

Waarde van toetsing

Kwaliteit van onderwijs, gericht op wat onderwijs weet te realiseren met leerlingen, was helemaal geen vraagstuk tot ver in de vorige eeuw. Tot die tijd ging het bij kwaliteit van onderwijs vooral over gebouwen, inrichting, leermiddelen. Dat wat voorwaardelijk is om überhaupt onderwijs te kunnen aanbieden. Pas in de tijd van de debatten over de invoering van de middenschool kwam het vraagstuk van kwaliteit van onderwijs expliciet op tafel. In die discussie kwam voor het eerst expliciet de vraag op tafel of die vernieuwing zou leiden tot betere resultaten van onze leerlingen. Resultaten als indicator van kwaliteit! Deze verschuiving van aandacht van voorwaarden voor het verzorgen van onderwijs naar opbrengsten van onderwijs als indicator van kwaliteit kwam met name uit de koker van psychologen die zich met het onderwijs gingen bemoeien met Adriaan De Groot als belangrijke aanjager van deze manier van kijken.

Met zijn boekje Vijven en Zessen had deze testpsycholoog in 1965 veel opzien gebaard door te stellen dat het schandalig was dat we in Nederland geen idee hadden van leerdoelen en dat onze manier van cijfers geven zeer onbetrouwbaar en daarmee ook heel oneerlijk was. [sheet doorklikken naar boekje vijven en zessen]. Het lezen van dit boekje liet menig lezer achter met het gevoel: 'er wordt maar wat aan gerotzood in het voortgezet onderwijs'. Hij pleitte voor een veel meer rationele benadering van onderwijs en de opbrengsten ervan. Zijn adagium was: 'wat niet objectief getoetst kan worden, doet er niet toe'.

De Groot was van origine een testpsycholoog. Hij zag testen als een vorm van voorspellen van toekomstig succes. En daar mocht volgens hem niet te licht over worden gedacht. Ook voor onderwijs waren volgens hem objectief scorebare vormen van testen nodig die door experts ontwikkeld moesten worden en wat je niet kon overlaten aan het onderwijs zelf. De Groot heeft met zijn aanwezigheid in het onderwijsdebat een aantal belangrijke effecten gehad. Hij heeft er in belangrijke mate aan bijgedragen dat:

- we de begrippen 'testen' en 'toetsen' gelijk zijn gaan stellen en door elkaar zijn gaan gebruiken;
- we toetsing voor een belangrijk deel uit het onderwijs gehaald hebben en als een externe controle buiten het onderwijs geplaatst hebben;
- het onderwijs verengd is tot datgene wat meetbaar gemaakt kan worden;
- onderwijs langzaam van kleur is verschoten: van onderwijs gericht op brede vorming naar trajecten met meetbare doelen, waarover later meer.

Daarmee heeft De Groot een niet te onderschatten invloed gehad op het onderwijsdebat in Nederland.

Het aangezette denken over kwaliteit van onderwijs en hoe dat te definiëren kreeg de wind in de zeilen vanuit een geheel andere, onvoorziene hoek. Namelijk vanuit het managementdenken dat vanuit de VS kwam onder de vlag van 'New Public Management'.

Dat denken was bedoeld om overheidsinstanties meer te laten functioneren als bedrijven zodat ze efficiënter en effectiever zouden gaan functioneren en daarmee de kwaliteit van de publieke dienstverlening omhoog zouden kunnen brengen. Het begrip 'accountability' deed zijn intrede. De hele overheid, maar ook het onderwijs, onderging een gedaanteverandering, ook het ministerie van onderwijs. Er moest door ambtenaren niet langer geroeid worden maar vooral gestuurd.

Dit sturen op resultaten bleek voor de denkers over kwaliteit van onderwijs een uitkomst omdat men er maar niet uit kwam wat nu precies kwaliteit van onderwijs zou zijn. Het debat kon nu verschoven worden van het zoeken van een inhoudelijke definitie naar een praktische oplossing: namelijk opbrengsten van het onderwijs definiëren zodat daar op kon worden gestuurd en toegezien. Scholen werd meer autonomie beloofd als ze maar lieten zien dat ze resultaten wisten te boeken. Op papier een aantrekkelijke uitruil.

