

In de schijnwerpers: Elise Bruijn-Cromptvoets

■ Carlijn van Herpt

C. van Herpt MSc MA is toetsdeskundige bij Stichting Cito, divisie CTE, sector primair onderwijs en lid van de redactie van Examens. E-mail: carlijn.vanherpt@cito.nl

In deze rubriek zetten we iemand in de schijnwerpers die via verschillende activiteiten bijdraagt aan het verhogen van de kwaliteit van toetsing in Nederland. In deze aflevering is het de beurt aan Elise Cromptvoets.

Wie ben je en wat is je achtergrond?

Ik ben Elise Bruijn-Cromptvoets. Ik heb een bachelor Psychologie en Gezondheid aan de universiteit van Tilburg afgerond. Daarna heb ik de Research Master Social and Behavioral Sciences met de specialisatie Methodology and Statistics gedaan, ook in Tilburg. Mijn promotieonderzoek, uitgevoerd bij Tilburg University en Cito, richtte zich op paarsgewijze vergelijking voor het meten in het onderwijs.

Ik heb twee jaar als universitair docent gewerkt bij Tilburg University bij het departement Methode en technieken van onderzoek. Sinds een aantal jaar werk ik als onderzoeker en adviseur bij Sardes, wat onderdeel is van Stichting CAOP. Sardes maakt zich sterk voor gelijke kansen voor kinderen in het onderwijs.

Hoe, wanneer of door wie ben je geraakt door het onderwerp toetsing?

Ik ben voor het eerst door het onderwerp toetsing geraakt binnen de bachelor Psychologie. Ik vond het heel interessant om te leren wat bij het opzetten van een goede psychologische test allemaal komt kijken. En het maakt concreter op basis waarvan je als psycholoog een diagnose stelt. Gelijke kansen vond ik altijd al erg belangrijk en testen inzetten voor gelijke kansen sprak me aan.

Daarom was het extra leuk dat Tilburg Univer-

sity een lijntje met Cito had, waardoor ik bij Cito mijn masterscriptie kon schrijven. Gedurende mijn stage verschoof mijn aandacht van testen bij psychologie naar toetsing in het onderwijs. Bij Cito ben ik ook in contact gekomen met Anton Béguin die het inzetten van paarsgewijze vergelijking in het onderwijs wilde gaan verkennen. Daarin zijn we de samenwerking aangegaan en vanuit deze samenwerking is het idee ontstaan om het onderwerp verder



te onderzoeken in een promotietraject, waarbij Anton als co-promotor optrad.

Waar heb je onderzoek naar gedaan en wat heb je ontwikkeld?

Ik heb onderzoek gedaan naar paarsgewijze vergelijking in het onderwijs. Het idee achter deze beoordelingsmethode is dat beoordelaars leerlingwerken in paren vergelijken. Daarbij bepalen ze welke van de twee leerlingen het meest heeft laten zien op een bepaalde eigenschap of vaardigheid. Deze methode kun je goed inzetten wanneer je complexe, kwalitatieve producten moet beoordelen, waarvan de kwaliteit lastig objectief te scoren is. Denk aan essays, schrijfp opdrachten of ontwerpwerk. Zo kun je bijvoorbeeld kijken welke van twee leerlingwerken het meest creatief is. Als meerdere beoordelaars dat voor meerdere leerlingwerken doen, krijg je uiteindelijk een rangorde in de leerlingwerken van minst naar meest creatief. Als je dat wilt kun je de rangorde weer koppelen aan een beoordelingsschaal.

Een belangrijk nadeel van deze methode is dat er doorgaans veel vergelijkingen nodig zijn om tot een betrouwbare meting te komen. Daarom heb ik een adaptief algoritme ontwikkeld dat steeds die vergelijkingen selecteert waarover nog de meeste onzekerheid bestaat en die dus de meeste informatie oplevert. Paren van leerlingwerken met een heel hoge en een heel lage kwaliteit worden minder geselecteerd en daardoor minder vergeleken. Van deze paren kun je namelijk uit de andere vergelijkingen afleiden welk leerlingwerk beter is. Hierdoor kan het aantal benodigde vergelijkingen worden verminderd zonder dat dit ten koste gaat van de meetkwaliteit.

