

Bijlage

Toetssystemen in Nederland: weinig samenhang

Er zijn in het Nederlandse basisonderwijs verschillende systemen in omloop om de rekenprestaties van leerlingen te meten (cq. toetssystemen). Zo zijn er de leerlingvolgsystemen (zoals dat van Cito bv), de systemen om de prestaties bij eindtoetsen te meten, en het systeem om het prestatieniveau in een (al dan niet behaald) referentieniveau vast te leggen. Deze systemen hebben doorgaans allemaal hun eigen achtergrond, methodiek en bedoelingen. Het is de vraag in hoeverre ze onderling samenhangen en vergelijkbaar zijn.

In alle gevallen maken de leerlingen een toets (of opeenvolging van toetsen, verspreid over een langere periode) en wordt vastgesteld hoe goed ze deze toets(en) maken. De toetsopgaven zijn 'geijkt' in de zin dat ze min of meer dekkend zijn voor de betreffende leerstofgebieden (hoewel hierover ook gereede twijfels bestaan; zie Gravemeijer & Van Galen, 2023) en dat ze een bepaald prestatieniveau representeren dat mede op grond van empirisch onderzoek is vastgesteld. In het geval van de eindtoetsen (cq. doorstroomtoetsen) wordt bovendien de onderlinge vergelijkbaarheid gewaarborgd door gebruik te maken van landelijk gebruikte referentiesets van opgaven die in alle toetsen dienen voor te komen. Door deze 'ijking' is het mogelijk om de ruwe toetsscore ('zoveel van de zoveel opgaven goed') om te zetten in een vaardigheidsscore die op een bijbehorende vaardigheidsschaal geplaatst kan worden (Hollenberg & Veerbeek, 2023). Een bekend voorbeeld van zo'n vaardigheidsschaal is die van het Cito-leerlingvolgsysteem. Door op een zeker moment een aantal vaardigheidsscores op een rij te zetten en grafisch weer te geven kan nagegaan worden op welk niveau een leerling globaal functioneert en in hoeverre hij/zij de gewenste groei in prestaties vertoont. Alle aldus verkregen gegevens zijn uiteraard 'momentopnames' die in zekere mate onderhevig zijn aan toevallige omstandigheden. Ze zijn vooral bedoeld als een (meer objectieve) aanvulling op de persoonlijke observaties van de leraar verkregen tijdens instructiemomenten, verwerkingsmomenten en het analyseren van schriftelijk werk.

Een belangrijke vraag is nu in hoeverre deze toetssystemen onderling samenhangen en tot elkaar herleidbaar zijn. Op zich zou dat voordelen kunnen hebben, bijvoorbeeld doordat het een school in staat stelt om via de vaardigheidsscores uit een leerlingvolgsysteem in te schatten of leerlingen op de goede weg zitten voor het bereiken van de referentieniveaus 1F en 1S medio groep 8. In de literatuur zijn aanzetten te vinden om vaardigheidsscores om te zetten in een (mogelijk) referentieniveau. Zo is er op de Cito-site bij het onderdeel over Passende Perspectieven ('Plaatsing in een leerroute') een verwijzing te vinden naar het feit dat een score van 250 ongeveer overeen zou komen met referentieniveau 1F. En Ooievaar (2022) presenteert, zich basierend op 'leerlingenrapporten uit het Cito-portal', zelfs een overzicht van 'bij benadering' omgerekende scores, waarbij een score van 247 overeen zou komen met niveau 1F en een score van 281 met niveau 1S.

Maar het is zeer de vraag hoe betrouwbaar dergelijke omzettingen zijn, en zelfs of het überhaupt wel mogelijk is om de ene soort aanduiding van een prestatieniveau om te zetten naar een andere soort. De betreffende toetssystemen zijn immers grotendeels of geheel los van elkaar tot stand gekomen, doorgaans via uiteenlopende methodieken en met verschillende doelstellingen. Voorts lijkt het hoogstens verantwoord ervan uit te gaan dat er een globale samenhang is: naarmate een leerling hoger scoort op een bepaalde vaardigheidsschaal, zal hij/zij ook in zekere mate hoger uitkomen wat betreft het behaalde referentieniveau; en omgekeerd. Maar voor het direct omrekenen van de ene soort prestatieaanduiding in een andere soort (zoals bij wijze van spreken bij temperatuurmeting graden Fahrenheit op een vaste wijze omgerekend kunnen worden naar graden Celsius, en omgekeerd), is er op dit moment weinig bewijs. Gezien de steeds toenemende aandacht voor het meten van leerprestaties (vergelijk ook de aankondiging van de Inspectie dat men bezig is om een nieuw 'resultatenmodel' te ontwikkelen met een centrale rol voor de referentieniveaus) zou het zeker aanbeveling verdienen om hier nader onderzoek naar te doen.