

Wat leren kinderen op school? (I)*

Verslag van een literatuurstudie en een studiereis ter verkenning van de mogelijkheden van een nationaal evaluatieproject in het funderend onderwijs

J. M. WIJNSTRA

Centraal Instituut voor Toetsontwikkeling
(Cito), Arnhem

Samenvatting

De laatste jaren is in algemeen-maatschappelijke, politieke en onderwijskundige discussies herhaaldelijk de vraag gesteld: Wat leren kinderen nu eigenlijk op school? (vgl. bijv. de recente nota van de Bewindslieden van Onderwijs en Wetenschappen, getiteld *Kwaliteit van het onderwijs*). In zijn algemeenheid betreft deze vraag de kwaliteit en het niveau van het onderwijs, die echter op verschillende manieren kan worden uitgelegd. Enerzijds wordt gevraagd naar de geldende, respectievelijk de feitelijk nagestreefde doelstellingen en het daarop afgestemde vormingsaanbod, anderzijds wordt gevraagd wat de leerlingen van het genoten onderwijs hebben opgestoken.

Het Cito heeft aan dit onderwerp in 1979 en 1980 een voorstudie gewijd waarvan in dit artikel (in twee delen) verslag wordt gedaan. Naast literatuurstudie is ook een studiereis naar de Verenigde Staten gemaakt om nader kennis te maken met het project *National Assessment of Educational Progress (NAEP)* en enkele state assessment-programs. Deze projecten zijn bedoeld om informatie te leveren over wat kinderen kunnen en weten, en het vaststellen van veranderingen daarin, ten dienste van het beleid op verschillende niveaus.

In dit eerste deel van het artikel wordt, na de algemene inleiding, de achtergrond en opzet van NAEP beschreven: de doelstellingen, de instrumentontwikkeling, de steekproeftrekking en de afname, de verwerking en de rapportage van de gegevens. Uitvoerig wordt ingegaan op een tweetal externe evaluaties van het project: een

onderzoek van de *Carnegie Corporation*, die in de beginjaren veel geld in het project heeft gestoken, en een onderzoek van de Amerikaanse rekenkamer. In beide onderzoeken wordt getwijfeld of de oorspronkelijke doelstellingen van het project, het vormen van een informatiebron voor het identificeren van problemen, het stellen van prioriteiten en het vaststellen van veranderingen ten behoeve van beleidsinstanties, curriculumontwikkelaars, onderzoekers en practici, voldoende gerealiseerd worden. Tenslotte wordt ingegaan op veranderingen die de laatste jaren hebben plaatsgevonden, zowel naar aanleiding van de kritiek als naar aanleiding van budgetbeperkingen (een vermindering van bijna een miljoen dollar per jaar op een budget van oorspronkelijk bijna vijf miljoen dollar).

In het tweede deel van het artikel zal aandacht worden geschonken aan de ontwikkelingen in de afzonderlijke Amerikaanse staten, met bijzondere en uitvoerige aandacht voor het state assessment-program in Minnesota. In de slotparagraaf van het artikel worden een aantal overwegingen m.b.t. een educational assessment-project in Nederland gepresenteerd. Een dergelijk project wordt geplaatst in het kader van (de beleidsbepaling m.b.t.) de basisvorming in het funderend onderwijs.

1 Algemene inleiding

De laatste jaren is in algemeen-maatschappelijke, politieke en onderwijskundige discussies herhaaldelijk de vraag gesteld: Wat leren kinderen nu eigenlijk op school? In zijn algemeenheid betreft deze vraag de kwaliteit en het niveau van het onderwijs, die echter op verschillende manieren kan worden uitgelegd. Enerzijds wordt gevraagd naar de geldende, respectievelijk de feitelijk nagestreefde doelstellingen (of eindtermen, zo men wil) en het daarop afgestemde vormingsaanbod, anderzijds wordt gevraagd wat de leerlingen van het genoten onderwijs hebben opgestoken, of anders gezegd: het rendement van het gegeven onderwijs.

Reeds in 1969 heeft de Stichting voor On-

* Met dank aan P. R. L. van Dam, Hoofd van de Afdeling Basisonderwijs van het Cito, tevens mijn gewaardeerde reisgenoot tijdens de studiereis naar de Verenigde Staten.

derzoek van het Onderwijs (SVO) in de nota Opstaan tegen het zittenblijven in verband met de vereiste waarborgen voor de kwaliteit en effectiviteit van het onderwijs gewezen op het belang van het tot ontwikkeling brengen van vormen van onderwijskundige voortgangs- en kwaliteitscontrole. Men dacht hierbij aan een centrale registratie van een periodieke vorderingeninventarisatie die grotendeels door het onderwijs zelf zou moeten worden uitgevoerd (Doornbos, 1969, p. 92). Hiernaar verwijzend heeft de Innovatiecommissie Basisschool in haar Innovatieplan basisonderwijs in verband met de verdere beleidsontwikkeling ten aanzien van doel en inhoud van de basisschool gepleit voor een periodiek onderzoek naar de kwalitatieve ontwikkeling van het basisonderwijs, waarbij het accent zou moeten liggen op een inventarisatie en publicatie van het vormingsaanbod (Innovatiecommissie Basisschool, 1978). In zijn beleidsnotitie bij het genoemde innovatieplan heeft de Minister van Onderwijs en Wetenschappen positief gereageerd op deze aanbeveling. In de Memorie van Toelichting bij de onderwijsbegroting voor 1980 is vervolgens een nota aangekondigd over de problematiek van de kwaliteit en het niveau van het onderwijs¹. Ter verkenning van de mogelijkheden van een periodiek onderzoek in het basisonderwijs is ongeveer tegelijkertijd een 'technisch vooroverleg' van de grond gekomen waarin naast het ministerie (met inbegrip van de inspectie) de SVO, de SLO en het Cito participeren.

