

Leren en Toetsen: Samen naar Succes!

Inaugurele lezing uitgesproken ter aanvaarding van de leerstoel Leren en Toetsen aan Maastricht University op vrijdag 8 november 2024 door prof. dr. Desirée Joosten-ten Brinke

Leren en Toetsen: Samen naar Succes!

Inaugurele lezing
Prof. dr. Desirée Joosten-ten Brinke

8 November 2024



Inhoudsopgave	
Introductie	3
Promovendi.....	4
De doelen in het onderwijs.....	5
Assessment <i>of</i> Learning	5
Assessment <i>for</i> Learning.....	8
Een oefentoets als leeractiviteit	11
Assessment <i>as</i> Learning.....	12
Erkennen en waarderen van competenties.....	14
Technologische ondersteuning van leren en toetsen	16
Ethiek bij AI bij leren en toetsen.....	19
Samenvatting	20
Tot besluit	21
Referenties.....	23
Verantwoording afbeeldingen	25

*Geachte pro rector,
Leden van de begeleidingscommissie,
Collega's,
Familie en vrienden,
Beste mensen,*

Introductie

Welkom bij mijn inaugurele lezing als hoogleraar Leren en Toetsen bij de Faculty of Health, Medicine and Life Sciences aan Universiteit Maastricht. Het is een eer om hier vandaag te mogen staan om mijn plannen voor de komende jaren met collega's, vrienden en familie te delen. Plannen die geïnspireerd zijn door velen, onder andere door collega's als Cees van der Vleuten, rond programmatisch toetsen, Paul Kirschner en Jeroen van Merriënboer, rond instructional design, en Theo Eggen, ten aanzien van psychometrie en digitaal toetsen. Helaas hebben we van de laatste twee onlangs te vroeg afscheid moeten nemen.

Ik schreef deze oratie grotendeels in de zomerperiode, in de maanden met een overdaad aan sport: De Tour de France, het EK-voetbal, en de Olympische spelen.



In al die sporten stond bij de deelnemers de zoektocht naar succes centraal. Daarbij is succes voor de een niet hetzelfde als succes voor een ander. Om te komen tot dat succes hebben al die sporters in zichzelf geïnvesteerd. Ze willen zich ontwikkelen, ze willen beter worden. En dat beter worden deden ze door keihard te trainen, oefeningen te herhalen totdat het routines werden en door zichzelf regelmatig in wedstrijden te vergelijken met anderen. Zichzelf te toetsen, waarbij de eigen prestaties vergeleken werden met anderen, met normen gesteld door de sportorganisaties, maar vooral met eerdere prestaties van zichzelf. Waarbij feedback van coaches, familie en van de sporters zelf, gebruikt werd om beter te worden en om te weten wat de volgende stappen in het leerproces zijn. Deden ze dat allemaal met veel plezier? Nee, beter willen worden vraagt om doorzettingsvermogen. En het vraagt om doelen stellen. Sifan Hassan stelde bijvoorbeeld als doel om de 5 km, 10 km en de marathon

te lopen tijdens de Olympische spelen dit jaar. Terwijl de meeste deskundigen aangaven dat dit gekkenwerk was, onderbouwde Sifan Hassan haar plannen met de uitspraak: “If your dreams don’t scare you, they’re not big enough”. Een uitspraak die toegedicht wordt aan Muhammad Ali, maar een originele bron ontbreekt.

En ik weet niet of de doelen die mijn collega’s zich stellen bij hun onderzoek of mijn promovendi in hun promotietraject angstaanjagend voor hen zijn. Het zijn wel uitdagende doelen die zij nastreven. En dat is dan niet eens het doel van de promotie of het afronden van een onderzoek, maar het doel om met hun onderzoek het onderwijs beter te maken. Impact hebben op de kwaliteit van leren en toetsen in hun organisatie, in lerarenopleidingen of in de gezondheidszorg. Het is daarom ook zeer kwalijk dat het huidige kabinet onder andere de promotiebeurs voor leraren wil afschaffen. Deze beurs maakte het mogelijk voor leraren om onderzoek te doen met als doel de kwaliteit van het onderwijs te verbeteren. En ook zorgde dit ervoor dat goede leraren behouden werden voor het onderwijs en niet de uitdagingen elders te zoeken. Wat zeker in deze tijd van het lerarentekort onwenselijk is. Leraren in het funderend onderwijs zorgen voor de basis waar het hoger onderwijs op voort moet kunnen bouwen.

Promovendi

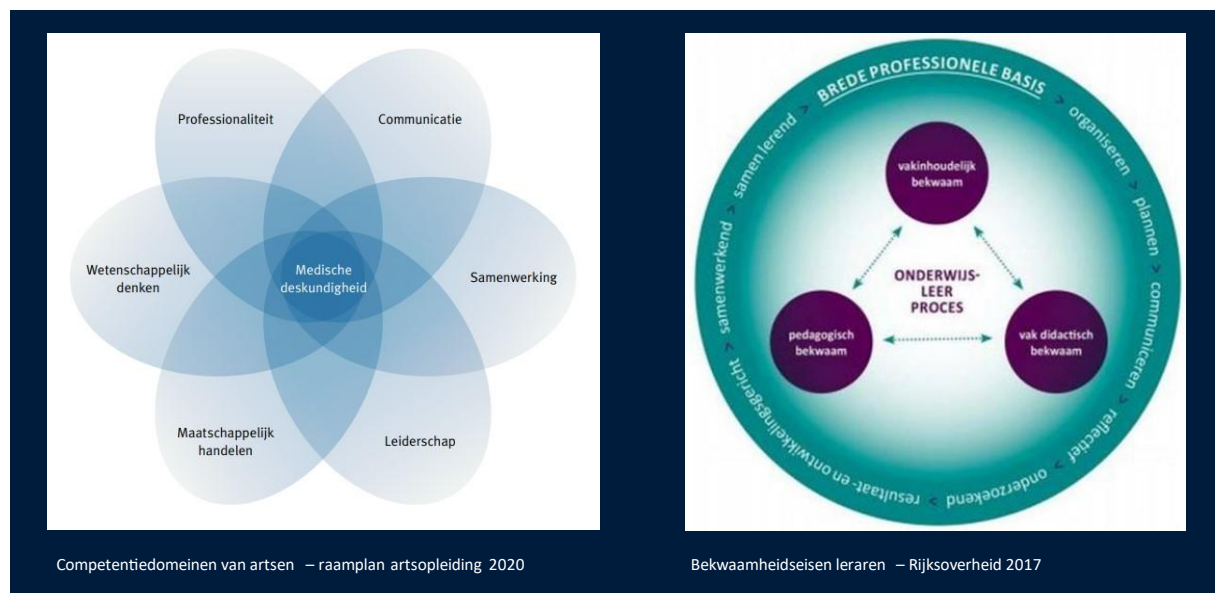
In deze lezing leg ik uit waarom het realiseren van doelen het succes is van een goede samenwerking tussen leren en toetsen, tussen onderwijswetenschappers en vakspecialisten, tussen docenten en studenten en tussen mensen en Artificial Intelligence. Met deze lezing geef ik inzicht in de thema’s van mijn leerstoel Leren en Toetsen, waar ik me samen met promovendi en vele collega’s (zie Tabel 1) de afgelopen jaren op heb gericht en de komende jaren mee verder ga, om samen met hen het onderwijs waarin we de professionals van de toekomst opleiden, beter te maken.