Wat zijn je belangrijkste bevindingen?

Een van de twee algoritmes die ik heb ontwikkeld kan gebruikt worden om paarsgewijs vergelijking efficiënter toe te passen zonder vertekend beeld.

Een van de nevenbevindingen in mijn eerste twee studies van het promotieonderzoek was

dat de maat die wordt gebruikt om te zien hoe betrouwbaar de meting was, dus of er goed onderscheid gemaakt kon worden tussen leerlingen, niet zo goed was. Daar heb ik een simulatiestudie naar gedaan, waaruit bleek dat de betrouwbaarheidsmaat in sommige situaties niet betrouwbaar was, bijvoorbeeld wanneer er weinig variatie in de vaardigheden van leerlingen is. Ook wanneer adaptieve algoritmes gebruikt worden die, in tegenstelling tot het door mij ontwikkelde algoritme, onvoldoende rekening houden met de meetonzekerheid, kan de betrouwbaarheidsmaat niet vertrouwd worden.

Het onderzoek naar paarsgewijs vergelijken in het onderwijs is nog vrij recent en er zijn nog zoveel meer aspecten te onderzoeken. Toen ik begon met mijn promotieonderzoek was er al wel wat onderzoek gedaan, zoals in Nieuw-Zeeland en Engeland. In Antwerpen waren ze het onderzoek (waaruit later Comproved is ontstaan) net aan het opstarten. Het onderzoek dat er was, was allemaal erg praktisch ingestoken. Ik had in eerste instantie bedacht om na het algoritme de praktijk in te gaan, maar toen bleek dat de fundamenteën nog niet zo breed onderzocht waren, was dat eerst nodig. Naar mijn weten ben ik de eerste en enige die op deze manier naar betrouwbaarheid bij paarsgewijze vergelijking heeft gekeken.

Het laatste onderzoek dat ik in mijn proefschrift beschrijf, gaat over het gebruiken van paarsgewijze vergelijking door experts voor het linken van toetsen: de scores op twee toetsen die hetzelfde meten vergelijkbaar maken, zonder dat in de toetsen identieke vragen (ankeritems) zitten, zodat de moeilijkheid van de toets niet uitmaakt voor welke score je krijgt. Denk hierbij bijvoorbeeld aan de examens van twee verschillende jaren voor hetzelfde vak. Ik heb een model ontwikkeld dat twee toetsen zonder identieke vragen op één schaal kan zetten door paarsgewijze vergelijkingen van toetsvragen door experts. Op deze manier is paarsgewijze vergelijking in te zetten voor het

meten van onderwijs, zonder de onbetrouwbare betrouwbaarheidsmaat nodig te hebben. Maar ook hierbij zijn er veel vergelijkingen nodig van experts. Bovendien werkt het model op dit moment alleen nog met dichotome items. Mogelijk wordt het daardoor, voor zover ik weet, nog niet zoveel gebruikt.

Wie heeft het meeste baat bij jouw resultaten?

In eerste instantie toetsontwikkelaars. Zij weten nu wat er wel en niet kan met paarsgewijs vergelijken ten aanzien van adaptieve algoritmes en de betrouwbaarheidsmaat. En ze kunnen mijn model gebruiken om toetsen aan elkaar te linken. Dit is niet beperkt tot één sector, maar geldt voor alle lagen van het onderwijs. Leerkrachten en andere professionals weten nu dat ze voorzichtig moeten zijn bij het gebruiken van paarsgewijs vergelijken voor het meten van eigenschappen of vaardigheden. Paarsgewijze vergelijking kan waardevol zijn binnen het leerproces, bijvoorbeeld voor formatieve doeleinden, maar is op dit moment door de beperkte betrouwbaarheid minder geschikt voor hoge-stakes beslissingen.

