

Optimalisering van het gebruik van onafhankelijke toetsen in een leerling- en onderwijsvolgsysteem: 'strategische' bijdrage aan duurzame onderwijsontwikkeling ...

Gerrit Staphorsius, Cito – mei 2008

De Kwaliteitsnota PO, *Scholen voor morgen*, zet in op een duurzame verbetering van het onderwijs en concentreert zich op de opbrengsten van het reken- en taalonderwijs. De nota ziet twee belangrijke wegen om de opbrengsten van het onderwijs op scholen te verhogen. De eerste is: leren van praktijken die zich bewezen hebben, en de tweede: het gebruiken van systematisch verzamelde harde gegevens over de kwaliteit van het onderwijs.

In deze notitie doen we een voorstel dat aansluit bij deze tweede pijler van de aanpak die *Scholen voor morgen* voorstelt. We koppelen daarin het gebruik van onafhankelijke toetsen voor de verbetering van de kwaliteit van het onderwijs aan de verantwoording van die kwaliteit door de school.

Volgen van de vorderingen van individuele leerlingen

In het basisonderwijs gebruikt naar schatting meer dan 90% van de scholen een **leerling- en onderwijsvolgsysteem**. Zo'n systeem bestaat uit onafhankelijke toetsen die voldoen aan de eisen die aan goede meetinstrumenten gesteld worden. Als we kijken naar de functies van een leerling- en onderwijsvolgsysteem is dat hoge gebruikspercentage niet zo verrassend.

Een leerling- en onderwijsvolgsysteem (verder kortheidshalve ook lovs genoemd) wordt in de onderwijspraktijk gebruikt naast dagelijkse informele evaluatiemiddelen. Zulke middelen zijn bijvoorbeeld vragen of oefeningen die leerlingen maken na een instructie om na te gaan of ze datgene wat is uitgelegd wel of niet kennen, kunnen of begrepen hebben, maar bijvoorbeeld ook methodegebonden toetsen, die met hetzelfde doel na een aantal lessen worden ingezet. Zonder deze, direct op het leerproces inwerkende, informele middelen zou onderwijs geen onderwijs zijn, of in ieder geval geen goed onderwijs. Immers: een systematisch gebrek aan directe feedback zal op de leerresultaten een negatieve invloed hebben!

Naast informele middelen worden 'formele' toetsen van een lovs (verder spreken we ook van lovs-toetsen) gebruikt. Met deze toetsen volgt de school individuele leerlingen met grotere tussenpozen; afhankelijk van de getoetste vaardigheden eens per half jaar of eens per jaar. Bijvoorbeeld: in het Cito-LOVS voor het basisonderwijs wordt vanaf groep 3 elk half jaar een toets Rekenen-Wiskunde afgenomen, omdat in het onderwijs in relatief korte tijd veel verschillende onderwerpen aan de orde komen. Het toetsen van Begrijpend Lezen gebeurt vanaf groep 5 eenmaal per jaar omdat deze vaardigheid zich na groep 4 betrekkelijk langzaam ontwikkelt. Een nieuwe meting na een half jaar zou niet of nauwelijks nieuwe informatie opleveren.

De lovs-toetsen voor de verschillende vaardigheden vormen in principe een doorgaande lijn vanaf groep 1 tot en met groep 8. In principe, want niet voor alle vaardigheden start het onderwijs in groep 1. De laatste toets in een lovs voor het basisonderwijs wordt in groep 8 afgenomen. In het geval van het LOVS is dat de Eindtoets Basisonderwijs, beter bekend als de Citotoets¹.

De waarde van onafhankelijke lovs-toetsen zit in de vergelijkingsmogelijkheden die deze toetsen bieden. Een eerste type vergelijking is de vergelijking met het landelijk gemiddelde. In het onderzoek dat de constructie van deze toetsen ondersteunt, is vastgesteld wat landelijk gezien de prestaties van leerlingen in hetzelfde leerjaar zijn. Daardoor is bekend wat de gemiddelde prestatie is en kan dus bepaald worden welke leerlingen landelijk gezien tot de zwakst presterenden behoren en wel eens het risico zouden kunnen lopen om een grote leerachterstand op te lopen. Die vergelijking met landelijke prestaties is uiteraard alleen zinvol als de toetsen op dezelfde manier worden afgenomen als in het onderzoek. Met het oog daarop zijn er richtlijnen voor de afname: de leerkracht is, als deze een lovs-toets afneemt, in de eerste plaats onderzoeker. In deze standaardisering schuilt voor een belangrijk deel het onafhankelijke karakter van lovs-toetsen.

