

Toetsen op niveau

Truus Dekker
Freudenthal Instituut, Universiteit van Utrecht
David C. Webb
University of Wisconsin-Madison

Samenvatting

In dit artikel bespreken we resultaten van het CATCH-project, een nascholingsprogramma voor wiskundeleraars dat gericht is op het verbeteren van de toetspraktijk. Omdat de toetspraktijk van invloed is op de manier van lesgeven verwachten we langs deze weg ook de onderwijspraktijk te beïnvloeden. Doel van het programma is het ontwikkelen van opvattingen en praktijken met de deelnemende docenten, en de verspreiding van deze ideeën naar andere scholen en schoolvakken. Het onderzoek werd uitgevoerd in internationale samenwerking tussen het Freudenthal Instituut en de universiteit van Wisconsin, Madison, USA. Het zogenaamde piramidemodel blijkt in dit onderzoek een belangrijk houvast te bieden bij het ontwikkelen en implementeren van de nieuwe toetspraktijk.

1. Inleiding

In dit artikel worden enkele resultaten besproken van het CATCH project, Classroom Assessment as a basis for Teacher Change, dat door het Freudenthal Instituut werd uitgevoerd in samenwerking met het Wisconsin Center for Education Research van de universiteit van Wisconsin-Madison in de Verenigde Staten. Belangrijkste doel van het project was het ontwikkelen en uitproberen van een nascholingsprogramma op het gebied van toetsen voor wiskundeleraars van Amerikaanse middle schools met leerlingen van 10 – 14 jaar. Door middel van het verbeteren van hun didactische (Eng.: formative) toetstechnieken wordt geprobeerd een fundamentele verandering te bewerkstelligen in hun manier van lesgeven. De manier waarop lesgegeven wordt en de manier waarop getoetst wordt beïnvloeden elkaar volgens onze opvatting wederzijds. Wanneer in de externe toetsen die deze docenten moeten afnemen alleen meerkeuzevragen voorkomen die vooral basiskennis en –begrippen toetsen, richten ze hun lessen voornamelijk op het oefenen van die basisvaardigheden, zoals we ook al in eerdere projecten vonden (Romberg, in press). Omgekeerd, als leraren vinden dat wiskundig redeneren, eigen oplossingsmethoden bedenken en werken met modellen belangrijk zijn, zullen in hun toetsen en in de klas eerder vragen gesteld worden waarmee leerlingen die vaardigheden kunnen demonstreren.

In Nederland bleek het bij de invoering van het nieuwe wiskundeprogramma in 1992 voor –destijds– vbo en mavoleerlingen van groot belang dat tegelijkertijd de eindexamens voor deze groepen werden aangepast. Voor de docenten en de schrijvers van schoolboeken was zowel de vorm van de examenvragen, namelijk uitsluitend open vragen binnen een al dan niet wiskundige context, als het scoringsvoorschrift met verschillende correcte oplossings-

methoden, een voorbeeld waar in de dagelijkse lespraktijk rekening mee gehouden moest worden. Naar deze invloed van de examens op het daaraan voorafgaande onderwijs is overigens voor zover wij weten in Nederland nog geen onderzoek gedaan. Uit de resultaten van het RAP project¹ dat aan het CATCH project voorafging bleek dat wanneer docenten vasthouden aan traditionele toetstechnieken, vooral gericht op het toetsen van de mate waarin basiskennis wordt beheerst en gekenmerkt door het gebruik van meerkeuzevragen, er ook in hun lessen weinig aandacht is voor verschillende oplossingsstrategieën van leerlingen, voor klassendiscussie en in het algemeen voor 'inzichtelijk leren'. Dat was zelfs het geval wanneer ze een lesmethode gebruikten die een dergelijke manier van lesgeven juist stimuleert en ondersteunt. Op grond van de hierboven genoemde ervaringen veronderstelden we dat het veranderen van de toetstechnieken van docenten die al een lesmethode gericht op 'inzichtelijk leren' hanteren een verandering in hun manier van lesgeven zou kunnen bewerkstelligen. Het belangrijkste middel om zo'n verandering te bereiken is het verbreden van de kennis en vaardigheden van de betrokken docenten over andere manieren van toetsen, gericht op het begrijpen en gebruiken van de wiskunde die in de lessen is geleerd. Door vragen te stellen en toetsopgaven te gebruiken waarbij van leerlingen verwacht wordt hun eigen wiskundegereedschap te kiezen, wiskundig te redeneren, en te generaliseren wordt duidelijk op welk niveau de leerling werkt en wat hij of zij kan in plaats van welke begrippen, algoritmes en definities hij of zij kent. Er wordt veel aandacht besteed aan het beoordelen van leerlingwerk en het gebruiken van door de leerlingen zelf bedachte oplossingsstrategieën tijdens de les. Docenten leren door deze aanpak om beter naar hun leerlingen te luisteren en beslissingen te nemen over de manier waarop (mis)concepties van leerlingen in de erna volgende lessen gebruikt kunnen worden. Een volgende stap is het verbeteren en aanpassen van eigen toetsopgaven of van de vragen uit het gebruikte boek. Daarna volgen het zelf maken van evenwichtige toetsen met vragen op diverse niveaus en het maken van een toetsplan. Toetsen wordt zo een continue activiteit, een integraal onderdeel van het lesgeven, in plaats van iets dat als een onderbreking van het leerproces, iets van buitenaf en louter ter afsluiting van dat leerproces wordt ervaren. We denken dat op die manier een meer leerling gerichte leeromgeving ontstaat waarin meer mogelijkheden zijn tot verbetering van de prestaties van leerlingen (Bransford, Brown & Cocking, 1999).

