

Meetbaarheid en cijferbaarheid - Roger Standaert

Impuls Blog, Roger Standaert

Het voorbije decennium heeft er een hele ontwikkeling plaats gevonden waarbij steeds meer belang wordt gehecht aan de inzet van centrale toetsen. Denk maar aan de toetsen op het einde van het zesde leerjaar, de KOALA-toets voor kleuters, de Columbusproef, de toename van ijkingstoetsen, de instapproef voor lerarenopleidingen, de geplande inburgeringstoets en niet te vergeten de grootscheepse en dure operatie van de Vlaamse toetsen. Daarnaast krijgen de resultaten op de PISA-toetsen, de TIMSS-toetsen en de PIRLS-toets een bijna heilige verering. De vraag is of dat groeiende geloof in de exacte meetbaarheid van leerprestaties niet in een ongewenste stroomversnelling is terecht gekomen. Daarom is het belangrijk te wijzen op de functies en dysfuncties van metingen van leerlingengedrag en leerlingenprestaties. Die prestaties worden met cijfers bekleed en geven daarmee een gevoel van betrouwbaarheid en exactheid. Maar is dat wel terecht?

Punten en contexten

Puntenkaarten zijn een vertrouwd beeld in ons onderwijs. Rapportcijfers hebben een waarde als communicatie. Als je een tien hebt, heb je zeer goed gewerkt. Zit je onder de vijf dan is er iets mis. Cijfers hebben het voordeel van duidelijkheid voor leerlingen en ouders. Maar je kan niet zeggen dat die cijfers echt op een nauwkeurige schaal zitten zoals meters, graden, frequenties, voltages of welke wetenschappelijke schalen ook. Leraren geven cijfers om duidelijkheid te verschaffen over wat ze in die klas met die leerlingen en die leerstof verwachten. In die bepaalde context geven punten een - weliswaar beperkte - feedback aan de leerlingen maar ook aan de leraar. Die verwachtingen kunnen verschillen bij andere leraren en vakgroepen. Dat kan ook, voor zo ver ze binnen de klijtlijnen van de eindtermen en leerplannen blijven. Een eindterm laat immers diverse operationalisering toe in de vorm van lesdoelen. Eindtermen kunnen nooit lesdoelen zijn omdat ze dan niet alleen

strijdig zijn met de grondwettelijke vrijheid van onderwijs, maar ook de professionaliteit van leraren ondergraven. Lesdoelen zitten steeds in een bepaalde context van een klasgroep of school. Overigens zeggen die punten helemaal niets over waar je echt goed of minder goed in bent. Dat is dan weer de zwakte ervan. Kort gezegd, punten op toetsen zijn een interessant en pedagogisch verantwoord hulpmiddel om met leerlingen in een bepaalde concrete context aan de slag te gaan. Ze zijn echter niet het doel van het onderwijs.

Het schaduwspel van de cijfers

Vanuit deze concrete situatie kan je overschakelen naar de tendens om steeds maar meer verplichte toetsen in te voeren. Toetsen zouden dan exact meten of en in welke mate de leerlingen de doelen hebben bereikt. De overheid pretendeert daarmee menselijk gedrag te meten, net zoals de natuurwetenschappen dat doen. Daar komt dan de psychometrie om de hoek kijken. Die tak van de psychologie kent cijfers toe aan intellectuele prestaties, zodat je dan met die cijfers alle mogelijke statistische operaties kan uitvoeren. Gekend zijn de becijferde gemiddeldes en standaardafwijkingen van leerlingenprestaties. In een verder stadium worden die dan in verband gebracht met andere cijfers via correlatieberekeningen, vaak met correcties voor allerlei factoren, die dan ook een cijferwaarde krijgen. Uiteindelijk heeft dit meetbaarheidsdenken geleid tot een bijna blind geloof in de waarde van toegekende cijfers en de steeds verder uitdijende bewerkingen ermee. De toetsvragen hebben echter geen een- op- een- relatie met de doelen van het onderwijs. Je kan die doelen immers op verschillende wijze operationeel maken en dus ook verschillend evalueren.

