

Stichting Cito onderzoekt de toetspraktijk

Cijfers, cijfers, cijfers

Of leerlingen daadwerkelijk het inmiddels mythische aantal van 150 toetsen per jaar moeten doorwerken is de vraag. Maar dat er tussen de schoolmuren veel cijfers geproduceerd worden staat vast. Als er inderdaad zoveel getoetst wordt, wat voor toetsen zijn dat dan, en wie maakt ze? Hoe bepalen docenten slaag-zakgrenzen? Onderzoekers van Stichting Cito hielden onlangs een survey onder docenten van het voortgezet onderwijs.

DOOR DE REDACTIE

De vragenlijst bestond uit drie onderdelen: vragen over de toetsconstructie en betrokkenheid bij toetsbeleid, vragen over het becijferen (hoe bepalen docenten de slaag-zakgrenzen en becijfering, in hoeverre is er sprake van collegiaal overleg) en vragen over de ervaren toetscultuur. Aan het onderzoek deden ervaren en minder ervaren docenten mee vanuit de taalvakken, exacte vakken, zaakvakken, en ook beroepsgerichte vakken. Het totaal aantal respondenten was 181, waarbij is aangetekend dat niet iedereen elke vraag beantwoordde. Met de onderzoekers Wilco Emons, Emmelien van der Scheer en Bas Hemker bespreken we een aantal uitkomsten.

Zullen we beginnen met de basis? Wie maken de toetsen en hoe komt men tot cijfers?

“We hebben twee aspecten specifiek bevraagd: wie de toetsen maken en of er gebruik wordt gemaakt van een toetsmatrijs. Op die vragen geeft 84% aan zelf de toetsen te maken, meestal in overleg met collega’s. Vervolgens geeft 10% aan altijd een toetsmatrijs te gebruiken, maar meldt 70% dit zelden of nooit te doen. Het is lastig om zo’n uitkomst te vertalen naar de praktijk, maar het is tenminste opmerkelijk en wellicht valt hier nog een aantal slagen te maken. Wel wordt een aantal keren aangegeven dat de RTTI-methode (wat een concrete uitwerking is van een toetsmatrijs) als basis voor toetsconstructie wordt gebruikt.

Een andere observatie is het toch relatief hoge percentage docenten dat aangeeft nooit - of hooguit soms - oude toetsen van collega’s te gebruiken, of oude toetsen aan te passen. Het is lastig om hier nu verregaande conclusies aan te verbinden, maar wellicht zouden de docenten nog meer van elkaars toetsen kunnen leren. In de voorbereiding van dit onderzoek zijn we overigens ook wel op praktijken gestuit waar je vraagtekens bij kunt plaatsen. Zo spraken we met een docent van een school, waar verschillende docenten voor hetzelfde vak en hetzelfde jaar hun eigen toetsen maken. Dit betekent dat leerlingen die precies hetzelfde vak volgen, op dezelfde school en met dezelfde lesmethoden, toch met andere toetsen (en dus mogelijk anders) beoordeeld worden.”

Jullie vroegen ook specifiek naar manieren waarop slaag-zakgrenzen tot stand komen en hoe de cijfers berekend worden?

“Ja. We zien dan dat de meeste docenten vrij consequent een bepaalde methode toepassen. Een veel genoemde methode is: ‘aantal punten/maximaal punten x 9 + 1’. Dit betekent dat leerlingen minstens 50% van de punten moeten halen om een voldoende te halen. De meeste docenten geven aan daarnaast rekening te houden met de inhoudelijke kenmerken van de toets (bijv. het beheersingsniveau van de vragen of de moeilijkheid van de vragen) en het aantal onvoldoendes die de basis slaag-zakgrens oplevert, zowel nu als ten opzichte van vorig jaar.

In de reacties zien we een grote mate van zorgvuldigheid: docenten



streven er echt naar om optimaal passende cijfers te geven. Tegelijkertijd is onze indruk dat vooral wordt gevaren op inschattingen en ervaringen. Daar is op zich helemaal niets mis mee, maar er zijn de loop der jaren technieken ontwikkeld om het bepalen van slaag-zakgrenzen en cijfers nog systematischer vorm te geven.”

Wordt er inderdaad zoveel getoetst?

“Daar hebben we naar gevraagd. We hebben docenten gevraagd om een bepaald jaar in gedachten te nemen en een inschatting te maken van de hoeveelheid toetsen die leerlingen per jaar moeten maken, en of zij denken gemiddeld genomen meer, minder of evenveel toetsen te geven vergeleken met hun collega's. De meeste docenten hebben deze vraag beantwoord voor een bovenbouwklas.