De testpsychologen en hun manier van denken sloten goed aan bij het zoeken naar opbrengsten van onderwijs: Kwaliteit kon volgens hen worden vastgesteld aan de hand van wat leerlingen wisten te presteren, gemeten met testen. En dan niet met de testen die in het onderwijs tot dan toe gebruikelijk waren maar met testen bestaand uit uitsluitend meerkeuze vragen omdat die handig waren om berekeningen uit te voeren met de gegenereerde data. De testen moesten bij voorkeur extern ontwikkeld en gevalideerd worden zodat de beoordeling van opbrengsten losgekoppeld kon worden van degenen die het onderwijs verzorgen. Want de slager die zijn eigen vlees keurt, paste niet in het denken over opbrengsten waarmee je kon sturen en toezien. Zo kon de klassieke testtheorie, die uitgaat van het vergelijken van leerlingen met elkaar op basis van eenduidig te scoren antwoorden, dus meerkeuzevragen, heel gemakkelijk binnendringen in het onderwijs en daar aan terrein winnen.

De verschuiving van pedagogen naar psychologen als toonzetters van het onderwijsdebat, is van enorme invloed geweest. Van onderwijs als plek waar onze toekomstige generatie gevormd moest worden, is onderwijs veel meer overgegaan naar de instantie die verantwoordelijk is voor het kwalificeren van nieuwe arbeidskrachten uitgaand van talenten of, als spiegel daarvan, hun deficiënties. Deze veranderingen zijn diepgaand, ingrijpend en nog altijd aan de gang. Tot op het niveau van ons taalgebruik. We zijn gaan praten over leren in plaats van onderwijzen en ontwikkelen, we praten steeds minder over 'de school' als instituut maar steeds vaker over de school als 'leeromgeving' en uitkomsten van toetsen zijn uitgegroeid tot indicatoren van kwaliteit, veel meer dan feedback of liever feedforward, bedoeld voor de groei voor de leerling. De bandbreedte van wat nu eenmaal individuele verschillen zijn en geaccepteerd behoren te worden als zodanig, is versmald: leerlingen en studenten die niet mee kunnen, worden zo snel mogelijk van een label voorzien waar specialistische hulp voor nodig is als differentiatie geen soelaas biedt. Zo zijn inhoud en plaats van leren van elkaar losgezongen, zijn proces en resultaten van elkaar ontkoppeld geraakt en zijn onderwijsinstellingen steeds meer kenmerken gaan krijgen van individuele 'behandelcentra' waar toetsen, vergelijkende toetsen, een prominente rol hebben gekregen en uitkomsten van toetsen bepalend zijn voor de 'behandeling' van onze leerlingen.

Met terugwerkende kracht is het verbazingwekkend dat er zo gemakkelijk gekomen is tot de slotsom dat leerlingresultaten het 'product' zijn van onderwijs. Hoezo is dat zo? In de gezondheidszorg is er veel genuanceerder gedacht over wat het 'product' is van de gezondheidszorg. Daar is de 'behandeling in relatie tot de patiënt' toch steeds overeind gebleven. In onderwijs is de 'behandeling' in de vorm van een onderwijsprogramma helemaal naar de achtergrond gedrongen en is het resultaat van de behandeling in termen van scores op toetsen de focus geworden voor het vaststellen van 'output' of resultaten of opbrengsten. Dat terwijl we weten dat resultaten van leerlingen niet één op één het resultaat zijn van het gegeven onderwijs maar dat er veel meer over te zeggen is. En we weten dat de inhoud van het onderwijs, het programma, wel degelijk van groot belang is en voortdurend onderwerp van gesprek zou moeten zijn.

