

Hoe werkt een adaptieve toets?

Over de AMN Doorstroomtoets

Inhoudsopgave

Een belangrijk toetsmoment vraagt om een doordachte keuze	2
Wat is een adaptieve toets?	3
De voordelen van adaptieve toetsing in het onderwijs	3
Zo werkt adaptiviteit in de AMN Doorstroomtoets	4
<i>Verdieping 1: hoe de toets het vaardigheidsniveau inschat</i>	5
<i>Verdieping 2: het statistisch model achter de toets</i>	5
<i>Voorbeeld: Hoe werken de parameters in de praktijk?</i>	6
Hoe valide en betrouwbaar is de AMN Doorstroomtoets?	7
Conclusie: De kracht van een adaptieve toets	8
Literatuur	9

Een belangrijk toetsmoment vraagt om een doordachte keuze

De doorstroomtoets is een belangrijk moment. De uitslag kan het definitieve schooladvies beïnvloeden en kan mede bepalen welke stap een leerling zet richting het voortgezet onderwijs. Dat maakt dit toetsmoment soms spannend – voor de leerling, de leerkracht én de ouders/verzorgers.

Hoe zorg je ervoor dat de toets recht doet aan de leerling? Dat hij niet te makkelijk is, of juist te moeilijk? En hoe weet je of de toets eerlijk meet, ook als leerlingen verschillende vragen krijgen?

De AMN Doorstroomtoets is een adaptieve toets op vraagniveau en biedt extra mogelijkheden, zoals de optie om opgaven over te slaan of later naar terug te keren. Scholen kiezen voor deze manier van toetsen – niet alleen vanwege de technologie, maar juist vanwege de onderwijskundige voordelen.

In dit whitepaper beantwoorden we de meest gestelde vragen over adaptieve toetsing. We leggen uit hoe het werkt, hoe het binnen de AMN Doorstroomtoets is toegepast, en hoe we zorgen voor betrouwbaarheid, validiteit en objectiviteit. Zo bieden wij je inzicht in de werking en techniek, en kun jij – als school – een doordachte keuze maken voor een doorstroomtoets die past bij jullie leerlingen.

Wat is een adaptieve toets?

Een adaptieve toets is een toets die zich tijdens de afname aanpast aan het niveau van de leerling. Dit betekent dat de toets geen vaste route volgt. In tegenstelling tot een lineaire toets, waarbij er één vaste versie wordt aangeboden aan alle leerlingen. Bij dit "one-size-fits-all"-model is de toets voor sommige leerlingen te makkelijk en voor anderen juist te moeilijk.

Bij adaptieve toetsing zijn er grofweg twee varianten:

- **Adaptief op groepsniveau**

Dit wordt ook wel een 'multistage test' (MST) genoemd (*Mead, 2006*). In deze variant maakt de leerling eerst een vaste set vragen, waarna zijn of haar vaardigheid wordt ingeschat. Op basis daarvan wordt een nieuw blok met vragen geselecteerd. Dit proces herhaalt zich een aantal keer tot het onderdeel stopt. Uit hoeveel vragen de blokken bestaan en hoeveel van deze blokken er zijn, verschilt per toets.

- **Adaptief op vraagniveau**

De leerlingen beginnen met een aantal startvragen waarna de vragen op een volledig adaptieve manier worden aangeboden. Dit houdt in dat na ieder antwoord dat de leerling geeft, het vaardigheidsniveau opnieuw wordt berekend en op basis daarvan een nieuwe vraag wordt geselecteerd.

Dus beide toetsvormen zijn adaptief, maar de mate waarin en de snelheid waarmee de toets zich aanpast verschilt.

De voordelen van adaptieve toetsing voor het onderwijs

Adaptieve toetsing is meer dan een technische innovatie; het zorgt ook voor een betere toetservaring van de leerling. Doordat de toets zich aanpast aan het vaardigheidsniveau van de leerling, worden leerlingen uitgedaagd op het juiste niveau – en niet ontmoedigd door te makkelijke of te moeilijke opgaven.