Bij deze toetsenbeweging wordt het moeilijk erop te wijzen dat cijfers telkens een mathematische weergave zijn van een bepaalde kwalitatieve werkelijkheid. Anders gezegd, je zet een cijfer op een leerprestatie. Maar hoe kom je aan dat cijfer? Waarom staan er zes punten op een vraag en geen tien? Waarom staat er op iedere vraag één punt, ook al zijn de

bevraagde inhouden niet gelijk van waarde? Waarom kies je een bepaalde vraag en geen andere om een doel te meten?

Belangrijk is dat de aannames waarop de statistiek gebaseerd is, telkens kwalitatief van aard zijn, vaak ideologisch gebonden en dus ook voor discussie vatbaar (cf discussies over eindtermen). Menselijk gedrag en dus ook leerlingenprestaties zijn niet te vatten in natuurwetenschappelijke schalen, zoals bijvoorbeeld graden, meters, frequenties of voltages. Er is geen kookpunt voor begrijpend lezen en geen vriespunt voor algebra. Je kan wel zien of de leerlingen de 'participe passé' beheersen of dat ze een vergelijking van de tweede graad kunnen oplossen. Maar hoeveel punten is die prestatie dan waard? Dat blijft een waardeoordeel. Wat je wel kan zien is of de leerlingen de eindtermen al dan niet bereikt hebben (bijvoorbeeld door dat ze de daartoe bestemde oefeningen hebben opgelost). Aan die prestatie een cijferwaarde toekennen is een bepaalde keuze, al dan niet volgens een bepaald algoritme. De statistische verwerking maakt vaak indruk door de uitgebreide en nogal eens ingewikkelde cijferacrobatie op de verkregen punten. Maar die punten verbergen de ijsberg van aannames onder de zeespiegel. De punten zijn zoals bij een schaduwspel schaduwprojecties van de werkelijkheid op een scherm. Hoe paradoxaal het ook klinkt: wanneer je cijfers zet op een bepaald gedrag of op leerprestaties maak je een keuze die per definitie niet objectief is.

De populariteit van toetsen

Het voorbije decennium is het begrip kwaliteit van het onderwijs meer en meer doorspekt geworden met de taal van de meetbaarheid. Maar, zoals gesteld, menswetenschappen volgen niet de exacte logica van schalen uit de natuurwetenschappelijke wereld. Punten zijn niet vergelijkbaar met schalen die een vast beginpunt en eindpunt hebben. Cijfers waren in de oude Latijnse scholen (de voorgangers van onze humaniora) niet aan de orde. Je moest aan bepaalde normen voldoen en dat tonen om geslaagd te zijn of een onderscheiding te halen. Het op cijfers zetten van leergedrag kwam later en volgde de logica van de natuurwetenschappelijke en technologische evolutie ook al is die niet toepasbaar op menselijk gedrag. Het is wat gissen over het ontstaan van cijfers in het onderwijs. De industriële logica van het meten van arbeidsprocessen via het Taylorisme in het begin van de twintigste eeuw, was zeker een gangmaker. Voordien, al in de negentiende eeuw, kwamen de eerste experimentele psychologen voor de dag met op de natuurwetenschappen gebaseerde metingen van mentaal gedrag. Die beweging culmineerde in de selectiepsychologie met intelligentietests en schoolvorderingentests, telkens met de aanname dat ze dekkend waren voor de bedoelde mentale activiteit. Dat kon echter niet zonder die activiteiten van een getal of cijfer te voorzien, zodat er allerlei statistische berekeningen konden mee worden gemaakt. Die stap van het meten of vaststellen van een gedrag (zoals gebruikelijk in het oude onderwijs) naar een cijfermatige waardetoekenning is in feite een kritisch punt in de evolutie van het evalueren op school. In de verdere evolutie van testmethodes en toetstechnologie wordt systematisch verdoezeld dat die stap van het waarderen met een getal in feite een subjectieve ingreep is en dus een beperkte of vervormde voorstelling is van een vaststaand en waarneembaar gedrag of leerresultaat. Zo kom je bijvoorbeeld tot de paradoxale en algemeen aanvaarde uitspraak: intelligentie is wat de intelligentietest meet.

