Eindtoets en schoolevaluatie

Samenvatting

*In dit statement ga ik uit van een principiële scheiding tussen de eindtoets als een proxy-beoordeling van de resultaten van een leerling (met het oog op oriëntering) en het evalueren van een school. Het verband tussen school en leerlingresultaten is zeer moeilijk aantoonbaar en betwistbaar. Toetsen kunnen een waardevolle bijdrage leveren als ze gebaseerd zijn op gelegitimeerde minimumdoelen en als ze een instrument zijn in de handen van het lerarenteam.*

Intro

Het is best de eindtoets en de evaluatie van een school van elkaar te onderscheiden. De eindtoets basisonderwijs is een evaluatie van leerlingresultaten. De resultaten op de eindtoets kunnen ook belang hebben als een robotfoto voor het macrobeleid. Het evalueren van een school (het mesoniveau) is van een andere orde. Leerlingresultaten zijn afhankelijk van hoofdzakelijk leerlinggerichte factoren: intelligentie, motivatie, sociaal-economische status, thuistaal, gezinskenmerken… Er is ook een impact van de leraar en wellicht ook van de organisatie waarin die leraar en leerling zich bevinden. Het bepalen van de grootte van die impact is een erg moeilijke opdracht. Het verband tussen de resultaten op de eindtoets en de kwaliteit van een school is dus betwistbaar.

Het evalueren van een school is het evalueren van een lerende organisatie. Een school wordt gekenmerkt door een groot aantal contextuele kenmerken. Daarvan zijn een aantal kenmerken onveranderlijk: ruraal of stedelijk gebied, etnische samenstelling, overwicht van kansrijke of kansarme leerlingen, gebouwen, juridische aspecten. Andere contextvariabelen zijn dan wel veranderlijk: de directie, de relatie tussen collega’s, het leiderschapsmodel, de inbreng van de ouders, het niveau van de leraren, het nascholingsbeleid, de samenstelling van de klassen, de didactiek, de gebruikte methodes enzovoort. In feite een complex geheel van factoren, die bovendien nog eens met elkaar kruisen. Ik ga in op de twee lijnen: de eindtoets en de schoolevaluatie.

De eindtoets

De eindtoets streeft ernaar een goed beeld te geven van de eindprestaties van een kind op het einde van groep acht. Het is een objectieve toets, die voor iedereen in de school dezelfde is. Dat is een voordeel. Maar bij nader toezien moet je de volgende vragen stellen:

-Wat meet die toets precies? Hoe bepaal je wat er wordt gemeten?

-Wat meet hij niet?

-Wie bepaalt wat er wordt gemeten?

-Waar ligt de grens tussen voldoende en onvoldoende?

-Wie bepaalt die grens?

-Hoever kan je gaan met meten?

Als je over deze vragen verder nadenkt, komen er heel wat relativeringen van een dergelijke eindtoets naar boven.

-Het is onduidelijk wat kinderen op het einde van groep acht precies moeten beheersen. Tussendoelen volstaan daartoe niet. De oplossing van een toetsconstructie met een statistische klokcurve als scheidsrechter, is een gemakkelijkheidsoplossing. Je vermijdt daardoor discussies over wat al dan niet belangrijk is. Het wiskundige model van de klokcurve geeft je namelijk uitsluitsel over wie gemiddeld is en wie eronder of erboven zit, eventueel zelfs met decielen. De vragen worden zodanig bewerkt en ‘gemasseerd’ dat ze een spreiding in resultaten moeten weergeven. Je moet dus steeds leerlingen met elkaar vergelijken en dus heb je altijd middelmaten, zwakkeren en sterkeren. Het gevolg is dat kinderen die in die spreiding onder het gemiddelde zitten, daar vrijwel nooit boven geraken, want de sterke leerlingen gaan ook vooruit. Volgens een zo geconstrueerd evaluatiesysteem of leerlingvolgsysteem, creëer je de bestendige zwakke groepen. Dat is erg ongunstig voor het zelfbeeld van die leerlingen en voor hun motivatie om te leren. Die toetsconstructie, dus volgens de statistische norm, geeft een relatieve maatstaf op basis van de soort vragen (wie bepaalt die?) en de populatie waarop je ze toepast. Je kan bijvoorbeeld perfect een toets construeren, die psychometrisch geijkt is op de normaalcurve met alleen vragen over spraakkunst bijvoorbeeld, of over geschiedenis, of over eender wat. Inhoudelijk is er geen duidelijkheid over wat precies moet worden bereikt of alleszins moet worden nagestreefd. De eindtermen waarover De Groot het jaren geleden had, ontbreken. Tussendoelen en streefdoelen voor enkele vakken zijn te beperkt.