Onderzoek laat zien dat adaptieve toetsing verschillende onderwijskundige voordelen heeft:

- ✓ **Een toets op het juiste niveau**

Een te moeilijke of te makkelijke toets kan leiden tot frustratie of verveling. Een adaptieve toets beweegt mee met de leerling, waardoor de afname als eerlijker en prettiger wordt ervaren. Leerlingen blijven gemotiveerd om door te werken (*Martin & Lazendic, 2018*).

- ✓ **Kortere en efficiëntere toetsafnames**

Omdat het systeem voortdurend de vaardigheid van de leerling inschat en bijstuurt, zijn er minder opgaven nodig om tot een betrouwbare score te komen. Dat maakt de toets korter en efficiënter. Leerlingen blijven daardoor beter betrokken en ervaren minder toetsangst (*Ling et al., 2017*).

- ✓ **Precieze schatting van de vaardigheid**

Adaptieve toetsen kunnen het vaardigheidsniveau preciezer inschatten voor leerlingen met een (zeer) lage of hoge vaardigheid (*Weiss, 2011*).

- ✓ **Minder kans op afkijken**

Elke leerling krijgt een persoonlijke route door de toets en zij maken dus niet dezelfde opgaven. Hierdoor is het risico op afkijken kleiner (*Kainz et al., 2015*).

✓ Bescherming van de opgaven

Elke leerling krijgt slechts een beperkt aantal opgaven uit de totale vragenbank. Daardoor is de geheimhouding van de vragen beter gewaarborgd (*He & Reckase, 2014*).

Er kunnen ook enkele nadelen aan adaptieve toetsen zitten. Zoals:

- Een adaptieve toets is digitaal. Dit houdt in dat scholen voorzieningen moeten treffen, denk aan laptops of tablets, om te kunnen toetsen.
- Bij een papieren toets kun je een vraag overslaan om er later op terug te komen. Bij veel adaptieve toetsen is dit niet mogelijk. In de AMN Doorstroomtoets is het wél mogelijk om opgaven over te slaan of later aan te passen. Dit biedt leerlingen extra rust en ruimte om zelf hun tempo te bepalen, zonder dat dit ten koste gaat van de betrouwbaarheid van de uitslag.



Benieuwd of deze manier van toetsen past bij jullie school?

Bekijk voorbeeldopgaven en -rapportages of plan een vrijblijvend gesprek met één van onze adviseurs op www.amn.nl/doorstroomtoets.

Zo werkt adaptiviteit in de AMN Doorstroomtoets

De AMN Doorstroomtoets past zich tijdens het maken van de toets aan aan het niveau van de leerling. Dat betekent dat niet alle leerlingen dezelfde vragen krijgen.

De toets gebruikt hiervoor een slim rekenmodel (een algoritme). Een algoritme beschrijft hoe de toets moet verlopen: hoe de vaardigheid van de leerling moet worden geschat, op welk punt de toets moet stoppen en de verhouding van toetsvragen uit de verschillende domeinen.

Simpel gezegd werkt dat zo: de leerling begint met startvragen en daarna wordt een eerste grove inschatting gemaakt van het niveau van de leerling. Op basis daarvan wordt de volgende vraag gekozen. Vervolgens wordt er na iedere vraag een nieuwe inschatting van het niveau gemaakt en wordt op basis daarvan de volgende vraag geselecteerd. Niet te makkelijk, niet te moeilijk – maar precies passend. Daardoor krijgt iedere leerling een persoonlijke route door de toets. Wanneer de schatting van de vaardigheid van de leerling zeer betrouwbaar is, krijgt de leerling geen vragen meer en stopt het desbetreffende onderdeel automatisch. Om te voorkomen dat de toets vragen blijft aanbieden, is er ook altijd een maximaal aantal vragen ingesteld.

Leerlingen maken dus niet meer vragen dan nodig, en de vragen die ze maken zijn op hun eigen niveau.

