

# De doorstroomtoets als kompas, maar wie ijk het?

Monika Vaheoja en Martin Ooijevaar, 1 juli 2026



Bron: NOS Jeugdjournaal

Sinds 2015 maakt elke leerling in groep 8 een verplichte toets, eerst de eindtoets, sinds 2024 de doorstroomtoets. De toets heet niet voor niets anders: hij markeert geen eindpunt, maar de overgang naar het voortgezet onderwijs, en hij is uitdrukkelijk bedoeld als second opinion bij het schooladvies. Tegelijkertijd is dit stelsel in tien jaar fors duurder geworden: de subsidie liep op van 1,6 miljoen euro in 2015 naar 8,4 miljoen in 2025 (van den Berg & Belt, 2026). Wie zoveel publiek geld in een toets steekt, mag verwachten dat de kwaliteit van die toets publiek navolgbaar is. Juist dáár wringt het.

In deze bijdrage brengen we drie observaties samen. Eén: het stelsel groeit en kost steeds meer. Twee: de psychometrische onderbouwing die de overheid van toetsaanbieders eist, blijft achter gesloten deuren. Drie: precies rond de cesuren: de grenzen tussen adviezen, valt een groot deel van de leerlingen, en dat is nu juist het punt waar een toets het minst zeker is. We sluiten af met de vraag die boven alles hangt: Als de voorspellende waarde van de gemeten domeinen niet is aangetoond, waarop rust dit stelsel dan eigenlijk?

## Een stelsel dat steeds meer publiek geld kost

Het aantal groep 8 leerlingen ligt al jaren rond de 175.000. De kosten van de toets niet; die zijn in tien jaar ruim vervijfvoudigd (van den Berg & Belt, 2026). Dat komt niet door meer leerlingen, maar door de manier waarop het stelsel is ingericht. De overheid koos bij de invoering bewust voor keuzevrijheid; scholen mogen kiezen tussen aanbieders en de overheid financiert die keuzevrijheid via een subsidieregeling. Inmiddels zijn er zes aanbieders met acht toetsvarianten, terwijl de overheidstoets, ooit goed voor meer dan driekwart van de leerlingen, is teruggevallen tot ongeveer 1 procent.

SEO Economisch Onderzoek liet zien dat het stelsel primair is ontworpen om die keuzevrijheid te waarborgen, en minder op kostenbeheersing (van den Berg & Belt, 2026). De keerzijde is een openeinderegeling: Er is geen plafond voor het aantal toegelaten aanbieders en kleine aanbieders ontvangen dezelfde vaste voet als grote, waardoor hun toetsen relatief duur uitvallen. De vraag die de Onderwijsraad en anderen nu stellen: “Is dit niveau van keuzevrijheid de prijs waard?” is legitiem. Maar het is niet de enige vraag die telt. Want de discussie gaat bijna volledig over de verschillen tussén toetsen, en nauwelijks over de betrouwbaarheid binnen één toets. En juist dat laatste raakt elke individuele leerling.

## Selectie op een smalle basis

De selectie van leerlingen voor het voortgezet onderwijs rust op een smalle basis. Voor veel scholen zijn de resultaten van het leerlingvolgsysteem, een gemankeerde vorm van begrijpend lezen, taalverzorging en rekenen-wiskunde, leidend voor het voorlopige advies. De doorstroomtoets fungeert daarbovenop als second opinion; scoort een leerling hoger dan het voorlopige advies, dan moet de school heroverwegen en in beginsel ophogen; naar beneden bijstellen mag nooit (Rijksoverheid, z.d.).

