



### Pieter Leenheer

Pieter Leenheer is lid van de redactie van DNM

E-mail: pieter.leenheer@planet.nl

Toetsing en beoordeling in de doordeweekse schoolpraktijk

# Een korte geschiedenis van **dagelijkse** **toetspraktijken** tussen 1801 en 2023

Schrijven over stukjes onderwijsgeschiedenis heeft iets confronterend als je zelf ooit lesgegeven hebt. Mij bevangt in elk geval nogal eens een lichte gêne. Zo heb ik ongetwijfeld indertijd maar al te vaak proefwerken of opstellen niet gebruikt als een serieuze kans voor de leerlingen om te leren. Eerder als een soort proefexamen, een test of de kinderen het nou eindelijk eens een beetje konden. Een excuus is hooguit dat ik begon met lesgeven in een tijd dat je didactische bagage op het punt van toetsing en beoordeling zacht gezegd nogal licht was en dat is inmiddels wel anders. Hoewel er ook nu nog sombere geluiden te horen zijn.

Een jaar of vier geleden schreef de Onderwijsraad dat de toetspraktijk uit balans is: beslissende toetsing kost zoveel tijd dat te weinig ruimte overblijft voor formatieve toetsing. Kwalitatieve toetsen zoals mondeling tentamen of presentatie worden te weinig gebruikt, en teveel toetsen worden niet door docenten zelf gemaakt maar extern vormgegeven (Onderwijsraad 2018). Nu is het sowieso moeilijk de toetspraktijk in balans te houden, maar dat ze momenteel uit balans is, is deels natuurlijk een pervers effect van het streven om eerlijker, dus minder subjectief te oordelen. Dat blijkt als neveneffect te hebben dat de toetsing als vanzelf onverantwoord versmald wordt tot een heel beperkt deel van het curriculum. Maar dat de toetspraktijk uit

balans is, is evengoed een gevolg van het feit dat leraren als beroepsgroep onvoldoende tegenwicht geven en bovendien lang niet altijd professioneel omgaan met toetsen en beoordelen.

In 1966 publiceerde Adriaan de Groot zijn spraakmakende *Vijven en zessen. Cijfers en beslissingen: het selectieproces in ons onderwijs*, waarin hij alle mogelijke misstanden op dit gebied aan de kaak stelde. Die eyeopener sorteerde echter zacht gezegd weinig effect zoals René Kneyber en Dominique Slujsmans in 2016 in de inleiding van *Toetsrevolutie. Naar een feedbackcultuur in het voortgezet onderwijs* constateerden: '[De Groot gaf] allerlei praktijkvoorbeelden die een tijdloos



karakter lijken te hebben: leerlingen die een 1 krijgen omdat ze zich niet goed gedragen; leerlingen die het laatste rapport 'inzakken' en daardoor het hele jaar opnieuw moeten doen; een leraar die bij toetsen veel hogere eisen stelt dan zijn collega's; en rapportvergaderingen waarbij gekeken wordt naar gemiddelden over vakken, met dubieuze rekenformules (Kneyber & Sluijsmans 2016). Een observatie die trouwens goed spooft met wat Janneke Sleenhof zag bij de rapportvergaderingen die ze op een aantal scholen volgde (zie elders in dit themadeel). Kortom, zoals Cito-medewerker Tom Erkens in 2013 in een artikel in het toenmalige Meso magazine opmerkte: '[De kwaliteit van toetsen is] dikwijls een lelijk probleem. Het denken over toetsing heeft sinds *Vijven en zessen* zo ongeveer stilgelegen' (Leenheer 2013).

De zojuist beschreven stand van zaken heeft zeer oude wortels en dat verklaart voor een deel waarom het handelen van leraren ten aanzien van toetsing en beoordelen nogal eens resistent blijkt tegen veranderin-

gen. Hieronder volgt een beknopt – en onvermijdelijk nogal generaliserend – beeld van de ontwikkelingen op dit terrein. En voor de goede orde: het gaat hierbij om toetsen die leraren zelf maken of kiezen uit methodes, en waarvoor ze zelf de normering bepalen; de eindtoets en examens blijven buiten beschouwing.

