In de schijnwerpers: Eva de Schipper

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 dr. Eva de Schipper. Zij is werkzaam als research & development specialist bij CitoLab, de onderzoeks- en innovatieafdeling van Stichting Cito.

Wie ben je en wat is je achtergrond?

Mijn naam is Eva de Schipper. Als research & development specialist bestaat mijn werk voor een groot deel uit het ontwikkelen van concepten en prototypes. We krijgen bijvoorbeeld een vraagstuk waarvan we denken dat we er met technologie of onderzoek iets aan kunnen verbeteren. Samen met het team ontwikkelen we dan een prototype, waar ik producteigenaar van word.



Ik studeerde psychologie en behaalde een onderzoeksmaster in methodologie en statistiek aan de Universiteit Utrecht en de Universiteit Twente. Tijdens een stage bij de afdeling epidemiologie van het VUmc onderzocht ik vragenlijsten in de gezondheidszorg. Dat vond ik erg interessant, maar mijn interesse verschoof naar ontwikkelings- en leerprocessen, waarna ik als stagiaire bij Cito terechtkwam.

Ik werkte meerdere jaren als psychometrisch onderzoeker, onder meer aan psychometrische analyses voor centrale examens, zoals het equivaleren van digitale vmbo-examens. Daarna stapte ik over naar het team Prototypes, waar we Examenkompas ontwikkelden, een gratis online oefentool met domeinrapportages die leerlingen gericht inzicht biedt welke onderwerpen nog aandacht verdienen. Later werkte ik aan Biologie+, een prototype voor gepersonaliseerde oefentoetsen, en CheckMate, een prototype voor nakijkondersteuning. Deze prototypes combineren onderzoek met directe meerwaarde voor scholen.

Mijn promotieonderzoek liep parallel aan deze werkzaamheden en had veel inhoudelijke overlap. Een groot goed van promoveren is dat je tijdens werktijd over van alles kunt en mag leren, dat leek me ontzettend leuk. Ik voerde mijn promotieonderzoek uit onder begeleiding van Prof. Bernard Veldkamp en Prof. Remco Feskens aan de Universiteit Twente.

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

De rode draad in mijn promotieonderzoek was de vraag hoe we bestaande toetsdata (scores, antwoorden, procesdata) kunnen benutten om meer waarde toe te voegen voor leerlingen en docenten.

De vier onderzoeken die ik heb uitgevoerd waren alle vier verbonden aan deze rode draad. De eerste twee onderzoeken gingen over gepersonaliseerde oefentoetsen. Daarnaast onderzocht ik of we uit procesdata konden afleiden hoe leerlingen het oplossen van een wiskundevraag aanpakken. Het vierde onderzoek richtte zich op digitale nakijkondersteuning voor docenten bij open toetsvragen. Dit onderzoek heeft de meeste overlap gehad met mijn werk bij Cito, omdat dit onderzoek werd uitgevoerd met het prototype CheckMate.

Wat zijn je belangrijkste bevindingen en wie heeft daar het meeste baat bij?

Gepersonaliseerde oefentoetsen stemmen vraagselectie af op individuele knelpunten, wat in theorie efficiënt oefenen mogelijk maakt. In mijn onderzoek onderzochten we een nieuwe methode om op basis van eerder gemaakte toetsvragen prestaties op nieuwe toetsvragen te voorspellen. Vervolgens heb ik onderzocht wat er in de praktijk gebeurt als leerlingen daadwerkelijk gepersonaliseerde oefentoetsen maken; groeien deze leerlingen meer dan leerlingen die een standaard oefentoets krijgen?