Mede naar aanleiding van een reisverslag van Bos (1977), die o.a. een bezoek had gebracht aan het Amerikaanse project National Assessment of Educational Progress, werd reeds in 1977 in de Beleidsgroep van het Cito een memorandum besproken over de peiling van de opbrengst van het basisonderwijs. Als uitvloeisel van deze bespreking werd in het werkplan voor 1979 een Voorstudie National Assessment² opgenomen om de mogelijkheden van een dergelijk project te bestuderen. Om uitvoering te geven aan de voorstudie werd het wenselijk geacht om – naast de bestudering van de beschikbare literatuur – in de Verenigde Staten kennis te nemen van het hier te lande tamelijk onbekende project National Assessment of Educational Progress (verder aangeduid als NAEP), met name ook van veranderingen die na 1976 zouden hebben plaatsgevonden. Het reisprogramma (6-20 juni 1979)

werd zodanig gepland dat in Denver kon worden deelgenomen aan de jaarlijkse door NAEP georganiseerde Large-Scale Assessment Conference. Omdat over de assessment-programs in diverse Amerikaanse staten nauwelijks informatie beschikbaar was, werd in het reisprogramma ook een bezoek aan een drietal state assessment-programs opgenomen (Minnesota, Connecticut en Rhode Island). Vanwege de federale staatsvorm en de grote schaalverschillen tussen de Verenigde Staten en Nederland werd verondersteld dat de state assessment-programs wel eens net zo relevant voor de Nederlandse situatie zouden kunnen zijn als NAEP. Ter afsluiting van de studiereis werd in Washington D.C. gesproken met medewerkers van het National Institute of Education dat sinds begin 1979 de federale supervisie over NAEP heeft.

In dit artikel wordt in par. 2 het Amerikaanse project National Assessment of Educational Progress beschreven, gevolgd (in het tweede deel) door een beschrijving van de ontwikkelingen in enkele Amerikaanse staten, en met name Minnesota. In de slotparagraaf wordt nader ingegaan op de betekenis die een assessment-project in Nederland zou kunnen hebben.

2 *National Assessment of Educational Progress*

2.1 *Inleiding*

De lancering van de eerste Russische kunstmaan (Sputnik) in 1957 veroorzaakte in de Verenigde Staten veel ongerustheid over het peil van het onderwijs. Hoewel de kritiek zich in eerste instantie voornamelijk richtte op het wis- en natuurkunde-onderwijs, richtte die zich later meer op het onderwijssysteem als geheel (zie bijv. Rickover, 1963). Wilsey & Schroeder (1974) noemen de volgende punten van kritiek:

- toenemende kosten, met weinig of geen aanwijzingen dat meer geld tot beter onderwijs leidt;
- onduidelijkheid over de doelstellingen van het onderwijs;
- inefficiëntie bij de uitvoering van het onderwijs;
- gebrek aan gegevens over de effectiviteit van het onderwijs.

Mede naar aanleiding van deze publieke be-

zorgdheid over de kwaliteit van het onderwijs, waarbij ook het financiële aspect ruime aandacht kreeg, ontwikkelde zich in de loop van de zestiger jaren de 'accountability movement', die in de zeventiger jaren zijn hoogtepunt bereikte. Deze beweging is erop gericht dat het onderwijs publiekelijk rekening en verantwoording aflegt over de nagestreefde doelstellingen, de effectiviteit van de ondernomen pogingen om die doelstellingen te bereiken en de efficiëntie waarmee dat gebeurt. In dit kader werden verscheidene, ten dele aan het bedrijfsleven ontleende technieken aangepast voor gebruik in het onderwijs, zoals Planning-programming-budgeting system (PPBS), Management by objectives, Performance contracting.

Het past in dit beeld dat ook de federale overheid zich zorgen ging maken over het niveau en de kwaliteit van het onderwijs, hoewel de regeling van het onderwijs vrijwel geheel was overgelaten aan de afzonderlijke staten³. Niettemin was het federale Office of Education in 1867 als taakstelling o.a. meegegeven 'to collect statistics and facts showing the condition and progress of education in the several states and territories' (geciteerd naar Greenbaum et al., 1977, p. 9). Dit was voor Francis Keppel (Commissioner of Education, 1962-1965) aanleiding om na te gaan of het mogelijk was ook gegevens te verzamelen over het inhoudelijk functioneren van het onderwijs, buiten de gebruikelijke gegevens over het aantal leerlingen per leerkracht, de kosten per leerling, etc.

Op verzoek van Keppel organiseerde de Carnegie Corporation eind 1963 en begin 1964 een tweetal conferenties met een aantal kopstukken uit de onderwijswereld. Hieruit kwam het Exploratory Committee on Assessing the Progress of Education (ECAPE) voort, onder voorzitterschap van de bekende Ralph Tyler. Al spoedig had men zich onder leiding van Tyler (en Tukey) een beeld gevormd van de opzet van een assessment-project. Greenbaum et al. (1977, p. 11, 12) vatten dit beeld als volgt samen:

- Het project richt zich op het meten van kennis en vaardigheden in het algemeen en niet alleen op wat men op school heeft geleerd, met als uitgangspunt dat de opzet zo breed moet zijn dat de Amerikaanse opvoedings- en onderwijsdoelen weerspiegeld worden, zonder dat het onderwijs wordt verstoord.
- In het project worden toetsen van verschil-

lende moeilijkheidsgraad afgenomen bij schoolkinderen van verschillende leeftijd (tegen het einde van de elementary school, junior highschool en senior highschool) en bij volwassenen.

