

Onderwijs op peil?

Een samenvattend overzicht van 20 jaar PPON

Frank van der Schoot
Projectleider PPON

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2008)

Alle rechten voorbehouden. Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Inhoud

1 Inleiding	5
Mijlpaal wordt peilstok	5
De PPON-geschiedenis in vogelvlucht	5
Een overzicht van de peilingonderzoeken	6
Doel van het peilingonderzoek	7
Het beschrijven van de vaardigheden van leerlingen	9
Doelgroepen voor peilingresultaten	13
2 Overzicht van peilingresultaten voor Nederlandse taal	15
3 Overzicht van peilingresultaten voor rekenen/wiskunde	19
4 Overzicht van peilingresultaten voor wereldoriëntatie,	23
Geschiedenis	23
Aardrijkskunde	25
Natuuronderwijs	25
5 Overzicht van peilingresultaten voor Engels in het basisonderwijs	27
6 Conclusie, terugblik en vooruitblik	29
Literatuur	33
Overzicht van PPON-onderzoeksrapporten	35

1 Inleiding

Mijlpaal wordt peilstok | Ruim twintig jaar PPON, Periodiek Peilingonderzoek van het Onderwijsniveau in Nederland. We kunnen gerust spreken van een mijlpaal voor dit project van Cito. Een goed moment dus voor een evaluatie. Wat is er allemaal gebeurd? Wat is er terechtgekomen van de uitgangspunten? Wat hebben onze inspanningen opgeleverd aan inzichten over de ontwikkelingen van kwaliteit? En hoe is de stand van zaken anno nu? In deze publicatie peilen we in zekere zin het peilen. Mijlpaal wordt peilstok. We beginnen met een stukje voorgeschiedenis en een chronologisch overzicht van belangrijke momenten. Dan volgt een overzicht van de peilingonderzoeken die we in de loop der jaren hebben uitgevoerd. Vervolgens belichten we het specifieke doel van ons peilingonderzoek en geven deze een plaats naast andere grootschalige evaluatieprojecten. Daarna volgt een overzicht van de resultaten van de peilingonderzoeken, gebaseerd op standaarden en jaarvergelijkingen. Hierbij beperken we ons tot jaargroep 8 voor de leer- en vormingsgebieden Nederlandse taal, rekenen/wiskunde, wereldoriëntatie en Engels: onderwerpen waaraan we in de verschillende peilingcycli structureel aandacht hebben besteed. We sluiten af met een samenvattende conclusie, een terugblik en een kijkje vooruit.

De PPON-geschiedenis in vogelvlucht | Het gestructureerde peilingonderzoek is eigenlijk begonnen met een studiereis in 1979. Toen gingen de Cito-medewerkers Van Dam en Wijnstra naar de Verenigde Staten om daar kennis te nemen van grootschalige onderwijs-evaluatieprojecten (Wijnstra & Van Dam, 1980; Wijnstra, 1982). In hun verslag deden zij het voorstel om een dergelijk evaluatieproject ook in Nederland van de grond te tillen. Die gedachte is geleidelijk verder ontwikkeld en in 1984-1985 heeft het toenmalige SCO onderzocht of zo'n project ook voor het Nederlandse basisonderwijs zinvol zou zijn (Wesdorp et al., 1986). De conclusie was bevestigend. De rest is geschiedenis die globaal in de volgende jaartallen gestalte kreeg.

1986 | Het Ministerie van Onderwijs en Wetenschappen verleent Cito de opdracht peilingonderzoeken te verrichten. Dit betreft in principe alle leer- en vormingsgebieden aan het einde van het basisonderwijs én – voor de gebieden Nederlandse taal en rekenen/wiskunde – ook halverwege het basisonderwijs aan het begin van jaargroep 5. Cito start het project PPON.

1987 | Het eerste peilingonderzoek wordt uitgevoerd in de maanden mei/juni voor rekenen/wiskunde in jaargroep 8 van het basisonderwijs. Voortaan voert PPON jaarlijks peilingonderzoeken uit voor wisselende leer- en vormingsgebieden.

1990 | We krijgen opdracht de peilingonderzoeken voor de vakken Nederlandse taal en rekenen/wiskunde ook uit te voeren in het speciaal basisonderwijs (toen LOM en MLK).

Dit betreft dan de leerlingen die even oud zijn als de leerlingen in jaargroep 8 van het reguliere basisonderwijs.

1993 | De eindtermen worden kerndoelen (Ministerie van OCW, 1993). Wij krijgen het verzoek om de resultaten van onze peilingonderzoeken te evalueren mede in het licht van de kerndoelen basisonderwijs.

1994 | De Commissie Evaluatie Basisonderwijs presenteert haar rapport over de ontwikkelingen in het basisonderwijs sinds 1985 toen de nieuwe wet op het basisonderwijs van kracht werd. De commissie maakt ter onderbouwing van haar conclusies uitgebreid gebruik van onze peilingresultaten (CEB, 1994).

1995 | Naar aanleiding van het CEB-rapport beginnen we met zogenaamde standaardonderzoeken om de uitkomsten van ons onderzoek te kunnen relateren aan de kerndoelen basisonderwijs.

1998 | De eerste herziening van de kerndoelen basisonderwijs ziet het licht (Ministerie van OCW, 1998). Op verzoek beperken wij het peilingonderzoek voor jaargroep 8 tot de leer- en vormingsgebieden Nederlandse taal, rekenen/wiskunde en wereldoriëntatie.

1999 | Peilingonderzoeken voor Nederlandse taal en wereldoriëntatie zijn zo veelomvattend dat we in overleg besluiten deze gebieden op te splitsen. Voor Nederlandse taal peilen we voortaan de vier hoofdvaardigheden – lezen, schrijven, luisteren en spreken – afzonderlijk en voor wereldoriëntatie de deelgebieden geschiedenis, aardrijkskunde en natuuronderwijs.

1999 | De Onderwijsraad presenteert *Zeker weten. Leerstandaarden als basis voor toegankelijkheid*. Dit advies is mede geïnspireerd door het standaardonderzoek. We leveren een bijdrage met werkdocumenten over leerstandaarden voor Nederlandse taal en rekenen/wiskunde.

2004 | Het project krijgt het verzoek opnieuw peilingonderzoek uit te voeren voor Engels en bewegingsonderwijs in jaargroep 8.

2005 | De presentatie van de tweede herziening van de kerndoelen basisonderwijs is een feit. Deze zal vanaf 2009 richtinggevend zijn en ook ons werk beïnvloeden.

Een overzicht van de peilingonderzoeken | Houd om de vijf jaar peilingonderzoeken voor de leerstofgebieden in het basisonderwijs. Dat was de opdracht waarmee we van start gingen. Inmiddels is deze opzet enigszins veranderd. Niet meer alle leerstofdomeinen zijn vertegenwoordigd en de cycli beslaan ook niet meer strikt vijf jaar. In onderstaande overzichtstabel met alle peilingonderzoeken wordt dat duidelijk. We onderscheiden daarin vier perioden van peilingonderzoeken, met als uitgangspunt de periodieke herhalingen van de peilingonderzoeken voor Nederlandse taal en rekenen/wiskunde.

Overzicht van door PPOON uitgevoerde peilingonderzoeken

		1e periode 1987-1991	2e periode 1992-1996	3e periode 1997-2003	4e periode 2003-2007
Peilingen einde basisonderwijs	Nederlandse taal (algemeen)	1988	1993	1998*	
	- Schrijfvaardigheid en handschrift			1999	
	- Spreekvaardigheid			2002/03	
	- Leesvaardigheid				2005
	- Luistervaardigheid				2007
	- Rekenen/wiskunde	1987	1992	1997	2004
	- Wereldoriëntatie (algemeen)	1990/91	1995		
	- Geschiedenis			2000	
	- Aardrijkskunde			2001	
	- Biologie			2001	
	- Natuurkunde / Techniek			2002	
	Engels	1991	1996		2006
	Verkeer	1991		1997	2006
	Muziek	1988		1997	
Beeldende vorming	1988	1996			
Lichamelijke opvoeding	1988	1994		2006	
Peilingen medio basisonderwijs	Nederlandse taal (algemeen)	1989	1994	1999*	
	- Schrijfvaardigheid en handschrift			1999	
	- Spreekvaardigheid			2002/03	
	- Leesvaardigheid				2005
	- Luistervaardigheid				2007
	- Rekenen/wiskunde	1987	1992	1997	2003
Peilingen speciaal basisonderwijs	Nederlandse taal (algemeen)		1993	1999*	
	- Schrijfvaardigheid en handschrift			1999	
	- Spreekvaardigheid			2002/03	
	- Leesvaardigheid				2005
	- Luistervaardigheid				2007
- Rekenen/wiskunde		1992	1997	2004	
* exclusief schrijfvaardigheid en spreekvaardigheid					

Doel van het peilingonderzoek | De systematische en periodieke peilinginspanningen moeten leiden tot inzicht in het leeraanbod en in de leeropbrengsten van het basisonderwijs. Dat is in algemene bewoordingen het PPOON-doel, zoals omschreven in de overheidsnotitie van juni 1985, die het ministerie van Onderwijs en Wetenschappen ons meegaf bij de opdrachtverstrekking.

Meer concreet wilde de overheid toen dat onze onderzoeksresultaten:

1. de school een referentiekader bieden voor de beoordeling van de eigen doelstellingen en resultaten;
2. indicaties leveren voor het verrichten van nadere (diepte-)studies en evaluatieonderzoeken die van belang kunnen zijn voor de kwaliteitsbeoordeling en -bevordering van het onderwijs;

3. bijdragen aan de toetsontwikkeling, waardoor scholen zich kunnen spiegelen aan landelijke onderwijstrends;
4. een empirische basis verschaffen aan de algemeen maatschappelijke discussie over inhoud, kwaliteit en niveau van het onderwijs.

Voor de doelstellingen 1 en 3, de typisch school- en leerlinggerichte evaluatiedoelstellingen, is daarna al vrij snel een afzonderlijk leerlingvolgsysteem ontwikkeld voor met name de cruciale leergebieden Nederlandse taal en rekenen/wiskunde. Hierdoor kunnen leraren van iedere leerling de ontwikkeling van de prestaties in en over de leerjaren van het basisonderwijs volgen. Daarnaast kan de school op basis van de leerlingresultaten ook analyses uitvoeren voor groep- en schoolevaluatie. Op dit moment kunnen scholen dan ook beschikken over een uitgebreid en samenhangend geheel van evaluatie-instrumenten, bestaande uit toetsen uit het leerlingvolgsysteem, entreetoetsen voor de verschillende jaargroepen en de eindtoets basisonderwijs.

Van alle hiervoor geformuleerde doelstellingen is de vierde wel de belangrijkste van onze inspanningen het onderwijsaanbod en de leeropbrengsten van leerlingen in het basisonderwijs in beeld te brengen. Niet voor niets benadrukt Wijnstra (1988), dat in het PPON-ontwerp een ruime dekking van de leer- en vormingsgebieden voorop heeft gestaan om op basis van de resultaten een steekhoudende inhoudelijke discussie over het onderwijs te kunnen voeren.

Dat inzicht heeft geleid tot de ontwikkeling van een evaluatie-instrument met een modulaire opbouw. PPON streeft er naar de vaardigheden van leerlingen te beschrijven op basis van compacte, didactisch betekenisvolle onderwerpen of modules.