Daarvoor hebben we het prototype Biologie+ ontwikkeld. Ik vond het erg belangrijk dat deelname aan het onderzoek direct meerwaarde had voor deelnemende scholen. We hebben met Biologie+ echt een programma gebouwd dat ze konden gebruiken in de aanloop naar de Centrale Examens. In de praktijk resulteerde het echter niet in een aantoonbare verbetering in de scores van leerlingen die gepersonaliseerde oefentoetsen maakten, ten opzichte van leerlingen die een standaard oefentoets kregen. Dat betekent niet dat personalisatie nooit werkt, maar in het geval van deze oefentoetsen bleek dat niet per se het geval.

In het derde onderzoek heb ik procesdata gebruikt van een Franse digitale wiskundepeiling. Procesdata, zoals gegevens over waar leerlingen klikken in een afnameomgeving, wat ze invullen en hoe ze zich door een digitale toetsomgeving bewegen, zijn veelbelovend. Je kunt er mogelijk uit afleiden welke oplossingsstrategieën een leerling gebruikt, en daar kun je wellicht je onderwijs op afstemmen. We konden inderdaad iets zeggen over de manier waarop leerlingen met de toets omgingen. Met name weinig interactieve leerlingen, die niet zo bezig waren met het maken van de toets, konden we goed herkennen. Dat kan verschillende oorzaken hebben: een leerling kan bijvoorbeeld afgeleid zijn of niet goed weten waar hij moet beginnen. Echter is er meer onderzoek nodig om goede inferenties te maken over wat er echt bij een leerling gebeurt. Daar worden we steeds beter in, maar procesdata vertellen niet vanzelf het hele verhaal.

In het laatste onderzoek hebben we het prototype CheckMate ontwikkeld, gericht op het ondersteunen van docenten bij het nakijken van open vragen. We hebben onderzocht hoe we de omgeving voor docenten zo prettig en efficiënt mogelijk kunnen inrichten. Zo hebben we vergelijkbare antwoorden van leerlingen bij elkaar gezet. Wie op papier nakijkt, moet soms terugzoeken naar een antwoord dat eerder op de stapel lag. Als vergelijkbare antwoorden bij elkaar worden gezet, hoeft een docent niet opnieuw na te denken over een antwoord dat



Op alle aspecten van toetsing gebeurt er al van alles wat met AI te maken heeft.

sterk lijkt op een eerder beoordeeld antwoord. Dat kan het nakijken overzichtelijker en consistenter maken. Daarnaast hebben we geëxperimenteerd met automatische scoresuggesties; bepaalde delen van een leerlingantwoord markeren die relevant zijn voor puntenverlening. Docenten reageerden positief, volgens mij vanwege de combinatie van een gebruiksvriendelijke omgeving en steeds betere analysetechnieken, ook met behulp van Large Language Models.

Uiteindelijk hoop ik dat zowel leerlingen als docenten van het onderzoek profiteren. Ik vind het erg belangrijk om me te richten op hoe we het schoolleven van docenten en leerlingen kunnen verbeteren. Hoe kunnen we ze meer geven, zonder dat we daarvoor meer van hen vragen?

Wat zie je als belangrijke aandachtspunten in het onderwijs/onderzoek op het gebied van toetsing?

In het onderwijs staan leerlingen onder druk om te presteren en hebben docenten vaak een hoge werklast. Daarom vind ik dat we zorgvuldig moeten omgaan met nieuwe toepassingen. Als we bestaande gegevens kunnen gebruiken om leerlingen beter inzicht te geven in hun ontwikkeling of docenten te ondersteunen bij het nakijken, kan dat waardevol zijn. Zeker als we daarmee iets toevoegen zonder meer belasting te creëren.

Ik zou het heel mooi vinden als we steeds meer richting laag-intrusief meten kunnen gaan: informatie halen uit bestaande activiteiten en gedrag in plaats van ze op één specifiek moment een summatieve toets te laten maken. We kunnen meer leren van het gedrag dat leerlingen al laten zien en de activiteiten die sowieso al worden ondernomen in de klas.