Een lovs levert ook longitudinale gegevens op. Zo maakt het LOVS het mogelijk de ontwikkeling van leerlingen te volgen, dat wil zeggen de prestaties van een leerling op de toetsen van moment tot moment met elkaar te vergelijken. Dat kan ondanks het feit dat de toetsen op verschillende momenten andere opgaven hebben. Door een kalibratieonderzoek zijn de scores op verschillende toetsen naar één schaal te vertalen (ongeveer zoals het mogelijk is prijzen in landen met verschillende muntsoorten te vergelijken door de bedragen in de eigen muntsoort om te zetten naar euro's). De waarde van deze longitudinale informatie is groot: eerder behaalde prestaties onderbouwen verwachtingen over latere.

¹ Als we LOVS gebruiken, spreken we over het leerling- en onderwijsvolgsysteem van het Cito, dat als merknaam LOVS heeft. Het LOVS wordt in groep 8 ook gebruikt door scholen die de Citotoets niet afnemen. Die gebruiken de LOVS-toetsen voor medio groep 8 als 'eindmeting'. Het gebruik van de Citotoets als eindmeting heeft een voordeel: er zijn grotere garanties voor onafhankelijkheid, omdat van de toets met het oog op de primaire functie elk jaar een splinternieuwe versie wordt gemaakt.

De informatie voorziet dus in een ontwikkelingsperspectief. Als de ontwikkeling van prestaties niet aan die verwachtingen blijkt te voldoen, gaat daarvan een sterk signaal uit dat wijst op de noodzaak van een nader onderzoek naar mogelijke oorzaken.

Een derde type vergelijking die veel informatie met een duidelijke signaalwerking kan opleveren, noemen we wel discrepantie- of profielanalyse: de vergelijking van de objectief vastgestelde prestaties van een leerling op verschillende onderdelen. Daarmee kan de leerkracht nagaan wat relatief sterke en zwakke 'kanten' van een leerling zijn. Natuurlijk kunnen daarbij gegevens over de prestaties op eerdere momenten (een goed lovs maakt het mogelijk jaren terug te kijken!) worden betrokken.

Bewaken van de kwaliteit van het onderwijs op school

Scholen kunnen met lovs-toetsen ook de kwaliteit van hun onderwijs volgen. Daarbij worden de prestaties van individuele leerlingen geaggregeerd naar schoolniveau: het gemiddelde van alle individuele leerlingen in bijvoorbeeld groep 6 op een lovs-toets Rekenen-Wiskunde wordt dan het 'Schoolgemiddelde Rekenen-Wiskunde Groep 6'.

Een database met daarin de toetsresultaten van Taal (onder meer technisch lezen, begrijpend lezen, spelling en woordenschat), Rekenen-Wiskunde en Studievaardigheden van verschillende 'cohorten' (in een bepaald jaar op een school in groep 1 groep gestarte leerlingen) maakt tal van relevante analyses mogelijk. Elk van de genoemde kennis- en vaardigheidsgebieden bestaat uit verschillende onderdelen. Over elk van deze onderdelen zijn gegevens opgenomen. In het LOVS zijn er bijvoorbeeld voor Taal toetsen beschikbaar voor vijf hoofdonderdelen, voor Rekenen-Wiskunde zijn dat er drie en voor Studievaardigheden vier. Een dergelijke database met toetsresultaten bevat voor het volgen van het onderwijs op school een schat aan gegevens. Daardoor is de informatiewaarde van zo'n database aanzienlijk. We zijn er van overtuigd dat er, als het om gebruik van deze gegevens voor de verbetering van de kwaliteit van het onderwijs gaat, nog veel te winnen valt.

Een schoolgemiddelde kan natuurlijk vergeleken worden met een landelijk gemiddelde. Van deze vergelijking kan een eerste signaal uitgaan (wat scoren we laag in vergelijking met andere scholen!). Voordat we verder spreken over het vergelijken van schoolgemiddelden, ontkomen we echter niet aan een kort intermezzo.