Het fundament van ieder peilingonderzoek is de domeinbeschrijving. Dat is een structurele beschrijving van het betreffende leergebied aan de hand van een geordende lijst van leer- en vormingsdoelen. Deze beschrijving is gebaseerd op de eindtermen respectievelijk de kerndoelen basisonderwijs, methodeanalyses en vakinhoudelijke en vakdidactische publicaties. De domeinbeschrijving vormt de basis voor de indeling in onderwerpen die voor een leergebied bij het peilingonderzoek wordt gehanteerd. De tabellen in de volgende hoofdstukken laten zien welke onderwerpen voor de verschillende leergebieden in opeenvolgende peilingen werden onderscheiden.

Het peilingonderzoek wordt dus niet uitgevoerd op basis van één breed samengestelde toets, maar op basis van een reeks van toetsen die zijn gerelateerd aan de diverse onderwerpen van een leergebied. Aldus kan vrij gedetailleerd worden beschreven waartoe leerlingen aan het einde van het basisonderwijs in staat zijn. De periodiciteit van het onderzoek maakt het vervolgens mogelijk gedetailleerd per onderwerp aan te geven welke ontwikkeling in de loop der jaren wordt doorgemaakt.

Samengevat richt het peilingonderzoek zich op het beantwoorden van vragen als:

- Wat proberen scholen hun leerlingen te leren?
- Wat steken de leerlingen van het onderwijs op?
- Welke wijzigingen voltrekken zich over de jaren in de onderwijsresultaten?

- In hoeverre worden de kerndoelen basisonderwijs gehaald?
- Welke verschillen zijn er in de leerresultaten tussen categorieën van leerlingen?

Het peilingonderzoek is een typische vorm van systeemevaluatie. We spreken ook wel van evaluatie op macroniveau, waarbij het object van evaluatie de kwaliteit van het onderwijs-systeem is. Er zijn meerdere vormen van systeemevaluatie binnen het onderwijs te onderscheiden. De evaluatie van de leeropbrengsten zoals PPOON dat uitvoert, heeft vooral betrekking op de inhoudelijke kwaliteit van de leerresultaten. Daarbij zoeken we antwoorden op vragen als ‘Wat leren de kinderen in het basisonderwijs?’ en ‘In hoeverre zijn de beoogde kerndoelen basisonderwijs gerealiseerd?’

Daarnaast kennen we de meer beleidsgerichte evaluatiestudies die zich vooral richten op de effecten op het onderwijs van overheidsrichtlijnen en de daaraan gekoppelde doelstellingen. Typische voorbeelden zijn: het onderzoek dat in het verleden is opgezet en uitgevoerd rond het onderwijsvoorrangbeleid, het Prima-cohort onderzoek en de onlangs gestarte cohort-studies in het kader van COOL 5-18. Leerlingprestaties op het gebied van Nederlandse taal en rekenen/wiskunde worden in deze studies voornamelijk gebruikt als indicatoren voor beleids-effecten.

Internationaal vergelijkend onderzoek, zoals PISA, PIRLS en TIMSS, is nog een derde vorm van systeemevaluatie. Dit onderzoek is erop gericht de leeropbrengsten van de deelnemende landen te vergelijken voor zover het gaat over gemeenschappelijke doelstellingen van het onderwijsprogramma.

Peilingonderzoek is dus niet gericht op leerling- en schoolevaluatie. Daardoor is het ook niet gebonden aan allerlei beperkingen die kleven aan dergelijke evaluaties. Zo kan nu ondermeer op basis van steekproeftrekking en matrixsampling een groot aantal evaluatie-instrumenten aan de leerlingen worden voorgelegd zonder dat dit hoeft te leiden tot een onaanvaardbare toetslast voor individuele leerlingen. Er is ook geen noodzaak tot directe feedback, iets wat bij leerling-evaluatie uiteraard van groot belang is. Hierdoor kunnen we ook evaluatievormen toepassen die meer tijdrovende beoordelingen en/of analyses noodzakelijk maken.

Het beschrijven van leeropbrengsten | In de PPOON-rapportages beschrijven we de vaardigheden van de leerlingen steeds aan de hand van voorbeeldopgaven.

De voorbeeldopgaven illustreren wat leerlingen op verschillende niveaus van vaardigheid goed, matig of in onvoldoende mate weten of kunnen. Geïnformeerde beoordelaars geven vervolgens aan wat naar hun oordeel het gewenste niveau van vaardigheid is in het licht van de kerndoelen basisonderwijs. Zij bepalen standaarden voor minimale en voldoende beheersing, die fungeren als referentieniveaus. De standaarden krijgen inhoudelijk betekenis doordat ze gerelateerd zijn aan de voorbeeldopgaven. Ten slotte wordt onderzocht hoe de leerresultaten van leerlingen zich in de loop van de tijd ontwikkelen.

De voorbeeldopgaven | Het primaire doel van peilingonderzoek is – als eerder gezegd – het beschrijven van de vaardigheden van leerlingen aan het einde van een bepaalde onderwijs-fase. De beschrijvingen zijn input voor een meer algemene discussie over de kwaliteit van de leeropbrengsten van het onderwijs, i.c. van het basisonderwijs. Maar, hoe beschrijf je de vaardigheden?

Voordat we daarop ingaan is het van belang vast te stellen dat peilingonderzoek zich bezighoudt met de vraag wat leerlingen weten over een bepaald onderwerp binnen de vastgestelde structuur van een leergebied. Doorgaans spreken we dan van 'de vaardigheid van de leerling' in relatie tot dat onderwerp, zonder verder onderscheid te maken in diverse niveaus van vaardigheid zoals kennis, inzicht of toepassing.

Vrijwel alle leerinhouden van de leergebieden – rekenen/wiskunde in het bijzonder – zijn tot op zekere hoogte hiërarchisch gestructureerd. Maar toch vaak onvoldoende om op basis daarvan te beschrijven in hoeverre de leerlingen bepaalde vaardigheden wel of niet beheersen. Daarom is de beschrijving van de vaardigheden van leerlingen vooral gebaseerd op voorbeelden van opgaven die aan de leerlingen zijn voorgelegd. De voorbeeldopgaven maken concreet welke vragen en opdrachten de leerlingen, gerelateerd aan hun vaardigheidsniveau, wel en niet kunnen beantwoorden of oplossen.

Voor rekenen/wiskunde kunnen we de vaardigheden van leerlingen op een iets hoger niveau beschrijven, omdat het mogelijk blijkt om – op basis van de uitgebreide opgavenverzamelingen per onderwerp – empirisch onderbouwde ontwikkelingslijnen op te stellen. Wanneer we de opgaven voor een onderwerp op moeilijkheidsgraad rangschikken, dan verkrijgen we een hiërarchische opbouw van rekeninhoudelijke opgavenkenmerken waardoor opgaven moeilijker worden. Dergelijke ontwikkelingslijnen vertonen uiteraard een sterke overeenkomst met de opbouw van de rekenvaardigheid in de gehanteerde onderwijsmethoden.

De opgavenverzamelingen zijn een uitwerking van de geformuleerde onderwijsinhouden uit de kerndoelen. Deze verzamelingen of overzichten zijn dus van essentieel belang voor het beschrijven van de leeropbrengsten van het onderwijs. Het gaat echter te ver om in deze beknopte publicatie dergelijke overzichten te presenteren. De geïnteresseerde lezer kan ze vinden in onze balansrapportages van de peilingonderzoeken. Op de PPO website publiceren we in de regel voor ieder vakgebied de twee meest recente rapporten integraal. In deze samenvattende publicatie over twintig jaar peilingonderzoek beschrijven we de leeropbrengsten op basis van de twee andere gegevens, namelijk op basis van de standaarden en op basis van de vergelijking over de tijd.

De standaarden als referentieniveaus | We beschrijven in de balansen de vaardigheden van leerlingen en illustreren die met voorbeeldopgaven. Maar daarmee zijn we er niet.

We willen ook graag weten of de leerlingen in kwalitatief opzicht het gewenste kennis- of vaardigheidsniveau bereiken. Met ander woorden: in hoeverre bereiken ze de doelen, die met het onderwijs worden nagestreefd? Peilingresultaten krijgen pas betekenis als we die vraag kunnen beantwoorden. Wat we peilen, willen we kunnen toetsen aan het gewenste niveau. Pas dan immers kunnen we aangeven of het onderwijs kwalitatief gezien voldoet. Daartoe staan ons de zogenoemde standaarden ter beschikking als referentieniveaus. Het standaardenonderzoek is ontwikkeld naar aanleiding van de rapportages door de Commissie Evaluatie Basisonderwijs (1994) en is verantwoord in Van der Schoot (2001). Voor een goed begrip beschrijven we in het kort de procedure die wordt gevolgd bij het vaststellen van de standaarden.

We moeten vaststellen dat een criterium waaraan de leeropbrengsten getoetst kunnen worden, ontbreekt. We nemen daarom onze toevlucht tot de oordelen van ter zake geïnformeerde deskundigen. Het gaat hier om leeropbrengsten in relatie tot de kerndoelen basisonderwijs, dus het ligt voor de hand het oordeel te vragen van mensen die goed bekend zijn met het basisonderwijs. Voor een standaardonderzoek nodigen we een panel van ongeveer vijftientig personen uit, bij voorkeur in de verhouding vijftien leraren basisonderwijs met minstens drie jaar onderwijservaring in jaargroep 8, vijf pabo-docenten en vijf medewerkers van schoolbegeleidingsdiensten. De samenstelling van het panel komt tot stand door steekproefsgewijze trekking uit adressenbestanden van de verschillende instellingen. Voor alle duidelijkheid: een standaardonderzoek wordt uitgevoerd nadat het peilingonderzoek heeft plaatsgevonden en de resultaten dus bekend zijn.

De centrale vraagstelling van het standaardonderzoek is steeds: Wat zouden de leerlingen aan het einde van het basisonderwijs omtrent een bepaald onderwerp moeten kennen, weten of kunnen wil er sprake zijn van voldoende beheersing van de kerndoelen?. Op aanbodniveau gelden de kerndoelen in principe voor alle leerlingen, maar het is gezien de spreiding in vaardigheid niet reëel te veronderstellen dat dan ook alle leerlingen een niveau zouden kunnen bereiken, dat we als voldoende kunnen aanmerken. De kerndoelen zouden dan al gauw op een vrij triviaal niveau gedefinieerd moeten worden. Wel dient – overeenkomstig het uitgangspunt – een meerderheid van de leerlingen in staat te zijn aan de kerndoelen te voldoen. Daarom vragen we in het standaardonderzoek aan beoordelaars om een vaardigheidsniveau te benoemen dat volgens hen door 70% tot 75% van de leerlingen bereikt zou moeten worden. Dit niveau wordt dan aangeduid met de standaard Voldoende.

Daarnaast wordt de beoordelaars gevraagd een vaardigheidsniveau aan te wijzen voor de standaard Minimum, een niveau dat door zeker 90% tot 95% van de leerlingen aan het eind van het basisonderwijs bereikt zou moeten worden.

De deelnemers aan het standaardonderzoek doorlopen voor ieder onderwerp drie beoordelingsfasen.