Het heeft niet mogen baten. Er heeft zich langzaam een omdraaiing van zaken voltrokken: Toetsen zijn geen middel om het leerproces te bevorderen maar het leerproces is een middel om toetsresultaten te bevorderen. Doordat resultaten in toezicht en verantwoording een prominente plek hebben gekregen, is er meer druk ontstaan op de prestaties van leerlingen: Leerlingen moeten met hun scores aantonen dat de kwaliteit van hun ontvangen onderwijs op orde is.

We zien deze omkering van zaken in alle vormen van onderwijs in eigen gedaantes terugkomen. Ik licht dit per vorm van onderwijs nader toe.

In het basisonderwijs worden leerlingen vergeleken met elkaar en in een onderlinge competitie wordt beslist wie met welke prijs in de vorm van een ticket tot vervolgonderwijs naar huis mag. Door steeds te benadrukken dat de wedstrijd eerlijk is, iedereen mag tenslotte meedoen aan dezelfde eindtoets en iedereen heeft kans om te winnen, zorgen we ervoor dat de winnaars van die onderwijswedstrijd juichen en de verliezers beschaamd wegstaren, in de volle overtuiging dat ze een eerlijke kans hebben gehad maar niet hebben weten te winnen. Op 12-jarige leeftijd worden zo in Nederland de kaarten geschud.

Die selectie stopt niet. Ook in het voortgezet onderwijs gaat de selectie door. Ook daar voelen leerlingen voortdurend de druk om op basis van toetsresultaten afgewezen te worden voor vervolg. De manier van toetsen zorgt ervoor dat leerlingen hun positie nooit veilig weten. Dit zelfde geldt voor scholen. Ook zij moeten op basis van hun resultaten aantonen dat zij de kwaliteit van hun onderwijs op orde hebben. Scholen moeten zich verantwoorden over leerlingen die afstromen en leerlingen hoger plaatsen die dat niet waar maken kan financiële gevolgen hebben. Dit maakt dat scholen voor voortgezet onderwijs huiverig zijn geworden voor het bieden van kansen aan leerlingen. Het streven naar en afgerekend worden op de mate van efficiëntie en effectiviteit van onderwijstrajecten staat op gespannen voet met de gedachte dat onderwijs de plek is waar de leerlingen de rust en de ruimte hebben om te ontdekken waar ze goed in zijn, wat bij hen past en wat ze willen. Zoals de filosofe Hannah Arendt dat bedoelt als ze spreekt over onderwijs als 'tussenruimte'.

In het hoger onderwijs heeft de toetscultuur een wat andere gedaante maar in de basis heeft ook daar zich een vergelijkbare verandering voltrokken in de laatste decennia. En dan bedoel ik niet de ontwikkeling van theorietoetsen naar competentietoetsen naar

assessments waarbij we proberen de vinklijstjes achter ons te laten omdat achter elk beoordelingspunt wel weer een beoordelingspunt schuil gaat en we dan gevangen zitten in wat ik de 'matroeska-val' noem. Of dat we van analytisch beoordelen naar meer holistisch beoordelen zijn gegaan. Dit zijn met alle respect voor de verschillen die echt wel bestaan tussen de verschillende vormen van toetsen en beoordelen toch minor points. Wat ik bedoel is dat ook in het hoger onderwijs de bewijslast voor succes van onderwijs bij de studenten en hun onderwijsresultaten is komen te liggen. De controleerbaarheid, de meetbaarheid en de standaardisatie vormt het kader waarbinnen over verantwoording moet worden nagedacht. Dit gaat ten koste van wat de bedoeling is van het volgen van hoger onderwijs, de rol van de professional in dat onderwijs en houdt ook de kwetsbaarheid van studenten in stand. Anders dan in het basisonderwijs en het voortgezet onderwijs is de manier waarop toetsing zich in het hoger onderwijs ontwikkelt in zijn uitwerking minstens zo heftig voor studenten. Ik zal hier wat dieper op ingaan.

