectiemoment in combinatie met de aard en opzet van LVS en doorstroomtoets maken dat meer dan ooit de wieg bepalend is voor de kansen op succes in onderwijs. Het belang van de toetsing zorgt daarbij voor verschraving van het curriculum en leidt niet tot een beter zicht op

kwaliteit van onderwijs. Dat is ernstig in het licht van de zorgen over de taal- en rekenvaardigheid van leerlingen in het basisonderwijs.

Op dit moment is er veel te doen over de vergelijkbaarheid van de verschillende doorstroomtoetsen. Op de website van het onderwijsvakblad Didactief heb ik uitgebreid beschreven waarom dit geen oplossing biedt.¹⁰ Het teruggaan naar één doorstroomtoets gaat aan het zicht onttrekken dat toetsen grove inschattingen zijn van wat ze beogen te meten en op individueel niveau de nodige 'foute' inschattingen maken. Een toets, welke doorstroomtoets dan ook gekozen zou worden, kan nooit meer zijn dan een aanvulling op alle andere informatie die scholen en leerkrachten kunnen gebruiken voor een inschatting van een advies voor voortgezet onderwijs. Daarvoor is het als eerste stap belangrijk dat de uitkomsten van de doorstroomtoets een beperktere rol gaan spelen zowel bij de selectie als bij het beoordelen van onderwijskwaliteit. Voor het realiseren van meer gelijkheid van kansen is het van belang de aandacht te richten op uitstel van selectie door de invoering van brede brugklassen, langere periodes van een gezamenlijke onderbouw en meer ruimte en mogelijkheden om te veranderen van track in het voortgezet onderwijs.

Literatuur

College voor Toetsen en Examens. (2015). Verantwoording Centrale Eindtoets PO. Utrecht: auteur.

Didau, D. (2018). Making Kids Cleverer: A manifesto for closing the advantage gap. Crown House Publishing Ltd.

Gemmink, M.M., Fokkens-Bruinsma, M., Pauw, I. & Van Veen, K. (2020). Under pressure? Primary school teachers' perceptions of their pedagogical practices. *European Journal of Teacher Education*, 43(5), 695-711.

Hebbink, P., Warrens, M. J., Fleur, E., Dijks, M. A., & Korpershoek, H. (2022). De voorspellende waarde van het initiële schooladvies, het toetsadvies en het definitieve schooladvies in het Nederlandse onderwijs. *Pedagogische Studiën*, 99(1).

Hemels, I. (2022). Welke onderzoek ondersteunt jouw school? *Kader AVS*, nr. 2. Heij, K. (2021). *Van de kat en de bel. Tellen en vertellen met de eindtoets basisonderwijs*. Tilburg University.

Heij, K. (2024). *Lezen in beweging. Ontwikkelingen in leesbevordering, lees- en literatuuronderwijs, digitaal lezen en toetsing. Tijd voor nieuwe muziek*. Stichting Lezen Reeks 38.

¹⁰ <https://didactiefonline.nl/blog/blog/doorstroomtoets-terug-naar-een-toets>

- Jolink, A., Tomesen, M., Hilte, M., Weekers, A., & Engelen, R. (2015). Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor groep 4.
- Luyten, H., Oomens, M. & Scholten, F. (2019). Evaluatie wet eindtoetsing PO: Eindrapportage. Universiteit Twente.
- Standaert, R. (2014). De becijferde school: Meetcultus en meetcultuur. Acco.