Tabel 1. Overzicht van de promotieteams waar ik bij betrokken ben.

Promovendi	Begeleiders
Janneke van der Steen	Tamara van Schilt-Mol Cees van der Vleuten
Sander Langbroek	Kim Dirkx
Tim Debets	Gideon Maillette de Buy Wenniger
Debby Weerlink	Tamara van Schilt-Mol Kelly Beekman
Marjolein Mantelaers	Kelly Beekman Tamara Schleepen
Agnes Brinks	Gerry Geitz
Anne Lohuis	Karin Gerritsen – van Leeuwenkamp
Zhien Li	Maryam Asoodar Diana Dolmans
Joost Godschalk	Tamara van Schilt-Mol Bas Hemker
Anne Le Roy	Piet Leroy
Marco Dumont	Xander Lub Roland Klemke
Linda Jakobs	Tamara van Schilt-Mol Helma Oolbekkink-Marchand
Jannie Lensen- Botter	Ruud Duvekot
Claudia Klarenbeek	Ruud Duvekot Jos Sanders

De doelen in het onderwijs

De doelen die we in het hoger onderwijs nastreven zijn niet de snelste tijd op de honderd meter sprint of een gele trui, maar richten zich specifiek op een beroepsprofiel, zoals de CanMed doelen in het medisch onderwijs of de bekwaamheidseisen voor leraren (zie Figuur 1). Overeenkomst tussen beroepsprofielen in verschillende domeinen is dat onze studenten complexe cognitieve vaardigheden moeten aanleren, zoals kritisch denken, het oplossen van complexe problemen, en het zelf kunnen reguleren en sturen van het leren. Vaardigheden en competenties die nodig zijn om effectief deel te nemen aan de toekomstige beroepspraktijk (Ama & Emetarom, 2020, p. 25). Ook al worden deze niet expliciet genoemd. De competenties vormen de basis voor het ontwerp van het onderwijs.

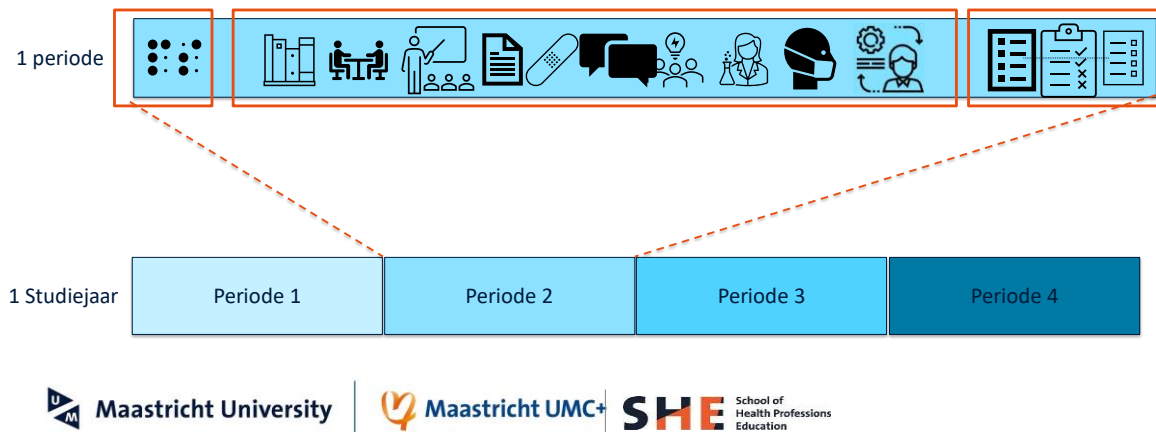


Figuur 1. Twee competentiemodellen

Assessment of Learning

Om dit te realiseren is in veel opleidingen het onderwijs nog vrij traditioneel ontworpen, waarbij toetsing is vormgegeven op basis van het perspectief van Assessment of Learning. Dat wil zeggen dat toetsing gebruikt wordt om aan het einde van een onderwijsperiode, die vaak zo'n tien tot twaalf weken beslaat, een uitspraak te doen over delen van een competentie. Worden de leerdoelen die centraal stonden in deze periode beheerst of niet? De nadruk ligt op de geloofwaardigheid van de beslissing die op basis van deze beoordeling wordt genomen (Van der Vleuten, Sluijsmans, & Joosten-ten Brinke, 2017). Na een periode van onderwijsactiviteiten wordt de onderwijseenheid afgesloten met een toetsperiode, bestaande uit een tentamen en eventueel een hertentamen twee weken later. In elke periode staat een afgebakende set leerdoelen centraal, die via de syllabus of in een eerste les aan de studenten wordt medegedeeld (zie Figuur 2).

Assessment of Learning



Figuur 2. Voorbeeld van een traditioneel onderwijsprogramma.

In dit onderwijsontwerp is de toets het sluitstuk van het leren, met als conclusie dat studenten met een bepaalde mate van zekerheid hebben aangetoond dat ze de stof van die onderwijseenheid beheersen. Als 70% van de studenten geslaagd is en de toets een betrouwbaarheid heeft van .80, zijn we over het algemeen heel tevreden. Maar helaas kunnen we op basis daarvan niet zeggen dat studenten die een 5 hebben gekregen terecht zijn gezakt, of dat degenen die een 6 hebben gekregen terecht zijn geslaagd voor die toets. Nee, op basis van enkele berekeningen kunnen we concluderen dat een deel van de studenten een onterechte uitslag krijgt. Hoe lager de betrouwbaarheid (en hoe hoger het percentage gezakten), hoe hoger het aantal onterechte uitslagen (Veldhuijzen, Goldebel, & Sanders, 1993). Alleen bij een betrouwbaarheid van 100% kunnen we uitgaan van geen enkele onterechte uitslag. Maar geen enkele toets is 100% betrouwbaar. Zelfs redelijk betrouwbare instrumenten zoals een keukenweegschaal of een koortsthermometer zijn niet 100% betrouwbaar, laat staan een toets waarbij allerlei subjectieve factoren tijdens de toetsconstructie, de toetsafname en de toetsbeoordeling verstorend kunnen werken.