De Franse filosoof Foucault gebruikt in zijn boek *Discipline, Toezicht en Straf*, de metafoer van een panopticum: een vorm van bouwen die veel gebruikt is voor gevangenis.. Er is voortdurend toezicht mogelijk wat het gevoel geeft dat de kans dat je wordt gezien en gesnapt groot is. Dit werkt disciplinerend: iedereen gaat in de pas lopen om te zorgen dat je doet wat je geacht wordt te doen en geen gedonder krijgt. Toezicht op onderwijs heeft een panoptisch karakter waarbij examinering volgens Foucault de ultieme vorm van disciplinerend is.

In het hoger onderwijs zien we zich een nieuwe vorm van disciplinerend ontwikkelen gebaseerd op het 'profiel', vastgelegd in het portfolio dat studenten vanaf de start van de opleiding moeten aanleggen. De pedagogen Simons en Masschelein spreken in dit verband over een nieuwe machtsconfiguratie. Zij definiëren profiel met portfolio als instrument als de manier waarmee iets eigens of bijzonders van iemand tot uitdrukking wordt gebracht dat publiekelijk kenbaar of (h)erkenbaar is. Het toont met andere woorden een publieke identiteit aan de hand van onderscheidende of bepalende kenmerken. En zo'n profiel vergt permanente monitoring in de betekenis van permanent meten/evalueren en permanent zichtbaar maken/tonen. Hier hoort, zo stellen zij, feedback bij als heersende machtstechniek.

Op het eerste gezicht zou je denken dat er bij het werken met profielen, vastgelegd in een portfolio, ruimte is voor wat het individu zelf belangrijk vindt en wil laten zien maar in de context van leren is niets minder waar. Het portfoliodenken ontpopt zich als een spel van 'snakes and ladders'.

Want het denken in termen van een leeromgeving, met leertrajecten, certificeringen op basis van kwalificaties of competenties heeft ervoor gezorgd dat de student permanent in observatie is en permanent bewaakt wordt voor wat betreft zijn ontwikkeling op basis van feedback in de richting van de volgens de opleiding relevante competenties, vastgelegd in zijn portfolio. De verwachting dat alle leerlingen dezelfde leerstof op dezelfde wijze en binnen hetzelfde tempo moeten verwerven, geeft automatisch aanleiding tot achterblijvers en achterstand: normalisering als uitgangspunt dus. Die leeromgevingen zijn er niet om de student zoveel mogelijk ruimte te bieden zich te ontwikkelen, wat ik nadrukkelijk onderscheidt van leren. Dat is niet de bedoeling want die leeromgevingen moeten effectief

en efficiënt zijn en met zo weinig mogelijk inspanningen zoveel mogelijk opbrengsten realiseren.

Daar komt bij dat studenten ook nog eens nauwelijks zeggenschap hebben over de waardering van wat ze zelf verkiezen te doen. Want feitelijk tellen alleen die competenties die samenvallen met wat is afgesproken als maatschappelijk relevante competenties of, in het geval van een opleiding, voor een opleiding geldende competenties. Daarmee ligt uniformering aan de basis van diplomering. De hele individualisering in termen van maatwerk leidt zo niet tot pluriformiteit maar juist tot een enorme eenzijdigheid, het beste samen te vatten als 'Worden zoals wij', de titel van het proefschrift van historicus Bram Mellink.

Zo lijkt het alsof studenten veel vrijheid hebben in het organiseren en structureren van hun leerweg maar is de werkelijkheid vooral dat ze nauwelijks bewegingsruimte hebben en zich voortdurend moeten blootgeven wat betreft de stand van wat ze geleerd hebben, vastgelegd in hun portfolio en besproken in assessments. Dat voortdurende reflecteren op wat ze geleerd hebben biedt studenten helemaal geen ruimte om te ervaren, te observeren, zich te verwonderen en rustig te laten neerdalen wat dat allemaal met ze doet.