2. Nascholing: Theoretische achtergrond, strategie en materiaal

De ideeën van waaruit binnen het CATCH project werd gewerkt en waarop zowel de nascholingsstrategie als het nascholingsmateriaal zijn gebaseerd, vinden hun achtergrond in het werk van het Freudenthal Instituut op het gebied van onderwijs en toetsen. Daarin zijn de volgende componenten te onderscheiden:

1. wiskundige inhoud verdeeld in domeinen
2. inzichtelijk leren
3. didactisch toetsen
4. gebruik van de *Piramide van Jan de Lange*

1. wiskundige inhoud verdeeld in domeinen

De opvatting over het organiseren van wiskundige inhoud op basis van domeinen in plaats van als een grote verzameling concepten en vaardigheden is niet nieuw. In twee bekende publicaties – *On the shoulders of giants: New approaches to numeracy* (Steen, 1990) en *Mathematics: The science of patterns* (Devlin, 1994) – is de wiskunde op deze manier beschreven. De nadruk ligt op hoe concepten en vaardigheden binnen een domein met elkaar verbonden zijn en vervolgens op de relaties met andere domeinen. Verder steunt de domeinbenadering op het gebruiken van wiskundetaal die mensen hebben bedacht om met elkaar te kunnen communiceren over de begrippen binnen de domeinen. In plaats van onderwerpen zoals algebra en meetkunde worden grotere domeinen als ‘verandering en groei’, ‘ruimte en vormen’ en dergelijke gebruikt.

2. inzichtelijk leren

In het Nederlandse TAL project² werden *longitudinale leerlijnen* voor de basisschool ontwikkeld (Treffers, 1999, Van den Heuvel-Panhuizen, 2000). Reken- en wiskundelessen bestaan volgens deze leerlijnen niet uit geïsoleerde stukjes leerstof. Van de docenten wordt verwacht dat zij een serie lessen plannen die de leerling helpt om de begrippen binnen een domein en de relaties tussen de domeinen te begrijpen; eerst op een informele, dan preformele en tenslotte op een formele manier.

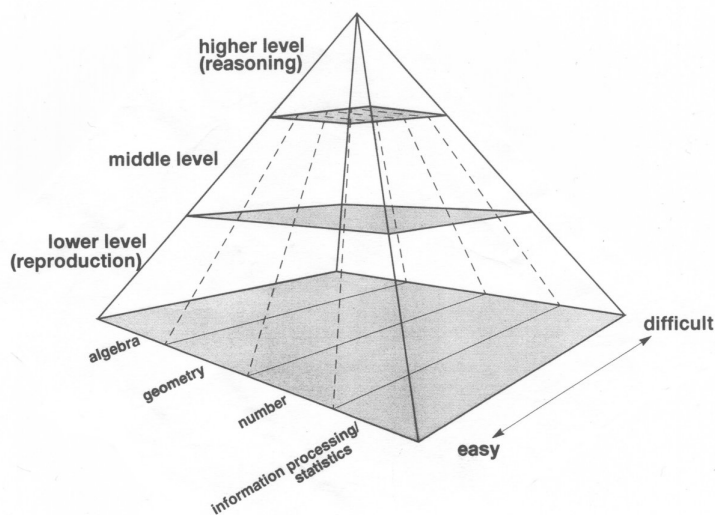
3. didactisch toetsen

Didactisch toetsen kan naar verwachting bereikt worden door het opstellen van een *hypothetical assessment trajectory* of toetsplan. De docent volgt de ontwikkeling van de leerling, gaat na in hoeverre wordt voortgebouwd op al aanwezige kennis en beoordeelt op grond van de antwoorden op de gestelde vragen hoe ver een leerling gevorderd is op de weg van informele, preformele en formele beheersing van de begrippen binnen een bepaald domein. Het toetsplan omvat een serie toetsen, van informele mondelinge toetsen tijdens het werken in de klas tot meer formele afsluitingstoetsen na een hoofdstuk, maar ook projectopdrachten, diagnostische toetsen en korte toetsen die aangeven of leerlingen de basisbegrippen uit een hoofdstuk voldoende onder de knie hebben om verder te kunnen gaan.