In de eerste fase maken en bestuderen de beoordelaars de opgaven, meestal een selectie van opgaven, die voor een bepaald onderwerp aan de leerlingen zijn voorgelegd. De opgaven worden hen daarbij in volgorde van moeilijkheidsgraad aangeboden. De beoordelaars geven voor zowel de standaard Minimum als de standaard Voldoende aan welke opgaven naar hun persoonlijke oordeel de leerlingen zouden moeten beheersen en daarmee dus ook welke opgaven ze niet hoeven te beheersen. Dat resulteert op het beoordelingsformulier in grensscores voor de beide standaarden. Het is uiteraard te verwachten dat beoordelaars verschillend zullen oordelen en dus ook verschillende grensscores zullen aanwijzen.

In de tweede fase krijgen de beoordelaars daarom de gelegenheid onderling over de gemaakte keuzen te overleggen; ze doen dat in kleine groepen van gemengde samenstelling. Het gaat om een uitwisseling en bespreking van de argumentaties bij hun persoonlijke oordelen uit fase 1. Elke individuele deelnemer krijgt aansluitend de gelegenheid het eerste oordeel bij te stellen.

De oordelen uit de eerste twee fasen zijn uitsluitend gebaseerd op de leerinhouden van de voorgelegde opgaven. De deelnemers weten dan nog niet hoe de vaardigheidsverdeling in feite is onder de betreffende populatie leerlingen. Pas in de derde fase worden de oordelen van de groep als geheel en van ieder afzonderlijk afgezet tegen de werkelijk bereikte resultaten. Dan zien de deelnemers de verschillen tussen het door hen gewenste resultaat en het bereikte resultaat. Daarna geeft iedere beoordelaar zijn of haar definitieve oordeel over het gewenste vaardigheidsniveau voor de beide standaarden Voldoende en Minimum. Informatie over dit eindoordeel wordt in de balansen gerapporteerd. We geven daarin voor iedere standaard aan binnen welke grensscores de middelste 50% van de oordelen liggen. Daarmee wordt zichtbaar gemaakt in hoeverre er tussen beoordelaars overeenstemming bestaat over het gewenste niveau: hoe kleiner dit gebied hoe groter de overeenstemming.

Het standaardenonderzoek kent over het algemeen twee trends. De discussiefase leidt doorgaans tot een afname van de spreiding in oordelen ten opzicht van de oordelen uit de eerste fase. En dus tot meer onderlinge overeenstemming over de gewenste niveaus. Nadat we in de derde fase deze oordelen hebben vergeleken met de werkelijk bereikte resultaten, stellen de beoordelaars hun oordeel soms bij in de richting van het bereikte resultaat. Dat zien we met name als het verschil tussen het eerdere oordeel en het werkelijke resultaat groot is. Uiteraard heeft dat tot gevolg dat de spreiding in oordelen weer toeneemt, doordat niet iedere beoordelaar in gelijke mate zijn oordeel wenst aan te passen aan het bereikte resultaat. Toch is de spreiding van de oordelen dan meestal nog kleiner dan na afronding van fase 1.

De standaarden bieden een kwalitatief referentiekader voor de interpretatie van de onderzoeksresultaten én voor de evaluatie van de onderwijsopbrengsten. In de volgende hoofdstukken zullen we daarom overzichten presenteren van het percentage leerlingen dat het niveau van de standaarden Minimum en Voldoende bereikt. Daarbij baseren we ons op de mediaan van de oordelen – dit is het middelste oordeel – voor deze standaarden. We zullen zien dat op tal van onderwerpen het niveau van de standaard Voldoende hoger ligt dan het bereikte resultaat. Dat wil zeggen, dat minder dan 70% van de leerlingen de standaard Voldoende bereikt. Percentages tussen 50% en 60% zijn daarbij lichtrood gemarkeerd, percentages beneden 50% zijn donkerrood gemarkeerd. De verschillen tussen het beoogde en het bereikte resultaat worden in één oogopslag zichtbaar. De standaarden krijgen inhoudelijk betekenis wanneer zij gerelateerd worden aan de opgaven die leerlingen verondersteld worden op het niveau van de standaard te beheersen. Vandaar dat in de balansrapportages van PPON op ruime schaal opgaven zijn opgenomen die inhoudelijk het niveau van de standaarden illustreren.

De vergelijking over de tijd | Naast de standaarden presenteren we in dit samenvattend overzicht gegevens waarmee we de prestaties van leerlingen uit verschillende peilingjaren met elkaar vergelijken. We willen graag weten of er in de loop der tijd veranderingen optreden in de prestaties van leerlingen. Een deel van de opgaven die PPON aan leerlingen voorlegt wordt geheim gehouden om in volgend peilingonderzoek opnieuw aan de leerlingen voor te leggen. Op basis van deze gemeenschappelijke opgaven construeren we

vaardigheidsschalen, waardoor peilingresultaten van verschillende jaren ook daadwerkelijk vergelijkbaar worden.

Verschillen tussen jaren worden uitgedrukt in termen van effectgrootten. Positieve effectgrootten wijzen op een verbetering en negatieve op een achteruitgang van de leerprestaties. Voor de interpretatie van de effectgrootten volgen we de in de literatuur gebruikelijke kwalificaties, waarbij effectgrootte 0.2 wordt aangemerkt als een klein effect, 0.5 als een matig en 0.8 als een groot effect.

In bijgaande overzichten worden de effectgrootten voor de jaarvergelijkingen steeds uitgedrukt ten opzichte van een basisjaar (= 0). Het basisjaar is het jaar dat het betreffende onderwerp voor het eerst in de peiling is opgenomen. Positieve effectgrootten tussen 0.2 en 0.5 zijn lichtgroen gemarkeerd, tussen 0.5 en 0.8 groen en boven 0.8 donkergroen. De negatieve hebben we in vergelijkbare gradaties rood gekleurd.

In de in deze publicatie opgenomen overzichten van effecten voor de jaarvergelijkingen is meestal sprake van gestapelde effectgrootten wanneer in de vergelijking drie of meer afnametijdstippen zijn opgenomen. De gestapelde effectgrootten worden verkregen door de effectgrootten van twee opeenvolgende peilingjaren samen te nemen.

De opzet van het peilingonderzoek voor de verschillende vakgebieden is in de afgelopen twintig jaar regelmatig aan verandering onderhevig geweest. Soms is er gekozen voor een andere indeling in onderwerpen, soms zijn onderwerpen weggelaten of zijn er nieuwe onderwerpen aan toegevoegd. Als gevolg daarvan zijn vergelijkingen over de tijd dan slechts over een beperkte periode mogelijk. Dankzij de stabiliteit van de peilingonderzoeken voor rekenen/wiskunde was het voor dit leergebied mogelijk de vergelijking over vier peilingjaren te baseren op twee elkaar deels overlappende effectschattingen over drie perioden.

Doelgroepen voor peilingresultaten | Het peilingonderzoek wordt uitgevoerd in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen, vanuit haar verantwoordelijkheid voor de kwaliteit van het onderwijs. Met de resultaten van peilingonderzoeken worden de diverse actoren in het onderwijsveld gedetailleerd geïnformeerd over de opbrengsten van het onderwijs.

Belangrijke doelgroepen voor de balansrapportages zijn:

- het Ministerie van OCW, de Vaste Commissie voor Onderwijs van de Tweede Kamer en de Onderwijsraad,
- de Inspectie van het Onderwijs,
- landelijke en regionale instituten voor onderwijs- en schoolbegeleiding,
- lerarenopleidingen voor het basisonderwijs,
- methodeontwikkelaars en uitgevers van methoden voor het basisonderwijs,
- vakinhoudelijke experts en onderzoekers,
- leraren basisonderwijs en mogelijk ook onderbouw voortgezet onderwijs.

Resultaten van peilingonderzoek publiceren we in de balansrapporten. Deze zijn ook integraal en voor iedereen toegankelijk te vinden op onze website. Daarnaast publiceren we van iedere balans een samenvatting met de belangrijkste resultaten in het katern 'PPON informeert', dat we verspreiden onder alle basisscholen en andere onderwijsinstellingen. De peilingresultaten zijn niet gericht op directe veranderingen in het onderwijs. Wel vestigen ze met regelmaat de aandacht op wat leerlingen feitelijk in het basisonderwijs leren en spelen ze een mogelijke rol bij het vaststellen van wenselijke veranderingen. Positieve trends vragen om ondersteuning en negatieve om bijsturende maatregelen. De kwaliteit van ons onderwijs staat centraal. En telkens weer moeten we ons bezinnen. Wat is het rendement van onze onderwijsinspanning? En sluit dat rendement aan bij wat we ervan verwachten? Resultaten van peilingonderzoek levert stof tot nadenken en input voor afwegingen over de kwaliteit en het niveau van het onderwijs.

2 Overzicht van peilingresultaten voor Nederlandse taal

Voor Nederlandse taal einde basisonderwijs hebben we peilingen verricht vanaf 1988 in alle vier de peilingperioden. De opzet is daarbij niet altijd dezelfde geweest. In de loop der tijd zijn op drie niveaus belangrijke wijzigingen doorgevoerd.

In de eerste plaats zijn de peilingonderzoeken vanaf 1998 niet meer gericht op het leer- en vormingsgebied in haar totaliteit, maar vinden er afzonderlijke peilingonderzoeken plaats voor de vier hoofdvaardigheden lezen, luisteren, schrijven en spreken. Een volledig peilingonderzoek bleek zo omvangrijk dat rapportages te lang op zich lieten wachten.

Een tweede ingrijpende wijziging betreft het uitgangspunt van de peilingonderzoeken.

In de eerste peilingen waren de rapportage-eenheden voor iedere hoofdvaardigheid ingedeeld naar teksttype, zoals bijvoorbeeld het lezen van rapporterende, beschouwende en argumentatieve teksten. In de laatste peilingen is niet meer het type tekst maar de te onderscheiden vaardigheden van leerlingen het uitgangspunt. Zo hanteren we voor lezen en luisteren een indeling op basis van vaardigheden als begrijpen, interpreteren en reflecteren. Het type tekst is daarbij ondergeschikt.

Een derde ingrijpende wijziging gaat over de evaluatie van de schrijfvaardigheid en de spreekvaardigheid. Aanvankelijk hebben de beoordelaars op basis van een beoordelingsmodel een globaal oordeel toegekend aan de schrijfprestaties en de spreekprestaties van de leerlingen. Bij de laatste peilingsonderzoeken hebben we gekozen voor een analytische beoordeling van inhoudelijke elementen in de lees- en schrijfprestaties. Er kunnen dan meerdere beoordelingsdimensies worden onderscheiden en er ontstaat een duidelijker en genuanceerd beeld van de vaardigheden van leerlingen op de verschillende dimensies. Bijkomend voordeel is dat de overeenstemming tussen beoordelaars daarmee aanzienlijk kan worden vergroot.

In dit overzicht hebben we voor Nederlandse taal de resultaten verwerkt uit de balansen 19, 28 en 29, 30 en 33.

De standaarden | Het eerste standaardenonderzoek voor Nederlandse taal is uitgevoerd in 1995 op basis van de resultaten van de peiling in 1993. De resultaten van dat standaardenonderzoek zijn ook gebruikt voor de peiling in 1998. Voor *Leesvaardigheid* zijn in bijgaand schema ook de standaarden opgenomen van de peiling in 2005.

De beoordelaars waren in 1998 vrij kritisch over de bereikte resultaten. Alleen voor *Taalbeschouwing* bereikte toen meer dan 60% van de leerlingen de standaard Voldoende. Voor zes onderwerpen ligt het scoringspercentage voor het gewenste niveau tussen 50% en 59%. En maar liefst voor acht onderwerpen is de toenmalige standaard Voldoende door minder dan 50% van de leerlingen bereikt. Voor de beoordelaars was het duidelijk: te weinig leerlingen beheersten de kerndoelen basisonderwijs op voldoende niveau. Het oordeel over de standaard Minimum daarentegen was doorgaans vrij mild.