Dat betekent niet dat toetsen overbodig zijn. De druk die samengaat met toetsing vind ik problematisch, maar door een leerling actief te bevragen, kan er juist geleerd worden. Ook summatieve toetsing en diplomering blijft nodig. Van een dokter of piloot verwachten we immers dat deze voldoende kennis en vaardigheden heeft aangetoond. Maar toetsen hoeven niet altijd los te staan van het leerproces.

Hoe zie je de toekomst van toetsing voor je in relatie tot de opkomst van AI?

Op alle aspecten van toetsing gebeurt er al van alles wat met AI te maken heeft. Ook bij Cito doen we onderzoek naar de vraag waar in het toetsproces AI goed kan worden ingezet, waar de grenzen liggen en op welke momenten de mens noodzakelijk blijft. AI biedt geweldige kansen, maar het is geen oplossing voor alles. Ik zou graag zien dat we als onderwijs blijven leren en zelf blijven nadenken. Welke vaardigheden precies belangrijk zullen zijn, is lastig te voorspellen. We weten niet hoe de wereld er

over tien jaar uitziet. Juist daarom moeten we ervoor waken dat technologische vooruitgang ten koste gaat van de ontwikkeling van het denkvermogen van leerlingen.

Wat is je ambitie voor de toekomst?

Ik zou het heel mooi vinden als nakijkondersteuning straks breed verspreid is in het onderwijs. Dat hoeft niet per se met onze eigen tool te zijn, maar ik zou het een heel mooie ontwikkeling vinden als ethisch gebruik van nakijkondersteuning beschikbaar is voor elke docent, en dat wat wij hierover hebben geleerd hierbij wordt gebruikt. Dat is niet alleen gunstig voor docenten, maar draagt ook bij aan een eerlijkere beoordeling van leerlingen. Wanneer vergelijkbare antwoorden op een consistente manier worden beoordeeld, kan dat helpen om verschillen in beoordeling te beperken.

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

Vooruitgang en technologie is prachtig en biedt veel kansen, maar blijf ook kritisch. Dat geldt voor AI, maar ook voor personalisatie en op andere gebieden. Mijn onderzoek naar gepersonaliseerde oefentoetsen laat juist zien dat een toepassing die in theorie veelbelovend is in de praktijk niet automatisch tot betere leerresultaten leidt.

En vergeet daarbij de mensen niet om wie het uiteindelijk gaat. Docenten hebben te maken met een vaak hectische werkpraktijk en leerlingen moeten de ruimte krijgen om zich te ontwikkelen. Wij kunnen vanuit onderzoek en innovatie mooie oplossingen bedenken, maar we moeten goed luisteren naar de mensen die ermee werken. Soms zit de grootste winst in een relatief eenvoudige verbetering. Juist daarom vind ik het belangrijk om stakeholders direct actief te betrekken bij onderwijskundig onderzoek en ontwikkeling door middel van co-design of consultatie.

De kern blijft voor mij: hoe kunnen we leerlingen en docenten zo goed mogelijk ondersteunen, zonder ze extra lasten op te leggen? ■

Verder lezen?

Het proefschrift van Eva de Schipper is te vinden op de website van Cito.

Gesignaleerd

Meer controle, meer werkdruk voor examencommissies?
30 oktober 2026 Hoger Onderwijs Persbureau en Noëlle van den berg

Het artikel bespreekt de spanning tussen de toenemende verwachtingen aan examencommissies en de beschikbare capaciteit. Meer toezicht op eindniveau en toetskwaliteit kan betekenen dat examencommissies extra werk krijgen zonder overeenkomstige ondersteuning.

Het eindniveau van opleidingen moet beter gecontroleerd worden door examencommissies. Dat wil minister Letschert, omdat studenten steeds vaker gebruikmaken van AI. Examencommissies blijken echter niet genoeg tijd te hebben voor dat werk. Hoe zit dat bij Fontys?
Lees verder: https://bron.fontys.nl/meer-controle-meer-werkdruk-voor-examencommissies/."