De Onderwijsraad pleit in haar advies “Anders kijken naar toetsing” voor één doorstroomtoets voor de selectie van leerlingen, en acht het gewicht dat de toets bij het uiteindelijke schooladvies krijgt nodig. De verplichting om het voorlopige advies met de toets te vergelijken en zo nodig te heroverwegen, draagt volgens de raad bij aan een meer gestandaardiseerde onderbouwing van de selectiebeslissing en aan meer kansengelijkheid. Tegelijk vindt de raad dat scholen toetsen uit het leerlingvolgsysteem niet langer voor selectiebeslissingen zouden moeten gebruiken. Omdat er nog geen duidelijke landelijke criteria zijn voor wanneer een leerling succesvol kan doorstromen naar een schoolsoort, is voor scholen onduidelijk op welke gronden zij de selectie precies beargumenteren. De kans bestaat dat de doorstroomtoets als second opinion in de toekomst een nóg grotere rol gaat spelen dan nu al het geval is.

Over de systematische verschillen tussen toetsvarianten is al veel gezegd en geschreven (Po-raad, 2024, 2025). Wat tot nu toe in de discussie onderbelicht blijft, is de betrouwbaarheid van de adviezen binnen één variant. Is het advies dat een leerling krijgt werkelijk het best passende advies, en met welke zekerheid valt dat eigenlijk vast te stellen? Om als toetsaanbieder een goedkeuring te krijgen, moet elke aanbieder voldoen aan onderwijskundige en psychometrische eisen (Staatscourant 2026, nr. 100). Deze uitwerkingen en verantwoordingen worden door het CvTE echter niet openbaar gemaakt. Daardoor is voor de buitenwereld niet na te gaan hoe betrouwbaar de toetsadviezen op individueel niveau zijn.

## Toetsadviezen in het hier en nu

Bij de huidige doorstroomtoetsen kan een leerling zes verschillende adviezen krijgen. Onderstaande tabel toont, voor de twee meest gemaakte toetsen (LIB van Cito en IEP van Bureau ICE), welke score bij welk advies hoort.

| Toetsadvies     | LIB     | IEP     |
|-----------------|---------|---------|
| pro-vmbo-b      | 151–154 | 750–751 |
| vmbo-b / k      | 155–165 | 752–768 |
| vmbo k / gl-t   | 166–172 | 769–776 |
| vmbo g-t / havo | 173–179 | 777–784 |
| havo - vwo      | 180–185 | 785–791 |
| vwo             | 186–200 | 792–800 |

Toetsen zijn nooit volkomen betrouwbaar; een afwijking van enkele punten is gebruikelijk en te verwachten. Voor de doorstroomtoetsen zijn de betrouwbaarheidsintervallen echter niet openbaar. Aannemelijk is dat een score rond een grens in feite op twee niveaus kan wijzen: Het niveau waarbinnen de score valt, én het niveau er net boven of onder. In de praktijk pakt dat zo uit.

**Voorbeeld 1.** Een leerling heeft als voorlopig advies vmbo-b / k en behaalt 165 op de LIB. Het toetsadvies is vmbo-b / k; de school hoeft niet bij te stellen. Maar 165 wijkt één punt af van vmbo-k / gl-t. De score valt binnen de verwachte foutmarge en had net zo goed één niveau hoger kunnen uitvallen.

**Voorbeeld 2.** Een andere leerling met hetzelfde voorlopige advies behaalt 166. Nu luidt het toetsadvies vmbo-k / gl-t en stelt de school bij. Ook hier: één punt verschil, ruim binnen de foutmarge: het advies had net zo goed vmbo-b / k kunnen zijn.

**LIB: 8 scholen, 244 leerlingen**

| LIB                                 | Score   | 1 punt van de grens | 1–2 punten van de grens |
|-------------------------------------|---------|---------------------|-------------------------|
| pro-vmbo-b                          | 151–154 |                     |                         |
| vmbo-b / k                          | 155–165 | 1                   | 4                       |
| vmbo k / gl-t                       | 166–172 | 11                  | 22                      |
| vmbo g-t / havo                     | 173–179 | 25                  | 48                      |
| havo – vwo                          | 180–185 | 26                  | 45                      |
| vwo                                 | 186–200 | 9                   | 19                      |
| <b>Totaal (8 scholen, 244 lln.)</b> |         | <b>72 (30%)</b>     | <b>138 (57%)</b>        |