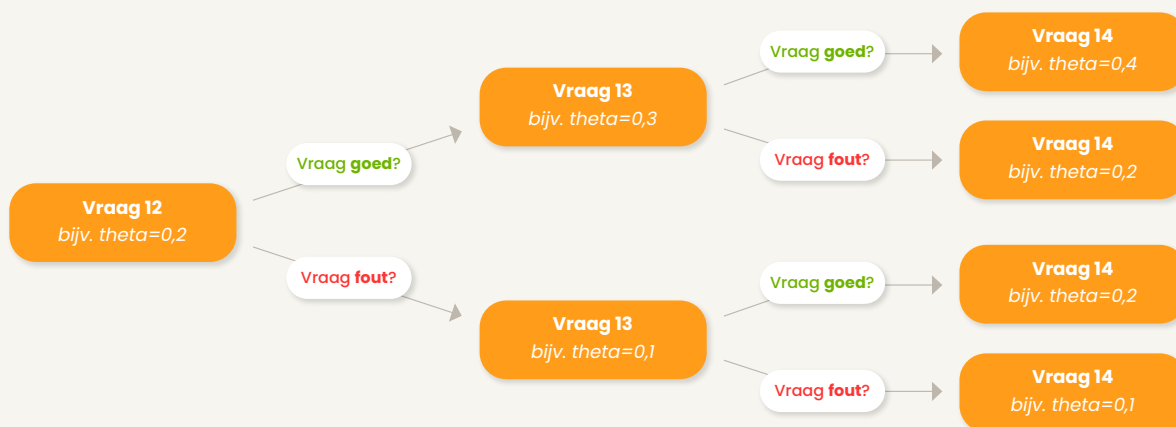
Wil je weten hoe het schatten van de vaardigheid van de leerling technisch in elkaar steekt? Lees dan Verdieping 1.

Wil je weten hoe adaptiviteit technisch in elkaar steekt? Lees dan Verdieping 2.

Verdieping 1: hoe de toets het vaardigheidsniveau inschat

De vaardigheid van de leerling wordt uitgedrukt in 'theta' (θ). Theta heeft meestal een waarde tussen -3 (lage vaardigheid) en $+3$ (hoge vaardigheid). Leerlingen beginnen met een aantal startvragen van gemiddeld niveau. Op basis van deze antwoorden maakt het algoritme een eerste inschatting: de initiële theta. Op basis daarvan biedt de toets een vraag aan die past bij deze theta.

Na elk nieuw antwoord berekent de toets een aangepaste theta. Is het antwoord goed? Dan gaat de theta omhoog en krijgt de leerling een iets moeilijkere vraag. Is het antwoord fout? Dan gaat de theta omlaag en volgt een makkelijkere vraag. Zie afbeelding 1 voor een schematische weergave van dit proces.



Afbeelding 1: schematische weergave van de steeds betere inschatting van de vaardigheid van de leerling.
Dit is een illustratie; de stappen die gemaakt worden zijn in werkelijkheid subtieler.

Verdieping 2: het statistisch model achter de toets

De AMN Doorstroomtoets maakt gebruik van Item Response Theory (IRT). Met opgaven meet je een bepaalde vaardigheid van een leerling, en met IRT krijg je inzicht in hoe goed elke vraag die vaardigheid meet.

AMN maakt gebruik van het 2-parameter logistisch model van Birnbaum (1968). In dit model zijn er twee parameters die van belang zijn: de discriminatieparameter en de moeilijkheidsparameter. Iedere toetsvraag heeft een waarde voor beide parameters.

- **De moeilijkheidsparameter (β -parameter of bèta-parameter)**

Deze parameter geeft aan hoe moeilijk een vraag is. De vragenbank bevat daarom vragen met lage, gemiddelde en hoge moeilijkheidswaarden, zodat elk leerlingniveau goed bediend kan worden.

Bij het kiezen van de volgende vraag kijkt de toets of de moeilijkheidsgraad (β) past bij de huidige theta-waarde van de leerling. Een moeilijke vraag voor een leerling met een lage theta levert weinig op, net als een te makkelijke vraag voor een vaardige leerling. De meeste informatie wordt namelijk verkregen als de moeilijkheid van de vraag, past bij de geschatte vaardigheid van de leerling.

- **De discriminatieparameter (α -parameter of alpha-parameter)**

Deze parameter geeft aan hoe goed een vraag onderscheid kan maken tussen het vaardigheidsniveau van leerlingen. Als deze parameter laag is, dan geeft de vraag relatief weinig informatie. Dit betekent dat als een leerling de vraag goed heeft, dat niet direct hoeft te betekenen dat die leerling ook beter is in die specifieke vaardigheid dan een leerling die de vraag fout had.