### **1801 – 1963: van Van der Palm tot Cals**

In 1801 neemt met de Onderwijswet van Van der Palm de overheid de zorg voor het lager onderwijs op zich. Nederland verkeerde op dat moment in diep verval en om uit het dal te komen waren geletterde arbeidskrachten nodig. En omdat alle kinderen onderwezen moesten worden en wel massaal en zo efficiënt mogelijk, moest het hoofdelijk onderwijs zoals dat tot dan toe gebruikelijk was, vervangen worden door klassikaal onderwijs waarbij de leerlingen naar leeftijd gegroepeerd werden (Boekholt & De Booy 1987; p.106). Nu stelt zo'n klassikaal systeem vanwege de overgang naar een volgende laag of sector hoge eisen aan

toetsing en beoordeling, maar daar lijkt weinig oog voor te zijn geweest. Misschien wel doordat de overheid onderwijs vooral zag als een instrument om het nationaal besef te bevorderen en de bestaande orde waarin iedereen zijn plaats kende, te handhaven. En zoals Boekholt en De Booy opperen: het lijkt erop dat het op de lagere school, zeker in eerste helft van de 19e eeuw, vooral ging om een goede ontwikkeling van het normbesef, minder om die van de verstandelijke vermogens. Dat laatste verandert in de tweede helft van de 19e eeuw, als de leerstof wordt vastgelegd en er naast de gebruikelijke vakken als lezen en schrijven vakken bij komen als geschiedenis en aardrijkskunde. Maar of de onderwijzers inmiddels beschikten over voldoende expertise om adequaat te toetsen en te beoordelen, daarover zijn weinig gegevens te vinden. Aan de andere kant, wellicht waren de problemen nog niet zo groot omdat het accent in het onderwijs toen (en ook nog tot ver in de 20e eeuw) sterk op kennisreproductie lag, iets dat wat makkelijker te toetsen valt dan bijvoorbeeld inzicht of toepassing.

## In 1801 neemt met de Onderwijswet van Van der Palm de overheid de zorg voor het lager onderwijs op zich.

Wat we nu het voorgezet onderwijs noemen, werd pas in de tweede helft van de 19e eeuw een voorwerp van - zoals artikel 23 van de Grondwet dat uitdrukt - aanhoudende zorg voor de overheid. Voordien waren er al wel vormen van voortgezet onderwijs, zoals de Latijnse scholen, maar die waren vooral een kwestie van particulier initiatief. Dat verandert in 1863, met de komst van Thorbeckes wet op het middelbaar onderwijs waarbij onder de hoede van de overheid (hogere) burgerscholen kunnen worden gesticht. In dit verband is van belang dat de overheid veel ruimte liet. De wet somde

bijvoorbeeld wel mogelijke vakken op, maar omschreef geen verplichte leerstof. De ratio hierachter was dat op die manier 'de verscheidenheid van inrichting van onderwijs, naar verscheidenheid van inzicht en bijzondere behoefte georganiseerd' tot zijn recht zou kunnen komen (Boekholt & De Booy 1987; p.183). Maar het gevolg van het feit dat er weinig geregeld was, was uiteraard dat iedereen deed wat goed was in zijn ogen.

Dat laatste betekende niet dat iedere leraar alle noodzakelijke wielen zelf uit moest zien te vinden. Scholen en individuele leraren zochten elkaar wel degelijk op en wisselden ideeën en ervaringen uit. En daarbij zullen veel leraren ook wel ideeën voor toetsvormen van elkaar hebben overgenomen. Maar – en dat is hier het punt – iedere leraar was koning in zijn eigen klas. En bij gebrek aan bijvoorbeeld een beroepsorganisatie die de kwaliteit van het beroep bewaakte, hield geen beroepsgenoot hem ooit een spiegel voor. En waar het bovendien ook vrijwel geheel ontbrak aan onderzoek naar het dagelijks handelen van leraren ten aanzien van toetsing en beoordeling, kon je het hem nauwelijks kwalijk nemen dat hij (want het was nu eenmaal gewoonlijk een hij) op dat terrein de nodige steken liet vallen.