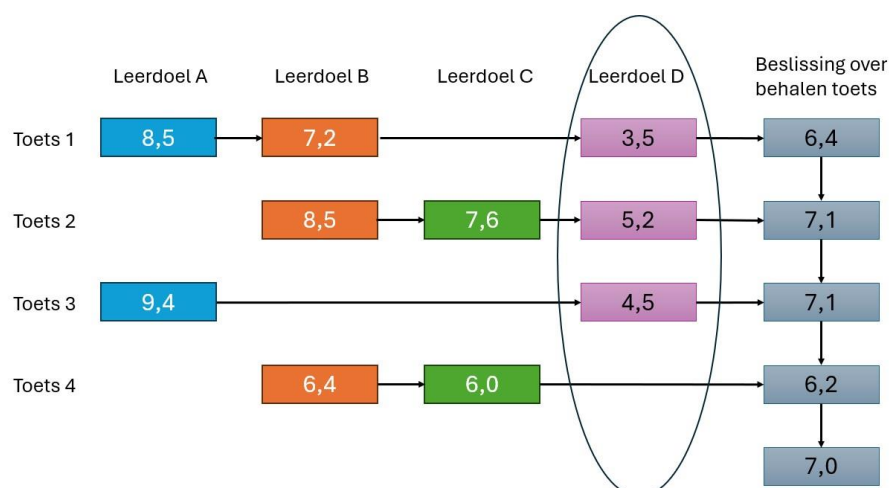
Dat geldt zeker ook bij het beoordelen van de eerdergenoemde complexe vaardigheden. Het beoordelen van deze vaardigheden vereist authentieke leeromgevingen, met toetstaken die kennis, vaardigheden en houdingen integreren (Van Merriënboer, Kirschner, & Frèrejean, 2025). Het aanleren van deze complexe vaardigheden heeft ook tijd nodig en vraagt om longitudinale leerlijnen (Carless & Boud, 2018), leerlijnen die zich over een langere tijd, zelfs over meerdere studie jaren, uitstrekken. Het beoordelen van complexe vaardigheden gebeurt vaak nog analytisch, aan de hand van checklists of rubrics met afzonderlijke criteria. Deze vooraf geselecteerde criteria kunnen echter lastig de volledige complexiteit van de gehele vaardigheid weerspiegelen (Sadler, 2009). Daarom wordt bij taken waarbij het geheel meer is dan de som van de afzonderlijke delen gepleit voor holistisch beoordelen in plaats van analytisch. In Figuur 3 is links analytisch beoordelen abstract weergegeven en rechts holistisch beoordelen.



Figuur 3. Analytisch beoordelen op basis van losse criteria (ingrediënten), rechts holistisch op basis van het eindproduct (het gerecht).

Promovenda Anne Lohuis onderzoekt in de context van deze traditionele onderwijsmodellen op welke wijze holistisch beoordelen kan bijdragen aan valide, transparante en betrouwbare uitspraken over authentieke prestaties en het leerproces van studenten. Dit nog wel in een context waar toetsing grotendeels is vormgegeven vanuit het perspectief van Assessment of Learning, waarbij op basis van toetsen beslissingen worden genomen over de som van de leerdoelen. De som van alle toetsen bepaald hierbij of de eindtermen behaald zijn (Figuur 4).

Traditioneel toetsen

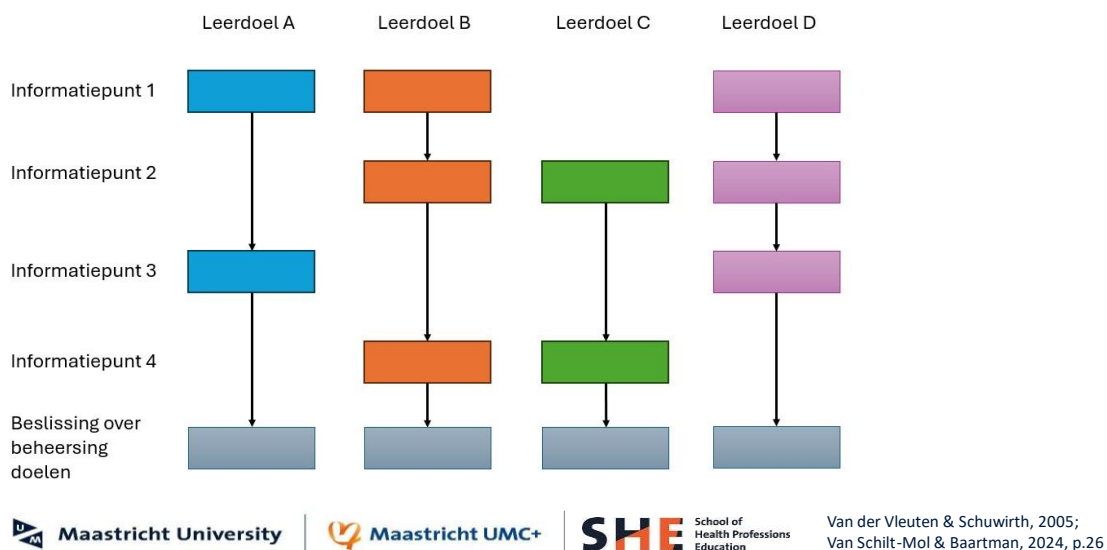


Figuur 4. Aggregatie van informatie uit toetsen bij conventionele manieren van toetsing. Overgenomen van slides gebaseerd op (Van Schilt-Mol & Baartman, 2024, p. 25)

Assessment *for* Learning

In het medisch onderwijs (internationaal) en ondertussen ook in vele andere opleidingen in Nederland, met name in het hbo, zien we dat afgestapt wordt van dit traditionele onderwijsmodel, waarbij toetsing gebruikt wordt om onderdelen van het onderwijs af te sluiten. Er wordt steeds meer ingezet op een geïntegreerde aanpak waarbij toetsing informatie oplevert die gebruikt kan worden om beslissingen te nemen én om het leren te stimuleren (Baartman, Van Schilt-Mol, & Van der Vleuten, 2021; Van Schilt-Mol & Baartman, 2024). Een benadering waarin Assessment *of* Learning en Assessment *for* Learning wordt geïntegreerd is aan het begin van deze eeuw hier aan deze universiteit geïntroduceerd door Cees van der Vleuten en Lambert Schuwirth (2005). In deze benadering, programmatisch toetsen genoemd, wordt niet alleen de uitvoering van een taak holistisch beoordeeld, maar wordt de ontwikkeling van de student tijdens de hele (of groter deel van de) opleiding holistisch beschouwd (zie Figuur 5). Belangrijke aspecten zijn continue feedback en besluitvorming op basis van geaggregeerde informatie uit meerdere bronnen. Beslissingen worden dus niet genomen op basis van afzonderlijke toetsen, maar op basis van de beheersing van de competenties op basis van informatie uit verschillende assessment activiteiten, zoals feedback, opdrachten of toetsen (Baartman, Van Schilt-Mol, & Van der Vleuten, 2021).