Ter bevrediging van de nieuwsgierigheid kan het voor een school aardig zijn om te weten of een schoolgemiddelde onder of boven een landelijk gemiddelde ligt – en meestal is men nieuwsgieriger als men vrijwel zeker weet dat men er ver boven zit! – maar voor het volgen van de kwaliteit van het onderwijs is de informatie die deze vergelijking oplevert niet zonder meer zinvol. Vanwege kenmerken die niet of nauwelijks door het onderwijs te beïnvloeden zijn, zijn niet alle leerlingen even eenvoudig onderwijsbaar en niet op alle scholen zijn de leerlingen 'gemiddeld' genomen in gelijke mate onderwijsbaar. En zelfs als dat laatste wel zo zou zijn, dan geldt nog dat verschillen in schoolgemiddelden verklaarbaar kunnen zijn door de spreiding om de 'gemiddelde onderwijsbaarheid'. In ons land zijn de verschillen tussen scholen, als het om de onderwijsbaarheid van hun leerlingen gaat, tamelijk groot en daarom moeten we schoolgemiddelden corrigeren voor verschillen in onderwijsbaarheid van de leerlingen. We raken hier een punt dat onder de aanduiding 'toegevoegde waarde van het onderwijs' de laatste jaren in de discussies rond de kwaliteit van het onderwijs nogal wat aandacht heeft gekregen. Haalt het onderwijs of haalt een school uit de leerlingen wat er in zit? En is dat wel genoeg als we letten op wat het had kunnen zijn? Op de vraag naar wat redelijkerwijs verwacht mag worden, is het antwoord: als we rekening houden met verschillen in onderwijsbaarheid, mogen we verwachten dat het schoolgemiddelde ongeveer overeenkomt met het landelijk gemiddelde. In het LOVS worden de correcties voor verschillen in onderwijsbaarheid van de leerlingen toegepast², zonder dat daarbij de 'begintoets' die in de discussies over 'de toegevoegde waarde' steeds weer opduikt, nodig is. We zien dat na correctie voor verschillen in onderwijsbaarheid de hoge schoolgemiddelden van scholen met gemakkelijk lerende leerlingen in het algemeen zakken in de richting van het landelijk gemiddelde (of daar zelfs onder komen) en dat gemiddelden van scholen met lastiger onderwijsbare leerlingen stijgen naar rond het gemiddelde (of zelfs naar hogere niveaus). Ook al wordt de spreiding om de landelijke schoolgemiddelden na correctie dus aanzienlijk kleiner, er blijven wel degelijk verschillen tussen scholen bestaan waarvan het niet uitgesloten is dat die verklaard kunnen worden door verschillen in de kwaliteit van het onderwijs. Voor meer informatie over de toegepaste correcties verwijzen we naar de internetsite van Cito (www.cito.nl).

Na dit intermezzo over de correctie voor verschillen tussen scholen op het punt van de onderwijsbaarheid van hun leerlingen, vervolgen we ons verhaal over het gebruik van

² In groep 7 en 8.

schoolgemiddelden met het oog op (de verbetering van) de kwaliteit van onderwijs op school. Zeer informatief zijn de vergelijkingen van de schoolgemiddelden van de verschillende vakken en van de onderdelen daarbinnen. Deze schoolgemiddeldenprofielen of dwarsdoorsneden maken het mogelijk na te gaan of er sterke, minder sterke en zwakke kanten aan het onderwijs zitten. Zo moet een school waarvan de leerlingen zeer goed presteren bij begrijpend lezen zich zorgen maken als het schoolgemiddelde van een of meerdere groepen op het gebied van spellen achterblijft. Zo'n constatering kan aanleiding zijn nog eens terug te kijken en een trendanalyse te maken: hoe zat dat een half jaar, een jaar, anderhalf jaar, twee jaar geleden met hetzelfde cohort en – ook zeer informatief – hoe zat het een jaar geleden met het cohort dat een jaar eerder op school startte? Dit soort analyses kan belangrijke vragen oproepen (Is er weinig aandacht voor het onderwerp in de leergang?; Is er door de leerkracht wel voldoende aandacht aan het onderwerp besteed? enzovoort) en zelfs zorgen voor een bruikbare hypothese over de mogelijke oorzaak. Het LOVS ondersteunt de school bij het maken van dit soort analyses met het Computerprogramma LOVS dat een database beheert en rapporten produceert.