-Je botst automatisch op de grenzen van meetbaarheid bij de menswetenschappen. Er zijn heel wat doelstellingen in het onderwijs (creatief denken, beoordelend gedrag, attitudes), die zich niet lenen tot natuurwetenschappelijke exactheid.

-De alleenzaligmakende toets bestaat overigens niet. Naargelang van je toetsdesign kan je verschillende toetsen hebben voor het meten van dezelfde doelen. Dat is trouwens nu het geval, omdat er naast de Centrale Eindtoets PO ook andere toetsen worden gebruikt. Ze kunnen andere klemtonen leggen. En daar is niets mis mee, gezien de grote diversiteit in de samenleving. Maar welke klemtonen? De oplossing is niet een strikt equivaleren aan vooropgestelde items, maar wel een ruime set van concrete en door de overheid gelegitimeerde minimumdoelen. De minimumdoelen zijn dan concreet, maar laten toe dat er diverse manieren zijn om die te meten.

-De eindtoets blijft een momentopname. Vorming is echter een proces dat continu verloopt in een bestendige pedagogische relatie tussen leraar en leerling, aangevuld met interactie met ouders en collega’s.

-Correctieformules om de resultaten te duiden, rekening houdend met contextgegevens worden toegepast: men corrigeert zo voor de sociaal-economische status (SES), voor etniciteit, voor thuistaal, intelligentie, sekse… Maar telkens moet je daarvoor een cijferwaarde toekennen aan ieder aspect dat je corrigeert. Er zijn niet alleen discussies mogelijk over de wijze waarop je bijvoorbeeld de SES meet, maar ook, over de weging die je daaraan toekent. Als je die wegingen verandert in je formule, verandert uiteraard ook je resultaat. De internationale praktijk wijst uit dat scholen met veel kansarme leerlingen er meestal bekaaid afkomen. Scholen met kansrijke leerlingen hebben de voordelen van het motiverende thuismilieu en komen er dus beter uit, zelfs met allerlei correctieformules.

Natuurlijk hebben de toetsenmakers hun oor te luisteren gelegd bij wat er in de praktijk in de klassen gebeurt, maar eenduidigheid daarover is er niet.

Het is anderzijds op zichzelf een goede zaak dat scholen gebruik maken van een gestandaardiseerde toets, die aan goede psychometrische eisen voldoet. Dat vermijdt ‘bedrijfsblindheid’ en overdreven denken in termen van het eigen gelijk. Maar de relativeringen ten opzichte van de validiteit en de scope van die toets geven hem een beperkte plaats als een waardevol middel bij het beoordelen van de leerlingen. Het is aan de professionals, het lerarenteam dus, om de hele context en het doorgemaakte proces mee in te calculeren. De eindtoets is een onderdeel van die globale context. In die zin is de toets dienstverlenend als ‘second opinion’. Het lerarenteam zal dan ook de andere vormingsinhouden, die niet gestandaardiseerd worden getoetst of die niet technisch kunnen worden getoetst, in een advies verwerken.

Of die calculatie nu, al dan niet goed gebeurt, blijft hier even terzijde. Het gaat om een principieel standpunt. Vorming als geheel is niet te meten met alleen een toets die afsluitend en eenmalig is. Ze gebeurt dicht bij de leerling en zeker in het basisonderwijs (met extensie naar leerplichtonderwijs). Brede vorming onderstelt een pedagogische relatie, met een groot belang van interactie en een bevorderend affectief klimaat. Professionals moeten in staat zijn om een dergelijk klimaat te creëren. Als dat niet zo is, moet daarin sterk worden geïnvesteerd.

Vanuit dit principiële standpunt, zou het een goede zaak zou zijn indien er duidelijke

minimumdoelen zouden worden bepaald voor het einde van groep acht. Op die doelen kan dan de toetsconstructie worden gebouwd. Op basis van die minimumdoelen hoeft dan niet een indeling in zwakkere en betere leerlingen te worden gemaakt, zoals dat bij een statistische norm het geval is. Hoe meer leerlingen die minimumdoelen bereiken, des te beter is dat, en op die manier is ook succesbeleving voor alle leerlingen mogelijk. Het ervaren van succesbeleving is een noodzakelijke voorwaarde om jongeren te motiveren voor het positief doorlopen van hun schooltraject en uiteraard ook voor het zo geroemde levenslang leren. De indruk is aanwezig dat het geloof in de exacte meetbaarheid van menselijk gedrag (want dat zijn toetsen) in het Nederlandse onderwijsbestel te nadrukkelijk aanwezig is. En dit gebeurt ten nadele van een positief vertrouwen in de professionaliteit van lerarenteams.