- Omdat het gaat om de vaststelling van het niveau van categorieën van personen, en niet van individuele personen of identificeerbare groepen van personen (bijv. steden of staten), kan een matrix sampling design – waarbij een deel van de proefpersonen een deel van de opgaven maakt – worden toegepast, zodat een gering beslag op de (onderwijs)tijd wordt gelegd.
- In plaats van een normgerichte interpretatie wordt een domeingerichte interpretatie toegepast.

De levensduur van ECAPE werd oorspronkelijk op vijf maanden geschat. Daarna zou het project onder federale supervisie moeten worden gebracht. De geschatte vijf maanden werden uiteindelijk vijf jaar. De oppositie tegen het project kwam vooral van de afzonderlijke staten en de American Association of School Administrators en had betrekking op de veronderstelde federale controle op het onderwijs die door het project mogelijk zou worden. Het kostte erg veel tijd om deze oppositie te overwinnen en een instantie te vinden die het project onder zijn hoede zou willen nemen (zie Greenbaum et al., 1977 en Hazlett, z.j.). Dit neemt niet weg dat in de periode na 1964 ook reeds aan de inhoudelijke voorbereiding werd gewerkt. Omdat in 1968 nog geen uitvoerder was gevonden, terwijl in 1969 het project operationeel zou worden, werd per 1 juli 1968 ECAPE omgezet in CAPE en de uitvoering van het project ging van start onder haar auspiciën en met geld uit privéfondsen (voornamelijk Carnegie Corporation en Ford Foundation). In de loop van 1969 werd de Education Commission of the States, een niet-federaal samenwerkingsverband van de afzonderlijke staten met een kantoor in Denver, bereid gevonden om het project over te nemen. De financiering van het project National Assessment of Educational Progress werd vanaf dat moment overgenomen door het U.S. Office of Education.

In deze paragraaf zullen we verder de opzet en ontwikkeling van NAEP van 1969-1979 beschrijven, waarbij we twee periodes onderscheiden: 1969-1974 en 1975-1979. Hoewel deze indeling tamelijk arbitrair is, is het voor

ons doel toch wel zinvol om in het midden van de zeventiger jaren een grens te trekken en wel om de volgende redenen:

- eind 1974 was de eerste ronde van nagenoeg alle tien vakken (learning areas) achter de rug;
- budgettaire beperkingen maakten het na die tijd erg moeilijk om nog langer volwassenen in het onderzoek te betrekken;
- de federale supervisie ging over van het Office of Education naar het National Center for Education Statistics;
- gebaseerd op de ervaringen tot en met 1974 verscheen in 1976 een rapport van het General Accounting Office (de Amerikaanse rekenkamer) onder de titel *The National Assessment of Educational Progress: its results need to be made more useful*, waaruit aanwijzingen voor veranderingen voortkwamen;
- in 1977 verscheen een rapport van de Carnegie Corporation (Greenbaum et al., 1977) afgesloten begin 1973, met een antwoord van NAEP uit 1975;
- in 1975 werd de tegenwoordige directeur, Roy H. Forbes, aangesteld.

Begin 1979 is de federale supervisie van het National Center for Education Statistics overgegaan naar het National Institute of Education. Tegelijkertijd werd een nieuwe 'grant competition' geopend. Hierdoor zal het jaar 1979 ook een grensjaar worden. Hoewel de Education Commission of the States het project wel zal voortzetten, zullen er toch veranderingen optreden vanwege budgettaire redenen. Het voorziene subsidie zal vanaf 1980 \$ 3.900.000,- bedragen, terwijl de voorgaande

jaren het budget bijna een miljoen dollar hoger was.

In par. 2.2 wordt de opzet van het project besproken zoals die er aan het eind van de eerste fase uitzag. We baseren ons hierbij vooral op het *General Information Yearbook* (1974). Vervolgens komen in par. 2.3 de rapporten van het *General Accounting Office* (1976) en *Greenbaum et al.* (1977) aan de orde. In par. 2.4 wordt aandacht geschonken aan recente ontwikkelingen.

2.2 De opzet van NAEP

NAEP ging van start met de volgende hoofddoelstellingen (*General Information Yearbook*, 1974, p. 2):

- '- to make available on a continuing basis comprehensive data on the educational attainments of young Americans;
- to measure change in the educational attainments of young Americans.'

Het totale te bestrijken vormingsgebied werd als volgt ingedeeld, gedeeltelijk overeenkomend met de gebruikelijke (Amerikaanse) schoolvakken; art; career and occupational development; citizenship; literature; mathematics; music; reading; science; social studies; writing.