4. piramidemodel

Dit model, gebaseerd op de verdeling van wiskundige competenties in drie niveaus (De Lange, 1987, Van den Heuvel-Panhuizen, 1996), vormt de theoretische basis van het ontwikkelde nascholingsmateriaal. Het model werd verder ontwikkeld door het Freudenthal Instituut (De Lange & Boertien, 1994, Verhage & De Lange, 1997, De Lange, 1999) en gebruikt voor het ontwerpen van de Nederlandse Nationale Optie bij het internationale onderzoek TIMSS³ (Kuiper, 2000).

De piramide dient twee doelen. In de eerste plaats is het een model voor het classificeren van toetsopgaven van verschillend niveau. Maar het model geeft ook een – globale – verdeling aan van het aantal opgaven van elk niveau dat in een evenwichtige schriftelijke toets zou moeten voorkomen. Omdat opga-



Figuur 1. Jan de Lange's Piramide⁴.

ven op niveau I meestal gesloten vragen zijn die minder tijd vragen dan de open vragen op niveau II en III, gebruiken we vaak de volgende verhouding:

$$\text{Niveau I} : \text{Niveau II} : \text{Niveau III} = 3 : 2 : 1$$

Door sommige docenten wordt overigens in plaats van de verdeling gemakkelijk \leftrightarrow moeilijk ook de verdeling informeel \leftrightarrow formeel gebruikt.

Hieronder wordt een korte beschrijving gegeven van de verschillende niveaus. De niveaus zijn niet duidelijk van elkaar gescheiden maar lopen in elkaar over.

Niveau I: reproductie, gebruik van standaard procedures, kennis van feiten en definities

Hierbij horen vragen als:

- *Hoeveel zijvlakken heeft een kubus?*
- *Bereken de BTW (19%) bij een bedrag van €250,-*
- *Wat is de naam van de grafiek van een tweedegraads functie?*

De vorm van de vraag is meestal een één-antwoord-vraag, er is maar één goed antwoord mogelijk. Meestal hoeft geen berekening of uitleg gegeven te worden.

Niveau II: verbanden leggen, integratie

Kenmerkend voor dit niveau is dat leerlingen informatie uit verschillende bronnen moeten kunnen integreren en in het algemeen hun eigen wiskundige "gereedschap" kiezen om een probleem op te lossen.

Vragen op dit niveau worden meestal gesteld binnen een context en zijn open vragen. Een voorbeeld:

Je kunt een kopieermachine gebruiken om een tekening te vergroten of te verkleinen. Ben heeft de machine ingesteld op verkleinen tot 60%. Hij is niet tevreden met het resultaat en maakt een nieuwe kopie op 140%. Heeft hij nu weer het oorspronkelijke formaat terug? Leg je antwoord uit.

De opgave kan op verschillende manieren worden gemaakt en zowel op een formeel als op een informeel niveau. De uitwerking van de leerling levert de docent daardoor veel informatie op.

Niveau III: mathematiseren, wiskundig redeneren, generaliseren, inzicht

Leerlingen moeten op dit niveau bijvoorbeeld inzien dat een hele serie (context)opgaven allemaal betrekking hebben op een lineaire functie. Ze maken, gebruiken en bekritisieren (wiskundige) modellen en kunnen een passende redenering opschrijven om hun beweringen te ondersteunen. We geven opnieuw een voorbeeld:



Stel je voor dat je buurman een muurtje heeft gemetseld tussen jullie voortuinen. Dan wil je wel graag zeker weten dat het muurtje recht staat en de verdeling van de voortuinen eerlijk is. Hoe zou je dat kunnen aantonen? Geef een duidelijke toelichting.

Van leerlingen wordt verwacht dat ze voor het beantwoorden van zo'n vraag hun eigen wiskundige gereedschappen kunnen kiezen. Maar ook dat ze eerst een (wiskundig) model maken van de situatie en het probleem binnen dat model oplossen. Veronderstellingen die gemaakt worden omdat niet alles uit de foto zichtbaar is, moeten worden vermeld.