Evaluatie van scholen

 Het verband tussen de resultaten op een centrale toets en de kwaliteit van een school is zeer moeilijk aan te tonen. Een school is namelijk een complexe organisatie, waarin een heleboel contexten aanwezig zijn. Die allemaal in één oordeel of één cijfer in rekening brengen, is in feite onbegonnen werk. Men tracht via meerniveauonderzoek de impact van de school te isoleren in verband met de prestaties op de toets. Maar ook hier geldt dat het gebeurt op basis van aannames. Allerlei onderzoekingen, waarbij schoolkenmerken in verband worden gebracht met (meetbare) schoolresultaten geven via correlaties geen causaliteit weer. En bij meerniveauonderzoek verschillen de percentages van verklaarde variantie in sterke mate, naargelang van het onderzoek en de genomen aannames.

Uit de talrijke onderzoeken hieromtrent blijkt wel één rode lijn: de school doet er gemiddeld gezien, veel minder toe dan men meestal denkt. Let wel, het gaat dan enkel om de klassieke meetbare aspecten als wiskunde, natuurwetenschappen en bepaalde meetbare taalaspecten. We weten niet of de school er veel toe doet bij de vorming van de andere niet- exact meetbare aspecten: creatief denken, waarden, burgerzin, de zogenaamde competenties van de 21ste eeuw…

Resultaten op de eindtoets worden dus voornamelijk door leerlingkenmerken veroorzaakt: intelligentie, voorkennis, inzet, gezinscultuur, SES, thuistaal, buurt, sekse… De aanpak van de leraar zal wel een rol spelen, maar hoe ga je die in rekening brengen, gezien de praktisch ontelbare contexten waarin er wordt les gegeven? Hetzelfde geldt nog meer voor een school als organisatie. Ook daar heb je een veelvoud van interveniërende contextfactoren. Vele onderzoekingen kennen het schema van : “onder die voorwaarden en binnen die context, krijg je de volgende resultaten”.

Wanneer men in een onderwijssysteem toch de school en/of de leraars in rekening wil brengen als causaal element, krijg je negatieve aspecten als ‘teaching-to-the-tests’, het verschralen van het curriculum naar wat meetbaar is, de opkomst van shadowonderwijs, en niet te vergeten een heleboel stress voor leerlingen, leraars, scholen en ouders. Lichtere en zwaardere vormen van corruptie zijn dan niet denkbeeldig, consequent met de wet van de econoom Goodhart: “ When a measure becomes a target, it ceases to be a good measure” .

Als anekdote vermeld ik hier de kritiek van de commissie onderwijs van het Britse Lagerhuis, die de inspectie verweet, dat haar oordelen niet verschilden van de rankings van scholen die via de centrale examens werden gepubliceerd. Als de inspectie geen meerwaarde zou kunnen bieden, zou men ze beter afschaffen, gezien het dubbele gebruik. Refererend aan een mogelijke situatie in Nederland, zou de inspectie zich in een dergelijke situatie bevinden, wanneer ze de schoolexamens als goed beoordeelt, wanneer ze aansluiten bij de resultaten op de centrale examens. Op die manier wordt variatie en eigenheid uitgesloten. Dat is niet evident in een democratische samenleving waarin verschillen er toe doen.

Deze illustratie maakt duidelijk dat de meerwaarde van een inspectie erin bestaat, een school door te lichten onder de vorm van een audit. Daarbij maakt ze gebruik van een raamwerk van een aantal indicatoren die een goede organisatie kenmerken. Dat raamwerk is best een raamwerk waar voldoende consensus over bestaat. In een audit komen de sterke en zwakke punten van een school naar voor. Sommige aspecten zijn achterhaalbaar via data; andere dan weer via observaties en gesprekken met participanten. De audit vormt een spiegel met aanbevelingen. Een dergelijk raamwerk volgt met zijn indicatoren een CIPO-model: het brengt de context, de input, de processen en de output samen in een genuanceerd oordeel. Daarbij zal blijken dat het vrij gemakkelijk is om zeer zwakke scholen en zeer sterke scholen te onderscheiden. Maar de grote massa van scholen zit daartussen met zowel goede aspecten als verbeterpunten.

Principieel zou ook hier kunnen gelden, dat naarmate een school beter is in beleidsvoerend en zelfevaluerend vermogen (inzake de overeengekomen indicatoren), de inspectie-intensiteit kan variëren. Ook hier geldt weer de bemerking: als veel scholen dat nog niet aankunnen, is dat geen wezenlijk argument om er niet in te investeren en om er prioriteit aan te geven in het onderwijsbeleid.

Roger Standaert

Ereprof. Comparatieve en Internationale Pedagogiek UGent