Niet alleen de cognitieve kant van deze vormingsgebieden zou aandacht moeten krijgen, maar ook de affectieve en psychomotorische. In overeenstemming daarmee werd het opgaventype (NAEP spreekt van *exercises*) niet beperkt tot alleen maar meerkeuze-opgaven. Wat de kennis- en toepassingsopgaven betreft zou gestreefd moeten worden naar een grote spreidingsbreedte in moeilijkheidsgraad (p-

Tabel 1 *National Assessment Timetable 1969-1980*

Learning area	Initial or baseline assessment	First measurement of change	Second measurement of change
Art	1974/75	1978/79	
Career and occupational development	1973/74		
Citizenship	1969/70	1975/76	
Literature	1970/71	1979/80	
Mathematics	1972/73	1977/78	
Music	1971/72	1978/79	
Reading	1970/71	1974/75	1979/80
Science	1969/70	1972/73	1976/77
Social studies	1971/72	1975/76	
Writing	1969/70	1973/74	1978/79

waarden van <.25 tot >.90). De peiling zou uitgevoerd moeten worden bij 9-, 13- en 17-jarigen, en jonge volwassenen in de leeftijd van 26 tot 35 jaar.

In 1969 werd gestart met citizenship, science en writing. Uit Tabel 1 is af te lezen hoe het tijdschema er verder uit zag.

Instrumentontwikkeling

De instrumentontwikkeling doorloopt in principe de normale stadia:

- opstellen van de lijst van leerdoelen;
- samenstelling van de opgaven;
- beoordeling van de opgaven;
- proefafname;
- definitieve selectie van opgaven.

In de beginfase werd het opstellen van de leerdoelenlijst en het samenstellen van de opgaven uitbesteed. Omdat dit niet bevredigend verliep - NAEP had er te weinig greep op - zijn deze onderdelen naderhand onder directe supervisie van NAEP gebracht. Zoals gebruikelijk worden bij de opstelling van de leerdoelenlijst en de beoordeling van de opgaven praktijkmensen en onderwijs- en toetsdeskundigen ingeschakeld. De praktijkmensen worden meestal voorgedragen door vakorganisaties van bijv. science of social studies teachers. Hun aantal is afhankelijk van het onderwerp. Voor social studies, een moeilijk onderwerp, zijn het er enkele honderden. Met de praktijkmensen worden regionale conferenties belegd om de leerdoelenlijsten en de opgaven te bespreken.

Gezien de doelstelling van NAEP worden ook mensen van buiten het onderwijs ingeschakeld (concerned lay citizens). Ook hiermee worden conferenties gehouden over de leerdoelenlijsten en de ontworpen opgaven. In de 'normale' toetsontwikkelingspraktijk is dit vrij ongebruikelijk, maar binnen de NAEP-uitgangspunten een essentieel onderdeel vanwege de maatschappelijke legitimering.

Steekproeftrekking en afname

De bedoeling van NAEP is om een schatting te maken van het beheersingsniveau van de totale populatie van 9-, 13- en 17-jarigen en jonge volwassenen, en van segmenten uit deze populaties. We zullen hier alleen de procedure voor de steekproeftrekking van leerlingen die nog op school zitten, kort bespreken. De procedure voor de steekproeftrekking van 17-jarigen die niet meer op de highschool zitten (early graduates en drop outs) en jonge volwassenen ver-

loopt analoog, maar verschilt in de tweede en derde fase wat betreft de lokalisatie van de in aanmerking komende respondenten.

Ten behoeve van NAEP zijn de Verenigde Staten verdeeld zg. primary sampling units (PSU's) gebaseerd op volkstellingsgegevens over geografische eenheden. In de eerste fase wordt hieruit een steekproef van ruim 100 PSU's getrokken, gestratificeerd naar regio in de Verenigde Staten (noordoost, zuidoost, centraal en west), en binnen de regio's naar grootte van de plaats. Bij enkele van deze categorieën wordt in de eerste fase ook al gestratificeerd naar sociaal-economisch niveau van de plaats. Anders gebeurt dit in de tweede fase. Hierbij wordt een oververtegenwoordiging van de laagste strata nagestreefd om voldoende waarnemingen in deze strata te krijgen.

In de tweede fase wordt uit de in de eerste fase getrokken PSU's een steekproef van scholen getrokken, waarna in de derde fase een steekproef uit de in aanmerking komende leerlingen wordt getrokken. De omvang van de steekproeven in de tweede en derde fase is afhankelijk van het aantal afnamepakketten (packages) dat voor een leeftijdsgroep beschikbaar is. Als de packages groepsgewijs worden afgenomen is het richtgetal ongeveer 2600 leerlingen en bij de individuele afname ongeveer 2200.

Ieder afnamepakket neemt ongeveer 35 minuten in beslag, terwijl daarnaast 15 minuten voor inleiding en verzameling van additionele gegevens beschikbaar zijn (opleidingsniveau van de ouders, geslacht, huidskleur, etc.). Normaal gesproken bevat ieder pakket opgaven uit verschillende learning areas, van verschillende moeilijkheidsgraad. Behalve bij lezen staan alle instructies ook op geluidsband met pauzes voor het invullen van de antwoorden. De afname, individueel of in groepen van twaalf, wordt geleid door de district supervisor of door hem aangestelde en getrainde tijdelijke krachten. De district supervisor verzorgt ook het contact met de scholen wat betreft de organisatie e.d., nadat deze vanuit Denver zijn aangeschreven. Om de anonimiteit te garanderen blijft de lijst met namen van leerlingen en NAEP-identificatienummers op de scholen. Na zes maanden wordt deze vernietigd.

De gehele procedure van steekproeftrekking en dataverzameling wordt uitbesteed aan het Research Triangle Institute.

Verwerking en rapportage

Na de dataverzameling wordt de scoring, tot en met de produktie van de data tapes, verzorgd door het Measurement Research Center, een onderdeel van de Westinghouse Learning Corporation. Voor de scoring van de opgaven met open antwoorden is hier een speciaal getrainde staf aanwezig. De eigenlijke verwerking gebeurt door de staf van NAEP, ten dele op de eigen computer, ten dele op de computer van Princeton University. In verband met de disproportionale steekproeftrekking levert het Research Triangle Institute na de afname de benodigde gewichten om gewogen nationale en regionale gemiddelden te kunnen berekenen.