Of het muurtje scheef staat kun je bepalen met de stelling van Pythagoras maar als je veronderstelt dat de voortuinen rechthoekig zijn kun je ook kijken of alle diagonalen even lang zijn. Een leerling die nog op een informeel niveau werkt zou kunnen schrijven: "Ik zou kranten naast elkaar neerleggen op het hele erf. En dan bij de buurman hetzelfde doen. Zie je zo of hij scheef heeft gemetseld!"

De in dit onderdeel genoemde theoretische componenten, uitgewerkt in het *Framework for Classroom Assessment in Mathematics* (De Lange, 1999) dat gebruikt werd als theoretische basis voor het CATCH project, werden ook gebruikt voor PISA⁵, het internationale onderzoek naar wiskundige vaardigheden van 15-jarige leerlingen, uitgevoerd door de Organisatie voor Economische Samenwerking en Ontwikkeling (OECD, 1999). Het piramide model vormde een belangrijk onderdeel in de nascholing van docenten en districtsbestuurders die deelnamen aan het CATCH project.

De nascholingsstrategie zoals gebruikt in het CATCH project is gebaseerd op het "train de trainer" model. Het Nederlandse CATCH team ontwikkelde een voorbeeld seminar voor een driedaagse startconferentie in Utrecht. Later werd dit seminar door de deelnemende docenten als uitgangspunt gebruikt voor het organiseren van hun eigen CATCH nascholingscursus, bedoeld om docenten van andere klassen en scholen uit het district na te scholen.

Aan de startconferentie in Utrecht namen 19 docenten en ander leidinggevend schoolpersoneel deel uit twee Amerikaanse schooldistricten. District A is een relatief klein schooldistrict in een voorstad, met meer dan 3000 leerlingen. Van de leerlingen is 85% van Europese afkomst en ongeveer 30% heeft recht op gratis schoolmaaltijden. District B is een groot stedelijk gebied waar de leerlingenpopulatie overwegend zwart of latino is en meer dan 75% gratis schoolmaaltijden krijgt. Sinds de start van het project hebben ongeveer 120 docenten uit beide districten deelgenomen aan een plaatselijke CATCH nascholingscursus. De kosten daarvoor werden door de schooldistricten zelf gedragen. Door mensen op districtsniveau bij het project te betrekken werd een snellere verspreiding van de ideeën uit het CATCH project bereikt. Docenten hadden deze steun op districtsniveau ook nodig, zo bleek uit gesprekken die met hen werden gevoerd.

Na het bijwonen van de startconferentie werd van de deelnemende docenten verwacht dat ze experimenten met andere toetsvormen zouden uitvoeren in hun klassen. Ze kregen ondersteuning ter plaatse van de coördinatoren en leden van het onderzoeksteam en konden verder informatie en ideeën uitwisselen via de CATCH website (www.fi.uu.nl/catch). De docenten kwamen tijdens het schooljaar regelmatig bij elkaar, om resultaten van hun experimenten te bespreken en om de organisatie van de plaatselijke nascholingscursus voor te bereiden, zowel in schooljaar 2000/01 als schooljaar 2001/02. Aan het eind van schooljaar 2002/03 kon geen nascholingscursus worden georganiseerd, vooral vanwege financiële redenen.

Door het Nederlandse team werd een boek geschreven dat als nascholingsmateriaal gebruikt werd. Dit boek, *Great Assessment Problems (GAP)* (Dekker & Querelle, 2002), werd ook op de CATCH website gepubliceerd. Het boek is de uitwerking voor docenten van het eerder genoemde *Framework for Classroom assessment in Mathematics*, eveneens gepubliceerd op de website. Het piramide model wordt in het GAP boek zeer uitgebreid besproken, met veel voorbeelden van opgaven op de verschillende niveaus. De docenten hadden al tijdens de startconferentie opgemerkt dat veel van hun toetsen ("quizzes") alleen opgaven zonder context, op niveau I, bevatten en dat ze nauwelijks andere toetsvormen gebruikten. Deze andere toetsvormen, de verschillende rollen van de context en het maken van een evenwichtige toets komen daarom in het boek aan de orde. Het opstellen van een scoringsvoorschrift, het beoordelen van leerlingenwerk en het geven van

feedback worden besproken. Het boek eindigt met het opstellen van een toetsplan en met het geven van een grote hoeveelheid didactische tips.

Verder werd een toetsenbank gebruikt bij de nascholing, *AssessMath!*. Het Freudenthal Instituut heeft eerder aan de ontwikkeling van deze toetsenbank meegewerkt (Cappo, De Lange & Romberg, 1999). Allerlei artikelen op het gebied van didactisch toetsen en relevante hoofdstukken uit het boek *Insight Stories* (Romberg, in press) werden als nascholingsmateriaal gebruikt. Diverse elementen uit de startconferentie en de plaatselijke nascholingscursussen vormden de basis voor een manuscript: *Elements for professional development* (Abels, 2003) dat als voorbeeld voor het geven van nascholing op het gebied van didactisch toetsen van wiskunde gebruikt kan worden. Viermaal werd een 'Newsletter' verzonden aan alle deelnemers aan het project.