**IEP: 8 scholen, 184 leerlingen**

| IEP                                 | Score   | 1 punt van de grens | 1–2 punten van de grens |
|-------------------------------------|---------|---------------------|-------------------------|
| pro-vmbo-b                          | 750–751 |                     |                         |
| vmbo-b / k                          | 752–768 | 0                   | 1                       |
| vmbo k / gl-t                       | 769–776 | 4                   | 11                      |
| vmbo g-t / havo                     | 777–784 | 23                  | 36                      |
| havo – vwo                          | 785–791 | 13                  | 25                      |
| vwo                                 | 792–800 | 10                  | 15                      |
| <b>Totaal (8 scholen, 184 lln.)</b> |         | <b>50 (27%)</b>     | <b>88 (48%)</b>         |

**Uitzondering of regel?**

Een eerste gevoel zegt dat scores rond de grenswaarden zeldzaam zijn, hooguit enkele leerlingen per school. Om te toetsen of dat klopt, zocht Martin Ooijevaar het voor zestien scholen uit: acht die de LIB afnamen (244 leerlingen) en acht die de IEP maakten (184 leerlingen). Hoeveel leerlingen scoren één punt onder of boven een grens? En hoeveel binnen één of twee punten?

Het valt op dat bij beide toetsen ongeveer drie op de tien leerlingen op of binnen één punt van een grens scoort. Bij ongeveer de helft (IEP) tot meer dan de helft (LIB) ligt de score binnen één of twee punten van een grens. Voor een substantieel deel van de leerlingen wordt het advies dus mogelijk één niveau te laag of te hoog aangegeven.

Grensgevallen zijn dus geen uitzondering, maar eerder regel. En dat is precies wat de meetfouttheorie voorspelt, zoals we hieronder laten zien.

## De doorstroomtoets onder de psychometrische loep

Dat zoveel leerlingen rond de cesuren landen, is geen toeval. Het is de plek waar elke toets het kwetsbaarst is. Om te begrijpen waarom dat ertoe doet, moeten we kijken naar wat de overheid eigenlijk van de aanbieders eist, en wat daarvan voor de buitenwereld zichtbaar is.

### De overheid vraagt veel, de buitenwereld ziet weinig

Het beoordelingskader voor de doorstroomtoetsen (Staatscourant 2026, nr. 100) is bepaald niet vrijblijvend. Het werkt binnen een complex psychometrische kader en vraagt niet om één klassieke betrouwbaarheidscoëfficiënt voor de hele toets, maar om meetnauwkeurigheid op specifieke punten van de schaal. Concreet eist het kader onder meer:

- **Meetnauwkeurigheid** groter dan 0,70 rond de cesuren van de referentieniveaus en groter dan 0,90 rond de cesuren van de toetsadviezen. Haalt een aanbieder die 0,90 niet, dan moet hij verantwoorden waarom (een ontsappingsclausule die het CvTE in 2028 wil schrappen).
- **Itemkwaliteit** met numerieke drempels ( $p$ -waarde, discriminatie  $r_{it} > 0,15$ , afleiderkwaliteit) en bij open items een interbeoordelaarsovereenstemming  $> 0,80$ .
- **DIF-onderzoek** voor alle items, om vertekening tussen groepen op te sporen, plus pretesten met representatieve steekproeven.
- **Classificatieconsistentie** via een verwarringsmatrix: niet meer dan 5 procent van de observaties mag buiten de diagonaal vallen ten opzichte van de landelijke normering.
- **Gezamenlijke ankeritems** voor equivalering tussen aanbieders, plus rapportage achteraf met herhaald DIF-onderzoek.

Dit is, op papier, een stevig kwaliteitsregime. Maar er zit een paradox in. Diezelfde regeling noemt de “wetenschappelijke en inhoudelijke verantwoording, inclusief de psychometrische aspecten” als document dat de aanbieder bij het CvTE indient. Het CvTE publiceert vervolgens alleen zijn besluit, met verwijzingen naar paragrafen, zonder de coëfficiënten zelf te tonen.