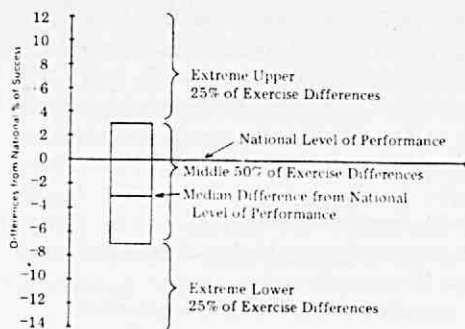
De eenheid van verwerking is de afzonderlijke opgave. De basisgegevens die opgeleverd worden zijn p-waarden voor de totale populatie en subpopulaties en Δp -waarden, verschillen tussen subpopulaties en de totale populatie. Een Δp -waarde wordt als significant aange merkt als deze tweemaal zo groot is als de standaardmeefout. In de technische rapporten worden voor alle opgaven de p-waarden en Δp -waarden vermeld voor de volgende rapportagegroepen:

- regio's (noordoost, zuidoost, centraal, west);
- geslacht;
- huidskleur;
- opleidingsniveau van de ouders;
- grootte en type van de woonplaats.

De helft van de opgaven wordt bij de rapportage vrijgegeven en in het technische rapport opgenomen. De andere helft wordt achtergehouden om in een volgende assessmentronde opnieuw gebruikt te worden, zodat verschillen tussen twee rondes aangegeven kunnen worden.

In de samenvattende rapporten wordt over groepen samenhangende opgaven gerapporteerd, waarbij de mediane Δp en eventueel de kwartieldeviatie van Δp wordt aangegeven zoals in Figuur 1 schematisch wordt weergegeven.

Naast de technische en samenvattende rapporten worden er thematische rapporten geproduceerd, bijv. over aspecten van lezen, vooral bedoeld voor leerkrachten en organisaties op dat gebied. NAEP zelf onthoudt zich van interpretaties, conclusies en aanbevelingen. Wel worden organisaties aangemoedigd om dergelijke commentariërende rapporten naar aanleiding van NAEP-gegevens te produceren. NAEP organiseert hiertoe eventueel



Figuur 1 Schematische weergave van de rapportage over afwijkingen van nationale gemiddelden over groepen samenhangende opgaven

conferenties om de mensen bij elkaar te brengen en zorgt voor reproductie van het rapport. Tenslotte vindt de verspreiding van de gegevens ook plaats via samenvattende populaire rapporten, persbulletins en de eigen Newsletter.

De assessment-cyclus

Voor een eerste assessment-ronde van een learning area werd oorspronkelijk vijf à zes jaar uitgetrokken: drie tot drie en een half jaar voor de voorbereiding, één tot anderhalf jaar voor de dataverzameling en één jaar voor de verwerking en de basisrapportage. De tweede assessmentcyclus van een learning area is uiteraard korter (drie à vier jaar), omdat de voorbereidingsperiode korter kan zijn.

Service aan externe organisaties

Enkele jaren na de start is binnen NAEP een Utilization and adaptation-sectie gecreëerd om diensten te kunnen verlenen aan met name de state departments of education, die in hun staat van de NAEP-materialen en -procedures gebruik wilden maken. Deze diensten omvatten consultatie, advisering, technische hulp en het organiseren van workshops en conferenties. Hieruit is ook de jaarlijkse Large-Scale Assessment Conference voortgekomen die in 1979 voor de negende maal werd gehouden.

Naast deze gratis service aan overheidsorganen zijn in de loop van de zeventiger jaren, ten dele in opdracht van andere organen, ook de zg. special probes op gang gekomen. In 1974 en 1975 is bijv. een onderzoek gedaan naar 'func-

tional literacy' bij 17-jarigen in opdracht van de organisatie Right to Read.

2.3 Externe evaluatie van NAEP na de eerste fase

In de voorbereidingsfase van NAEP heeft de Carnegie Corporation ruim 2.5 miljoen dollar beschikbaar gesteld. Toen de Carnegie Corporation in 1971 besloot om een aantal grotere projecten te gaan evalueren, viel de eerste keus op NAEP. De studie werd in 1972 en 1973 uitgevoerd door William Greenbaum van Harvard University, in samenwerking met Michael Garet en Ellen Solomon. Hun verslag, *Measuring educational progress*, verscheen uiteindelijk in 1977, voorzien van een antwoord van NAEP. Voor hun conceptuele analyse maakten Greenbaum et al. (1977) gebruik van documenten uit de voorbereidingsfase en de eerste rapporten van NAEP. Een belangrijke constatering met betrekking tot de voorbereidingsfase is, dat men ervan uitging dat meer informatie over het onderwijs zonder meer nuttig zou zijn, maar dat men bijzonder weinig aandacht heeft besteed aan het nader specificeren van de doelstellingen. Daardoor liepen ook binnen ECAPE de gedachten over de gebruikswaarde en de toepassingsgebieden van de informatie uiteen (beschrijving, resp. verklaring van de stand van zaken, zwakke plekken opsporen, passende beleidsmaatregelen nemen, etc.). We zullen de kritiek van Greenbaum et al. (1977) verder kort samenvatten.