We verwachtten dat deelnemers aan het project het volgende traject zouden afleggen tijdens hun nascholing:

- docenten bekritisieren bestaande toetsinstrumenten en –opgaven, samen met het team en ze maken kennis met andere toetsvormen;
- docenten leren een nieuw toetsmodel kennen, gebaseerd op *Jan de Lange's piramide* zoals besproken in het *GAP*-boek;
- docenten bekijken en beoordelen leerlingenwerk dat hoort bij opgaven op de verschillende niveaus;
- docenten kiezen bestaande opgaven en vragen binnen het piramidemodel, ze passen die zonodig aan;
- docenten ontwerpen hun eigen opgaven die voldoen aan de eerder beschreven theoretische componenten;
- docenten leren hoe ze de mondelinge en schriftelijke uitwerkingen van hun leerlingen kunnen gebruiken bij het nemen van didactische beslissingen;
- docenten spreken met hun collega's over het geleerde tijdens informele bijeenkomsten;
- docenten ontwerpen een toetsplan volgens een *hypothetical assessment trajectory* en maken evenwichtige toetsen;
- docenten vinden een balans tussen het gebruik van formele toetsen, informele toetsen en externe toetsen en beoordelen de relatie tussen toetsresultaten op klassenniveau en op districtsniveau;
- docenten geven nascholing aan collega's tijdens plaatselijke nascholingscursussen en helpen op die manier mee aan het verder verspreiden van de ideeën uit het project.

Zelf nieuwe opgaven maken op niveau II of III is niet simpel als je dat nooit eerder gedaan hebt en er ook weinig voorbeelden van gezien hebt. Daarom begonnen we met het wijzigen van bestaande opgaven zodat die meer inzicht geven in de manier waarop de leerling denkt. Een voorbeeld daarvan is te zien in figuur 2.

Niveau I: Vul op de open plaats het juiste teken in < > of =

$$\frac{4}{5} \dots\dots \frac{4}{7}$$

Niveau II: Je jongere broer wil weten waarom $\frac{4}{5}$ groter is dan $\frac{4}{7}$

Schrijf op hoe jij hem dat zou uitleggen.

Figuur 2. Voorbeeld van het aanpassen van een opgave.

3. Onderzoeksvragen en methode

In het CATCH project werden de volgende onderzoeksvragen geformuleerd:

1. Welk nascholingsmateriaal is er nodig om de ideeën voor het verbeteren van didactische (Eng.: formative) toetsen te verspreiden onder een grotere groep docenten en scholen?
2. Welke ondersteuning hebben schoolpersoneel en docenten in verschillende scholen nodig wanneer ze de ideeën opgedaan in het project aanpassen aan de plaatselijke omstandigheden?
3. Hoe beslissen docenten welke toetsinstrumenten ze zullen gebruiken, wanneer ze die gaan gebruiken en welke redenen geven ze om die keuzes te motiveren?
4. Hoe verandert de toetspraktijk van de docenten die deelgenomen hebben aan het geboden nascholingsprogramma?
5. Hoe worden veranderingen in de toetspraktijk van de docenten zichtbaar in de resultaten van hun leerlingen?

Om deze vragen te kunnen beantwoorden werden zowel kwantitatieve als kwalitatieve gegevens verzameld. De kwantitatieve gegevens bestaan uit de resultaten van leerlingen behaald op zogenaamde "standardized tests", externe toetsen op districtsniveau. Ze zijn gebruikt om vraag 5 te beantwoorden maar worden in dit artikel verder niet besproken.

Voor het onderzoeksdeel van het project werden verder kwalitatieve gegevens verzameld door middel van het regelmatig houden van interviews met de deelnemende docenten, het doen van klassenobservaties en het houden van postobservatie-interviews. De deelnemers aan de startconferentie vulden allemaal een vragenlijst in, bedoeld om de situatie vóór deelname aan het project vast te leggen. Aan de docenten werd gevraagd een didactisch toets-experiment uit te voeren in de eigen klas en daarover verslag te doen. Tenslotte werden door de docenten toetsen met bijbehorend leerlingewerk verzameld tijdens de duur van het project. Aan het eind van het project werd een afsluitende vragenlijst ingevuld door de betrokken docenten. Aan alle deelnemers werd gevraagd het ontwikkelde nascholingsmateriaal te bekritisieren.

De klassenobservaties en de ingeleverde toetsen werden gebruikt om de individuele ontwikkeling van de docenten te analyseren en om de onderzoeksresultaten die uit de interviews werden verkregen te valideren.