Programmatisch toetsen



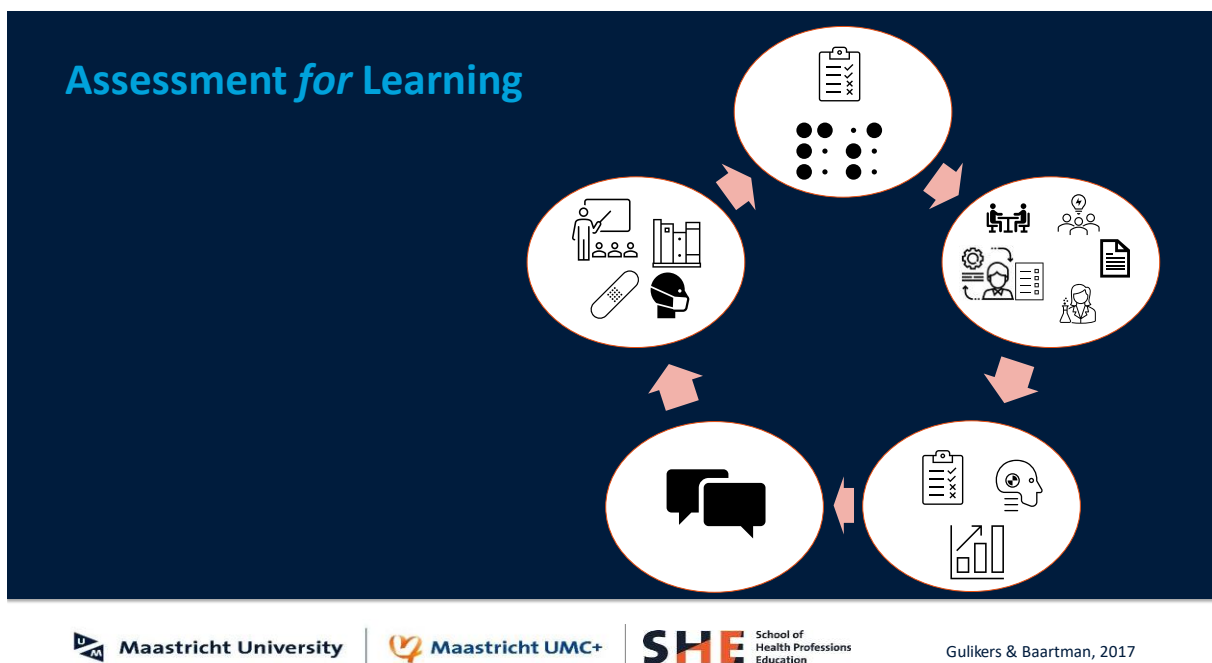
Figuur 5. Aggregatie van informatie uit informatiepunten bij programmatisch toetsen. Overgenomen van slides gebaseerd op (Van Schilt-Mol & Baartman, 2024, p. 26)

Het omzetten van een traditioneel onderwijsontwerp naar dit onderwijsontwerp is niet eenvoudig, kost tijd en moet alleen gedaan worden als het ook past bij de onderwijsvisie van een opleiding.

Als een volledige transitie niet mogelijk of wenselijk is, is het toch mogelijk om ervoor te zorgen dat er beter gebruik gemaakt wordt van de leerfunctie van toetsing, zodat leren wordt gestimuleerd. Onderzoek dat zich richt op Assessment for Learning, dus toetsen om het leren te stimuleren, ook wel formatief toetsen of formatief evalueren genoemd, geeft hiervoor vele handreikingen. Black en Wiliam (1998) introduceerden in 1998 het concept van formatief toetsen. In hun eerste definitie was formatief toetsen omschreven als 'het doelgericht uitlokken van reacties van studenten om informatie over hun leren te verzamelen en geïnformeerde beslissingen te nemen over vervolgstappen'. Daarmee kreeg

formatief toetsen vooral een connotatie van een instrument: een beoordelingsinstrument of -methode met een formatieve functie. Echter, formatief toetsen is vooral een proces dat onderdeel is van de dagelijkse onderwijspraktijk van studenten, docenten en peers, waarbij zij informatie over het leren zoeken, erop reflecteren en verwerken op een manier die het leren bevordert (AssessmentReformGroup, 1999; Broadfoot et al., 2002). In deze definitie is het proces nadrukkelijker zichtbaar.

Judith Gulikers en Liesbeth Baartman (2017) maakten dit zichtbaar in hun *Assessment for Learning*-cyclus (zie Figuur 6). Deze cyclus beschrijft het proces aan de hand van vijf fasen: 1) het verduidelijken van instructiedoelen en beoordelingscriteria; 2) het selecteren van geschikte technieken om bewijsmateriaal te verzamelen over het leren van studenten; 3) het interpreteren van verzamelde informatie om tot conclusies te komen over de ontwikkeling van de student en de effectiviteit van de instructie; 4) het formuleren en geven van feedback en 5) het ondernemen van concrete acties om het leren of de instructie aan te passen. Alle fasen zijn afgestemd op de definitie van Black en Wiliam (1998) en zijn bedoeld om duidelijk te maken waar het leren naartoe gaat (feed up), waar het leren nu is (feed back) en hoe het leren aangepast zou kunnen worden om het leren van studenten daadwerkelijk te verbeteren (feed forward) (Black & Wiliam, 1998). Uit vervolgonderzoek (Veugen, Gulikers, & den Brok, 2022) werd duidelijk dat docenten inderdaad reacties van studenten uitlokken (Fase 2), deze antwoorden analyseren (Fase 3) en feedback geven (Fase 4), maar dat zelden de leerdoelen verhelderd worden (Fase 1), noch dat het onderwijs of het leren van studenten werd aangepast (Fase 5), terwijl feedback verwerkt moet kunnen worden door studenten en docenten moeten de ruimte hebben om hun onderwijs nog aan te passen aan de behoeften van de studenten (Black & Wiliam, 2009). Om het volledige potentieel van Assessment for Learning te benutten, moeten docenten echter alle vijf fasen in hun onderwijs worden doorlopen (Hawe & Parr, 2014).