Onlangs is een boek verschenen van Jos Geerligts. Hij introduceert daarin het begrip 'curriculair bijpraten'. Wat hij daarmee bedoelt is dat gesprekken over wat je meemaakt en wat dat met je doet vóórwaarde zijn voor reflectie en evaluatie. Zo werkt dat voor ons allemaal in het dagelijks leven ook: praten over wat je meemaakt en beleeft, zorgt voor structuur en ordening van je gedachten en pas dan kun je er goed over nadenken en komen tot reflectie en nieuwe inzichten. Dit slaan we allemaal over in onderwijs en vormt geen expliciet onderdeel van het onderwijs. Wat ik zie en hoor zijn studenten die proberen de reflecties te doen volgens de instructies omdat ze weten dat ze door die hoepeltjes moeten maar vooral stress hebben van deze aanpak.

En het kan nog gekker zo bleek een paar weken geleden. Had ik net nog met plezier gekeken naar het filmpje van Van Kooten en De Bie over Pretpong, een vorm van tafeltennis waarbij winnen niet de bedoeling was, hoorde ik op radio 1 een uitgebreid bericht dat Artificiële intelligentie zou kunnen gaan zorgen voor betrouwbare, automatische beoordelingen van studenten. Een volgende fase staat voor de deur. De dominantie van de klassieke testpsychometrie staat op het punt ingeruild te worden voor AI. Het is de hoogste tijd voor reflectie. Reflectie op wat we allemaal aan het optuigen zijn, hoe we dat doen en wat de effecten ervan zijn op onze leerlingen en ons onderwijs.

Het willen verbinden van een verantwoordingscyclus aan een voortdurende monitoring van de student ten opzichte van zijn graad van inzetbaarheid heeft een situatie gecreëerd die haaks staat op hoe onderwijsfilosofen en pedagogen onderwijs zien. Onderwijs is een proces waarvan de uitkomst per definitie ongewis is: een prachtig risico, zoals Gert Biesta ons met zijn boek heeft voorgehouden.

Hoewel in de verschillende vormen van onderwijs het denken in termen van output, leerwinst of opbrengsten een ander gezicht heeft, zijn er wel degelijk zorgen die gelden voor

alle gedaantes van deze manier van denken over grip op kwaliteit zoals we die in alle vormen van onderwijs kunnen waarnemen.

Overall leidt het denken over onderwijs als leerinstelling met meetbare opbrengsten tot verschraling van het onderwijs, met in de basis een streven naar uniformiteit als uitkomst en leerlingen en studenten die zich voortdurend publiekelijk moeten laten zien om feedback te krijgen om te bewijzen waar ze zich bevinden richting hun weg naar vervolgonderwijs of arbeidsmarkt.

Zij ervaren in alle geledingen van het onderwijs druk en stress om te moeten presteren om daarmee aan te tonen dat zij zelf succesvol zijn, daartoe aangezet door hun leeromgeving die daar grote belangen bij heeft. En voor het bewijzen van de eigen kwaliteit als instelling daarvan zelfs in hoge mate afhankelijk is.

Leerlingen en studenten zijn zo gezien geworden tot pionnen op het schaakbord van een leerinstelling in de verantwoordingssystematiek. Deze manier van denken en werken heeft ons afgehouden van nadenken over wat de bedoeling is van onderwijs, wat een goed programma is, wat we bedoelen met kwaliteit van onderwijs en hoe dat vastgesteld kan worden.

We hebben nu zo'n dikke 50 jaar geëxperimenteerd op alle mogelijke manieren met toetsen als basis voor denken over onderwijsresultaten, ingegeven door een testpsychologische kijk op onderwijs. We komen daar niet uit. Hoewel we altijd spreken over validiteit en betrouwbaarheid kan de conclusie niet anders zijn dan dat we toetsen niet betrouwbaar en valide kunnen maken op zo'n manier dat we daar high stakes beslissingen mee kunnen nemen.