De analyse van de interviews maakte het mogelijk om de onderzoeksvragen 4 en 2 te beantwoorden. De analyse gaf ook inzicht in de verspreiding van ideeën uit het project, zoals bedoeld in onderzoeksvraag 1. Vraag 3 tenslotte werd beantwoord door middel van het bestuderen van de ontvangen toetsen en leerlingengerwerk en het beoordelen van de observaties (Webb & Dekker, 2002, Dekker & Feijs, 2003).

Voor het analyseren van de interviews werden codes geformuleerd die gebaseerd waren op het eerder genoemde traject dat naar verwachting door de docenten tijdens de nascholing zou worden doorlopen. Er zijn drie groepen hoofdcodes. De eerste heeft betrekking op de verschillende typen veranderingen in de opvattingen van de betrokken docenten ten aanzien van didactisch toetsen en het gebruik van het piramide model. De tweede hoofdgroep spitst zich toe op de veranderingen in de klassenpraktijk van de docenten. De derde groep tenslotte heeft betrekking op de verschillende soorten steun die de docenten hebben ervaren. De hoofdcodes werden vervolgens onderverdeeld in subcodes. Er werd gebruik gemaakt van het programma MEPA⁶, geschikt voor het coderen van verbale gegevens, in ons geval de uitgeschreven interviews. De gecodeerde tekstfragmenten kunnen vervolgens geteld worden. Maar er kunnen ook selecties gemaakt worden op basis van de toegekende codes zodat bij elke code voorbeelden gevonden kunnen worden die een dieper inzicht geven dan de aantallen alleen.

4. Enkele resultaten

In het algemeen werden docenten zich beter bewust van hetgeen ze doen (en vooral ook niet doen) om te toetsen in hoeverre hun leerlingen de kennis en vaardigheden die ze tijdens de wiskundelessen hebben geleerd kunnen toepassen in nieuwe situaties en of de leerlingen hun antwoorden (nog) op een informeel niveau geven of al op een formeler niveau. Uit onze observatie van de door de docenten zelf georganiseerde nascholingscursussen bleek dat de docenten zich de gepresenteerde theorie en praktijk van didactisch toetsen in hoge mate eigen hadden gemaakt. Zo gebruikten ze voorbeelden uit de eigen klassenpraktijk om het piramidemodel te verduidelijken.

Het piramidemodel (zoals een docent in een interview opmerkte: *sweet, short and simple*) bleek een belangrijk instrument om verandering in de opvattingen van de docenten te bewerkstelligen. Enkele citaten uit de interviews:

Analyzing the levelling of questioning that is important to me.

I am always thinking about a type 3 question, a type 2 question. And for years we used type 1. And I am weaning my stuff away from it. That is the most difficult part for me to do now, to get those type 3 questions. Two is pretty easy but I still haven't made that transitional to type 3 questions.

Zodra het echter gaat om het toepassen van deze ideeën in de praktijk, dus bij het zelf maken van toetsopgaven, gaven de docenten aan dat ze dat erg moeilijk vonden en aan hun zelfgemaakte toetsen was het ook te zien. Bijna alle docenten worstelden met het maken van goede opgaven op niveau II of III maar gaven tegelijkertijd aan dit onderwerp wel heel belangrijk te vinden. Een expert in het maken van goede toetsopgaven op een hoger niveau word je kennelijk niet zomaar, daar is tijd en aandacht voor nodig. Docenten gaven

aan dat ze de ideeën uit het project wel toepasten tijdens hun lessen, door andere vragen te stellen tijdens klassendiscussies of door meer aandacht te geven aan verschillende oplossingsstrategieën van hun leerlingen.

Docenten veranderden in hun opvatting over didactisch toetsen en deze verandering was, volgens de gegevens uit de interviews en observaties, ook in de praktijk zichtbaar. Enkele citaten:

I think that I am more aware of looking at student work. I may have looked at the answer last year or I am much more aware of looking at the work the student has put into the answer. And really make it a part of my assessment.

Trying to document the informal [assessment]. That was the big thing.

This year I used a combination of a lot of different things. Observation in the classroom. Just watching them as they work. Listening.

I think over the last year to this year I think I am getting away from as many tests, as many quizzes.

Steun van de collega's, het bespreken van de resultaten tijdens informele en formele bijeenkomsten werd heel belangrijk gevonden, met steun vanuit het CATCH team op de tweede plaats:

I think the more you get together and talk about it the more it keeps you, it certainly keeps you involved in what your goals are supposed to be with this.