Voor de onderwerpen uit de peiling Leesvaardigheid in 2005 is een nieuw standaardonderzoek uitgevoerd. De beoordelaars waren wederom kritisch. Bij vier van de vijf onderwerpen scoorde minder dan 60% van de leerlingen onder het niveau van de standaard Voldoende. Maar nergens werd het minder dan 50%.

Vaak vallen de prestaties van de leerlingen dus lager uit, dan de beoordelaars wensen.

Wie de balansen raadpleegt, verkrijgt meer inzicht in hun beoordelingen. Illustratief zijn daarbij de voorbeeldopgaven in de balans. Deze laten zien, waaraan de leerlingen volgens de standaarden zouden moeten voldoen.

Percentage leerlingen dat in 1998 en 2005 de standaarden Minimum en Voldoende bereikt voor onderwerpen Nederlandse taal

Domein	Onderwerp	MIN '98	VOL '98	MIN '05	VOL '05
Nederlandse taal algemeen	lezen van rapporterende teksten	66	29		
	lezen van beschouwende teksten	89	55		
	lezen van argumentatieve teksten	94	53		
	lezen van fictie	83	44		
	hanteren van naslagwerken	93	52		
	technisch lezen				
	luisteren naar rapporterende teksten	89	56		
	luisteren naar beschouwende teksten	91	48		
	luisteren naar argumentatieve teksten	86	36		
	luisteren naar fictie	88	35		
	stijl	85	32		
	grammaticaliteit	84	54		
	interpunctie	90	53		
	spelling				
	zinsontleding woordbenoeming	60	37		
	taalbeschouwing	94	66		
woordbegrip	76	29			
Leesvaardigheid	begrijpen van geschreven teksten			92	66
	interpreteren van geschreven teksten			84	50
	reflecteren op geschreven teksten				
	studerend lezen			84	50
	opzoeken van informatie			92	58
	woordenschat			92	58
	voorspellend lezen				
Legenda					
 50-60% bereikt het niveau van de standaard Voldoende					
 minder dan 50% bereikt het niveau van de standaard Voldoende					

Vergelijking over de tijd | De opzet van de peilingonderzoeken in 1988, 1993 en 1998 was in principe dezelfde; ook al zijn er wijzigingen doorgevoerd. De peilingresultaten uit die jaren zijn dus voor de meeste onderwerpen onderling vergelijkbaar. Over het algemeen vinden we geringe verschillen tussen de peilingjaren. Opmerkelijk zijn echter de scores op het onderwerp *Lezen van argumentatieve teksten*. Ten opzichte van 1988 registreerden we in

1993 een groot positief effect, dat in 1998 weer gereduceerd werd tot een matig effect. Voor deze schommeling hebben we geen bevredigende verklaring kunnen vinden. De afwijking spoort ook niet met de gemiddelde effectgrootte over alle onderwerpen voor lezen die in 1998 +0.10 bedraagt; en dat is een zeer gering positief effect.

Verder hebben we in 1998 significante negatieve effecten gevonden voor *Luisteren naar fictie*, *Stijl* en *Grammaticaliteit* en een negatief maar niet significant effect voor *Zinsontleding en woordbenoeming*.

Voor twee onderwerpen op het gebied van luistervaardigheid – *Luisteren naar beschouwende teksten* en *Luisteren naar argumentatieve teksten* – hebben we in 1993 kleine tot matig negatieve effecten gevonden ten opzichte van de eerste peiling in 1988, die in de volgende peiling weer sterk zijn gereduceerd.

Peilingen van schrijfprestaties uitgevoerd vóór 1999 zijn altijd beoordeeld op 'globale kwaliteit'. Vanaf dat jaar hebben we echter voor een andere aanpak gekozen. En dat werd als eerste een afzonderlijk peilingonderzoek voor schrijfvaardigheid, waarbij we met een analytische blik specifiek gekeken hebben naar verschillende aspecten van schrijfvaardigheid. Voor een vergelijking over de tijd zijn schrijfproducten uit 1993 opnieuw beoordeeld conform deze analytische benadering. Op drie onderwerpen was er geen verschil. Maar op drie andere hebben we significant positieve effecten geregistreerd ten opzichte van 1993.

De effectgrootten kunnen als matig worden gekwalificeerd. Het is echter niet duidelijk welke factoren aan deze positieve wending ten grondslag liggen. Er worden in taalmethodes geen aanwijzingen gevonden op grond waarvan deze positieve ontwikkeling op juist deze drie onderwerpen – *Formele structuur*, *Doelgerichtheid* en *Publieksgerichtheid* – verwacht zou kunnen worden. In de toekomst zal moeten blijken in hoeverre hier sprake is van een structureel effect.

De schrijfproducten van de leerlingen zijn in 1999 ook op een aantal aspecten van taal-kwaliteit onderzocht. Enkele resultaten: 19% van de schrijfproducten van leerlingen uit jaargroep 8 was volledig correct geïnterpungeerd, 17% bevatte geen spelfouten en 21% van de teksten bleef verschoond van ongrammaticale zinsstructuren. (zie balans 29).

In 2002-03 peilden we de spreekvaardigheid op een geheel nieuwe wijze. Hierdoor is vergelijking met eerdere peilingresultaten niet mogelijk. Daarover is gerapporteerd in balans nr. 30. Het onderzoek leverde wel enkele verrassende resultaten op. In principe zou jaargroep 8 bijvoorbeeld in staat moeten zijn op meerdere aspecten van spreekvaardigheid beter te scoren dan jaargroep 5. Daar is echter niets van gebleken. De variaties binnen de beide jaargroepen blijken groter dan tussen de jaargroepen.

Wel is er een verschil in positieve zin te constateren tussen de spreekprestaties van jaargroep 8 leerlingen op het inhoudsaspect van de korte responstoets en de prestaties op dit punt van jaargroep 5 leerlingen.

In 2005 is ten slotte de vierde peiling voor leesvaardigheid uitgevoerd. Deze onderscheidt zich van voorgaande peilingen, doordat de algemene leesvaardigheid gekoppeld werd aan deelvaardigheden als begrijpen, interpreteren en reflecteren. Voorheen stond het type tekst centraal. Daarnaast zijn in 2005 ook aan lezen gerelateerde vaardigheden gepeild.

Voor enkele onderwerpen hebben we ondanks de wijzigingen nog een vergelijking kunnen trekken met de peilingresultaten van 1998. Maar effecten van betekenis zijn daarbij niet gevonden.

Effectgrootten voor de jaarvergelijkingen voor Nederlandse taal over de periode 1988-2005

Domein	Onderwerp	1988	1993	1998	2005
Nederlandse taal algemeen	lezen van rapporterende teksten	0	-0,08	-0,08	
	lezen van beschouwende teksten	0	-0,14	-0,17	
	lezen van argumentatieve teksten	0	0,83	0,68	
	lezen van fictie		0	0,03	
	hanteren van naslagwerken				
	technisch lezen				
	luisteren naar rapporterende teksten	0	0,1	-0,15	
	luisteren naar beschouwende teksten	0	-0,36	-0,18	
	luisteren naar argumentatieve teksten	0	-0,32	-0,19	
	luisteren naar fictie		0	-0,41	
	stijl	0	-0,05	-0,48	
	grammaticaliteit	0	-0,18	-0,33	
	interpunctie		0	0,19	
	spelling	0	-0,02		
	zinsontleding woordbenoeming	0	-0,02	-0,40	
	taalbeschouwing	0	-0,09	-0,12	
	woordbegrip		0	-0,04	
Schrijfvaardigheid	gegeven inhoud		0	-0,04	
	zelfgegenereerde inhoud		0	0,15	
	organisatie		0	0,17	
	formele structuur		0	0,65	
	doelgerichtheid		0	0,48	
	publiekgerichtheid		0	0,58	
Leesvaardigheid	begrijpen van geschreven teksten			0	-0,04
	interpreteren van geschreven teksten			0	0,11
	reflecteren op geschreven teksten				
	studerend lezen				
	opzoeken van informatie				
	woordenschat			0	0,01
	voorspellend lezen			0	-0,04
Legenda					
■ Positieve effectgrootte > 0,8					
■ Positieve effectgrootte > 0,5					
■ Positieve effectgrootte 0,2 – 0,5					
■ Negatieve effectgrootte -0,2 – -0,5					

3 Overzicht van peilingresultaten voor rekenen/wiskunde

Rekenen/wiskunde peilden we in jaargroep 8 in 1987, 1992, 1997 en 2004 op telkens nagenoeg dezelfde wijze. En dat maakt vergelijkingen over alle peilingjaren mogelijk. Toch zijn er allengs enkele veranderingen doorgevoerd, waarvan de belangrijkste:

- Eerst vanaf 1992 rapporteerden we afzonderlijk over de onderwerpen *Basisoperaties optellen/afrekken* en *Basisoperaties vermenigvuldigen/delen*.
- In 1992 hanteerden we een andere procedure voor *hoofdrekenen* en *schattend rekenen* dan in de overige jaren. De resultaten van dit peilingjaar kunnen we dus niet betrekken in een vergelijking over meerdere jaren.
- Het onderwerp *Hanteren van de zakrekenmachine* maakte in 1987 nog geen deel uit van het peilingsonderzoek. En het onderwerp *Tabellen en grafieken* hoorde toen nog bij Nederlandse taal (Informatieverwerking). Vanaf 1997 valt dat onderwerp onder de rekenpeiling.
- Bij de onderwerpen *Verhoudingen*, *Breuken* en *Procenten* hanteerden we in de eerste peilingen nog een indeling in basiskennis en begrip én toepassingen of bewerkingen. Bij de laatste peiling is die onderverdeling weggefallen.
- Door de komst van de euro hebben we de peiling in 2004 op het onderwerp *Geld* buiten de vergelijking over de tijd gelaten.

Het hier gepresenteerde overzicht van standaarden is gebaseerd op de onderzoeken voor de peilingen in 1997 en 2004. Verder geldt, dat we voor ieder onderwerp zoveel mogelijk de effectgrootten vermelden ten opzichte van 1987 (= 0).

De standaarden | In 1997 én in 2004 zijn de geraadpleegde deskundigen veelal kritisch over het bereikte resultaat. Voor niet meer dan zes van de 22 onderwerpen stelden we vast, dat 60% of meer leerlingen de standaard Voldoende bereikte. Op zes andere onderwerpen voldeed tussen 50% en 59% van de leerlingen aan de standaard Voldoende en voor tien onderwerpen was dat percentage minder dan 50%. Er is reden tot zorg zeggen de geraadpleegde experts. In het bijzonder waar het onderwerpen betreft die betrekking hebben op de bewerkingen: *Bewerkingen: optellen en afrekken*, *Bewerkingen: vermenigvuldigen en delen* en *Samengestelde bewerkingen*. Onder 'bewerkingsopgaven' worden in het peilingonderzoek opgaven verstaan die niet 'uit het hoofd' opgelost behoeven te worden. Het zijn veelal opgaven met wat grotere getallen, waarbij in principe van de leerling wordt verwacht dat hij de opgave oplost door het noteren van tussenoplossingen of door middel van een algoritmische bewerking. De leerlingen worden daar ook expliciet op gewezen en er is daarvoor voldoende ruimte in de toetsboekjes. Samengestelde bewerkingen zijn opgaven waarbij de leerlingen twee of meer verschillende typen bewerkingen moeten toepassen om de opgave op te lossen. De resultaten bij juist deze onderwerpen blijven ver achter bij het gewenste niveau: minder dan 30% van de leerlingen haalt dan het niveau van de standaard Voldoende. Ook over de resultaten bij het onderwerp *Meten* zijn de beoordelaars bezorgd.