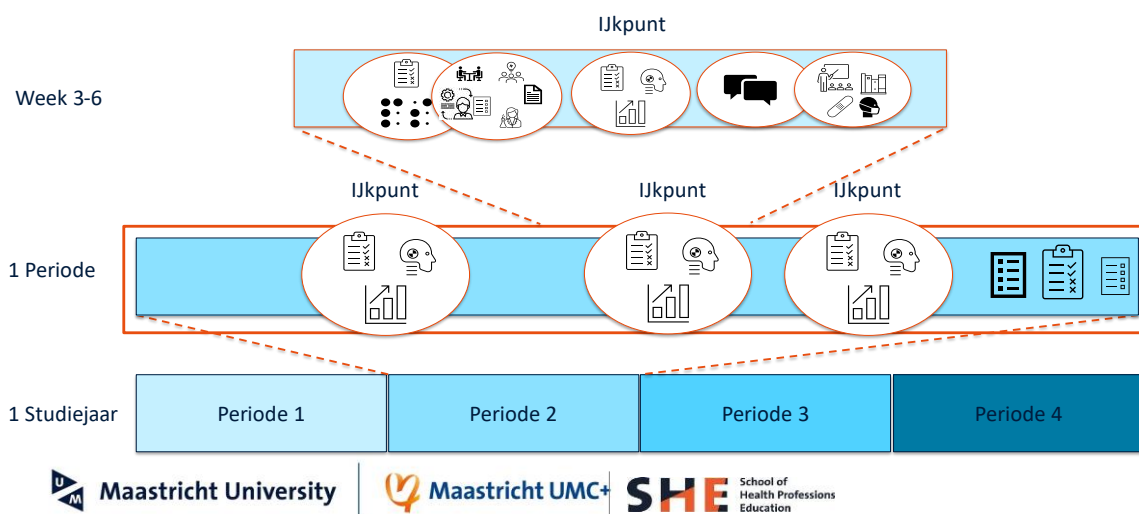


Figuur 6. De cyclus voor formatief evalueren (Gulikers & Baartman, 2017)

Promovenda Janneke van der Steen heeft in haar onderzoek (Van der Steen et al., 2022; Van der Steen et al., 2023) in het voortgezet onderwijs onderzocht hoe leraren hun onderwijs zodanig kunnen (her)ontwerpen dat er meer aandacht is voor Assessment for learning en de docent en leerlingen dus

tijdens een lessenserie op de hoogte zijn van waar ze staan in hun ontwikkeling en wat er gedaan moet worden om door te ontwikkelen, waarbij nadrukkelijk ook aandacht is voor het verhelderen van leerdoelen en ruimte bieden voor vervolgactiviteiten. Verder moet er sprake zijn van constructive alignment en moeten de leeractiviteiten informatie opleveren die gerelateerd is aan de beslissing die aan het einde van een periode genomen moet worden (zogenoemde beslissingsgerichte informatieverzameling). Uit haar onderzoek blijkt dat het ontwerpen van Assessment for Learning begint met het vaststellen van de momenten tijdens een onderwijsperiode waarop een docent wil weten waar studenten zich in hun leerproces bevinden, zogenoemde ijkpunten. Die ijkpunten zorgen ervoor dat er tijdens de onderwijsperiode beslissingen kunnen worden genomen die het leren sturen. Bijvoorbeeld als blijkt dat een theorie nog niet begrepen wordt door een grote groep studenten, dan is een extra instructie zinvol, als blijkt dat die theorie door slechts enkele studenten niet begrepen wordt is een andere beslissing nodig. Bijvoorbeeld, een verwijzing naar een kennisclip voor deze studenten of deze studenten een opdracht geven om bijvoorbeeld toetsvragen over deze theorie te schrijven. Dus bij het plaatsen van deze ijkpunten moet er al nagedacht worden over wat er gedaan kan worden met de informatie die op dat moment verzameld gaat worden. Als dat ijkpunt bepaald is, kan bepaald worden welke leeractiviteiten rijke informatie kunnen opleveren. Tevens kan dan al ruimte worden ingepland voor het verwerken van die informatie en opvolging te geven aan de verzamelde informatie (zie Figuur 7).

Assessment for Learning



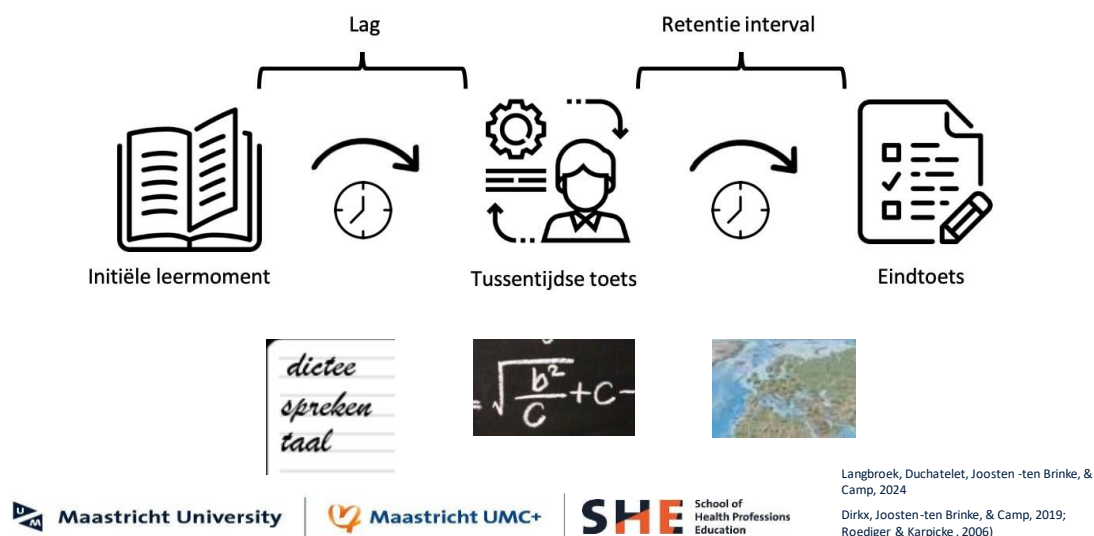
Figuur 7. Ontwerpen van onderwijs met Assessment for Learning.