But the CATCH, the team approach of having a person to talk with about what is happening at this time, is so valuable rather than waiting until the end of the year. And you maybe have made mistakes that you wish you could have discovered sooner or other ways of thinking about things or other methods of teaching or, so having that immediate feedback is very helpful. Talking with where my students need to be at the end of the year is extremely important because that has always sort of been a dead end.

Soms gaven docenten aan dat ze de theorie weliswaar belangrijk vonden maar dat deze in hun situatie niet toegepast kon worden. Ze gaven daar verschillende redenen voor. Enkele voorbeelden:

I haven't had a chance to really implement any of that outside what we did in the summer. I really feel like my hands are tied this year. And I really haven't been able to get a grip on what direction I go in.

And our children have... historically, they know the information, but they have a difficult time explaining that information.

Here in this environment, in this situation, they just don't have the basic reading skills.

I love these kids. They are a lot of great kids. And I hate to keep sounding so negative. But, they just can't move to that, this higher level of questioning and thinking.

But I am really finding it difficult for them to do it unassisted.

Onverwacht, ook voor de leden van het onderzoeksteam was de snelheid waarmee de ideeën uit het project zich verspreidden. Docenten gingen ze gebruiken voor andere vakken en praatten erover met collega's uit andere klassen.

The balanced assessment idea is like primary. It is a kind of thing we, you get all levels, all kinds of thinking, and in all, like, you know, there is that Level I, Level II, and the different kinds of levels, of Level I. And all that. So there is like so much that can be pulled. And that is kind of like a model, that it kind of has guided last year. And actually that has kind of even leaked into other curricular areas. I do that with my reading.

Dat de ideeën uit het project zich ook op districtsniveau verspreidden weten we niet uit de interviews maar uit andere bronnen. Docenten uit het project droegen ook daaraan bij. Omdat de beide districten het organiseren van CATCH nascholingscursussen bekostigden konden docenten uit andere klassen en van andere scholen in het district ook nageschoold worden. De docenten uit het project gaven presentaties tijdens diverse conferenties en seminars. In district A werden de CATCH docenten betrokken bij het schrijven van een wiskunde Framework op districtsniveau. Ze ontwierpen ook diagnostische toetsen die verder ontwikkeld en in het hele district gebruikt gaan worden. In district B is door de bestuurders een aanvraag gedaan voor een nieuw project dat onder andere een uitbreiding van het CATCH project naar alle klassen (K-12) en naar het vak science omvat.

In district B vonden ten tijde van de tweede ronde interviews grote organisatorische veranderingen plaats. Alle aan het project deelnemende docenten kwamen terecht op geprivatiseerde scholen waar ze moesten werken met een nieuwe lesmethode en vaak met verplichte toetsen. Dat leidde tot veel frustraties en maakte het moeilijk voor de docenten om hun veranderde ideeën te implementeren, maar zoals blijkt uit het volgende citaat is er hoop:

Interviewer: How close do you think you are to using exemplary classroom practices?

Docent: In honesty, I think this year, not very close at all. Because I feel compelled to, well, they tell me I have to follow their design. But I know in time I will be able to resurrect something out of the rubble.

5. Conclusies

Als docenten andere toetsvormen leren kennen die meer gericht zijn op didactisch toetsen en toetsen van begrip in plaats van uitsluitend op reproductie zien ze vaak ook de noodzaak om hun lespraktijk zodanig aan te passen dat leerlingen inzichtelijk gaan leren in plaats van 'leren om te reproduceren'. Het gehanteerde model van Jan de Lange's piramide leent zich er goed toe om deze verandering in houding ten opzichte van toetsing tot stand te brengen. Zodra docenten dit model in praktijk gaan brengen denken ze ook opnieuw na over de leerdoelen van hun lesmethode en de vragen die ze tijdens de les stellen. Om de wiskundige representaties die leerlingen gebruiken bij het

beantwoorden van de vragen op hogere niveaus goed te kunnen interpreteren en beoordelen moeten de docenten die kunnen relateren aan een toetsplan en aan modellen die toenemende, geleidelijke formalisatie beschrijven.

De ervaringen in het CATCH project wijzen erop dat zowel docenten als hun leerlingen kunnen profiteren van een beter inzicht in didactische toetspraktijken. De docenten moeten echter zelf die behoefte tot verandering voelen, ze moeten zich realiseren dat hun huidige toetspraktijk hen (te) weinig informatie verschaft over de vorderingen van hun leerlingen en ze moeten inzichtelijk leren als een belangrijk doel zien. Door het integreren van beoordelen en lesgeven kunnen docenten beter gefundeerde didactische beslissingen nemen. Deze ontwikkelingen leiden op hun beurt tot het verbeteren van de prestaties van de leerlingen bij de verplichte toetsen op districtsniveau.