Gebruik van een lovs ook voor de verantwoording van de kwaliteit van het onderwijs aan de hand van onderwijsopbrengsten

Een goede evaluatiepraktijk mag dan een belangrijk aspect van de kwaliteit van het onderwijs zijn, op de opbrengsten van een school zijn ook andere factoren van invloed. Verantwoording van de kwaliteit aan de hand van onderwijsopbrengsten door scholen zal alleen daarom al nodig blijven. En dan komt het toezicht om de hoek kijken!

Kunnen scholen de kwaliteit van hun onderwijs wel verantwoorden met de door een lovs bepaalde onderwijsopbrengsten, als het toezicht die opbrengsten mede gebruikt om de kwaliteit van het onderwijs op scholen te beoordelen? Er bestaat in de onderwijskundige wereld op z'n minst twijfel aan die mogelijkheid. Scholen zouden hun onderwijs eenzijdig gaan richten op de onderdelen die door het toezicht beoordeeld worden en zouden zelfs bereid zijn tot frauduleuze manipulaties. Daardoor zou de primaire functie van een lovs, het volgen van de vorderingen van leerlingen en van de kwaliteit van het onderwijs, worden ondergraven. Men adviseert daarom wel een afzonderlijke, al dan niet verplichte opbrengsten-beoordelingstoets in te zetten (denk aan de recente discussie rond een verplichte eindtoets). Als we op dit punt zouden worden geraadpleegd, zouden we het volgende overwegen: (1) als een lovs nauw aansluit bij de inhoud van het onderwijs en de toetsen zijn adaptief (de moeilijkheid van de opgaven die een leerling te maken krijgt, kan aangepast worden aan het niveau van de leerling die de toets maakt) dan zouden leerkrachten weleens minder uitgedaagd kunnen worden over te gaan tot het fenomeen dat we tegenwoordig ook in Nederland 'teaching to the test' noemen; (2) we zijn niet naïef, maar ook niet defaitistisch. We nemen de mogelijke negatieve effecten op het gebruik van een lovs dat ook bij de verantwoording van de kwaliteit van het onderwijs een functie heeft, serieus. Maar we gaan er ook vanuit dat de professionele integriteit van leerkrachten in ons onderwijs groot is en dat er draagvlak zal zijn voor de beoordeling van de onderwijskwaliteit op basis van lovs-resultaten door het toezicht als die beoordeling transparant is en als leerkrachten die als fair ervaren; (3) het is de vraag of we wel buiten een multifunctioneel gebruik van een lovs kunnen. Tenminste als we aan de verantwoording van de kwaliteit via opbrengsten en een beoordeling daarvan willen vasthouden, want daar zit uiteraard de crux. De onderwijspraktijk zal een afzonderlijke opbrengsten-beoordelingstoets waarschijnlijk als overbodig en belastend ervaren. Bovendien sluiten we niet uit dat een (verplichte) afzonderlijke toets geen oplossing is voor de verwachte negatieve effecten van multifunctioneel gebruik van een lovs. In de eerste plaats vragen we ons af of door de inzet van een afzonderlijke opbrengsten-beoordelingstoets de verwachte negatieve effecten op de inhoud van het onderwijs en op het gedrag van leerkrachten zouden verdwijnen. In de tweede plaats valt te verwachten dat als 'frauduleuze' manipulatie van de uitslag inderdaad een effect zou zijn, ook van een dergelijke toets, dit onmiddellijk doorwerkt in het gebruik van welk lovs dan ook. Immers, door manipulatie veroorzaakte discrepanties tussen een afzonderlijke toets voor de bepaling en beoordeling van opbrengsten en een lovs zullen voor het bevoegd gezag duidelijk zichtbaar zijn.

We gaan er vooralsnog vanuit dat een lovs ook voor de verantwoording van de kwaliteit van het onderwijs aan de hand van opbrengsten kan worden ingezet. Het draagvlak onder en de instelling van leerkrachten zal erg afhangen van de transparantie van de beoordeling van de opbrengsten.