Voor de twee grootste toetsen is het resultaat:

- **LIB (Cito):** de publieke pagina “Betrouwbaar beeld” bevat marketingtaal en een kwalitatieve aansluitingsstudie. Geen betrouwbaarheidscoëfficiënten, geen betrouwbaarheidsintervallen, geen publiek downloadbare wetenschappelijke verantwoording.

- **IEP (Bureau ICE):** marketing en een inhoudelijke toetswijzer. De enige publiek vindbare verantwoording met cijfers betreft de verouderde IEP Eindtoets 2015 op een externe site, niet de huidige doorstroomtoets.

*De cijfers worden dus geproduceerd en bij het CvTE ingediend, maar het CvTE maakt ze niet publiek. Wij vermoeden dat de aanbieders deze gegevens zelf graag zouden willen delen, maar dat de huidige procedure dat niet toelaat. Dat is nadrukkelijk onze inschatting en geen vaststaand feit; de aanbieders hebben zich hiervoor zover ons bekend niet publiek over uitgesproken. Hoe het ook zij, de uitkomst staat op gespannen voet met het transparantiebeginsel dat zowel de internationale Standards (AERA/APA/NCME, 2014) als het COTAN-systeem (Evers e.a., 2010) voorschrijven.*

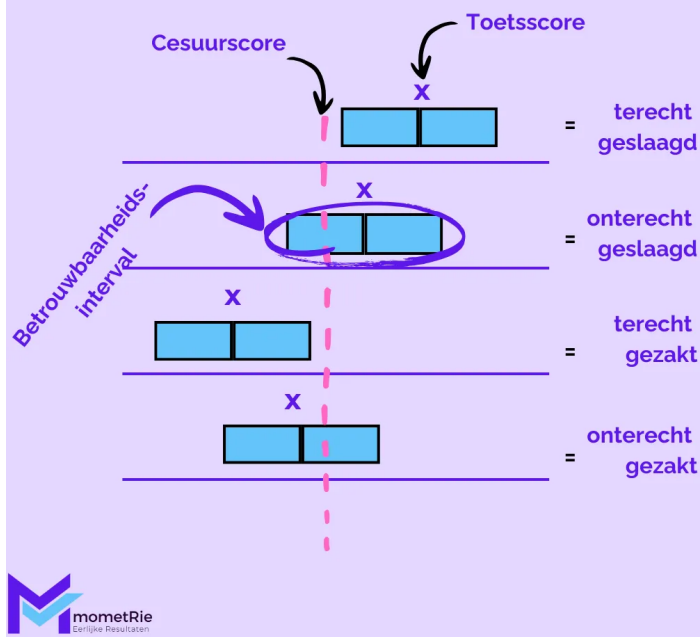
### **Waarom de cesuur het kritieke punt is**

De internationale literatuur wijst hier vrij consistent dezelfde kant op: cesuren midden in de leerlingverdeling kennen doorgaans de laagste 'decision consistency'. Niet de algehele betrouwbaarheid van de toets bepaalt hoe zeker een classificatie is, maar de standaardmeetfout precies op de cesuur. Daarom schrijven de Standards (AERA/APA/NCME, 2014) voor dat juist de conditionele standaardmeetfout (CSEM) bij elke cesuur wordt gerapporteerd, en niet enkel een globaal betrouwbaarheidsgetal.

Hier komen de twee helften van dit verhaal samen. De CvTE eis van meetnauwkeurigheid > 0,90 bij de toetsadvies-cesuren erkent impliciet dat dáár het risico zit. Martins data laten zien dat dáár ook de leerlingen zitten: bij dicht op elkaar liggende cesuren in een leerlingvaardigheidsverdeling valt een groot deel van de leerlingen binnen één standaardmeetfout van een grens. Een verschil van één of twee punten, ruim binnen de meetfout, kan het verschil maken tussen vmbo-t en havo. Voor die grote groep is de classificatie in technische zin onbetrouwbaar: bij een denkbeeldige her-afname zou een substantieel deel een ander advies krijgen.