Van alle respondenten schat 21% in dat leerlingen tussen de 25 en 50 toetsen per jaar krijgen voorgelegd. Een merendeel schat dat hoger in: tussen 51 en 75 toetsen (24%) en tussen de 76 en 100 toetsen (23%). Verder geeft ruim de helft van de docenten aan minder (33%) of veel minder (21%) dan collega's te toetsen. Een derde zegt ongeveer evenveel te toetsen en zo'n 1% veel meer. Het beeld is dus gedifferentieerd. Het is lastig om te duiden in hoeverre het hier gaat om perceptie, of dat de steekproef inderdaad uit docenten bestaat die relatief weinig toetsen. Het zou voor scholen interessant kunnen zijn om eens te kijken in hoeverre de eigen perceptie en de praktijk overeenkomen.”

In artikelen en blogs worden regelmatig vierkante uitspraken gedaan over toetsing in ons onderwijs. Jullie hebben een aantal uitspraken voorgelegd?

“Ja. In het onderzoek konden de respondenten reageren op een aantal stellingen. We zien dat docenten in de onderzoeksgroep de toetsdruk hoog vinden en de invloed van rendementsdenken ervaren. Op de vraag of er sprake is van ‘toetsterreur’ wordt wisselend gereageerd.”

Voeren scholen een centraal toetsbeleid?

“We zien een wisselend beeld. Een aantal docenten geeft aan dat er op hun school een gemeenschappelijk gedragen toetsingsbeleid is, terwijl anderen juist heel veel individuele vrijheid ervaren. We hebben ook gevraagd of er deskundigheidsbevordering is ten aanzien van toetsen. Uit de reactie proeven we dat scholen hierin grote stappen zetten en dat toetsing steeds meer wordt gezien als een gemeenschappelijke activiteit. De scholen die hier het meest actief in zijn, zijn de scholen die zich erg inzetten voor het formatief toetsen.”

De interesse voor formatief toetsen blijft groot?

“Ja, men geeft aan behoefte te hebben aan meer en verdiepende kennis en handvatten met betrekking tot formatief toetsen. Maar docenten geven ook hun dilemma's aan. Het gaat dan bijvoorbeeld om verwachtingen die leerlingen en de school hebben. Zoals een van de respondenten het verwoordde: 'Helemaal zonder [toetsen] kan denk ik niet, maar het zou fijn zijn als we leerlingen meer kunnen helpen met formatieve toetsen, zodat ze kunnen zien waar ze staan en wat al goed gaat en wat niet. Ik probeer dit

	N	Helemaal oneens	oneens	neutraal	eens	Helemaal mee eens
Het aantal toetsen is te hoog	126	6%	21%	13%	37%	22%
Het onderwijs kan prima zonder becijferde toetsen	128	20%	39%	16%	16%	8%
Cijfers zijn het resultaat van rendementsdenken	121	7%	15%	23%	40%	15%
Toetsing vergt onevenredig veel onderwijstijd	126	6%	36%	21%	26%	11%
Er is sprake van 'toetsterreur'	126	12%	29%	24%	18%	17%
De toetscultuur bevordert de zesjes-mentaliteit	124	10%	18%	20%	27%	24%

ook, maar het lukt me nog niet om op deze manier de leerlingen zover te krijgen. Ze willen toetsen zodat ze kunnen scoren.' Uit de voorbereidende gesprekken kwam deze competitie om aandacht ook naar voren. Dit onderstreept het belang van een gezamenlijke aanpak wil je iets aan de toetsintensiteit veranderen.

In het onderzoek hebben we ook specifiek gevraagd of docenten door corona anders tegen toetsing zijn gaan aankijken. Een van de docenten geeft daarbij aan: 'Ik ben cijfers tegelijkertijd belangrijker (als meetpunt) en minder belangrijk gaan vinden. Corona heeft laten zien wat een effect thuis leren heeft (negatief) en dat dat op school op zich weer snel bij te spijkeren is. Voor leerlingen is het cijfer nog steeds de hoepel, maar na het gratis overgaan zoeken ze meer naar wat het niveau zou moeten zijn. En daardoor zijn ze nu juist wat procesgerichter, lijkt het. Formatief is interessant, maar alleen formatief werkt denk ik niet.' Dat is een mooie weergave hoe verschillende vormen elkaar versterken; het is niet of-of, maar en-en."

Wat stel je vast, op basis van deze eerste survey?

"We zien heel veel goede ontwikkelingen. Er is op veel scholen aandacht voor toetsbeleid en ondersteuning (nascholing) bij toetsing. Tegelijkertijd zien we specifieke aspecten waar ruimte is voor verbetering. Een belangrijke punt is bijvoorbeeld het goed bepalen van de zak-slaaggrenzen. Er zijn technieken die daarbij kunnen helpen, maar die lijken nog onbekend. Een ander punt is dat er nog weinig samenwerking tussen collega-docenten lijkt te zijn op het vlak van toetsconstructie of het leren van elkaars toetsen. Inzetten op zo'n samenwerking kan écht voordelen opleveren."