Strategische keuze: 'verplichting' tot Schoolrapport Evaluatiepraktijk (SEP)

We verwachten dat van een gerichte overdracht van kennis over de mogelijkheden (en beperkingen) van een lovs een positieve invloed uitgaat op het inzicht en de instelling van leerkrachten. Meer inzicht

zal leiden tot een gemakkelijker acceptatie van beoordelingen van de kwaliteit van het onderwijs en een grotere weerbaarheid tegen als niet rechtvaardig beschouwde oordelen.

De noodzaak tot professionalisering zal nog groter worden en een goede evaluatiepraktijk zal duurzaam in het onderwijs verankerd worden als het bevoegd gezag van de scholen jaarlijks een rapport verlangt. Zo'n Schoolrapport Evaluatiepraktijk (korthedshalve aangeduid met SEP) kan een onderdeel of een hoofdstuk zijn van al bestaande verplichtingen tot rapportage. Het rapport moet inzicht geven in het gebruik van 'harde' gegevens in de evaluatiepraktijk op school.

Het toezicht moet zich op basis van het SEP een indruk van het systematisch gebruik van onafhankelijke toetsen in groep 1 tot en met 8 kunnen vormen. De laatste toets van een lovs in groep 8 kan beschouwd worden als 'eindtoets'³. Die hoeft niet verplicht gesteld te worden, want de toets heeft zijn vanzelfsprekende plaats in een lovs dat op zich weer een even natuurlijk onderdeel van de evaluatiepraktijk is.

De verplichting tot het jaarlijks leveren van een dergelijk rapport zal naar onze overtuiging stimulerend werken op een door kennis en inzicht geschraagd gebruik van een lovs. Cito merkt dat de cursussen die gericht zijn op een verdieping van het gebruik van LOVS-gegevens zeer gewaardeerd worden, doordat de cursisten na de training meer grip op de interpretatie en het gebruik van die gegevens hebben. We werken op het ogenblik onder de werktitel 'Evaluatiespecialist' intensief aan cursussen die leerkrachten op het punt van het gebruik van betrouwbare, harde LOVS-gegevens verder professionaliseren. We overwegen leerkrachten die deze cursus met succes gevolgd hebben te certificeren. Een onderdeel van de afsluiting van de cursus zou een praktijkopdracht kunnen zijn: het schrijven (en verantwoorden) van een SEP.

Suggesties voor de inhoud van het SEP

De inhoud van een Schoolrapport Evaluatiepraktijk in het primair onderwijs zou in goed overleg met onder meer de Inspectie van het Onderwijs, de PO Raad en het PKI kunnen worden gespecificeerd.

Om de gedachten te bepalen doen we een aantal suggesties:

- beschrijving van de evaluatiepraktijk voorzover het gaat om de inzet van onafhankelijke toetsen, inclusief een overzicht van gebruikte instrumenten⁴: hoe is er in de organisatie van de school voor garanties gezorgd dat toetsen inderdaad worden afgenomen en dat de resultaten worden gebruikt in het onderwijsproces? (Wie hebben daarbij een rol? Is er periodiek overleg? Wie bewaakt genomen besluiten? enzovoort)
- verslag van opbrengsten. Gedacht kan worden aan onder meer:
 - per leerjaar, per vak en per vakonderdeel de gemiddelde scores en longitudinale gegevens (trendanalyses) en indien referentieniveaus beschikbaar zijn: per leerjaar, per vak het aantal of het percentage leerlingen dat deze referentieniveaus bereikt heeft. Ook kan gedacht worden aan de rapportage van het aantal of het percentage leerlingen dat op basis van vorige afnames het geraamde niveau niet gehaald heeft;
 - per vak 'evidence based' conclusies over het onderwijsniveau en een overzicht van besluiten ter handhaving of verbetering van dat niveau en, indien van toepassing, een planning die aangeeft wanneer de genomen besluiten tot verbeteringen gerealiseerd moeten zijn;
 - per leerjaar een opgave van het aantal of percentage leerlingen dat op basis van de analyse van transversale en longitudinale gegevens in aanmerking komt voor een nader onderzoek en het aantal of percentage leerlingen dat inderdaad nader onderzocht is.

Als het door een school gebruikte lovs ondersteund wordt door een computerprogramma kan dat zo worden aangepast dat het alle benodigde gegevens genereert.