In de eerste opzet van NAEP was vastgelegd dat het niet alleen om schoolse kennis moest gaan. Desalniettemin werd bij de ordening van de leer- en vormingsgebieden voornamelijk uitgegaan van de traditionele indeling in schoolvakken. Om moeilijke onderwerpen als self concept, sociale redzaamheid, 'good' citizenship, e.d. onder te brengen, werd aan de indeling citizenship toegevoegd, terwijl career and occupational development een uitgebreider betekenis kreeg dan in het Amerikaanse onderwijs gebruikelijk is. Toch werd de blik van NAEP daarmee minder ruim dan oorspronkelijk bedoeld, en vooral gericht op wat de scholen proberen te bereiken. Dit punt heeft voor veel verwarring gezorgd, met name ook bij de ontwikkeling van de doelstellingenlijsten en de opgaven, omdat niet altijd duidelijk was of het om actuele, en feitelijk nagestreefde doelstellingen moest gaan of om onderwijskundig en maatschappelijk wenselijke. Een

aardig voorbeeld hiervan is de eerste writing assessment. Hoewel dat niet gepland was, drongen de conferenties met belangstellende burgers erop aan om ook de mechanics (spelling, interpunctie, e.d.) op te nemen. Dit is in eerste instantie niet overgenomen, maar vijf jaar later is er wel alsnog over gerapporteerd.

Greenbaum et al. (1977) zetten ook verder nogal wat vraagtekens bij de ontwikkelingsprocedures van de doelstellingenlijsten en de opgaven. De inbreng van de vakdeskundigen, onderwijskundigen en leken was volgens hen niet al te groot. NAEP zocht vooral instemming en gebruikte deze mensen niet als informanten. Het is in dit verband veelbetekenend dat de onderwijsmensen voor een groot deel vakbondsbonzen waren, die vrijgesteld waren van hun onderwijstaak. Hoewel een aantal doelstellingenlijsten (bijv. reading) en opgaven (bijv. de scienceproeven) volgens Greenbaum van uitstekend gehalte zijn, kleven er aan de instrumenten ook bezwaren. Er heeft geen weging van objectives en subobjectives plaatsgevonden, waardoor er vrij willekeurige opgavenverzamelingen van 'objectives related exercises' zijn ontstaan.

NAEP levert in eerste instantie een beschrijving van de stand van zaken, hoewel sommige mede-initiatiefnemers wel wat meer verwachtten, volgens Greenbaum et al. (1977). In dit verband wordt een kritische kanttekening geplaatst bij de methode van matrix sampling en de samenstelling van de afnamepakketten. Door de gekozen opzet is het uitgesloten om over series verwante opgaven verbanden met achtergrondvariabelen na te gaan. Dit zou van belang zijn in verband met het genereren van vraagstellingen en hypothesen voor dieptestudies.

Alles bij elkaar zijn Greenbaum et al. (1977) nogal sceptisch over de hele operatie en zij verwachten van de vergelijking van een volgende ronde met een voorafgaande niet veel. Het opleveren van 'meaningful data' is huns inziens slecht gelukt. Over de rapportage merken zij in dit verband op (Greenbaum et al., 1977, p. 142): 'Yet the several substantive detailed reports are quite unhelpful reading for either the lay public or politicians. If they are to be useful to anyone, it will be to curriculum developers and perhaps certain other education specialists.' NAEP heeft echter wel de ontwikkeling van assessment activiteiten op staats- en plaatselijk niveau gestimuleerd en zij

vragen zich af of het op een kleinere schaal misschien beter uitvoerbaar is. In de aanbevelingen van Greenbaum et al. (1977) wordt daarom, naast verbeteringen in het project (betere materialen en procedures ontwerpen, de matrix sampling aanpassen, meer becommentariërende en interpreterende rapporten leveren), vooral aandacht besteed aan overdracht van het projectmodel.

Terwijl Greenbaum et al. (1977) op een aantal plaatsen nogal fundamentele en substantiële kritiek leveren, gaat het antwoord van NAEP, dat in het verslag is opgenomen, hier grotendeels aan voorbij. De gesignaleerde problemen worden grotendeels als overwonnen aanloopmoeilijkheden afgedaan. Dit moet o.a. blijken uit de opgenomen positieve reacties van lezers van NAEP-rapporten.

Het onderzoek van het General Accounting Office

Omdat de federale overheid sinds 1969 jaarlijks bijna vijf miljoen dollar in NAEP had geïnvesteerd, hield het General Accounting Office, de Amerikaanse rekenkamer, in 1975 een onderzoek om de effectiviteit en de efficiëntie van het project te kunnen beoordelen. Onder alle state departments of education werd een enquête gehouden, evenals onder een steekproef van ruim 800 (van de ruim 12000) schooldistricten. Verder werden interviews gehouden met medewerkers van NAEP, de Education Commission of the States, het Office of Education, het National Center for Education Statistics en het National Institute of Education.

De centrale conclusie uit het onderzoek is, dat – hoewel NAEP de assessment-technologie heeft bevorderd en de state departments heeft gestimuleerd tot het toepassen daarvan – 'national assessment results have been of limited usefulness to education decisionmakers, researchers and practitioners. This has prevented the project from achieving its basic goal.' (General Accounting Office, 1976, p. ii).