Cijfers hebben het voordeel dat ze de werkelijkheid eenvoudig maken. Ze geven een gevoel van veiligheid omdat je dan je brein niet moet vermoeien met de vraag te stellen waarop die cijfers slaan en waarom ze die waarde kregen. Vaak wordt aangenomen dat iedere vraag op een toets één punt waard is. Zo krijg je uiteraard vanuit kennisfilosofisch standpunt een

enorme verarming van de realiteit. Iets simplistisch gesteld: de guldensporenslag is evenveel waard als de Franse revolutie. Of: de wet van Ohm is evenveel waard als de fotosynthese.

Die evolutie heeft zo geleid tot een bijna blind geloof in de exacte meetbaarheid van leerprestaties.

Links en rechts verenigd

Die evolutie naar toetsbaarheid wordt gevoed door twee ideologische stromingen.

De eerste is de stroming van het afrekenen (accountability), waarbij je leraren en scholen kan afrekenen op de behaalde resultaten. Daarbij is het vergelijken tussen scholen erg aantrekkelijk, zodat je via de verplichte toetsen ook de goede van de minder goede scholen kan onderscheiden. Die ideologie past perfect in een marktvisie op het onderwijs. Ze wordt ook vaak verdedigd door economen, die het gemakkelijk hebben om met de meetbare indicator 'geld' allerlei winst en verliesrekeningen te maken.

De tweede ideologische grondslag is in wezen emancipatorisch. Door met exacte cijfers te werken, zie je meteen waar kansarme leerlingen niet voldoende aan hun trekken komen. Op die wijze kan je dan druk uitoefenen op leraren en scholen om die resultaten te verhogen.

Paradoxaal zie je op die wijze dat zowel links als rechts zich vinden in een toetsencultus met allerlei cijfers in tabellen, taartdiagrammen, gemiddeldes, standaarddeviaties, correlaties en noem verder maar op.

Cijfers zijn aantrekkelijk en verleidelijk. Een complexe realiteit van een leerlingprestatie wordt in een cijfer gegoten, waardoor de realiteit ogenschijnlijk maximaal transparant wordt. Cijfers worden populair, zijn de eenvoud zelf en laten maximale communicatie in allerlei media toe.

Cijfers kunnen begeleiden, leiden, verleiden en misleiden.

Cijfermatige toetsen kunnen een nuttig hulpmiddel zijn, maar ze mogen niet de doelstellingen worden. Scholen en leraren moeten niet aan de leiband lopen van de toetsenmakers. Zoals in de aanhef van deze blog bleek, kunnen cijfers een nuttige hulp zijn voor een (eerder beperkte) vorm van feedback voor leerlingen, maar ook voor de leraar. Ze hebben dan een begeleidende en pedagogische functie. Centrale toetsen die via steekproeven van scholen worden afgenomen kunnen ook een begeleidende rol spelen. Ze doen nadenken over de bekomen resultaten, maar dienen niet om scholen te vergelijken. Dat was de bedoeling van de peilingsproeven, die nu echter afgeschaft worden.

Verplichte toetsen hebben daarnaast een leidende functie. Ze vervangen het curriculum door toetsen, zodat de doelen van dat curriculum zodanig worden geoperationaliseerd, dat slechts één variante ervan geldig is. Verplichte toetsen verslinden als het ware de doelstellingen en vormen op die manier het enige valabele curriculum.

Ze zijn bovendien verleidend omdat ze toelaten scholen, klassen leerlingen te vergelijken en op die wijze rangschikkingen op te stellen. Want wie wil er nu niet hoog staan in een rangschikking? Een verarming van het curriculum ligt dan voor de hand en door de druk op de leraren zal die toetsversie dan ook het curriculum vormen. Bovendien zullen dan de resultaten ook stijgen. Maar die stijging is dan congruent met het verarmde curriculum. De leergebieden of onderdelen die niet op die manier worden getoetst, zullen daar onder lijden.