Waar zie je de grootste uitdagingen op het gebied van toetsing en hoe ziet het ideale toetslandschap er voor jou uit?

Ik denk dat er op het gebied van toetsing twee grote uitdagingen zijn. Een is dat er nu heel veel aandacht gaat naar datgene wat er goed gemeten kan worden. Ik denk dat belangrijke kennis en vaardigheden die niet goed gemeten kunnen worden, zoals sociale of emotionele vaardigheden, daardoor soms in het gedrang komen. Ik vind het belangrijk dat daar genoeg aandacht naar uit blijft gaan.

Een andere uitdaging van toetsing is de balans houden tussen willen weten waar leerlingen staan in hun ontwikkeling, maar niet te veel gaan toetsen. In het ideale toetslandschap worden toetsen ingezet op die momenten waar het waarde toevoegt, zoals op dit mo-



Belangrijke kennis en vaardigheden die niet goed gemeten kunnen worden, zoals sociale of emotionele vaardigheden, komen daardoor soms in het gedrang.

ment in mijn ogen bij de doorstroomtoets. Hierbij heb je een objectievere second opinion naast het oordeel van de leerkracht. Maar ook op andere momenten kan toetsen zeker van meerwaarde zijn. Je wilt weten of de lesstof goed is overgekomen en waar er nog groeimogelijkheden zijn voor leerlingen. Er moet vooral niet te veel getoetst worden. Dit omdat toetsdruk een negatieve invloed kan hebben op de gezondheid van leerlingen en je met toetsen mogelijk ook tijd ontnemt van het leren zelf.

Een oplossing hiervoor ligt misschien in vertrouwen dat leerlingen leren wat de docent ze aanleert, zonder te toetsen of ze het daadwerkelijk geleerd hebben. We moeten dan wel

de condities creëren waarbij dat vertrouwen realistisch is.

Wat is je ambitie voor toekomstig onderzoek?

Kansengelijkheid heb ik altijd al erg belangrijk gevonden. Waar dat precies vandaan komt weet ik niet, maar het is altijd al een drive geweest die in me zat. In mijn ogen is toetsing een onderdeel dat ingezet kan worden voor kansengelijkheid. In mijn werk bij Sardes nu kijk ik breder naar kansengelijkheid dan alleen bij toetsing, bijvoorbeeld ook naar beleid, onderzoek in opdracht van het ministerie van OCW of het ondersteunen van gemeenten bij het opstellen, uitvoeren en/of evalueren van onderwijsachterstandenbeleid, dat vind ik een mooie verbreding van onderzoek naar kansengelijkheid.

Wat is de belangrijkste boodschap die je de lezers mee wil geven?

Als we het hebben over mijn onderzoek, zou ik willen meegeven: wees voorzichtig met het gebruik van paarsgewijze vergelijking voor het meten van vaardigheden. In ieder geval totdat het overschatten van de betrouwbaarheid is opgelost. Wat betreft toetsing: gebruik toetsen met mate en hou ook aandacht voor datgene wat niet getoetst kan worden. ■

Verder lezen?

Het proefschrift van Elise Bruijn-Crompvoets is te vinden op [Behind the scenes of pairwise comparison for educational measurement - Tilburg University Research Portal](#)

Gesignaleerd

Onterechte kritiek op rekenprestaties schaadt onderwijs

Aob, Kees Buijs en Koeno Gravemeijer - 7 september 2026

Dat de leerresultaten voor rekenen en wiskunde in het basisonderwijs gestaag zouden dalen is een misvatting, schrijven Koeno Gravemeijer en Kees Buijs. Hoog tijd, vinden zij, voor een herziening van de criteria die de Onderwijsinspectie hiervoor gebruikt.

Lees verder op: <https://www.aob.nl/onderwijsblad/columns-en-opinie/onterechte-kritiek-op-rekenprestaties-schaadt-onderwijs>