Zien jullie in de survey trends die voor gestandaardiseerde toetsen het onderzoeken waard zijn?

"In zekere zin. Neem de feedbackfunctie: hoe kunnen we ervoor zorgen dat gestandaardiseerde toetsing voldoende inhoudelijke feedback geeft? Een andere verdiepende onderzoeksvraag kan zijn hoe de gestandaardiseerde toetsen nog beter kunnen integreren in de hele toetscyclus, waarbij de verschillende vormen van toetsen elkaar als het ware een spiegel voorhouden."

In de inleiding van jullie vragenlijst schreven jullie: 'Om te weten hoe een leerling ervoor staat – tussentijds of als afsluiting – zijn toetsen onmisbaar.' Feitelijk was dat een aanname. Is dat wel zo?

"Ja. Toetsen – van goede kwaliteit – zijn nodig om tot een afgewogen oordeel te komen over hoe een leerling ervoor staat en maken het ook inzichtelijk voor leerlingen en hun ouders en verzorgers. In onze ogen hebben toetsen twee wezenlijke, onderscheidende kenmerken ten opzichte van andere vormen van evalueren en feedback. Ten eerste is sprake een 'systematisch, transparant en uitlegbaar onderzoek naar de leerprestaties van leerlingen met betrekking tot concrete leerdoelen'. Een goede toets begint dus bij een zorgvuldige taakgerichte constructie. Ten tweede wordt bij toetsing een normering gebruikt. Het gaat er daardoor niet alleen om wat een leerling wel of niet beheerst, maar ook hoe zich dat verhoudt tot de relevante leerdoelen, en de kwaliteit van het geleverde werk. Met andere woorden: aan een toets hangt in onze

ogen een kwalitatief (normatief) oordeel vast. Dit komt tot uiting wanneer we cijfers geven: 5 = 'onvoldoende' en een 8 = 'goed'. Wij denken dat het belangrijk is dat docenten en leerlingen op gezette momenten inzicht krijgen in hoe het ervoor staat. Niet alleen ten aanzien van de te leren stof en van wat inhoudelijk verwacht wordt, maar ook in termen van een waardering.

We hebben er ook naar gevraagd. Onder de respondenten wordt hierover verschillend gedacht, maar 52% onderschrijft de stelling."

Welk cijfer geven jullie de Nederlandse toetspraktijk?

"Het zou niet fair zijn die vraag te beantwoorden, want ons onderzoek heeft niet die reikwijdte, en bovendien is de toetspraktijk daarvoor te divers. Met die vraag illustreer je wel hoe ingeburgerd het geven van cijfers is, zelfs voor allerlei niet-schoolse waarderingen. Het illustreert ook een van de fundamentele functies van cijfers. Het geven van cijfers is ooit mede ontstaan om op een efficiënte manier aan ouders duidelijk te maken hoe de docent vond dat een leerling het deed op school in een tijd dat er nog sprake was van een hoge mate van laaggeletterdheid. Helaas was dat ook in een tijd dat het oordeel van de docent allesbepalend was, en dat was niet zelden geladen met allerlei vooroordelen. Dat laatste is gelukkig veranderd, maar de communicatieve functie van cijfers is gebleven."

Dit was de eerste keer dat Cito een online survey onder docenten hield. Het begin van een traditie?

"We merken dat de interesse groot is en zien dat docenten het graag goed willen doen en meer willen leren. We gaan dit onderzoek zeker een vervolg geven. Daarbij zullen we dan meer inzoomen op specifieke vragen, zoals hoe docenten hun eigen deskundigheid op het gebied van toetsen ervaren, zowel vanuit de praktijk als vanuit hun opleiding. We gaan hierover graag met scholen in gesprek" ●



Wilco Emons



Emmelien van der Scheer



Bas Hemker

Dr. Wilco Emons is senior wetenschappelijk onderzoeker binnen het team Onderwijskundig Onderzoek van Stichting Cito en houdt zich bezig met onderzoek naar toetsgebruik, technologische ontwikkelingen en de maatschappelijke en ethische kanten ervan. Dr. Emmelien van der Scheer is wetenschappelijk onderzoeker binnen het team Onderwijskundig Onderzoek van Stichting Cito en gespecialiseerd in het meten van spreekvaardigheid bij kinderen, empathisch handelen, en leestoetsen in het mbo. Dr. Bas Hemker is senior wetenschappelijk medewerker binnen team Psychometrisch Onderzoek en Dienstverlening van Stichting Cito. Hij houdt zich bezig met onder andere peilingsonderzoek, toetsgebruik, en de normering van eindtoetsen.

	N	Helemaal oneens	oneens	Neutraal	eens	Helemaal mee eens
Cijfers geven goed inzicht in hoe de leerling ervoor staat.	126	4%	21%	23%	46%	6%