Ten slotte kunnen verplichte cijfermatige toetsen ook misleiden wanneer ze gebruikt worden als legitimering bij ideologisch gekleurde opvattingen over wat kwaliteit is. Dan is de weg open voor influencers en in bepaalde media opgevoerde 'experten', die vaak door bepaalde organisaties of partijen worden ingeschakeld om de media te beroeren.

Doemdenken is dan nooit ver af.

Cijfergeletterdheid gewenst

De school moet best geen toetsinstituut worden met ingewikkelde berekeningen over prestaties die over elkaar buitelen. Die zien er telkens indrukwekkend uit, maar ze verbergen een heel palet aan opeenvolgende aannames. Een grondige reflectie over de kennisfilosofische waarde van meetbaarheid van menselijk gedrag dringt zich op. Leer- en onderwijsprocessen zijn per definitie interactief en contextueel van aard en vragen een flexibele en vaak onvoorspelbare aanpak.

Het is vaak bevreemdend om te zien hoe cijfers of toetsresultaten de ronde doen in de media. De opstellers van de toetsen krijgen steeds een breed forum, maar worden zelden gecounterd door kritische stemmen over de relativiteit van cijfers, de methodologische gebreken en de aannames die vaak verborgen worden onder een statistisch jargon, dat er voor leken erg exact uitziet. Het is daarom belangrijk dat leraren, schoolbesturen en zelfs politici enig inzicht krijgen in cijfergeletterdheid, toegepast op het onderwijs. Het gaat dan op de eerste plaats over statistische basiskennis maar ook over inzicht in de psychologische, commerciële, economische en cultureel bepaalde mechanismen die cijfers versluieren of verdraaien.

Discussie

Voor wie er na lezing van deze blog zin in heeft, volgt een aanzet voor een bezinning over iets dat op cijfergeletterdheid lijkt. Het zijn stellingen en daar kan uiteraard over worden gediscussieerd.

Stellingen in verband met cijfergeletterdheid

-We moeten aanvaarden dat we in de menswetenschappen niet kunnen meten zoals in de natuurwetenschappen.

- We geloven zelfverzekerde experts op hun woord, zonder echt inzicht te hebben in hun argumentaties.
- Een onderwijsexpert is die naam pas waard wanneer hij/zij zelf ooit voor de klas heeft gestaan.
- Als een expert spreekt, moet je niet je hersens op non-actief instellen.
- Als we onzeker zijn, zoeken we zekerheid bij experts waarvan we veronderstellen dat hun zekerheden gepaard gaan met dito competentie.
- Cijfers moet je altijd bekijken in een bepaalde context vooraleer je reageert.
- Als er in statistisch materiaal schokeffecten en extreme uitschieters voor komen, is wantrouwen gewettigd.
- Je kan ook zoeken naar mogelijke cijfers die niet gegeven zijn.
- In de veelheid van cijfers en statistisch gegevens moet je nutteloze cijfers kunnen elimineren omdat ze de cijfers die ertoe doen ondersneeuwen.
- Als assenstelsels worden gebruikt voor diagrammen of overzichten moet de y-as met nul beginnen om echte effecten te kunnen zien.
- Als de feiten veranderen stel dan je vertrouwde meningen in vraag.
- Breedsprakerigheid en ingewikkeld jargon imponeren mensen en wekken de indruk betrouwbaar te zijn.
- Resultaten bij grote onderzoeken en steekproeven moet je bekijken vanuit absoluut gestelde normen en niet op een vergelijking via een normaalverdeling.
- Heel wat cijfers en a fortiori hun beoordeling zijn ofwel uit een context gerukt, zwak ondersteund, nep ofwel gepubliceerd tegen betaling.
- Bij het interpreteren van cijfers is het aan te bevelen verschillende gezichtspunten met verschillende achtergronden aan bod te laten komen.
- Als er geklaagd wordt over daling van de kwaliteit, vraag dan over welke kwaliteit het gaat. DE kwaliteit bestaat immers niet.