Toetsen kunnen de werkelijkheid slechts zeer gebrekkig weergeven. Toetsen blijken in de praktijk vooral weerbarstige en veelkoppige monsters die tal van ongewenste effecten sorteren, vooral op de lange termijn, ondanks de soms best goede bedoelingen waarmee eraan begonnen is. We kunnen niet anders dan vaststellen dat toetsen hooguit een rol kunnen vervullen in een formatieve context, waarbij de resultaten stevig in handen zijn van een onderwijsprofessional die de uitkomsten op waarde kan schatten. Aan toetsen moet niet meer dan een zeer bescheiden plek aan de rand van het onderwijs toegekend worden, als onderdeel van het eigen didactisch repertoire van docenten. Die bescheiden, en vooral andere, plek is nodig omdat toetsen meer niet kunnen waarmaken.

Toetsen die niet bij machte zijn de werkelijkheid goed weer te geven in combinatie met studenten die er op alle mogelijke manieren last van hebben, moeten aanzetten tot verandering.

Afstand nemen van wat we kennen en al jaren doen kan op het eerste gezicht misschien stress geven maar zie vooral de ruimte die dit kan bieden. Als gesprekken die jullie met studenten hebben weer echte gesprekken kunnen worden waarin je in verbinding met studenten in alle openheid dieper in kan gaan op hoe zij het vakgebied zien, wat zij meemaken in de praktijk, wat dat met hen doet, hoe ze dat ervaren. Dat curriculaire bijpraten waar ik eerder aan refereerde.

Als verbinding en betrokkenheid met elkaar weer centraal komt te staan dan kan dat ook gelden voor de groei en ontwikkeling van het individu tot professional. De professionele dialoog als uitgangspunt. De professionals die we zo gaan vormen, zullen vele gezichten en kleuren kunnen hebben en de diversiteit vormgeven waar zo dringend behoefte aan is. Voor docenten betekent dit dat zij vanuit hun eigen professionaliteit, als vertegenwoordigers van het vakgebied, nieuwe generaties beter voorbereid de wereld in de wereld kunnen gaan. Met een gerust hart.

Het is de hoogste tijd dat het gesprek weer gevoerd kan worden over de bedoeling van onderwijs. Niet over de mate van validiteit van toetsen maar over de waarde, de validiteit, van onderwijs zelf. Dan kan vervolgens nagedacht worden hoe we verantwoording kunnen afleggen die past bij de visie op onderwijs. Verantwoording vanuit professionaliteit. En vanuit vertrouwen in de professionals die met elkaar het onderwijs vormgeven in lijn met een doordacht onderwijsprogramma. Want wat studenten doen met het onderwijs dat we bieden, hebben we niet in de hand, dat kan geen feitelijk resultaat zijn waarop onderwijs wordt beoordeeld en afgerekend.

Het programma zelf, de manier waarop het wordt gerealiseerd en door wie, dát zijn de indicatoren waarover we verantwoording kunnen afleggen. Of nog liever verantwoordelijkheid voor gaan nemen. Dit lijkt een semantisch verschil maar is het niet. Verantwoording afleggen veronderstelt een hiërarchische relatie waarbij je je verantwoordt naar boven toe. Bij verantwoordelijkheid nemen hoort gelijkwaardigheid, professionaliteit, gaan staan voor waar je in onderwijs mee bezig wil zijn: het bijdragen aan de ontwikkeling van een nieuwe, sterke generatie professionals.

Hoe mooi zou dat kunnen zijn? Praat er vandaag eens over, deel je gedachten erover hier met collega's, laat het eens bezinken. Van eerst maar eens curriculair bijpraten naar, op termijn, reflecteren: Dat werkt beter 😊

Karen Heij

Keynote Landelijke Assessorendag, 6 april 2023

Literatuur

Foucault, M. (1989) [1975]. *Discipline, toezicht en straf: De geboorte van een gevangenis*. Groningen: Historische Uitgeverij.

Geerligts, J. (2022). *Leren van betekenis. De kracht van curriculair bijpraten*. Gouda: Raccent.

Heij, K. (2021). *Van de kat en de bel. Tellen en vertellen met de eindtoets basisonderwijs*. Steenwijk: Bijzonderdruk.

Simons, M., & Masschelein, J. (2017). *De leerling centraal in het onderwijs. Grenzen van personalisering*. Leuven/Den Haag: Acco.