De eerlijkste maat voor wat er voor grensgevallen op het spel staat, is dus niet "is de toets betrouwbaar?" maar "welk percentage leerlingen zou bij herafname een ander toetsadvies krijgen?" Dat percentage is per toets niet publiek. Martins analyse maakt zichtbaar wat de regelgeving wél eist, maar wat in de openbare verantwoording ontbreekt. Dat roept de vraag op waarom deze informatie niet beschikbaar is.

# Onterecht geslaagd?



Bron: <https://mometrie.nl/post/doorstroomtoets-als-kompas>

En dan de fundamentele vraag: voorspelt de toets eigenlijk wel?

Een van de gestelde doelen van de doorstroomtoets is het voorspellen van het best passende onderwijspad. In psychometrische termen is dat een claim over criteriumvaliditeit, en zo'n claim hoort onderbouwd te worden met bewijs. Wie zoekt naar Nederlands, peer-reviewed bewijs dat juist de domeinen rekenen, begrijpend lezen en taalverzorging het vo-succes voorspellen, vindt dat bewijs vooralsnog niet. Sterker nog: recente onderzoek laat zien dat zelfs het referentieniveau zelf wankel is. Bakker e.a. (2025) tonen aan dat het streefniveau 1S geen stevig of eenduidig meetpunt is, dat verschillende doorstroomtoetsen sterk uiteenlopende percentages 1S opleveren bij vergelijkbare leerlingen, en dat 1S nooit als harde norm voor losse scholen bedoeld is. Als de meetlat zelf al niet eenduidig is, is een voorspellende claim die daarop rust des te lastiger hard te maken (luister ook de podcast met Arthur Bakker en Marian Hickendorff, [tijpcast.nl](https://www.tijpcast.nl)).

De huidige wetenschappelijk onderbouwing is met name bekend bij de eindtoetsen. Hieronder in hoofdlijnen de uitkomsten:

**1.Hebbink e.a. (2022), Pedagogische Studiën:** voor alle niveaus en cohorten voorspellen het initiële en definitieve schóóladvies de onderwijspositie in leerjaar 3 beter dan het toetsadvies. De auteurs bevelen expliciet méér onderzoek naar de voorspellende waarde van eindtoetsen aan, een teken dat dit bewijs ontbreekt.

**2.CPB (2019):** de eindtoets draagt bij aan een beter passend advies wanneer hij wordt benut, maar dit is correlatieel en betreft de oude eindtoets.

Een belangrijke nuance: ál dit onderzoek gaat over de oude centrale eindtoets, niet over de huidige zes doorstroomtoetsen. Voor de huidige doorstroomtoetsen is er geen gepubliceerd onderzoek dat de specifiek gemeten domeinen aan vo-succes koppelt.

Laten we hier precies zijn, want het is een subtiel maar wezenlijk onderscheid. We zeggen niet dat lezen, rekenen en taal er niet toe doen, dat zou een te grote claim zijn. We zeggen: voor de huidige doorstroomtoetsen is de voorspellende waarde van de gemeten domeinen niet aangetoond en niet publiek toetsbaar, en voor de oude eindtoets bleek het schooladvies de betere voorspeller. Dat verschuift de bewijslast. Bij een toets met wettelijke gevolgen voor de schoolloopbaan van een kind hoort die onderbouwing geleverd én openbaar te zijn, niet aangenomen.