Pas in 1940 verschijnt in De Gids een belangrijk artikel dat hem de ogen had kunnen openen: *Middelbaar onderwijs en schifting* van de hand van Kees Posthumus (op dat moment leraar aan een school in Bandung, later hoogleraar aan de TH Eindhoven). In de loop van zijn lange loopbaan als leraar had Posthumus gezien dat leraren zich onwillekeurig zo aan het prestatieniveau van hun klassen aanpassen dat ze, ongeacht dit niveau – dus zowel bij 'goede' als bij 'slechte' klassen - altijd ongeveer uitkomen op hun persoonlijke gemiddelde (en gemiddeld percentage onvoldoendes). Met andere woorden: leraren werken met rekbare maat-staven die zich aan het te meten object – d.w.z. de prestaties van de klas – aanpassen. Anders gezegd, in de praktijk van cijfers geven bestond (en bestaat) nogal wat ruimte voor alle mogelijke subjectieve factoren (De Groot 1966; p.114).

Overigens, voor zover er in de eerste helft van de twintigste eeuw gediscussieerd werd over toetsing, had dat

vooral betrekking op de forse uitval bij het toelatings-examen voor hbs en gymnasium of op de grote uitval bij de hbs. Typerend voor het onderwijskundig denken van die tijd is dat het daarbij vooral ging om de vraag hoe de selectie verbeterd kon worden, en niet over de vraag in hoeverre hbs en gymnasium daar zelf verantwoordelijk voor zijn. Leraren gingen er, zoals De Groot in *Vijven en zessen* opmerkt, kennelijk van uit dat altijd een deel van de leerlingen het niet zal kunnen, terwijl de vraag of het onderwijs wel geschikt is voor de leerlingen, niet gesteld werd (De Groot 1966; p. 44).

### Toetsing en beoordeling na de Mammoet

In de eerste helft van de twintigste eeuw stond mede door de schoolstrijd de onderwijsontwikkeling nagenoeg stil. Dat verandert snel na WO II, met als markeringspunt de Mammoetwet die in 1963 van kracht wordt en waarbij het hele onderwijsstelsel op de schop gaat. Vlak daarvoor had Philip Idenburg (toentertijd directeur van het CBS) zijn *Schets van het Nederlandse schoolwezen* gepubliceerd, een uitvoerige beschrijving van het toenmalige onderwijsstelsel, in het volle besef dat er een grote reorganisatie aan zat te komen. Maar, schreef Idenburg in zijn *Ten geleide*, 'ik geloof nauwelijks dat als gevolg van een eventuele nieuwe wetgeving revolutionaire ontwikkelingen in de structuur en functionering van een enorm en verspreid apparaat als ons schoolwezen te verwachten zijn'. En dat had hij goed gezien. Evenals trouwens een paar jaar later De Groot in *Vijven en zessen*. Een onderwijssysteem, aldus De Groot, is complex, omvangrijk en massief, dus is er heel wat voor nodig om verandering aan te brengen. En al helemaal als het om dingen gaat die direct te maken hebben met het dagelijks werk: dan gaat men vaak de verandering neutraliseren (De Groot 1966; p.173). En dat is exact wat dan op het terrein van toetsing en beoordeling in de schoolpraktijk van alledag gaat gebeuren.