Ter illustratie: als een vraag een hoge discriminatieparameter heeft, dan wordt die vaak juist beantwoord door leerlingen met een hogere vaardigheid (bijv. 90%) en minder vaak door leerlingen met een lagere vaardigheid (bijv. 10%). De vraag kan dan dus goed onderscheid maken tussen leerlingen op basis van hun vaardigheid.

Als een vraag een lage discriminatieparameter heeft, dan beantwoorden vaardige leerlingen de vraag ongeveer even vaak goed (bijv. 60%) als minder vaardige leerlingen (bijv. 50%). De vraag discrimineert dan nauwelijks. De oorzaak hiervan kan een verwarrende vraagstelling zijn of er wordt eigenlijk een andere vaardigheid gemeten dan de beoogde vaardigheid.

Hoe hoger de discriminatieparameter van een vraag, hoe beter, maar de vaardigheid van de leerling moet wel ongeveer gelijk zijn aan de moeilijkheid van de vraag. Of te wel: als de vraag niet past bij het niveau van de leerling, heb je nog niets aan een hoge discriminatieparameter.

Voorbeeld: Hoe werken de parameters in de praktijk?

Sam, Lisa, Luca en Ayla maken de AMN Doorstroomtoets en zijn bij het onderdeel rekenen. De eerste inschatting is dat Sam een lage rekenvaardigheid heeft. Als de toets een vraag zou selecteren met een gemiddelde moeilijkheid, dan zal Sam de vraag hoogstwaarschijnlijk fout maken. Het stellen van deze vraag levert dan weinig extra informatie op over hoe vaardig Sam nu precies is. Je krijgt alleen de bevestiging dat hij geen gemiddelde rekenvaardigheid heeft, maar het geeft geen informatie of zijn vaardigheid meer past bij vmbo-bb of vmbo-kb. Deze vraag is dus minder geschikt voor Sam. Lisa heeft een gemiddelde rekenvaardigheid. Als dezelfde vraag aan haar gesteld wordt, levert de vraag wél veel informatie op.

Luca en Ayla hebben ongeveer dezelfde rekenvaardigheid, maar Ayla is net iets vaardiger. Zij krijgen allebei dezelfde vraag aangeboden, die redelijk aansluit bij hun vaardigheid. Ayla maakt hem goed, Luca niet. Als de vraag een hoge discriminatieparameter heeft, dan zou je kunnen zeggen: Ayla is wat vaardiger dan Luca. Maar in dit geval heeft de vraag een lage discriminatieparameter. De rekenvraag bevat veel moeilijke woorden. Ayla is goed in lezen en snapt de vraag direct. Luca heeft moeite met lezen en ondanks dat hij prima kan rekenen, beantwoordt hij de vraag fout. In dit geval wordt er dus niet alleen de rekenvaardigheid gemeten, maar speelt leesvaardigheid ook een grote rol. De discriminatiewaarde van deze rekenvraag is daardoor laag; het kan de sterkere rekenaars niet heel goed onderscheiden van de zwakkere rekenaars. Ondanks dat de moeilijkheid van de vraag dus goed aansluit bij Luca en Ayla, levert het weinig informatie op over hun rekenvaardigheden.

We doen zeer uitgebreide analyses van onze toetsvragen. Een taalkundig complexe rekenvraag, zoals beschreven in het voorbeeld, zal nooit worden opgenomen in de AMN Doorstroomtoets.

Hoe valide en betrouwbaar is de AMN Doorstroomtoets?

Hoe weet je zeker dat een toets echt meet wat hij moet meten? En dat de uitslag betrouwbaar is om een leerling een vervolgadvis te geven? Hieronder leggen we uit hoe de AMN Doorstroomtoets dit garandeert – en waarom je er als school op kunt vertrouwen.

- **Een valide toets: meten wat je moet meten**

Een toets is valide als hij daadwerkelijk meet wat hij hoort te meten. In de Toetswijzer doorstroomtoetsen PO (*College van Toetsen en Examens, 2023*) worden eisen gesteld aan de inhoud van de doorstroomtoets. Jaarlijks wordt onze doorstroomtoets zeer streng gecontroleerd op deze eisen door het College voor Toetsen en Examens.