De beoordelaars zijn duidelijk niet zo kritisch geweest bij het bepalen van de standaard Minimum. Bij 16 onderwerpen bereikt meer dan 80% van de leerlingen het niveau van deze standaard en bij 10 onderwerpen zelfs 90% of meer. Een duidelijk uitzondering zijn ook hier de bewerkingsopgaven. Ongeveer 60% van de leerlingen bereikt dan het niveau van de standaard Minimum, terwijl deze standaard het niveau aanduidt dat voor 90% tot 95% haalbaar zou moeten zijn.

Percentage leerlingen dat in 1997 en 2004 de standaarden Minimum en Voldoende bereikt voor onderwerpen Rekenen/wiskunde

Domein	Onderwerp	MIN '97	VOL '97	MIN '04	VOL '04
Getallen en bewerkingen	getallen en getalrelaties	92	47	96	42
	basisoperaties: +/-			92	76
	basisoperaties: x/:			90	66
	hoofdrekenen: +/-	88	42	92	50
	hoofdrekenen: x/:	93	57	92	66
	schattend rekenen	82	22	84	42
	bewerkingen: +/-	75	49	62	27
	bewerkingen: x/:	82	52	50	12
	samengestelde bewerkingen	91	56	50	16
	zakrekenmachine	82	39	73	34
Verhoudingen, breuken en procenten	breuken: basiskennis en begrip	88	26		
	breuken: optellen en aftrekken	86	53		
	breuken: vermenigvuldigen en delen	80	41		
	procenten: basiskennis en begrip	84	41		
	procenten: toepassingen	82	41		
	verhoudingen: basiskennis en begrip	94	67		
	verhoudingen: toepassingen	81	40		
	verhoudingen			92	66
	breuken			95	60
	procenten			88	58
tabellen en grafieken			84	50	
Meten	meten: lengte	82	44	79	38
	meten: oppervlakte	75	35	67	21
	meten: inhoud	85	48	84	42
	meten: gewicht	78	37	88	58
	meten: toepassingen	86	50	92	50
	meetkunde	96	71	95	62
	tijd	90	60	92	50
	geld	93	69	84	42
Legenda					
 50-60% bereikt het niveau van de standaard Voldoende					
 minder dan 50% bereikt het niveau van de standaard Voldoende					

In 2004 hebben we met een ander panel beoordelaars een nieuw standaardenonderzoek uitgevoerd. Voor de meeste onderwerpen heeft dit opnieuw geresulteerd in niveaus voor de standaard Voldoende die door minder dan 60% van de leerlingen worden bereikt, vaak zelfs door minder dan 50% van de leerlingen. Wat uit deze cijfers niet direct blijkt is dat het niveau van de standaarden vrij stabiel is gebleven, vergeleken met de standaarden uit 1997. Bijgevolg loopt een toename dan wel afname van het percentage leerlingen dat de standaard in 2004 bereikt parallel aan positieve en negatieve ontwikkelingen in het vaardigheidsniveau van de leerlingpopulatie ten opzichte van 1997. Zo leiden de positieve jaareffecten voor de onderwerpen *Hoofdrekenen: optellen en aftrekken*, *Schattend rekenen* en *Procenten* ertoe dat meer leerlingen het niveau van de standaard Voldoende bereiken dan in het vorige peilingjaar. En, omgekeerd, leiden de negatieve effecten voor de bewerkingsonderwerpen tot de conclusie, dat minder leerlingen voldoen aan het niveau van de standaarden. Voorbeeldopgaven én vaak ook empirische ontwikkelingslijnen illustreren de standaarden die gepubliceerd zijn in de balanspublicaties. Aldus is een inschatting te maken van de verwachtingen over de leerprestatie conform het niveau van de standaarden.

Vergelijking over de tijd | De leerinhouden voor rekenen/wiskunde zijn in de laatste peiling van 2004 onderverdeeld in tweeëntwintig onderwerpen. Voor eenentwintig onderwerpen kunnen we de resultaten van vrijwel alle peilingjaren – dus vanaf 1987 – vergelijken. Bij tien onderwerpen zijn in die periode nauwelijks verschuivingen waargenomen. Bij *Basisoperaties optellen en aftrekken*, *Zakrekenmachine*; *Meten: gewicht* en ten slotte *Meten: toepassingen* blijken de verschillen erg klein. Voor deze in totaal veertien onderwerpen zijn de prestaties van leerlingen tussen 1987 en 2004 dus nagenoeg stabiel gebleven. Zij zijn niet slechter gaan presteren, maar ook niet beter; het effect ten opzichte van 1987 is minimaal. Een significant positief effect hebben we gevonden bij het onderwerp *Procenten* en bij de basale vaardigheden *Getallen en getalrelaties*, *Hoofdrekenen: optellen en aftrekken* (echter niet bij *Hoofdrekenen: vermenigvuldigen en delen*) en *Schattend rekenen*. De drie zogenaamde bewerkingsonderwerpen *Bewerkingen: optellen en aftrekken*, *Bewerkingen: vermenigvuldigen en delen* en *Samengestelde bewerkingen* scoorden betekenisvolle negatieve resultaten.

De effecten, zowel de positieve als de negatieve, zijn geen incidenten. We kunnen spreken van een tweetal trends in de ontwikkeling van de rekenvaardigheid van de leerlingen. De basale rekenvaardigheden vertonen op onderdelen een opgaande lijn. Dat is duidelijk in lijn met de ontwikkelingen binnen het reken-wiskundeonderwijs in Nederland in de laatste decennia. Negatieve effecten worden vooral gevonden bij de diverse bewerkingsopgaven. Maar hoe valt die neerwaartse lijn te verklaren naast een betere beheersing van de basale vaardigheden? Een begin van een verklaring zoeken wij in de uitvoering van de opgaven. We verwachten dat de leerlingen tussenoplossingen noteren of gebruikmaken van een algoritme. Leerlingen zijn vrij in de keuze van een oplossingsstrategie. Maar onderzoek wijst uit, dat de leerlingen meer en meer de opgaven zonder notaties en dus blijkbaar uit het hoofd proberen op te lossen. En dat is gegeven de grotere getallen geen succesvolle strategie. Ook blijken nieuwe algoritmische oplossingsstrategieën (de zogenaamde kolomsgewijze algoritmen voor optellen, aftrekken, vermenigvuldigen en delen) minder vaak te leiden tot een correct antwoord dan traditionele strategieën.

De volgende vraag is dan natuurlijk 'Waarom noteren de leerlingen minder tussenoplossingen of maken ze minder gebruik van algoritmen?' Die vraag is niet direct te beantwoorden, er is waarschijnlijk sprake van een complex aan factoren, maar ook die hebben ongetwijfeld te maken met de veranderingen die zich de laatste decennia in het reken-wiskundeonderwijs hebben voltrokken. Belangrijker is de vraag hoe deze tendens kan worden omgebogen. Uit de oordelen voor de standaard Voldoende die de geraadpleegde deskundigen hebben vastgesteld kan men afleiden dat deze ontwikkeling toch als niet gewenst wordt gezien.

Effectgrootten voor de jaarvergelijkingen voor Rekenen/wiskunde over de periode 1987-2004

Domein	Onderwerp	1987	1992	1997	2004
Getallen en bewerkingen	getallen en getalrelaties	0	0,28	0,46	0,94
	basisoperaties: +/-		0	-0,11	0,24
	basisoperaties: x/:		0	-0,3	-0,2
	hoofdrekenen: +/-	0		0,49	0,53
	hoofdrekenen: x/:	0		-0,12	-0,11
	schattend rekenen	0		0,94	1,04
	bewerkingen: +/-	0	-0,12	-0,17	-0,53
	bewerkingen: x/:	0	-0,17	-0,43	-1,16
	samengestelde bewerkingen	0	-0,4	-0,44	-0,78
	zakrekenmachine		0	0,29	0,26
Verhoudingen, breuken en procenten	verhoudingen	0	0,11	0,26	0,14
	breuken	0	0,09	0,23	0,15
	procenten	0	0,12	0,28	0,51
	tabellen en grafieken			0	0,1
Meten	meten: lengte	0	0	-0,03	-0,13
	meten: oppervlakte	0	-0,32	-0,04	0,05
	meten: inhoud	0	0,1	0	-0,03
	meten: gewicht	0	0,02	0,2	0,33
	meten: toepassingen	0	-0,05	-0,21	-0,25
	meetkunde	0	0	0,12	-0,08
	tijd	0	0,17	0,23	0
	geld	0	-0,21	-0,31	

Legenda

- Positieve effectgrootte > 0,8
- Positieve effectgrootte > 0,5
- Positieve effectgrootte 0,2 – 0,5
- Negatieve effectgrootte -0,2 – -0,5
- Negatieve effectgrootte -0,5 – -0,8
- Negatieve effectgrootte > 0,8

4 Overzicht van peilingresultaten voor wereldoriëntatie

Wereldoriëntatie omvat de leergebieden geschiedenis, aardrijkskunde en natuuronderwijs (waaronder biologie, natuurkunde en techniek). In de periode 2000-2002 peilden we deze leergebieden voor het laatst. Eerdere peilingonderzoeken voor dit leerdomein vonden plaats in 1990/1991 en in 1995. Maar het peilingonderzoek werd zo veelomvattend dat we met ingang van 2000 hebben besloten afzonderlijk te rapporteren over de peilingonderzoeken voor de verschillende leergebieden. Meestal ging dat ook gepaard met een herindeling van onderwerpen. Dat betekent dat we niet alle onderwerpen kunnen vergelijken met resultaten uit eerder gehouden peilingonderzoek.

Het curriculum voor wereldoriëntatie in het basisonderwijs is minder duidelijk gedefinieerd dan voor Nederlandse taal of rekenen/wiskunde. Dat is de aanleiding geweest om uitvoerig onderzoek te doen naar de relevante leerinhouden. Een cultuurpedagogische discussie onder experts heeft uiteindelijk geresulteerd in domeinbeschrijvingen voor de drie deelgebieden geschiedenis, aardrijkskunde en natuuronderwijs. Deze nieuwe domeinbeschrijvingen liggen ten grondslag aan onze peilingonderzoeken voor wereldoriëntatie die in de periode 2000-2002 zijn uitgevoerd. In 2002 zijn deze domeinbeschrijvingen gepubliceerd en daarmee voor een breed publiek beschikbaar (zie Wagenaar, 2002; Notté, 2002; Thijssen, 2002).

Geschiedenis | *De standaarden* | In 1995 waren de geraadpleegde experts kritisch over het bereikte kennisniveau van de leerlingen voor de standaard Voldoende. Alleen op het onderwerp *Sociaal-culturele verschijnselen* voldeden voldoende leerlingen aan deze standaard. In 2000 was het niet anders. Ook toen oordeelden de beoordelaars opnieuw kritisch. Op vrijwel alle onderwerpen voldeed minder dan 60% van de leerlingen aan het niveau van de standaard Voldoende. En voor drie onderwerpen – *Middeleeuwen*, *Nieuwe tijd: 19^e eeuw* en *Nieuwste tijd: 20^e eeuw* – zelfs minder dan 50%. Alleen het onderwerp *Samenleving* bleef overeind. Het percentage leerlingen dat de standaard Voldoende bereikt komt overeen met het voor deze standaard beoogde percentage.