In de jaren na WO II stapt de overheid over van distributief onderwijsbeleid (de overheid verstrekt geld en laat het verder aan de scholen over) naar constructief beleid (de overheid verstrekt geld en bemoeit zich actief met de vormgeving van het onderwijs). In dat

kader komen er steeds meer door onderwijskundigen bemande beleidsafdelingen bij, die van bovenaf trachten verandering en verbetering aan te brengen in het onderwijs. En verder wilde de minister met allerlei innovatiecommissies en beleidsnota's richting geven aan het denken over onderwijs. De minister en zijn ministerie probeerden op deze wijze te fungeren als centrum van waaruit de ontwikkelingen in het onderwijs gestuurd werden (Boekholt & De Booy 1987; p.239).

Bij de vormgeving van deze lineaire sturing speelde het denken van De Groot een grote rol. De oplossing van De Groot lag, zoals oud-Inspecteur-Generaal Ferdinand Mertens stelde in een aan *Vijven en zessen* gewijd boekje, 'in een meer 'rationele en wetenschappelijke benadering van onderwijs, een empirisch-analytische benadering, kort samengevat in "meten is weten". Leraren moeten af van hun persoonlijke en individuele manier van denken waarbij immers als consequentie elke leraar zijn eigen regime heeft.' (Mertens 2016)

#### Kritiese leraren en toetsing

Typerend voor de geest van de tijd, was hoe de Kritiese Leraren over toetsing oordeelden. In 'het rode boekje voor scholieren' schrijven ze: "Een diploma is (...) in de eerste plaats een bewijs dat je je kunt aanpassen. Daarom vragen de bedrijven naar diploma's. (...) Men zegt dat het examen bedoeld is als een test van je kennis. Maar in werkelijkheid werkt het als een test van je aanpassing en je blinde gehoorzaamheid aan wat anderen zeggen.' (Claartje Hülsenbeck e.a., *het rode boekje voor scholieren*. Bruna, 1970; p. 52/53).

De Groot had gepleit voor de oprichting van een instituut dat objectieve toetsen zou ontwikkelen en dat werd het Cito. En hoe dominant het 'meten is weten' zou worden, was eigenlijk van begin af aan duidelijk. 'Dit instituut', aldus Piet de Rooy in zijn geschiedenis van het Nederlandse onderwijs, 'presenteerde zich bij de oprichting in 1968 op een persconferentie als een 'Instituut

contra cijfer-dwang'. Het effect van het onderwijs zou voortaan op een objectieve wijze kunnen worden vastgesteld: "De docent zal zichzelf gaan bezinnen op de stof die hij geeft. Hij kan door de uniforme testen precies constateren welk deel van zijn stof zinvol en welk deel nutteloos is". Dit betekende dat ook "de kwaliteit van de docent" kon worden 'becijferd' (cursivering PL) en dan was het "best mogelijk" dat hierdoor 'de leerlingen minder en de leraren meer onvoldoendes zullen behalen'. (De Rooy 2018; p.232).

In de jaren daarna wordt het gevoel steeds sterker dat het onderwijs lijdt onder een meetcultuur. En afgaande op wat bijvoorbeeld Cito-onderzoekers zeiden in een interview in *Van Twaalf tot Achttien* van november 2021, is dat niet onbegrijpelijk: per jaar, aldus de onderzoekers, krijgen leerlingen 75 toetsen plus of min 25 voorgeschoteld (Valk 2021). Maar in een blog op Didactiefonline vinden Jaap Scheerens en Ton van Haperen dat het onzin is om van een meetcultuur te spreken: op systeemniveau zijn er immers maar drie