In de toetswijzer wordt onder meer beschreven dat de toets de onderdelen lezen, taalverzorging en rekenen moet meten uit het Referentiekader Taal en Rekenen (*Meijerink et al., 2009*). Elke toetsvraag in de AMN Doorstroomtoets is gekoppeld aan dit referentiekader. Dat betekent: de inhoud van de toets sluit aan bij wat leerlingen aan het eind van het primair onderwijs zouden moeten beheersen, volgens het Referentiekader.

In het algoritme van de AMN Doorstroomtoets zijn ‘regels’ opgenomen die beschrijven welke domeinen bevraagd moeten worden en in welke verhoudingen. Op die manier ontstaat er inhoudelijk een representatieve afspiegeling van de te toetsen lesstof. Je wilt bijvoorbeeld niet dat een leerling 21 vragen uit het domein ‘meten en meetkunde’ krijgt en maar 1 vraag uit het domein ‘getallen’. Daarom is in het algoritme voor ieder subdomein vastgelegd hoe groot het aandeel moet zijn, bijvoorbeeld “het percentage toetsvragen uit het domein ‘getallen’ ligt tussen de 30% en 40% van het totaal aantal rekenvragen”. Zo zijn er ook regels voor de onderliggende subonderdelen. Op basis van al deze regels, selecteert het algoritme de juiste vragen uit de grote vragenbank. Op die manier krijgt iedere leerling een inhoudelijk gebalanceerde toets.

- **Een betrouwbare uitslag: elk jaar opnieuw getest**

De mate van betrouwbaarheid van een toets geeft weer in hoeverre je dezelfde uitkomst zou krijgen als een leerling de toets opnieuw zou maken. De betrouwbaarheid wordt uitgedrukt met een waarde tussen 0 en 1 (of 0 en 100%). Hoe dichterbij 1, hoe betrouwbaarder. Vanaf 0,7 (70%) spreek je van een hoge mate van betrouwbaarheid (*Taber, 2018*).

Elk jaar onderzoeken we de betrouwbaarheid van de AMN Doorstroomtoets gebaseerd op de methode van *Glas en Emons (2017)* en *Bechger et al. (2003)*. De AMN Doorstroomtoets scoort hier zeer goed op. In 2025 was de betrouwbaarheid van de volledige toets maar liefst **97,1%**. Ook de subdomeinen van de toets scoren hoog.



Taalverzorging
91,2%



Lezen
88,5%



Rekenen
95,6%

- **Objectieve scoring: altijd dezelfde standaard**

De opgaven in de AMN Doorstroomtoets worden automatisch gescoord. De antwoorden van de leerlingen worden allemaal gescoord op basis van dezelfde standaarden. Het is onmogelijk dat eenzelfde antwoord de ene keer goed wordt gerekend en de andere keer niet. De scoring is daarmee objectief.

- **Efficiënt: toetsen zonder onnodige belasting**

De AMN Doorstroomtoets meet alleen de voorgeschreven domeinen: lezen, taalverzorging en rekenen. Er zijn geen extra onderdelen of opgaven die de leerling onnodig belasten. En dankzij het adaptieve karakter van de toets worden er niet meer vragen gesteld dan nodig is voor een betrouwbare inschatting. Dat maakt de toets efficiënt én aangenaam voor de leerling.

Conclusie: De kracht van een adaptieve toets

In dit whitepaper hebben we laten zien hoe adaptieve toetsing werkt binnen de AMN Doorstroomtoets en waarom deze aanpak onderwijskundig én psychometrisch waardevol is. De AMN Doorstroomtoets past zich namelijk aan aan het niveau van de leerling – op een manier die betrouwbaar, valide en inhoudelijk doordacht is. De kracht van adaptieve toetsing is dat de toets zich richt op het individu. De toets maakt gebruik van een slim algoritme op basis van Item Response Theory. Zo bepaalt het systeem na elk antwoord welke volgende vraag de meeste informatie geeft over de vaardigheid van de leerling. Dit zorgt voor een kortere afname, een prettigere toetsbeleving én een uitslag die recht doet aan het werkelijke niveau.