NAEP zou een belangrijke informatiebron moeten zijn voor het identificeren van problemen, het stellen van prioriteiten en het vaststellen van vooruitgang ten behoeve van beleidsinstanties, curriculumontwikkelaars, onderzoekers en practici. Volgens de informanten van GAO zijn deze doelstellingen onvoldoende uit de verf gekomen. De bezwaren kunnen als volgt worden samengevat:

- De informatie van NAEP is te weinig afgestemd op de vragen van beleidsinstanties of komt niet op tijd.
- De aangedragen informatie wordt niet door NAEP zelf voorzien van conclusies en aanbevelingen, terwijl ook andere instanties hieraan niet veel aandacht besteden. Dit beperkt de bruikbaarheid van de informatie. Een van de aspecten die met name te weinig aandacht krijgt, is het stellen van criteria waaraan een acceptabel resultaat moet voldoen.
- De opzet van NAEP is niet afgestemd op het gewenste gebruik van de gegevens, die te algemeen en onvoldoende uitgesplitst zijn. De gegevens zijn daardoor onvoldoende bruikbaar op het niveau van de afzonderlijke staten, terwijl juist daar de eerstverantwoordelijke beleidsinstanties zetelen, en niet op federaal niveau.
- De verspreiding van de informatie, met name naar instanties op plaatselijk niveau, is onvoldoende, hetgeen blijkt uit de relatieve onbekendheid van NAEP bij functionarissen van schooldistricten.

In de aanbevelingen wordt – uiteraard – gepleit voor verbeteringen die binnen de geldende opzet mogelijk zijn: de dataverzameling beter afstemmen op de informatiebehoefte, meer interpreterende en becommentariërende rapporten, verbetering van de verspreiding, etc. Echter ook de opzet van NAEP wordt ter discussie gesteld. Er worden twee mogelijke alternatieven gesuggereerd: enerzijds uitbreiding van de dataverzameling, zodat fijnere uitsplitsingen, bijv. tot op het niveau van de afzonderlijke staten mogelijk zijn, anderzijds beperking van de dataverzameling en in plaats daarvan vergroting van de service aan staten en schooldistricten door assessment-modellen te ontwikkelen, assistentie en begeleiding te bieden bij assessment op regionaal en plaatselijk niveau en te fungeren als leverancier van doelstellingen en opgaven. Het eerste gesuggereerde alternatief zou een merkwaardige koerswijziging zijn, gezien het feit dat bij de start van NAEP uitsplitsing tot op het niveau van de afzonderlijke staten nadrukkelijk afgewezen werd. De staten waren hier zelf tegen, vanwege de greep die de federale overheid daarmee zou kunnen krijgen op het onderwijs.

Uit de in het rapport opgenomen reacties van het Office of Education en NAEP zelf blijkt dat men geneigd is om veel van de sug-

gesties van GAO op te volgen, zonder overigens direct in te stemmen met een fundamentele wijziging in de opzet van NAEP.

2.4 NAEP in 1979 en daarna

Juist de uitvoerige en fundamentele kritiek van zowel Greenbaum et al. (1977) als het General Accounting Office (1976) was voor ons aanleiding om een reis naar Denver te ondernemen, met name om na te gaan wat er feitelijk was veranderd sinds 1976.

Tijdens ons bezoek omvatte de NAEP-staf in Denver ongeveer 75 medewerkers, afgezien van de managementsectie, als volgt verdeeld over de departments:

- Development, analysis and research: 30;
- Data processing: 22;
- Publications: 5;
- Public information and dissemination: 5;
- Field and research services: 5.

Het totaal aantal bij NAEP betrokken personen is echter belangrijk groter. In de hier genoemde aantallen zijn de medewerkers van de contractors en subcontractors niet opgenomen, zoals de district supervisors en de testassistenten.

Uit de medewerkers van het Department of development, analysis, and research worden teams samengesteld die belast worden met de voorbereiding en uitvoering van een assessment-cyclus. Deze teams bestaan meestal uit twee professionals, academici met een achtergrond in educational psychology of vakspecialisten, en één of twee research assistants. Tijdens ons bezoek opereerden er vijf van deze teams (het department beschikt verder over eigen technische en administratieve ondersteuning). In samenwerking met het Department of data processing werkt een team de cyclus af tot en met de technische rapportage. Daarna wordt de rapportage overgenomen door het Department of publications. De medewerkers van dit department hebben zich toegelegd op het samenvattend en begrijpelijk weergeven van de resultaten. Deze opzet lijkt te passen in de aanbevelingen uit het GAO-rapport (1976), maar deze opzet dateert al van enige jaren daarvoor. Naar aanleiding van het GAO-rapport is het Department of public information and dissemination wel versterkt. Er werd benadrukt dat men juist via algemene en gevarieerde publicaties de beleidsrelevantie van de gegevens wil verhogen. Het verband tussen de assessment-resultaten en eventuele

beleidsmaatregelen is niet zo direct als men in het begin wel eens had verondersteld.

Het Department of field and research services is voortgekomen uit de vroegere sectie Utilization and adaptation. Een van de wijzigingen die na 1976 door dit department zijn doorgevoerd, is het losbladig uitgeven van de vrijgegeven opgaven, om de overname en reproductie te vergemakkelijken.

Uit onze gesprekken kregen wij niet de indruk dat er na 1976 zo veel veranderd was. De kritiek van Greenbaum et al. (1977) beschouwde men inderdaad als achterhaald, terwijl het GAO-rapport (1976) volgens NAEP-medewerkers te veel gebaseerd was op meningen van de verkeerde informanten: bij de state departments of education zouden de vragenlijsten zijn ingevuld door mensen die te weinig op de hoogte waren. Dit neemt niet weg dat NAEP had toegestemd in het aanbrengen van veranderingen. Naast enkele al genoemde wijzigingen, zijn er inderdaad meer commentariërende rapporten van panels, bestaande uit vakdeskundigen, verschenen, maar NAEP heeft zichzelf nog steeds niet gewaagd aan het stellen van criteria waarmee de resultaten kunnen worden vergeleken.

