
Periodieke peiling van het rekenen

J. Bokhove

Cito Arnhem

opmerkingen vooraf

Sommige zaken hebben een hardnekkig leven. Eén van die zaken is het denkbeeld dat mensen die bij het Cito werken een bijzondere voorliefde voor de meerkeuzevraag hebben. Wel die hebben we niet. Dat we voor omvangrijke projecten meerkeuzevragen gebruiken hangt samen met de mogelijkheid van snelle en machinale scoring. Wanneer die noodzaak er niet is geven we meestal de voorkeur aan open vragen. Bij het periodiek peilingsonderzoek zal daarom overwegend de open-vraagvorm gehanteerd worden. Dit als geruststelling voor de deelnemers aan de Panama-conferentie die me reeds van hun afkeer voor meerkeuzevragen op de hoogte brachten.

Vorig jaar vertelde ik op de Panama-najaarsconferentie over 'eindtoetsen'. Het grootste deel van dat verhaal was gewijd aan het periodieke peilingsonderzoek. Ik zal daarom nu slechts zeer kort ingaan op opzet en doel van het peilingsonderzoek in het algemeen en uitgebreider op de concrete uitwerking bij het onderdeel rekenen.

1 wat houdt PPON in?

Eerst even de betekenis van de letters PPON. PPON staat voor Periodieke Peiling Onderwijs Niveau. Het 'periodiek' houdt in dat de peiling niet eenmalig is, maar om de zoveel tijd herhaald zal worden. Gedacht wordt aan een peiling eens in de vier jaar.

Om te beginnen zal de peiling plaatsvinden in de basisschool en in het speciaal onderwijs. De peiling in het basisonderwijs gebeurt bij de leerlingen van de achtste groep (èn de leerlingen die op grond van hun leeftijd in de achtste groep hadden kunnen zitten) en bij de leerlingen van de vijfde groep (èn de leerlingen die op grond van hun leeftijd in die groep hadden kunnen zitten).

Circa een jaar na de afname in het basisonderwijs volgt een peiling in het speciaal onderwijs met een daartoe aangepast instrument.

2 doel van het peilingsonderzoek

Doel van het periodiek peilingsonderzoek is een beeld te krijgen van de stand van zaken. Die betreft dan zowel de leeropbrengst als het onderwijsaanbod. Kennis van de stand van zaken is een van de elementen nodig voor het uitzetten van beleid en is bovendien nuttig voor de discussie over wat we met ons onderwijs willen bereiken.

3 wat wordt er gepeild?

In principe het gehele curriculum van de basisschool per cyclus van vier jaar.

Dat houdt in alles wat in de wet op het basisonderwijs onder de volgende namen is beschreven:

- zintuiglijke en lichamelijke oefening
- Nederlandse taal
- rekenen en wiskunde

- Engelse taal
- enkele kennisgebieden waarbij in elk geval aan de orde komen:
 - aardrijkskunde
 - geschiedenis
 - de natuur, waaronder biologie
 - maatschappelijke verhoudingen, waaronder staatsinrichting
 - geestelijke stromingen
- expressie-activiteiten, waarbij in elk geval aandacht besteed wordt aan:
 - de bevordering van het taalgebruik
 - tekenen
 - muziek
 - handvaardigheid
 - spel en beweging
- bevordering van sociale redzaamheid, waaronder gedrag in het verkeer
- bevordering van gezond gedrag

4 opzet van de peiling

De opzet van de peilingen hangt sterk samen met de aard van het gebied waarvan de opbrengst gepeild moet worden.

De voorkeur gaat uit naar een rechtstreekse peiling van de opbrengst van onderwijs en vorming bij leerlingen.

Waar dat mogelijk is worden dus opdrachten aan leerlingen voorgelegd. Aangezien er geen machinale verwerking nodig is, daar correctie plaatsvindt door testassistenten, worden de opgaven zoveel mogelijk in open vorm aan de leerlingen voorgelegd. De open vraag is natuurlijker dan de meerkeuzevraag en sluit beter aan bij wat de leerlingen gewend zijn. In een aantal gevallen zal vaststelling van wat het onderwijs bij de leerling opgeleverd heeft niet of nauwelijks mogelijk zijn. In dat geval zal de peiling zich richten op het onderwijsaanbod.

Daar waar de opbrengst bij leerlingen gepeild wordt, worden ook kenmerken van leerlingen en onderwijsaanbod verzameld. Die gegevens kunnen van belang zijn bij de interpretatie van de uitkomsten van het onderzoek en voor aansluitend onderzoek ter verklaring van waargenomen effecten.

Het peilingsonderzoek zal het karakter krijgen van een steekproefonderzoek. Dat kan omdat het niet de bedoeling is gegevens te verzamelen over individuele leerlingen en over individuele scholen, maar over de schoolsoort.

5 opzet van het onderdeel rekenen/wiskunde

Voor rekenen vindt de peiling plaats in 1987.

De peiling voor groep acht en de daarbij behorende leeftijdscategorie vindt plaats aan het einde van het schooljaar. De peiling voor groep vijf en de daarbij behorende leeftijdscategorie aan het begin van het schooljaar 1987-'88.

Bij rekenen staat het onderzoek naar de leerresultaten centraal. Daarnaast worden kenmerken van leerlingen en onderwijsaanbod verzameld.

Ik beperk me hier tot de peiling van de onderwijsleerresultaten in groep acht.