Wat maakt de AMN Doorstroomtoets de beste keuze voor iedere leerling?

- ✓ **Altijd op het juiste niveau**

Elke leerling krijgt een persoonlijke route door de toets, gebaseerd op zijn of haar eigen antwoorden. Niet te makkelijk, niet te moeilijk – maar precies afgestemd op wat de leerling op dat moment laat zien. Dat maakt de toets motiverender en effectiever.

- ✓ **Betrouwbare uitslagen waarop je kunt vertrouwen**

Onderbouwd door een bewezen statistisch model (Item Response Theory) en jaarlijkse betrouwbaarheids- onderzoeken weten we zeker: de AMN Doorstroomtoets geeft een valide en betrouwbare inschatting van de vaardigheid van de leerling.

- ✓ **Slim opgebouwd, zonder onnodige belasting**

De toets is efficiënt en doordacht: hij stopt zodra er voldoende zekerheid is over het niveau, vraagt niet méér dan nodig, en zorgt toch voor een inhoudelijk evenwichtige verdeling van alle domeinen. Zo krijg je een compleet en representatief beeld, zonder leerlingen onnodig te belasten.

Een adaptieve doorstroomtoets op vraagniveau past bij iedere leerling – zowel in het regulier als het speciaal basis- onderwijs. Want alle leerlingen verdienen een eerlijke kans om te laten zien wat ze in huis hebben, en dat bereik je met een toets op eigen niveau. Dat is wat de AMN Doorstroomtoets biedt.

Literatuur

- Birnbaum, A. (1968). Some latent trait models. In: F.M. Lord, & M.R. Novick. Statistical theories of mental test scores (pp. 397–424). Reading: Addison-Wesley
- College van Toetsen en Examens (2023). Toetswijzer doorstroomtoetsen PO. <https://wetten.overheid.nl/BWBR0047245/2023-12-07>
- Glas, C.A.W., & Emons, W.H.M. (2017). Blueprint voor psychometrische verantwoording normering toetsadviezen en ijking op de referentieniveaus. EPO: Utrecht
- He, W., & Reckase, M. D. (2014). Item Pool Design for an Operational Variable-Length Computerized Adaptive Test. Educational and Psychological Measurement, 74(3), 473–494. <https://doi.org/10.1177/0013164413509629>
- Kainz, O., Cymbalák, D. and Jakab, F. (2015) 'Adaptive web-based system for examination with cheating prevention mechanism', Lecture Notes on Software Engineering, 3(2), pp. 90–94. <https://doi.org/10.7763/LNSE.2015.V3.172>
- Ling, G., Attali, Y., Finn, B., & Stone, E. A. (2017). Is a computerized adaptive test more motivating than a fixed-item test? Applied Psychological Measurement, 41(7), 495–511. <https://doi.org/10.1177/0146621617707556>
- Martin, A. J., & Lazendic, G. (2018). Computer-Adaptive Testing: Implications for students' achievement, motivation, engagement, and subjective test experience. Journal of Educational Psychology, 110(1), 27–45. <https://doi.org/10.1037/edu0000205>
- Mead, A. D. (2006). An Introduction to Multistage Testing. Applied Measurement in Education, 19(3), 185–187. https://doi.org/10.1207/s15324818ame1903_1
- Meijerink, H. P., Letschert, J. F., Streun, A. van, Bergh, H. H. van den, & Rijlaarsdam, G. C. W. (2009). Referentiekader taal en rekenen. www.slo.nl/publish/pages/5901/referentiekader_taal_en_rekenen_referentieniveaus.pdf
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Weiss, D. J. (2011). Better Data From Better Measurements Using Computerized Adaptive Testing. Journal of Methods and Measurement in the Social Sciences, 6(1), 1–27. https://doi.org/10.2458/azu_jmmss.v2i1.12351

Scan voor
meer informatie

Meer weten of meteen aan de slag?

Op onze website lees je meer over de mogelijkheden voor extra ondersteuningsbehoeften, voorbeeldopgaven en voorbeeldrapportages. Ook kun je vrijblijvend contact opnemen met één van onze adviseurs. Ga naar: amn.nl/doorstroomtoets.