Onderwijs meer ontwikkelingsgericht maken betekent dus niet het toevoegen van leeractiviteiten aan al bestaande lesplannen; het vraagt om een herontwerp waarbij bewust nagedacht wordt over de leer- en instructieactiviteiten die het leren stimuleren. De beslissing die aan het einde van een periode genomen moet worden is het uitgangspunt voor de informatie die je wilt verzamelen tijdens het leerproces. De periode waarna een beslissing genomen moet worden kan variëren. Dat kan een periode van 10 weken zijn, zoals in het onderzoek van Janneke, maar het zou ook kunnen gaan over langere leerlijnen, bijvoorbeeld een schrijflijn bij Gezondheidswetenschappen. Het is belangrijk om de beslissing goed te omschrijven. Welke beslissing wil je nemen, op welke leerdoelen is de beslissing gebaseerd, waaraan kan je zien dat de leerdoelen behaald zijn?

Een oefentoets als leeractiviteit

Een leeractiviteit waar de meesten wel bekend mee zullen zijn, is het inzetten van oefentoetsen of tussentijdse toetsen. In sommige opleidingen is zo'n oefentoets standaard beschikbaar in de leeromgeving en kan een student deze toets maken op een moment dat het voor de student het beste uitkomt. Dit moment kan echter niet het beste moment zijn om een optimaal leerrendement uit zo'n toets te halen. De toets maken op de avond voor een tentamen is fijn als uit die toets blijkt dat je de stof beheerst, maar als blijkt dat je een groot deel nog niet beheerst, is er geen ruimte meer om nog te leren. Het inzetten van oefentoetsen zou dus een bewust gekozen leeractiviteit moeten zijn, zodat oefentoetsen worden gebruikt om het leren te bevorderen (Day et al., 2018). Vanuit de cognitieve psychologie is te onderbouwen waarom het inzetten van een oefentoets een effectieve leerstrategie is. Met oefentoetsen wordt namelijk informatie uit het geheugen opgehaald, wat zorgt voor het beter onthouden van die informatie op de lange termijn (Figuur 8). Dit effect wordt het 'retrieval practice effect' of 'toets-effect' genoemd (Roediger & Karpicke, 2006). Het toetseffect is nog effectiever als de oefentoetsen in de tijd worden gespreid (distributed practice) (Donovan & Radosevich, 1999). Dit soort oefentoetsen zijn geschikt in verschillende domeinen, voor verschillende soorten leermateriaal en voor verschillende cognitieve niveaus. Feedback op de gegeven antwoorden versterkt het leereffect (Dirkx, Joosten-ten Brinke, & Camp, 2019; Dirkx et al., 2024).

Leeractiviteit 'oefentoets'



Figuur 8. Het inzetten van een tussentijdse toets versterkt het onthouden van informatie (Dirkx, Joosten-ten Brinke, & Camp, 2019)

Retrieval practice en distributed practice zijn effectieve leerstrategieën, maar niet alle docenten weten wat de strategieën inhouden of zijn niet overtuigd van de effecten van deze leerstrategieën. Promovendus Sander Langbroek heeft onderzocht voor het voortgezet onderwijs welke leerstrategie het beste past bij welk schoolvak (Langbroek et al., 2024) en is nu bezig om te onderzoeken of een training aan docenten gericht op hun kennis en opvattingen over deze leerstrategieën hen helpt om de leerstrategieën op een goede manier aan te leren aan hun leerlingen. Ook promovenda Linda Jakobs gaat in haar onderzoek aan de slag met het optimaliseren van de professionalisering van docenten,

maar dan op basis van de evaluatieve functie van toetsing. In hoeverre gebruiken docenten de uitkomsten van toetsen voor hun eigen professionalisering?

Assessment *as* Learning

In het onderwijsontwerp zoals ik dat hiervoor beschreven heb, is het de docent die het onderwijs ontwerpt, met als doel studenten actiever te betrekken bij hun leerproces, hun zelfregulatie en hun zelfsturing te bevorderen. Deze zelfsturing is noodzakelijk om de juiste keuzes te kunnen maken. Promovenda Debby Weerlink onderzoekt dit in de context van flexibele leeromgevingen en alternatieve leerroutes. Ondanks dat onderzoek aantoonde dat Assessment *for* learning al in het primair onderwijs bijdraagt aan de ontwikkeling van zelfregulatie en zelfsturing van leerlingen (Beekman, Joosten-ten Brinke, & Boshuizen, 2021), vinden we in het hoger onderwijs nog steeds dat veel studenten hier niet goed in zijn. Een mogelijk oorzaak daarvan is dat docenten bepalen *welke* leeractiviteiten plaatsvinden en *wanneer* deze leeractiviteiten plaatsvinden en dat docenten te weinig daadwerkelijk studenten de mogelijkheid bieden om een betekenisvolle rol te spelen in hun eigen leerproces (Hawe & Parr, 2014). In hoeverre heeft de student zelf een actieve rol in het ontwerp van het leertraject. Om dat perspectief te benadrukken wordt het begrip Assessment *as* Learning gebruikt.

Assessment *as* Learning (AaL) is een aspect van Assessment *for* Learning, maar benadrukt de actieve rol van de student in leeractiviteiten. De leeractiviteiten dagen studenten uit tot assessment-gerelateerde activiteiten zoals het identificeren en interpreteren van succescriteria, het vragen en gebruiken van feedback, het reflecteren op hun vooruitgang, het aanpassen of bijsturen van strategieën indien nodig, en het beoordelen van de kwaliteit van hun werk en leerresultaten. Door na te denken over wat en hoe ze hebben geleerd en wat dit betekent voor verder leren, krijgen studenten een beter begrip van wie ze zijn als lerenden en hoe ze hun eigen leren kunnen sturen om competent te worden en te blijven (Govaerts & Joosten – ten Brinke, submitted; Yan & Boud, 2021). Door expliciet aandacht te besteden aan Assessment *as* Learning, krijgen studenten inzicht in wat bedoeld wordt onder kwaliteit van werk, wat zeker in de huidige tijd, met producten die door generatieve AI ontwikkeld kunnen worden, cruciaal is. Elke professional moet binnen de eigen context weten wanneer producten, diensten of handelingen van goede kwaliteit zijn en moet de eigen prestaties, en die van collega's, kunnen controleren en beoordelen aan de hand van relevante criteria en normen. Alleen als studenten inzicht hebben in wat kwaliteit is, is het mogelijk dat zij zichzelf beoordelen en kunnen zij goede feedback geven. Het vermogen om te evalueren en beslissingen te nemen over de kwaliteit van het eigen werk (en dat van anderen) en om te reflecteren op wat dit betekent voor het eigen leertraject wordt *evaluative judgement* genoemd (Naidoo, Tai, & Penman, 2021; Tai et al., 2018). Deze vaardigheid en feedbackgeletterdheid in algemene zin zijn ook complexe vaardigheden die onze studenten moeten leren; aanvullend dus op allerlei complexe vaardigheden die ik in mijn inleiding al had genoemd.