Vergelijking over de tijd | In de peilingen van '91 en '95 ging het ondermeer over de brede deelgebieden *Oude geschiedenis* en *Nieuwe geschiedenis*. Ten opzichte van het eerste peilingjaar werd in het tweede onderzoek een klein negatief effect gevonden voor het onderwerp *Oude Geschiedenis* en een klein positief effect voor *Nieuwe geschiedenis*. In 2000 hebben we zeven onderwerpen gepeild. Vijf meer dan in de eerste jaren. Om een vergelijking mogelijk te maken, hebben we die zeven onderwerpen ondergebracht in de twee onderwerpen die er aanvankelijk al waren. Wat blijkt? Het eerder gevonden negatieve effect voor het onderwerp *Oude geschiedenis* is stabiel gebleven. En het eerder geconstateerde kleine positieve effect voor *Nieuwe geschiedenis* is geheel verdwenen.

Tijdsbesef was pas in 1995 een afzonderlijk onderwerp. In 2000 vinden we voor dit onderwerp een significant negatief effect, waarbij de effectgrootte als matig gekwalificeerd kan worden. Het is daarom des te opvallender, dat de experts wat het onderwerp *Tijdsbesef* betreft, in 2000 positiever oordeelden over het bereikte resultaat dan in 1995.

Percentage leerlingen dat in 1995 en 2000-2002 de standaarden Minimum en Voldoende bereikt voor onderwerpen Wereldoriëntatie

Domein	Onderwerp	MIN '00/'02	VOL '95	MIN '00/2	VOL '00/'02
Wereldoriëntatie algemeen	politieke verschijnselen	91	56		
	soc. culturele verschijnselen	94	69		
	economische verschijnselen	94	36		
Geschiedenis	tijdsbesef / tijdsbegrip	95	42	92	58
	oude geschiedenis	79	35		
	nieuwe geschiedenis	88	50		
	prehistorie			84	58
	oudheid			91	56
	middeleeuwen			86	38
	nieuwe tijd: republiek			84	39
	nieuwste tijd: 19e eeuw			84	43
	nieuwste tijd: 20e eeuw			92	56
	samenleving			93	75
Aardrijkskunde	aardrijkskunde – nederland	77	25		
	aardrijkskunde – europa	87	32		
	aardrijkskunde – wereld	92	59		
	natuurkundige aardrijkskunde	90	44		
	kaartlezen	88	42	73	34
	topografie Nederland	71	19	58	16
	topografie Europa	72	34	73	27
	topografie Wereld	67	28	58	21
	aarde en landschappen			82	50
	bevolking			84	50
	bestaansmiddelen			92	66
Natuuronderwijs	oecologie en gedrag	88	54		
	bouw en systematiek	92	64		
	organismen, waarnemen en reageren			73	46
	stofwisseling en kringloop			95	66
	voortplanting en ontwikkeling			82	42
	biotopen			58	21
	natuurkunde	86	50	93	54
	techniek			84	34
Legenda					
 50-60% breikt het niveau van de standaard Voldoende					
 minder dan 50% bereikt het niveau van de standaard Voldoende					

Aardrijkskunde | De kennis van *Nederland, Europa, De wereld* en *Natuurkundige aardrijkskunde* stond centraal in de eerste twee peilingonderzoeken voor het leergebied Aardrijkskunde. In 2001 veranderde dat. Toen werd op basis van de nieuwe domeinbeschrijving een indeling gehanteerd naar de aardrijkskundige kennisvelden *Aarde en landschappen, Bestaansmiddelen* en *Bevolking*. De onderwerpen over kaartlezen en topografie zijn hetzelfde gebleven.

De standaarden | De geraadpleegde deskundigen zijn in de meeste gevallen zeer kritisch over het bereikte kennisniveau. Dat geldt voor 1995 én 2001. Ze vinden het bereikte niveau ver onder de maat. Niet meer dan 50% van de leerlingen bereikte het gewenste niveau voor de standaard Voldoende. Uitgesproken negatief valt het oordeel uit voor de onderwerpen die betrekking hebben op kennis van topografie.

Vergelijking over de tijd | In 1995 vonden we ten opzichte van de peiling in 1991 negatieve effecten voor kennis van *Europa* evenals voor *Natuurkundige aardrijkskunde* en een positief effect voor kennis van *de wereld*. Ondanks de herdefiniëring van de onderwerpen voor de peiling in 2001 bleef er voor de onderwerpen *Aarde en landschappen* en *Bestaansmiddelen* (niet echter voor *Bevolking*) voldoende overeenkomst met de peilingen in 1995 om een vergelijking te kunnen maken. Conclusie: de kennis en inzichten van leerlingen op deze gebieden zijn op hetzelfde niveau gebleven.

Het kennisgebied *Topografie* is over alle drie de peilingen inhoudelijk ongewijzigd gebleven. Dat maakt een vergelijking van de onderdelen topografie van Nederland, van Europa en van de wereld eenvoudig. Over de gehele periode 1991-2001 blijkt het kennisniveau van de leerlingen op deze gebieden vrij stabiel. Met uitzondering van een niet relevante, lichte negatieve tendens in 2001.

In 1991 peilden we *Kaartlezen* nog niet. De vergelijking over de jaren bleef dus beperkt tot de laatste twee peiljaren. We vonden op dit onderdeel een opmerkelijk, matig negatief effect in 2001.

Natuuronderwijs | Natuuronderwijs omvat de leerdomeinen biologie, natuurkunde en techniek. De opzet voor natuurkunde is vanaf de eerste peiling niet veranderd. *Techniek* is in 2002 als een nieuw onderwerp aan het peilingonderzoek toegevoegd. Het leergebied Biologie is in datzelfde jaar uitgebreid van twee naar vier onderwerpen.

De standaarden | In 1995 hebben de beoordelaars redelijk positief geoordeeld over het bereikte niveau bij het onderwerp *Bouw en systematiek*. Over de resultaten bij de onderwerpen *Oecologie en gedrag* en *Natuurkunde* was men minder te spreken, want minder dan 60% van de leerlingen bereikte het gewenste niveau.

In 2002 waren de beoordelaars aanmerkelijk kritischer over de bereikte resultaten. Op drie van de vier onderwerpen voor biologie voldeed minder dan 50% van de leerlingen aan het gewenste niveau voor de standaard Voldoende. In het bijzonder de kennis van biotopen was volgend de beoordelaars ver beneden de maat. De opgaven bij dit onderwerp gingen over de samenhang tussen organismen onderling en tussen organismen en hun omgeving in drie biotopen: *Stad en omgeving, In en langs de plas* en *Het bos*.

De beoordelaars kwamen voor *Natuurkunde* in 2002 net als hun collega's in 1995 tot een gematigd negatief oordeel. De kennis van de leerlingen op het gebied van *Techniek* achtten ze sterk onvoldoende.

De vergelijking over de tijd | Over de periode 1991-1995 zijn geen jaareffecten van betekenis gevonden. Bij *Biologie* is de indeling naar onderwerpen in 2002 veranderd. Toch zijn er voor twee onderwerpen voldoende gemeenschappelijke opgaven gevonden, waardoor een vergelijking met 1995 mogelijk bleek. Er is een klein positief effect gevonden voor het onderwerp *Voortplanting en ontwikkeling*. Maar over het algemeen is het kennisniveau van de leerlingen nauwelijks veranderd.

Ook het kennisniveau van *Natuurkunde* is nauwelijks gewijzigd.

Effectgrootten voor de jaarvergelijkingen voor Wereldoriëntatie over de periode 1991-2000

Domein	Onderwerp	1991	1995	2000/02
Wereldoriëntatie algemeen	politieke verschijnselen	0	0,19	
	soc. culturele verschijnselen	0	0,28	
	economische verschijnselen	0	-0,08	
Geschiedenis	tijdsbesef / tijdsbegrip		0	-0,44
	oude geschiedenis	0	-0,34	-0,35
	nieuwe geschiedenis	0	0,28	-0,11
	samenleving		0	0,11
Aardrijkskunde	aardrijkskunde – nederland	0	0,13	
	aardrijkskunde – europa	0	-0,43	
	aardrijkskunde – wereld	0	0,48	
	natuurkundige aardrijkskunde	0	-0,28	
	kaartlezen		0	-0,68
	topografie Nederland	0	-0,01	-0,28
	topografie Europa	0	0,03	0,01
	topografie Wereld	0	0,15	-0,09
	aarde en landschappen		0	-0,04
	bevolking		0	0,05
Natuuronderwijs	oecologie en gedrag	0	-0,02	
	bouw en systematiek	0	-0,03	
	organismen, waarnemen en reageren		0	-0,04
	stofwisseling en kringloop		0	0,22
	voortplanting en ontwikkeling		0	0,22
	biotopen		0	-0,11
	natuurkunde	0	-0,1	-0,11
	techniek			

Legenda

- Positieve effectgrootte 0,2 – 0,5
- Negatieve effectgrootte -0,2 – -0,5
- Negatieve effectgrootte -0,5 – -0,8

5 Overzicht van peilingresultaten voor Engels in het basisonderwijs

Engels in het basisonderwijs hebben we gepeild in de jaren 1991, 1996 en 2006. In het laatste jaar lag het zwaartepunt op de hoofdvaardigheden lezen, luisteren en spreken, aangevuld met woordenschat. De beoordelaars hebben zich in 2006 niet uitgesproken over spreekvaardigheid. Een jaarvergelijking was evenmin mogelijk, doordat we de leerlingenprestaties anders zijn gaan waarderen.

Het onderwerp *Hanteren van een woordenboek* is opgenomen in de kerndoelen basisonderwijs, maar is niet gepeild. Dat komt, doordat het woordenboek Engels in het basisonderwijs nauwelijks gebruikt wordt. Meestal zit er een woordenlijst bij de gehanteerde methode. Bovendien is het aantal aangeleerde woorden gering

De standaarden | In 2006 hebben de beoordelaars minder negatief geoordeeld over het bereikte resultaat dan in 1996. Misschien komt dat door het aanvankelijke optimisme. Engels was begin jaren negentig voor het basisonderwijs een relatief jong leergebied. En dan hoop je wellicht, dat je de vruchten kunt plukken van de nieuwe methodes die gebaseerd zijn op de meest recente inzichten. En dat viel tegen. De beoordelaars hebben zich in 2006 gerealiseerd, dat de positie van het Engels nog bescheiden is. De beschikbare tijd is beperkt en de prioriteit betrekkelijk laag. De inspanningen voor het vak zijn de laatste tijd vooral gericht op meer aandacht voor Engels in de lagere leerjaren van het basisonderwijs. Je kunt nu dus nog geen wonderen verwachten in groep 8. En dat inzicht hebben de experts expliciet in hun oordelen betrokken.

Maar de milde houding houdt niet automatisch in, dat ze kritiekloos waren. In het bijzonder betreft dat de resultaten op de onderwerpen *Lezen* en *Woordenschat*.