Rol toezicht

Het toezicht zou, als het om de beoordeling van de kwaliteit van het onderwijs op een school gaat, primair kunnen kijken naar het adequate gebruik van lovs-gegevens voor de sturing van leer- en onderwijsprocessen. Bij het periodiek inspectiebezoek aan de school zou de directie het SEP kunnen

³ De Citotoets is een mogelijke afsluiting van het LOVS, zowel voor het volgen van leerlingen als van het onderwijs (ook de LOVS-toetsen voor medio groep 8 kunnen worden gebruikt door scholen die voor de schoolkeuzeadviesing andere instrumenten dan de Citotoets gebruiken). Aan de Citotoets kunnen alle leerlingen deelnemen (voor achterstandsleerlingen en bijvoorbeeld voor kinderen met een visuele handicap zijn speciale versies van de toets). In de Citotoets is een stevige garantie ingebouwd voor de onafhankelijkheid: de toets is elk jaar compleet nieuw!

⁴ De aan de instrumenten door het bevoegd gezag te stellen eisen zouden betrekking kunnen hebben op de te toetsen 'vakken', vaardigheden of ontwikkelingsgebieden (bijvoorbeeld cognitieve en sociaal-emotionele ontwikkeling), de frequentie van toetsing, de onafhankelijkheid, interpretatiemogelijkheden (bijvoorbeeld mogelijkheid tot omzetting van scores naar referentieniveaus en correctie voor verschillen in onderwijsbaarheid).

verantwoorden. We merkten hierboven al op dat behalve processen ook opbrengsten bij beoordeling van de kwaliteit een rol moeten spelen. We kunnen ons voorstellen dat de inspectie via de onderwijsopbrengsten in groep 8 de kwaliteit van het onderwijs op een school monitort. We sluiten zeker niet uit dat we scholen die de Citotoets gebruiken aan kwaliteitsindicatoren kunnen helpen, die door het toezicht kunnen worden gebruikt ter detectie van scholen waar de kwaliteit dreigt achter te blijven.

In de discussie rond de beoordeling van de kwaliteit wordt door scholen niet zelden naar voren gebracht dat het 'afrekenen' op alleen cognitieve vaardigheden niet rechtvaardig is. Het pedagogisch klimaat en de mogelijke invloed daarvan op de sociaal-emotionele ontwikkeling van kinderen zou in de beoordeling betrokken moeten worden. Los van de vraag of dat nu al niet gebeurt, zouden we daarover het volgende willen opmerken. Scholen die dit punt inbrengen lijken te willen zeggen: 'Kijk, omdat we andere accenten leggen, blijven we op het gebied van rekenen en taal een stukje achter, maar onze leerlingen 'scoren' wel hoog op zelfstandigheid, creativiteit, sociaal gedrag, assertiviteit enzovoort.' Voor het draagvlak zou het daarom goed kunnen zijn in de beoordeling van onderwijsopbrengsten ook de sociaal-emotionele ontwikkeling te betrekken. Blijft de vraag of een min bij de 'cognitieve' opbrengsten kan worden weggestreept tegen een 'sociaal-emotionele' plus. Misschien nog wel meer dan bij de cognitieve ontwikkeling lopen we bij de sociaal-emotionele ontwikkeling aan tegen eigenschappen van leerlingen die hun ontwikkeling bevorderen of belemmeren zonder dat de school die eigenschappen eenvoudig kan beïnvloeden. Dat neemt niet weg dat het belangrijk is dat de school de sociaal-emotionele ontwikkeling volgt, ook omdat die ontwikkeling van invloed is op de vorderingen bij bijvoorbeeld rekenen en taal.

Tot slot: strategische keuze met het oog op duurzame onderwijsontwikkeling

Wij denken dat, als er met het oog op de opbrengsten van het onderwijs iets verplicht moet worden, het de verplichting tot het opleveren van een SEP moet zijn, een rapport aan de hand waarvan de school zijn evaluatiepraktijk kan verantwoorden. We noemen de keuze voor de verplichtstelling van een dergelijk rapport 'strategisch' omdat zo wat vanzelfsprekend is nog vanzelfsprekender wordt. Er kan een stimulans van uitgaan tot professionalisering van het gebruik van 'harde' gegevens in de evaluatiepraktijk. Op den duur zal de verplichting tot het SEP kunnen leiden tot een optimalisering van de opbrengsten van het basisonderwijs.