Ruimte voor sturing door de student?



- Vragen om feedback
- Reflecteren op eigen voortgang
- Aanpassen van leerstrategieën
- Beoordelen van kwaliteit van eigen werk en werk van anderen
- Feedback geven

Evaluative judgement

Leeractiviteiten die dit stimuleren zijn (Carless & Boud, 2018; De Kleijn, 2023):

- a. Het analyseren en bespreken van voorbeelden uit de beroepspraktijk en met de studenten de kwaliteitscriteria te identificeren en te definiëren. Door voorbeelden uit de beroepspraktijk te bekijken, wordt duidelijk dat kwaliteit in de ene context net wat anders is dan kwaliteit in een andere context. Het inzicht van studenten in hoe de context kwaliteitscriteria en -normen kan beïnvloeden, ondersteunt de ontwikkeling van hun beoordelingsvermogen.
- b. Studenten vragen om de kwaliteit van bijvoorbeeld rubrics te evalueren en te bespreken in relatie tot beroepsproducten.
- c. Self-assessment en peer-assessment. Door het werk van zichzelf of van medestudenten te evalueren, zijn studenten bezig met het interpreteren en toepassen van criteria en met het aanreiken van oplossingen of suggesties voor problemen die hun medestudenten in hun werk tegenkomen. Het beoordelen van het werk van anderen kan ervoor zorgen dat studenten een beter begrip krijgen van de kenmerken van kwalitatief goed werk. Dit kan studenten helpen om hun eigen werk effectiever te beoordelen. Sleutelfactoren in effectieve peer assessment zijn authenticiteit van de beoordelingstaken, wederzijds vertrouwen en waardering voor peer-feedback.

Het zelf vragen om feedback door studenten en op basis van deze feedback plannen maken voor vervolgacties biedt studenten de kans om te laten zien dat ze begrijpen wat er moet worden bereikt en welke stappen er moeten worden genomen om de kloof tussen huidige en verwachte prestaties te dichten. Promovenda Agnes Brinks wil met haar onderzoek meer inzicht krijgen in feedbackdialogen en hoe die geoptimaliseerd kunnen worden.

Voor het functioneren in de huidige maatschappij is het dus van wezenlijk belang om studenten naast vakspecifieke kennis ook vaardigheden aan te leren die een leven lang leren mogelijk maken: zelfsturingsvaardigheden, zelfregulatie-vaardigheden, beoordelingsvaardigheden, feedback vaardigheden. Ondanks dat we deze vaardigheden belangrijk vinden, en ze nadrukkelijk aandacht krijgen in het formatieve toetsproces, zijn het vaak onderdelen van een competentie, zoals professioneel gedrag of communicatie, en zijn ze niet van doorslaggevend belang op het moment dat er een beslissing wordt genomen over de voortgang van studenten of het afronden van een opleiding.

Erkennen en waarderen van competenties

Het is ook wel veel wat we willen in het onderwijs. Als alle competenties geleerd moeten worden vanaf nul, en we activiteiten moeten opnemen om vast te stellen of er geleerd is, om het leren te stimuleren en om de student zelf aan het roer te zetten, ziet het eruit dat we hele volle onderwijsprogramma's krijgen. De vraag is dan ook of opleidingen op nul moeten beginnen of dat studenten kunnen voortbouwen op eerder verworven kennis, vaardigheden en competenties (Joosten – ten Brinke, 2008). Dit zou ook passen bij de meer actieve rol van studenten door actief na te gaan wat ze al geleerd hebben in relatie tot de eindtermen van een opleiding waar ze net aan beginnen om zodoende voort te bouwen op competenties die studenten elders of op een andere manier ontwikkeld hebben. Zeker bij flexibele opleidingen waar studenten hun eigen route door het curriculum kunnen bepalen is dit een cruciale stap. Het zelf sturen van je leren begint bij weten waar je staat.

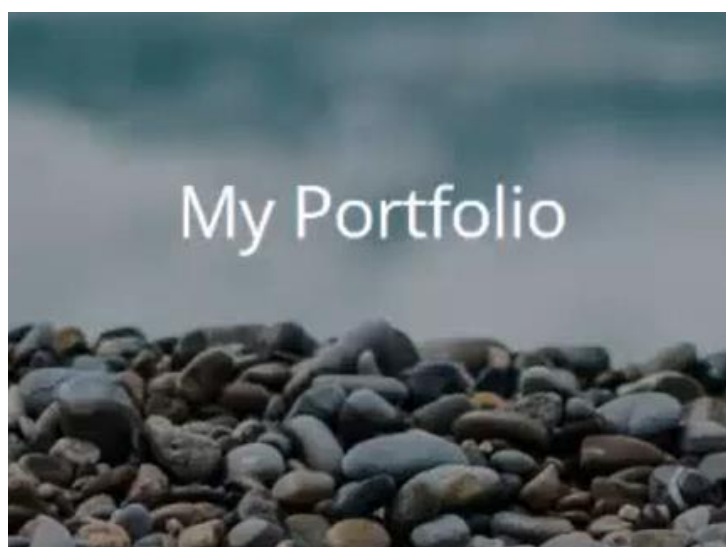
Het waarderen en erkennen van competenties, kennis of vaardigheden die opgedaan zijn tijdens het werk is niet eenvoudig. Vanuit een opleidingsperspectief wordt vaak gekeken naar eenheden van onderwijs die vrijgesteld kunnen worden of dat de leerdoelen van een opleiding een-op-een terugkomen in de bewijzen van een student uit het werkveld. Dat is vaak niet het geval. En ondanks dat de werkomgeving beschouwd kan worden als een van de krachtigste leeromgevingen, zeker voor complexe vaardigheden vanwege de authentieke taakervaring, toch is de werkomgeving als leeromgeving niet altijd optimaal (Poortman & Visser, 2009). Gebruik maken van wat er al op de werkplek geleerd is in het formele curriculum vraagt om een bredere kijk op het waarderen en erkennen van leren, waarbij ervaringsleren en academisch leren samengebracht worden, waarbij ervan uitgegaan wordt dat mogelijke verschillen tussen ervaringsleren en academisch leren kunnen worden overwonnen (Harris, 2006).