Er zit bovendien een dieper methodologisch probleem onder. Echt predictief-validiteitsonderzoek zou vereisen dat je de toets afneemt, de uitslag geheimhoudt, leerlingen onafhankelijk van die uitslag plaatst en jaren later de uitkomst meet. Zo'n experiment is ethisch onmogelijk. De beschikbare “validatie” is dus correlatieel binnen een systeem waarin het toetsadvies de plaatsing zélf beïnvloedt: padafhankelijkheid. En er bestaat geen inhoudelijke standaard voor wat een leerling moet kennen om “klaar voor theoretische leerweg/havo/vwo” te zijn. De toets rangschikt leerlingen vooral ten opzichte van elkaar, zonder inhoudelijk referentiepunt voor het niveau waarin ze terechtkomen. Daarmee functioneert hij eerder als een sorteerinstrument dan als een criteriumgerichte meting.

## Wat dit betekent voor de discussie over één toets?

De Onderwijsraad pleit voor één doorstroomtoets van één aanbieder. Dat lijkt een reëel probleem op te lossen: gelijk presterende leerlingen krijgen nu bij verschillende toetsen verschillende adviezen, wat de vergelijkbaarheid en uitlegbaarheid ondermijnt. Eén toets en één norm nemen dat verschil weg, en een publieke aanbieder zou commerciële belangen uit een instrument met een kansengelijkheidsfunctie halen.

Maar één toets lost het meetfout- en validiteitsprobleem niet op. Ook bij één landelijke toets blijven leerlingen rond de cesuren liggen, blijft de classificatie daar het minst zeker, en blijft de voorspellende claim onbewezen zolang die niet inhoudelijk wordt geijkt. Sterker nog: bij één toets met één norm worden de gevolgen van dezelfde meetfout uniform over alle leerlingen verdeeld. Eén toets is daarom een verbetering op het punt van vergelijkbaarheid en kosten, maar geen antwoord op de vraag of we aan het meten zijn waarover we een advies willen geven, en of we eerlijk genoeg zijn over de onzekerheid van

de resultaat. En mogen we wel conclusies trekken op basis van de toetsscore die verder reiken dan waarvoor die score bedoeld is?

Er is ook een argument dat zelden valt: een papieren, vaste toets is steeds minder van deze tijd. Een vaste set items kan een breed vaardigheidsspectrum van pro tot vwo nooit overal even nauwkeurig meten; de meeste informatie zit rond het midden van de toets, terwijl leerlingen over de hele schaal verspreid zitten. Adaptieve afname, waarbij de moeilijkheid zich aanpast aan de leerling, meet juist aan de uiteinden en rond de individuele cesuur preciezer. Wie de classificatie rond de grenzen wil verbeteren, exact waar Martins data het probleem laten zien, ontkomt op termijn niet aan die vraag.

### Een lange verhaal kort: wat wij voorstellen:

Onze inzet is niet “afschaffen” en niet “alles bij het oude laten”. Het is: Maak het stelsel navolgbaar, en wees eerlijk over wat een toetsadvies wel en niet kan.

1. **Publiceer de psychometrische verantwoording.** Voor een high-stakes toets met wettelijke gevolgen hoort de betrouwbaarheid, de conditionele standaardmeetfout bij elke cesuur, de classificatieconsistentie en de DIF-uitslag openbaar te zijn, en dat per toets. Zijn die data niet binnen afzienbare tijd publiek beschikbaar, dan staat de transparantie van het stelsel naar onze mening onder druk.
2. **Vraag naar de meetfout bij de cesuur, niet naar één globaal getal.** De relevante vraag is hoe groot de standaardfout precies is op de grenzen waar de meeste leerlingen zitten, en welk percentage leerlingen bij herafname een ander advies zou krijgen.
3. **Behandel het toetsadvies als wat het is: een indicatie met onzekerheidsmarge, een momentopname.** Rapporteer per leerling een betrouwbaarheidsinterval en markeer een leerling expliciet als grensgeval zodra die binnen één standaardmeetfout van een cesuur valt. Scholen, ouders en leerlingen verdienen dat inzicht.
4. **Onderschrijf het Onderwijsraad-advies kritisch.** Eén toets lost de vergelijkbaarheid op, maar verschuift tegelijk van rangschikken naar inhoudelijke standaarden voor wat “klaar voor niveau X” betekent. Anders blijft ook één toets een sorteermachine zonder referentiepunt.
5. **Bepleit écht predictief-validiteitsonderzoek binnen ethische grenzen.** Bijvoorbeeld via brede, meerjarige brugklassen waarin plaatsing minder padafhankelijk is, zodat de voorspellende claim eindelijk getoetst wordt in plaats van aangenomen.