Hoewel het erop lijkt dat er wezenlijk betrekkelijk weinig veranderd is, is sinds 1976 de kritiek voor een groot deel verdwenen. In de subsidie-aanvraag voor de periode 1979-1983 (Education Commission of the States, 1979) zijn zelfs aanbevelingsbrieven opgenomen van wat voorheen tegenstanders van het project waren. De veranderde houding ten opzichte van NAEP hangt waarschijnlijk ook samen met het beschikbaar komen van 'interessante' rapporten over ontwikkelingstrends en special probes. Naast de al genoemde studie naar functional literacy zijn special probes uitgevoerd op het gebied van de basic skills, consumer mathematics, health awareness en energy awareness. Deze rapporten worden ondersteund door een goede en gevarieerde publiciteit. Womer & Martin (1979, p. 31) schrijven in dit verband dat NAEP langzamerhand volledig geaccepteerd is als nationale 'onderwijsensus' en dat NAEP-gegevens in toenemende mate geciteerd worden, zowel in de dag- en weekbladpers als in vaktijdschriften.

Ook uit de nieuwe subsidie-aanvraag, die in september 1979 is toegekend, blijkt niet dat er veel zal veranderen, afgezien van enkele technisch-organisatorische zaken en de duur van de

assessment-cycli. In verband met de beperktere financiële mogelijkheden zullen reading/literature, writing en mathematics voortaan om de vier jaar aan de orde komen en de overige learning areas om de acht jaar, waarbij eens in de twee jaar leerlinggegevens verzameld zullen worden. De hierdoor extra vrij komende ruimte zal gebruikt worden om – aansluitend bij de aanbevelingen uit het GAO-rapport (1976) – meer accent te kunnen geven aan het verhogen van de gebruikswaarde (o.a. via user needs studies) en het beter toegankelijk maken van de databestanden voor secundaire analyse.

Het National Institute of Education (NIE, de Amerikaanse SVO) heeft begin 1979 de supervisie overgenomen van het National Center for Education Statistics. De medewerkers van NIE, die nog maar kort bij NAEP waren betrokken, vonden NAEP voor beslissingen op federaal niveau (nog) niet erg belangrijk, hoewel zij ook wel zagen dat een zo directe koppeling tussen een bestandsopname en beleidsbeslissingen ook niet goed mogelijk is. In dit opzicht brachten zij twee bezwaren naar voren: in de eerste plaats het feit dat NAEP niet altijd aansluit bij categorisering van achtergrondvariabelen die in andere projecten gebruikt worden, en verder het ontbreken van het research element (dieptestudies). Zij hoopten met betrekking tot het laatste punt via NIE de komende jaren meer activiteiten te kunnen initiëren, o.a. via verbetering van de toegankelijkheid van de databestanden voor secundaire analyse.

Inmiddels zijn in 1980 door NIE negen subsidies voor secundaire analyses van NAEP data toegekend tot een totaalbedrag van ruim 200.000 dollar.

Noten

1. De eerste deelnota *Kwaliteit van het onderwijs* is inmiddels in de zomer van 1981 verschenen (Bewindslieden van Onderwijs en Wetenschappen, 1981).

2. Naar analogie van de Amerikaanse terminologie werd de term 'assessment' gebruikt (to assess = peilen, schatten). In dit artikel worden een aantal Amerikaanse termen onvertaald gebezigd, omdat niet direct equivalente Nederlandse termen voorhanden waren.
3. Aan het begin van de zestiger jaren kwam bijv. nog slechts 5% van de totale onderwijskosten van ongeveer 30 miljard dollar uit de federale kas.

Literatuur

- Bewindslieden van Onderwijs en Wetenschappen, *Kwaliteit van het onderwijs*. 's-Gravenhage: Staatsuitgeverij, 1981.
- Bos, D. J., *Verslag van een studiebezoek aan NIE, NAEP en CSE in de U.S.A.* Amsterdam: Research Instituut voor de Toegepaste Psychologie, 1977.
- Doombos, K., *Opstaan tegen het zittenblijven*. 's-Gravenhage: Staatsuitgeverij, 1969.
- Education Commission of the States, *Grant application for the National Assessment of Educational Progress*. Denver: Education Commission of the States, 1979.
- General Accounting Office, *The National Assessment of Educational Progress: its results need to be made more useful*. Washington D.C.: U.S. General Accounting Office, 1976.
- General information yearbook*. Denver: National Assessment of Educational Progress, 1974.
- Greenbaum, W., M. S. Garet & E. R. Solomon, *Measuring educational progress*. New York: McGraw-Hill, 1977.
- Hazlett, J. A., *A history of the National Assessment of Educational Progress, 1963-1973*. Denver: National Assessment of Educational Progress, z.j.
- Innovatiecommissie Basisschool, *Innovatieplan basissonderwijs*. 's-Gravenhage: Staatsuitgeverij, 1978.
- Rickover, H. G., *American education: a national failure*. New York: Dutton, 1963.
- Wiley, C. E. & G. B. Schroeder, *Roles of participants in educational accountability*. Denver: Cooperative Accountability Project, 1974.
- Womer, F. B. & W. H. Martin, *The National Assessment of Educational Progress, Studies in Educational Evaluation*, 1979, 5, 27-37.

Manuscript aanvaard 24-11-'81