Correspondentie over dit artikel aan Truus Dekker, Freudenthal Instituut Postbus 9432, 3506 GK Utrecht. Email: T.Dekker@fi.uu.nl. Dit project is gesubsidieerd door het Office of Educational Research and Improvement (OERI PR/awrd R305A60007) 2000 – 2003. Zie ook <http://www.fi.uu.nl/catch>.

Noten

1. Research in Assessment Practices project van het National Center for Improving Student Learning and Achievement in Mathematics and Science (NCISLA), gesubsidieerd door OERI, 305A60007-98, 1996-2000.
2. TAL Tussendoelen Annex Leerlijnen voor het basisonderwijs.
3. TIMSS Third International Mathematics and Science Study, 1995.
4. Overgenomen uit het boek *Great Assessment Problems* (Dekker & Querelle, 2002).
5. PISA Programme for International Student Assessment.
6. MEPA, Multiple Episode Protocol Analysis, ontwikkeld door Gijsbert Erkens, Universiteit Utrecht

English summary

Assessing at the right level

In this paper we will share research results from the implementation of the Classroom Assessment as a basis for Teacher Change (CATCH) professional development program in mathematics with respect to changes in teachers' conceptions, changes in teachers' classroom practice and the dissemination of ideas from the project to other subjects and other schools. The research team includes a cross-national collaboration between mathematics educators from the Freudenthal Institute, Utrecht, The Netherlands and the University of Wisconsin, Madison, USA.

The CATCH professional development program seeks to bring about fundamental changes in teachers' classroom instruction by helping them change their formative assessment practices. The so called pyramid model, developed by researchers from the Freudenthal Institute forms a crucial element in the program.

Literatuur

- Abels, M. (2003). *Elements for professional development* (CATCH professional development guide). Intern document. Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.
- Cappo, M., Lange, J. de & Romberg, T. A. (1999). *AssessMath!* Santa Cruz, CA: Learning in Motion.
- Dekker, T. & Querelle, N. (2002). *Great Assessment Problems*. Utrecht: Freudenthal Instituut. Beschikbaar via: <http://www.fi.uu.nl/catch>.
- Dekker, T. & Feijs, E. (2003). Scaling Up Strategies for Change. In N.A. Pateman et al. (Eds.), *Proceedings of the 2003 Joint Meeting of PME and PMENA, Vol. 2*, (pp. 317-324). Honolulu, HI: CRDG, College of Education, University of Hawai'i.
- Devlin, K. J. (1994). *Mathematics, the science of patterns: the search for order in life, mind, and the universe*. New York: Scientific American Library.
- Heuvel-Panhuizen, M van den (1996). *Assessment and realistic mathematics education*. Proefschrift. Utrecht: Freudenthal Instituut.
- Heuvel-Panhuizen, M. van den, Buys, K. & Treffers, A. (Ed) (2000), *Kinderen leren rekenen. Tussendoelen Annex Leerlijnen. Hele Getallen Bovenbouw Basisschool*. Groningen: Wolters Noordhoff.
- Kuiper, W.A.J.M., Bos, K.Tj. & Plomp, Tj. (2000). The TIMSS national option mathematics test. *Studies in Educational Evaluation*, 26, 43-60
- Lange, J. de (1987). *Mathematics, Insight and Meaning*. Proefschrift. Utrecht: OW&OC.
- Lange, J. de & Boertien, H. (1994). *Model voor wiskundetoets en nationale optie TIMSS 1995, intern rapport*. Enschede: Universiteit van Twente, OCTO.
- Lange, J. de (1995). *Assessment: No change without problems*. In T. A. Romberg (Ed.), *Reform in School Mathematics*, (pp 87-172). Albany, NY: SUNY.
- Lange, J. de (1999). *Framework for classroom assessment in mathematics*. Utrecht: Freudenthal Instituut. Beschikbaar via: <http://www.fi.uu.nl/catch/products>.
- Organization for Economic Cooperation and Development. (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: OECD.
- Romberg, T.A. (Ed). (in press) *Insight Stories: Assessing Middle Grades Mathematics*. New York: Teachers College Press.
- Steen, L. A. (1990). *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.
- Treffers, A., Heuvel-Panhuizen, M. van den, & Buys, K. (Ed) (1999), *Kinderen leren rekenen. Tussendoelen Annex Leerlijnen. Hele Getallen Onderbouw Basisschool*. Groningen: Wolters Noordhoff.
- Webb, D. & Dekker, T. (2002). Classroom assessment as a basis for teacher change (CATCH). In Mewborn, D. et al. (Eds.), *Proceedings of the North American Chapter of the International Group for the Psychology of*

Mathematics education, Vol. II, (pp.599-609). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.