Tot slot

De doorstroomtoets wordt graag een “betrouwbaar kompas voor de toekomst van elk kind” genoemd. Een kompas is alleen betrouwbaar als het geijkt is, en als die ijking navolgbaar is voor wie erop vaart. Op dit moment vraagt de overheid een stevige ijking, maar houdt ze de uitkomst daarvan buiten beeld, terwijl de kosten oplopen en de voorspellende waarde van wat gemeten wordt nooit publiek is aangetoond.

Onze oproep is bescheiden en concreet tegelijk. Maak de kwaliteit zichtbaar. Wees eerlijk over de marge rond elke grens. En durf de vraag te stellen of we, met een vaste papieren toets en zonder bewezen voorspellende waarde, wel het juiste meten. Pas dan kan een gesprek over één toets, of over de rol van toetsing als geheel, op een eerlijke basis worden gevoerd. In het belang van elke leerling die rond zo'n grens scoort, en dat zijn er meer dan we dachten.

## Literatuurlijst

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bakker, A., Boels, L., Hickendorff, M., Jonker, V., Van Luit, H., Veldhuis, M., & Willems, W. (2025). Reken het rekenonderwijs niet af op percentages leerlingen die 1S behalen. *Volgens Bartjens – Ontwikkeling en Onderzoek*, 45(2), 41–51.

Beoordelingskader doorstroomtoetsen, *Staatscourant* 2026, nr. 100. (2026). <https://zoek.officielebekendmakingen.nl/stcrt-2026-100.html>

College voor Toetsen en Examens. (2025a). *Titel van de eerste CvTE rapportage/document*. [Vul hier eventueel de specifieke titel in].

College voor Toetsen en Examens. (2025b). *Titel van de tweede CvTE rapportage/document*. [Vul hier eventueel de specifieke titel in].

Centraal Planbureau. (2019). *De waarde van eindtoetsen in het primair onderwijs* (CPB Policy Brief).

Dienst Uitvoering Onderwijs & Ministerie van Onderwijs, Cultuur en Wetenschap. (z.d.). *Monitor schooladvies en doorstroomtoets 2023–2024*. [Vul eventueel het jaar van publicatie in als dit bekend is; 'z.d.' betekent 'zonder datum'].

Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2010). *COTAN-beoordelingssysteem voor de kwaliteit van tests*. Nederlands Instituut van Psychologen (NIP).

Hebbink, A., & anderen. (2022). [Titel van het artikel invullen]. *Pedagogische Studiën*, 99(1), 20–35.

Onderwijsraad. (2026, 28 mei). *Kijk anders naar toetsing*. <https://www.onderwijsraad.nl/documenten/2026/05/28/kijk-anders-naar-toetsing>

PO-Raad. (2024). *Themarapportage Doorstroomtoets 2024*.

PO-Raad. (2025). *Themarapportage Doorstroomtoets 2025*.

Rijksoverheid. (z.d.). *Verplichte doorstroomtoets basisonderwijs*. Geraadpleegd op 19 juni 2026, van <https://www.rijksoverheid.nl/themas/onderwijs/schooladvies-en-doorstroomtoets-basis-school/verplichte-doorstroomtoets-basonderwijs>

Van den Berg, E., & Belt, T. (2026). *Keuzevrijheid met een prijskaartje. Evaluatie van de subsidieregeling andere eind-/doorstroomtoetsen po* (SEO-rapport 2026-01). SEO Economisch Onderzoek.