De opzet van het onderdeel rekenen/wiskunde vereist dat nagedacht wordt over de wijze van rapporteren. Die rapportage zou kunnen gebeuren door te vermelden: x procent van de leerlingen heeft y procent van de opgaven goed gemaakt. Die bewering heeft alleen betekenis wanneer daarbij de opgaven die de basis voor deze uitspraak vormen alle bekeken worden. Een alternatief zou zijn per opgave te rapporteren. Van iedere afzonderlijke

opgave zou dan aangegeven kunnen worden hoeveel procent van de leerlingen die opgave goed had. Ofschoon die laatste methode wel zeer specifieke informatie geeft staat daar als nadeel tegenover dat het moeilijk is de samenhang tussen al die losse opgaven te zien.

Gekozen is daarom voor een vorm die redelijk specifiek is maar toch ook een redelijk overzicht geeft. De rekenstof is in zo'n dertig onderdelen uiteengelegd. De rapportage vindt plaats per onderdeel. Voorbeelden van onderwerpen zijn:

- Elementair rekenen, optellen en aftrekken.
- Elementair rekenen, vermenigvuldigen en delen.
- Schattend rekenen.
- Procenten, begrip en basiskennis.

Bij ieder onderwerp komt een beschrijving van wat daaronder gerekend wordt. Bij cijferend vermenigvuldigen zal bijvoorbeeld vermeld worden dat minstens de helft van de opgaven contextopgaven zullen zijn.

Per onderwerp komen zestien à twintig opgaven. Die opgaven vormen de meest concrete operationalisering van het onderwerp in de trits: onderwerp-beschrijving-opgaven. De opgaven die samen een operationalisatie van het onderdeel vormen, worden afgebeeld op een schaallijn. Dit schalen van de items is mogelijk wanneer het opgaven betreft die betrekking hebben op een zelfde onderliggende vaardigheid. Dit schalen is niet mogelijk en niet zinvol wanneer die opgaven op zeer uiteenlopende vaardigheden betrekking hebben.

Dit schalen moet men zich als volgt voorstellen:

De opgaven worden op een lijn (een schaal) in volgorde van moeilijkheid geplaatst. Dit zou natuurlijk met alle opgaven kunnen. Maar doordat we dit alleen doen bij opgaven die een zelfde onderliggende vaardigheid betreffen geeft de plaats op de schaal tot waar men ongeveer komt tevens aan in welke mate men die vaardigheid beheerst.

Het is duidelijk dat dit alleen opgaat wanneer die volgorde voor verschillende leerlingen gelijk is. En dat is alleen bereikbaar wanneer per schaal relatief kleine gebieden genomen worden. Hoe groter de gebieden des te minder voldoet dit model. Voor het rekengebied als geheel is dit model onbruikbaar.

De statistische technieken die voor deze werkwijze gebruikt worden staan bekend onder de naam: items-respons theorieën. Het voordeel van deze statistische technieken is dat het mogelijk is een bestaande geschaalde verzameling (men spreekt hierbij wel over gecalibreerde items) aan te vullen, uit te zuiveren, uit te breiden en in te perken zonder dat een geheel nieuw normeringsonderzoek nodig is.

Wanneer met genormeerde toetsen wordt gewerkt dan is de vergelijking met een gewijzigde genormeerde toets veel moeilijker. De normering is namelijk afhankelijk van de steekproef waarop genormeerd is. Dat is bij itemresponsmodellen niet zo. De schaling van de opgaven is onafhankelijk van de groep met behulp waarvan geschaald wordt.

Om te schalen zouden we dus niet noodzakelijkerwijze met een representatieve steekproef behoeven te werken. We werken echter wel met een representatieve steekproef van voldoende omvang om uitspraken over het rekenniveau uitgesplitst naar de verschillende schalen of onderwerpen te kunnen doen die geldig zijn voor de gehele basisschoolbevolking. Dat schalen van de items heeft allerlei voordelen bij een herhaling van de peiling over vier jaar. Bij die herhaalde peiling wordt een deel van de items uit het peilingsonderzoek 1987 opnieuw gebruikt, een ander deel van de opgaven zal nieuw zijn. Door die gedeeltelijke vernieuwing van de toetsen is het mogelijk te voorkomen dat het rekenonderwijs te veel vastgelegd wordt. Verschuivingen in accenten door ontwikkelingen in het onderwijs zijn mogelijk op deze manier. Nieuwe en eventueel anderssoortige opgaven kunnen worden ingepast in de bestaande schalen, terwijl de verbinding met de vorige toetsing bewaard blijft. Nieuwe schalen kunnen eenvoudig toegevoegd worden.

Te denken valt daarbij aan mogelijk toekomstige ontwikkelingen rond het gebruik van zakrekenmachines, aan onderwerpen uit ruimtelijke oriëntatie en meetkunde. Die flexibiliteit is nodig om verstarring tegen te gaan.

Naast het onderzoek waarbij de leerlingen een groot aantal schriftelijke opgaven maken wordt een selectie van daarvoor geschikte opgaven gemaakt, waarbij wordt nagegaan hoe de leerling die oplost. De testassistenten zullen na de schriftelijke toetsing aan een steekproef van leerlingen vragen op welke wijze ze de opgaven hebben gemaakt. Het gaat daarbij om het achterhalen van de oplossingsstrategie. Bij de omvang van die steekproef denk ik aan zo'n tweehonderd leerlingen.

Het is duidelijk dat de rapportage over dit onderdeel van het onderzoek een meer beschrijvend karakter zal krijgen. Het is een poging om ook enige procesinformatie te verzamelen als bijdrage aan de discussie die na de rapportage van de uitkomsten van het PPON-onderzoek naar ik hoop op gang zal komen.