Percentage leerlingen dat in 1996 en 2006 de standaarden Minimum en Voldoende bereikt voor onderwerpen Engels

Domein	Onderwerp	MIN '96	VOL '96	MIN '06	VOL '06
Engels	luisteren	84	46	95	66
	lezen	70	32	92	50
	woordenschat / woordbegrip	68	34	84	50
	spreken	64	33		
	woordenschat spreken	63	25		

Legenda

- 50-60% bereikt het niveau van de standaard Voldoende
- minder dan 50% bereikt het niveau van de standaard Voldoende

De vergelijking over de tijd | Voor zover vergelijking mogelijk was, kunnen we vaststellen dat de prestaties van de leerlingen over de gehele periode nauwelijks zijn veranderd waar het de vaardigheden *Luisteren, Lezen* en *Woordenschat* betreft.

Effectgrootten voor de jaarvergelijkingen voor Engels over de periode 1991-2006

Domein	Onderwerp	1991	1996	2006
Engels	luisteren	0	-0,18	-0,12
	lezen	0	-0,16	-0,1
	woordenschat / woordbegrip	0	0,05	-0,04
	spreken	0	-0,05	
	woordenschat spreken	0	0,02	
	uitspraak	0	-0,42	
	hanteren woordenboek	0	-0,62	
Legenda				
			Negatieve effectgrootte -0,2 – -0,5	
			Negatieve effectgrootte -0,5 – -0,8	

6 Conclusie, terugblik en vooruitblik

In deze publicatie is een beknopte presentatie gegeven van de wijze waarop we de afgelopen twintig jaren de kwaliteit van het onderwijs hebben gepeild. Beknopt in die zin dat we een globale samenvatting hebben gegeven van het onderwijspeil op basis van standaarden en op basis van effectgrootten voor vergelijkingen over de jaren voor enkele cruciale leergebieden. De standaarden laten zien wat ter zake deskundige en geïnformeerde beoordelaars vinden van de bereikte leerresultaten en de jaarvergelijkingen brengen de ontwikkeling van vaardigheden van leerlingen in het Nederlandse basisonderwijs tot uitdrukking.

Er is veel buiten beschouwing gebleven. De onderzoeken bijvoorbeeld voor Nederlandse taal en rekenen/wiskunde die zijn uitgevoerd bij leerlingen in jaargroep 5 (halverwege het basisonderwijs) en in het speciaal basisonderwijs. Ook de resultaten van peilingen voor Verkeersonderwijs, Bewegingsonderwijs, Muziek en Beeldende vorming ontbreken, evenals de meer specifieke onderzoeken naar Historische beeldvorming, Leesstrategieën en Handschriftkwaliteit. Het zou te ver voeren in het bestek van dit overzicht ook alle resultaten van deze peilingen op te nemen.

We komen tot een samenvatting, een korte terugblik en een blik vooruit.

De standaarden | We moeten constateren dat beoordelaars vaak kritisch, soms zeer kritisch oordeelden over de bereikte leerresultaten, met name voor de standaard Voldoende.

Zij hebben deze standaard vastgesteld in het licht van de kerndoelen basisonderwijs en vanuit het standpunt dat de meerderheid van de leerlingen – dat wil zeggen 70% tot 75% – aan het einde van het basisonderwijs het niveau van deze standaard bereikt zou moeten hebben. Vaak hebben we moeten constateren dat minder dan 60% en zelfs minder dan 50% van de leerlingen dit niveau haalt. Voor de standaard Minimum komen de beoordelaars wat meer tegemoet aan het feitelijke niveau van de zwakke leerlingen.

Deze kritische houding vinden we terug in de voortdurende discussie over de kwaliteit van het onderwijs. In feite ligt zij aan de basis van het project PPO, vanuit de behoefte aan die discussie over de onderwijskwaliteit een empirische basis te geven.

Peilingonderzoek verschaft die discussie – conform onze opdracht – die empirische basis.

Doordat de experts de lat hoog leggen, onderstrepen ze de ambitie om de kwaliteit van het onderwijs te verbeteren. De standaarden zijn in dit kader relevante referentieniveaus, want de deskundige experts kennen uit eigen ervaring en directe betrokkenheid het onderwijscurriculum van het basisonderwijs. Zij hebben onderling uitvoerig kunnen overleggen aan de hand van concrete opgaven. En uiteindelijk hebben ze hun oordeel kunnen toetsen aan de werkelijke leerresultaten.

Zijn de beoordelaars te ambitieus? Die vraag kan ieder voor zich beantwoorden door de opgaven te bekijken en te beoordelen. In de diverse balansen wordt uitgebreid met voorbeelden geïllustreerd wat van leerlingen op het niveau van de standaarden wordt verwacht.

De jaarvergelijkingen | Er is weinig of geen ontwikkeling over de laatste twintig jaar; niet in positief en niet in negatief opzicht voor het kennis- en vaardigheidniveau van de leerlingen. Dat is het algemene beeld, dat opdoemt uit onze jaarvergelijkingen. De enige systematische uitzondering op dit algemene beeld vinden we bij het domein rekenen/wiskunde. Daar zien we een negatieve trend bij de bewerkingsopgaven en een positieve trend bij de basale vaardigheden en procenten.

De signalering van zo'n negatieve trend onderstreept het belang van onze opzet en aanpak: een modulaire opbouw van het evaluatie-instrumentarium. Dat levert zinvolle en hanteerbare resultaten op. Een globale en brede toetsing van een leerstofgebied, die ter beperking van de toetslast en ter wille van directe feedback imperatief is voor leerlingenevaluatie, kan dergelijke differentiële effecten op onderdelen niet zichtbaar maken.

Terugblik | PPON is zo'n 20 jaar geleden gestart vanuit de opdracht systematisch bij te dragen aan het verkrijgen van een beeld van het leeraanbod en de leeropbrengsten van het onderwijs. We zijn van mening dat we aan die doelstelling ruimschoots inhoud hebben gegeven. We beschikken nu over onderbouwd cijfermateriaal voor de leeropbrengsten in het basisonderwijs en dat materiaal levert een empirische basis voor de maatschappelijke discussie over onderwijskwaliteit.

We zijn niet de enigen die op systeemniveau onderzoek verrichten naar de leeropbrengsten van het basisonderwijs. De resultaten van de onderzoeken van de verschillende projecten lijken niet altijd onderling consistent en eenduidig te zijn. Dat is vanzelfsprekend ongemakkelijk, maar te verklaren doordat er tussen de verschillende projecten geen enkele verbinding bestaat. Er zijn ook geen directe verklaringen voorhanden wanneer bepaalde discrepanties worden geconstateerd. Zo blijken we volgens internationale maatstaven goed te presteren op het gebied van rekenen/wiskunde, terwijl uit het peilingsonderzoek blijkt dat de vaardigheid van leerlingen op bepaalde onderwerpen achteruit gaat. Uit peilingsonderzoek op nationaal niveau blijkt dat de leesvaardigheid van leerlingen over de jaren vrij stabiel blijft, terwijl onze positie in internationaal onderzoek sterk fluctueert.

In steeds meer landen blijkt de behoefte te bestaan aan een nationaal peilinginstrument, dat fijnmaziger en gedetailleerder informatie verstrekt over de opbrengsten van het onderwijsstelsel, een toetssteen voor de onderwijsdoelen en -ambities op nationaal niveau.

Vooruitblik | Om de vijf tot tien jaren voeren we peilingonderzoek uit. Dat zijn eigenlijk te lange termijnen om de effectschommelingen voor de taal- en rekenprestaties te monitoren en af te zetten tegen het onderwijsbeleid. De recente kwaliteitsagenda 'Scholen voor morgen' signaleert dat terecht en biedt meer ruimte om deze leerdomeinen intensief te peilen op de ontwikkeling van de leeropbrengsten in het basisonderwijs.

De bestaande PPON-onderzoeken blijven van belang; denk alleen maar aan het inzicht over de lange termijn. Om een juiste invulling te geven aan die onderzoeksagenda zullen we leerdomeinen periodiek blijven peilen: Nederlandse taal en rekenen/wiskunde om de vijf jaar en de vakgebieden binnen wereldoriëntatie om de zeven à tien jaar.

Komend voorjaar verschijnen balansen over Engels en bewegingsonderwijs, over leesvaardigheid halverwege de basisschool (jaargroep 5) en over rekenen/wiskunde in het

speciaal basisonderwijs. In het najaar zal de balans verschijnen over de peilingonderzoeken voor Luistervaardigheid die in 2006 zijn uitgevoerd in jaargroep 5, in jaargroep 8 en in het speciaal basisonderwijs. In voorbereiding is een dieptestudie over het hanteren van oplossingsstrategieën van leerlingen halverwege het basisonderwijs op het gebied van rekenen/wiskunde.

In 2008 zijn nieuwe peilingonderzoeken gepland voor geschiedenis met specifieke aandacht voor de Tweede Wereldoorlog en de historische canon die door de commissie Ontwikkeling Nederlandse Canon (OCW, 2006) is opgesteld. Voor aardrijkskunde peilen we ondermeer de kennis van het thema 'Nederland leeft met water'.

In 2009 staat een nieuw peilingonderzoek voor Natuuronderwijs op de rol met extra aandacht voor het thema 'Voeding en gezondheid'. Intussen werken we aan peilingen voor burgerschapsvorming (2009) en schrijf- en spreekvaardigheid Nederlandse taal in 2009 en 2010.

Binnenkort publiceert de Expertgroep Doorlopende Leerlijnen Taal en Rekenen haar eindrapport. Daarin zullen de basiskennis en basisvaardigheden worden beschreven die nodig zijn om met succes het onderwijs te doorlopen. Ongetwijfeld ligt daarin een nieuwe uitdaging om op basis van peilingonderzoeken de kwaliteit van het onderwijs te monitoren.

Literatuur

Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit. Evaluatie van het basisonderwijs*. Z.p.: Inspectie van het Onderwijs.

Ministerie van Onderwijs, Cultuur en Wetenschappen (1993). *Besluit van 4 mei 1993, inhoudende vaststelling van de kerndoelen basisonderwijs (Besluit kerndoelen basisonderwijs)*, Staatsblad 264. 's-Gravenhage, Sdu.

Ministerie van Onderwijs, Cultuur en Wetenschappen (1998). *Kerndoelen basisonderwijs 1998. Over de relaties tussen de algemene doelen en kerndoelen per vak*. 's-Gravenhage, Sdu.

Ministerie van Onderwijs en Wetenschappen (1989). *Advies over de voorlopige eindtermen basisonderwijs*. 's-Gravenhage, OCW.

Ministerie van Onderwijs, Cultuur en Wetenschappen (2006). *Entoen.nu – De canon van Nederland*. Rapport van de commissie Ontwikkeling Nederlandse canon. 's-Gravenhage, OCW.

Notté, H. (red.) (2002). *Aardrijkskunde voor de basisschool. Een domeinbeschrijving als resultaat van een cultuurpedagogische discussie*. Arnhem, Cito.

Onderwijsraad (1999). *Zeker weten. Leerstandaarden als basis voor toegankelijkheid*. Adviesnr. 990449/352, 6 oktober 1999.

Schoot, F. van der (2001). *Standaarden voor kerndoelen basisonderwijs. De ontwikkeling van standaarden voor kerndoelen basisonderwijs op basis van resultaten uit peilingsonderzoek*. (Proefschrift Universiteit van Amsterdam). Arnhem, Cito.

Thijssen, J. (red.) (2002). *Natuuronderwijs voor de basisschool. Een domeinbeschrijving als resultaat van een cultuurpedagogische discussie*. Arnhem, Cito.