### **Idenburg over onderwijskwaliteit**

Deze ontwikkeling moet Idenburg een gruwel geweest zijn. Het zou mooi zijn, stelt hij aan het slot van zijn *Schets*, als we een winst- en verliesrekening van het onderwijsbedrijf konden opstellen, maar geen enkele weegschaal is daartoe toereikend: 'Weegt men kennis dan doet men aan de ontplooiing der verstandelijke vermogens tekort, meet men intelligentie dan komt de vorming van persoonlijkheid en karakter niet tot haar recht. De schade, welke de school aan haar leerlingen aandoet, valt evenmin te bepalen. (...) [De] rekening van het schoolwezen wordt nooit afgesloten en nimmer openbaar gemaakt. (Idenburg 1960; p.517)

verplichte toetsmomenten, dus als er teveel wordt getoetst, is dat de keuze van de school. De vraag is echter of de school op dat punt wel zoveel te kiezen heeft. De



druk om te toetsen is groot doordat de overheid, aldus Monique Marreveld en Bea Ros in *Van toets naar toets*, een artikel in de *Groene Amsterdammer* van 30 september 2021, sterk stuurt op onderwijskwaliteit aan de hand van onder meer toetsen. De inspectie let bijvoorbeeld op doorstroompercentages, op de mate van vertraging. En dat dan terwijl die toetsen daarvoor absoluut ongeschikt zijn. Maar dat niet alleen: ze maken van de leraar ook nog eens een simpele uitvoerder: hij neemt de toets af, maar het Cito maakt de uitslag. En daarmee zijn we dan weer terug bij de constatering van de Onderwijsraad dat de toetspraktijk uit balans is geraakt.

### Ten slotte

Eerder in dit artikel haalden we de uitspraak aan van Tom Erkens aan dat het denken over toetsen en beoordelen sinds 1966 zo ongeveer stil lijkt te hebben gelegen. Maar inmiddels lijkt er wel meer beweging gekomen. De Cito-onderzoekers van hierboven zagen een paar jaar geleden in een overigens bescheiden onderzoek in het vo (181 respondenten) dat er weliswaar op het gebied van toetsconstructie nog een wereld te winnen valt, maar dat in elk geval het aantal scholen met toetsbeleid en professionalisering ten aanzien van toetsing geleidelijk toeneemt (Valk 2021). En mede door de publicatie van Kneyber en Sluijsmans' *Toetsrevolutie* is her en der de aandacht toegenomen voor formatief toetsen en daarmee het denken vanuit de doelen van het onderwijs. Dus wie weet zitten we op een kantelpunt. Maar of dat werkelijk het geval is, is meer iets voor een terugblik in een DNM van jaargang 20. Oftewel over een jaar of tien. ■

### Referenties

- Boekholt, P.Th.F.M. en De Booy, E.P. (1987). *Geschiedenis van de school in Nederland vanaf de middeleeuwen tot aan de huidige tijd*. Van Gorcum
- Groot, A.D. de (1966). *Vijven en zessen. Cijfers en beslissingen: het selectieproces in ons onderwijs*. Wolters-Noordhoff
- Idenburg, PH.J. (1960). *Schets van het Nederlandse schoolwezen*. J.B. Wolters
- Kneyber, R. en Sluijsmans, D. (2016). *Toetsrevolutie*. Phronese
- Leenheer, P. (2013). De paraplu van de hardnekkige illusies. Kanttekeningen bij de erfenis van A.D. de Groots Vijven en zessen. In: *Meso magazine* 191, p. 3-7
- Marreveld, M. en Ros, B. (2021), *Van toets naar toets*. Het georganiseerde wantrouwen in het onderwijs. In: *De Groene Amsterdammer*, 30-9-2021
- Mertens, F. (2016). 'Vijven en Zessen' van Prof. Adriaan de Groot een boekje dat geschiedenis maakte. Wolf Productions
- Onderwijsraad (2016). *Toetswijzer. Naar een eigen(tijdse) wijze van toetsen en examineren*. Onderwijsraad
- Rooy, P. de (2018). *Een geschiedenis van het onderwijs in Nederland*. Wereldbibliotheek
- Scheerens, J. en Haperen, T. van (2022). *De mythe van de afrekencultuur*. Blog op Didactiefonline d.d. 13-5-2022
- Valk, R. (2021). *Cijfers, cijfers, cijfers*. Stichting Cito onderzoek de toetspraktijk. In: *Van Twaalf tot Achttien*, november, p. 24-26