Wagenaar, H. (red.) (2002). *Geschiedenis voor de basisschool. Een domeinbeschrijving als resultaat van een cultuurpedagogische discussie*. Arnhem, Cito.

Wesdorp, H., H. van den Bergh, D.J. Bos, J.B. Hoeksma, R.J. Oostdam, J. Scheerens & B. Triesscheijn (1986). *De haalbaarheid van periodiek peilingsonderzoek; een voorstudie op het gebied van taalonderwijs in de lagere school*. Lisse, Swets & Zeitlinger.

Wijnstra, J.M. & P.R.L. van Dam (1980). *National and statewide assessment of educational progress in the USA; verslag van een literatuurverkenning en een studiereis*. Arnhem, Cito. Documentatiereeks nr. 24.

Wijnstra, J.M. (1982). Wat leren kinderen op school? *Pedagogische Studiën*, 59, 223-232 en 277-287.

Wijnstra, J.M. (red.) (1988). *Balans van het rekenonderwijs in de basisschool. Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs*. Arnhem, Instituut voor toetsontwikkeling (Cito). PPON-reeks nr. 1.

Overzicht van PPON-onderzoeksrapporten

Wijnstra, J.M. (red.) (1988). *Balans van het rekenonderwijs in de basisschool. Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs*. PPON-reeks nr. 1. Arnhem, Instituut voor Toetsontwikkeling.

Zwarts, M. (red.) (1990). *Balans van het taalonderwijs aan het einde van de basisschool. Uitkomsten van de eerste taalpeiling einde basisonderwijs*. PPON-reeks nr. 2. Arnhem, Cito Instituut voor Toetsontwikkeling.

Sijtstra, J. (red.) (1992). *Balans van het taalonderwijs halverwege de basisschool. Uitkomsten van de eerste taalpeiling medio basisonderwijs*. PPON-reeks nr. 3. Arnhem: Cito Instituut voor Toetsontwikkeling.

Vinjé, M. (red.) (1993). *Balans van het Engels aan het einde van de basisschool. Uitkomsten van de eerste peiling Engels einde basisschool*. PPON-reeks nr. 4. Arnhem, Cito Instituut voor Toetsontwikkeling.

Weerden, J. van (red.) (1993). *Balans van het wereldoriëntatieonderwijs aan het einde van de basisschool. Uitkomsten van de eerste peiling wereldoriëntatie einde basisonderwijs*. PPON-reeks nr. 5. Arnhem: Cito Instituut voor Toetsontwikkeling.

Balans van de verkeerseducatie aan het einde van de basisschool. Beknopte schets van de eerste peiling verkeerseducatie einde basisonderwijs. (1993) PPON-reeks nr. 6. Arnhem, Cito Instituut voor Toetsontwikkeling. (Is verschenen als brochure.)

Weerden, J.J. van & N. Veldhuijzen (1995). *Balans van het muziekonderwijs aan het einde van de basisschool. Uitkomsten van de eerste muziekpeiling einde basisonderwijs met een uitbreiding naar LOM- en MLK-scholen*. PPON-reeks nr. 7. Arnhem: Cito Instituut voor Toetsontwikkeling.

Bokhove, J, F. van der Schoot & T. Eggen (1996). *Balans van het rekenonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling rekenen/wiskunde einde basisonderwijs*. PPON-reeks nr. 8a. Arnhem: Cito Instituut voor Toetsontwikkeling.

Bokhove, J, F. van der Schoot & T. Eggen (1996). *Balans van het rekenonderwijs halverwege de basisschool 2. Uitkomsten van de tweede peiling rekenen/wiskunde medio basisonderwijs*. PPON-reeks nr. 8b. Arnhem: Cito Instituut voor Toetsontwikkeling.

Kraemer J.M., F. van der Schoot & N. Veldhuijzen (1996). *Balans van het rekenonderwijs in LOM-en MLK-scholen. Uitkomsten van de eerste peiling rekenen/wiskunde*. PPO-reeks nr. 8c. Arnhem: Cito Instituut voor Toetsontwikkeling.

Schoot, F. van der, C. Leijten, H. Verstralen & T. Bouw (1997). *Balans van de lichamelijke oefening in het basis- en speciaal onderwijs. Uitkomsten van de eerste peiling lichamelijke oefening aan het einde van de basisschool en in LOM- en MLK-scholen*. PPO-reeks nr. 9. Arnhem: Instituut voor Toetsontwikkeling.

Sijstra J. (red.) (1997). *Balans van het taalonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede taalpeiling einde basisonderwijs*. PPO-reeks nr. 10a. Arnhem: Cito Instituut voor Toetsontwikkeling.

Roosmalen, W., N. Veldhuijzen & G. Staphorsius (1999). *Balans van het taalonderwijs halverwege de basisschool 2. Uitkomsten van de tweede taalpeiling medio basisonderwijs*. PPO-reeks nr. 10b. Arnhem: Cito Instituut voor Toetsontwikkeling.

Roosmalen, W., N. Veldhuijzen & G. Staphorsius (1999). *Balans van het taalonderwijs in LOM- en MLK-scholen. Uitkomsten van de eerste taalpeiling in LOM- en MLK-scholen*. PPO-reeks nr. 10c. Arnhem: Cito Instituut voor Toetsontwikkeling.

Schoot, F. van der & N. Verhelst (1998). *Balans van het verkeersonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling voor verkeersonderwijs*. PPO-reeks nr. 11. Arnhem: Cito Instituut voor Toetsontwikkeling.

Wijnstra, J.M. (red.) (1999). *Balans van het wereldoriëntatieonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling wereldoriëntatie einde basisonderwijs*. PPO-reeks nr. 12. Arnhem: Cito Instituut voor Toetsontwikkeling.

Janssen, J., F. van der Schoot, B. Hemker & N. Verhelst (1999). *Balans van het reken-wiskunde-onderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1997*. PPO-reeks nr. 13. Arnhem: Cito Instituut voor Toetsontwikkeling.

Kraemer, J.M., F. van der Schoot & R. Engelen (2000). *Balans van het reken-wiskunde-onderwijs op LOM- en MLK-scholen 2. Uitkomsten van de tweede peiling in 1997*. PPO-reeks nr. 14. Arnhem: Cito Instituut voor Toetsontwikkeling.

Noteboom, A., F. van der Schoot, J. Janssen & N. Veldhuijzen (2000). *Balans van het reken-wiskunde-onderwijs halverwege de basisschool 3. Uitkomsten van de derde peiling in 1997*. PPO-reeks nr. 15. Arnhem: Cito Instituut voor Toetsontwikkeling.

Weerden, J. van & N. Veldhuijzen (2000). *Balans van het muziekonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling in 1997*. PPO-reeks nr. 16. Arnhem: Citogroep.

Edelenbos, P., F. van der Schoot & H. Verstralen (2000). *Balans van het Engels aan het einde van de basisschool 2. Uitkomsten van de tweede peiling in 1996*. PPON-reeks nr. 17. Arnhem: Citogroep.

Hermans, P., F. van der Schoot, C. Sluiter & N. Verhelst (2000). *Balans van de peiling Beeldende Vorming aan het einde van de basisschool. Uitkomsten van de tweede peiling in 1996*. PPON-reeks 18. Arnhem: Citogroep.

Sijtstra J., F. van der Schoot & B. Hemker, B (2002). *Balans van het taalonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1998*. PPON-reeks 19. Arnhem: Citogroep.

Berkel, S. van, F. van der Schoot, R. Engelen & G. Maris (2002). *Balans van het taalonderwijs halverwege de basisschool 3. Uitkomsten van de derde peiling in 1999*. PPON-reeks 20. Arnhem, Citogroep.

Weerden, J. van, T. Bechger & B. Hemker (2002). *Balans van het taalonderwijs in het speciaal basisonderwijs 2. Uitkomsten van de tweede peiling in 1999*. PPON-reeks 21. Arnhem, Citogroep.

Schoot, F. van der, T. Bechger (2003). *Balans van handschriftkwaliteit in het primair onderwijs. Uitkomsten van de peilingen in 1999*. PPON-reeks nr. 22. Arnhem, Citogroep.

Wagenaar, H., F. van der Schoot & B. Hemker (2003). *Balans van het geschiedenisonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 2000*. PPON-reeks nr. 23. Arnhem, Citogroep.

Notté, H., F. van der Schoot & B. Hemker (2003). *Balans van het aardrijkskundeonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 2001*. PPON-reeks nr. 24. Arnhem, Citogroep

Thijssen, J., F. van der Schoot & B. Hemker (2003). *Balans van het biologieonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 2001*. PPON-reeks nr. 25. Arnhem, Citogroep.

Thijssen, J., F. van der Schoot, N. Verhelst & B. Hemker (2004). *Balans van het natuurkunde- en techniekonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling 2002*. PPON-reeks nr. 26. Arnhem, Citogroep.

Wagenaar, H. & F. van der Schoot (2004). *Balans van de historische beeldvorming aan het einde van de basisschool. Uitkomsten van het peilingsonderzoek in 2000*. PPON-reeks nr. 27. Arnhem, Citogroep.

Krom, R., N. Verhelst & N. Veldhuijzen (2004). *Balans van het schrijfvaardigheidsonderwijs op de basisschool. Uitkomsten van de peilingen in 1999: halverwege en einde basisonderwijs en speciaal basisonderwijs*. PPON-reeks nr. 28. Arnhem, Citogroep.

Gein, J. van de (2005). *Balans van taalkwaliteit in schrijfwerk uit het primair onderwijs*. PPON-reeks nr. 29. Arnhem, Citogroep.

Weerden, J. van, K. Heesters, I. Jongen, F. van der Schoot, B. Hemker, N. Veldhuijzen, N. Verhelst (2006). *Balans van het spreekonderwijs op de basisschool. Uitkomsten van de peilingen in 2002 en 2003 halverwege en einde basisonderwijs en speciaal basisonderwijs*. PPON-reeks 30, Arnhem, Cito.

Kraemer, J.-M., J. Janssen, F. van der Schoot & B. Hemker (2005). *Balans van het reken-wiskundeonderwijs halverwege de basisschool 4. Uitkomsten van de vierde peiling in 2003*. PPON-reeks nr. 31. Arnhem, Citogroep.

Janssen J., F. van der Schoot, B. Hemker (2005) *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Uitkomsten van de vierde peiling in 2004*. PPON-reeks nr. 32. Arnhem, Citogroep.

Schoot, F. van der (2006). *Inventarisatie verkeersonderwijs op de basisschool in 2006*. PPON-rapport. Arnhem, Cito.

Heesters, K., S. van Berkel, F. van der Schoot & B. Hemker (2007). *Balans van het leesonderwijs aan het einde van de basisschool 5. Uitkomsten van de vierde peiling in 2005*. PPON-reeks nr. 33. Arnhem, Cito.

Heesters, K., S. van Berkel, R. Krom, F. van der Schoot & B. Hemker (2007). *Balans van het leesonderwijs in het speciaal basisonderwijs 3. Uitkomsten van de derde peiling in 2005*. PPON-reeks nr. 34. Arnhem, Cito.

Moelands, F., I. Jongen, F. van der Schoot & B. Hemker (2007). *Balans over leesstrategieën in het primair onderwijs. Uitkomsten van de eerste peiling in 2005*. PPON-reeks nr. 35. Arnhem, Cito.