Drie promovendi richten zich op dit vraagstuk, vanuit drie verschillende perspectieven. Zo gaat Jannie Lensen-Botter onderzoeken hoe langstudeerders bij lerarenopleidingen geholpen kunnen worden om alsnog hun lerarenopleiding af te ronden door het valideren van hun praktijkervaring. Een groot aantal

studenten is soms al jaren werkzaam in het onderwijs en laten zien dat ze competent zijn. Deze uitspraak over hun competentie heeft echter geen waarde omdat de uitspraak niet gedaan is in de context van een formele beoordeling, maar informeel door collega's in de school (die gelijktijdig ook als schoolopleider werkzaam zijn en vanuit die rol studenten tijdens stages beoordelen). Deze studenten hebben vrijwel het hele curriculum gevolgd, maar door verschillende redenen enkele vakken niet kunnen afronden. Het gevolg hiervan is dat zij onbevoegd voor de klas staan, maar door het tekort aan leraren voltijds worden ingezet. Hierdoor ontbreekt tijd om de opleiding volgens de regels van de opleiding af te ronden. Jannie Lensen-Botter gaat in haar onderzoek na of een meer holistische beoordelingsbenadering gebruikt kan worden voor het beoordelen van de competenties van deze langstudeerders zonder afbreuk te doen aan de inhoud en kwaliteitseisen van een lerarenopleiding. Marjolein Mantelaers gaat in haar onderzoek in op elders verworven competenties van zij-instromers in het onderwijs. Dit zijn leraren die eerst een ander beroep hebben uitgeoefend en overstappen naar het onderwijs. In het Engels worden dit second-career teachers genoemd. Marjolein onderzoekt de relatie tussen de kennis, ervaringen en competenties die opgedaan zijn in het eerdere beroep met de self-efficacy voor pedagogisch-didactisch handelen in de klas. Als derde onderzoek in de context van het waarderen van eerder leren, wil Claudia Klarendijk gaan onderzoeken of het gebruik van een vaardighedenpaspoort (skills passport) in de bouwsector ervoor kan zorgen dat werknemers makkelijker kunnen switchen naar een andere functie op basis van hun vaardigheden, bijvoorbeeld als de huidige werkzaamheden te zwaar geworden zijn. Hierdoor zouden werknemers langer behouden kunnen worden voor de sector of in een andere sector een beter passend functie kunnen vinden.

Een instrument dat cruciaal is zowel bij deze valideringstrajecten (Bjornavold, 2001), bij werkplekleren (Van der Schaaf et al., 2017) als bij formele onderwijsprogramma's gericht op competentieontwikkeling (Driessen et al., 2006; Tigelaar et al., 2005), zoals bijvoorbeeld bij programmatisch toetsen (Torre et al., 2021), is het e-portfolio: een digitale verzameling van bewijzen die doelbewust worden samengevoegd als bewijs van leren. Feedback is daarbij een belangrijke bewijsbron. Het interpreteren van grote hoeveelheden feedback is echter zeer uitdagend en tijdrovend. Joyce Moonen en collega's onderzochten daarom of het mogelijk is om met geautomatiseerde tekstanalysetechnieken de interpretatie van de feedback te ondersteunen. Ze kwamen tot de conclusie dat de meest relevante onderwerpen in de feedback en de emotie van de feedback op deze manier zichtbaar gemaakt kon worden (Moonen-van Loon et al., 2022; Nair et al., 2024).



Bij het ontwerpen van onderwijs waarbij portfolio's worden ingezet moet nagedacht worden over de functie van het portfolio. Een ontwikkelportfolio geeft een inzicht in het leerproces (Kicken et al, 2009). De succescriteria en feedback van docenten, peers en werkgevers zijn zichtbaar voor de lerende. De lerende is bij voorkeur eigenaar van het portfolio en kan zelf aangeven wie welke informatie kan bekijken. Daardoor kan een ontwikkelportfolio ook ingezet worden als assessmentportfolio. Op het moment dat een lerende bepaalt dat er een beoordeling kan plaatsvinden, kunnen de bewijzen daarvoor getoond worden aan een beoordelaar. Afhankelijk van de situatie kan een beoordelaar dan inhoudelijk alle bewijzen doornemen en een beslissing nemen of kan een beoordelaar een beslissing nemen op basis van feedback van anderen die in hetzelfde systeem is opgeslagen. Dit laatste is bijvoorbeeld het geval bij programmatisch toetsen, waarbij belangrijke beslissingen worden genomen op basis van meerdere informatiebronnen (datapunten) (Baartman, Van Schilt-Mol, & Van der Vleuten, 2021). Het e-portfolio biedt verschillende mogelijkheden om deze informatie zo overzichtelijk en toegankelijk mogelijk te presenteren.

Technologische ondersteuning van leren en toetsen

Technologische ontwikkelingen, zoals deze, zorgen voor grote veranderingen in het onderwijslandschap en kunnen ingezet worden om de leerervaring van studenten te verbeteren. Promovendus Zhien Li onderzoekt bijvoorbeeld hoe studenten via virtuele patiënten hun vaardigheid van klinisch redeneren kunnen verbeteren.

2023 BA2 2.6 GEN2615 - Personalized medicine, Case A – Oropharyngeal carcinoma [Export PDF](#) | [Close preview](#)

Sore throat, Oncological diseases, other, Pain associated with cancer

General Introduction > 1 - Initial Assessment > 2 - Diagnostics > 3 - Diagnosis and Staging > 4 - Personalized Treatment Plan > Case Reflection and Learning Points

1.1 - Physical Examinations > 1.2 - Physical Examination > 1.3 - Display of Physical Examinations

1.2 - Physical Examination

[Save](#) [Previous](#) [Next](#)

Basic clinical characteristics

Patient name:	John
Age	56 Year
Gender	Male
Height	180 cm
Weight	95 kg
Job	Web Engineer

Page links

No link found.

Anamnesis

No restrictions in physical activities (WHO PS 0).

Complaints: The patient noticed a mass in the left side of his neck, since several weeks. Swallowing is painful with referred otalgia to his left ear. He lost 3 kg of weight in the last 3 months.

Smoking history: Never.
Alcohol: 1-2 units per day.

Medical History

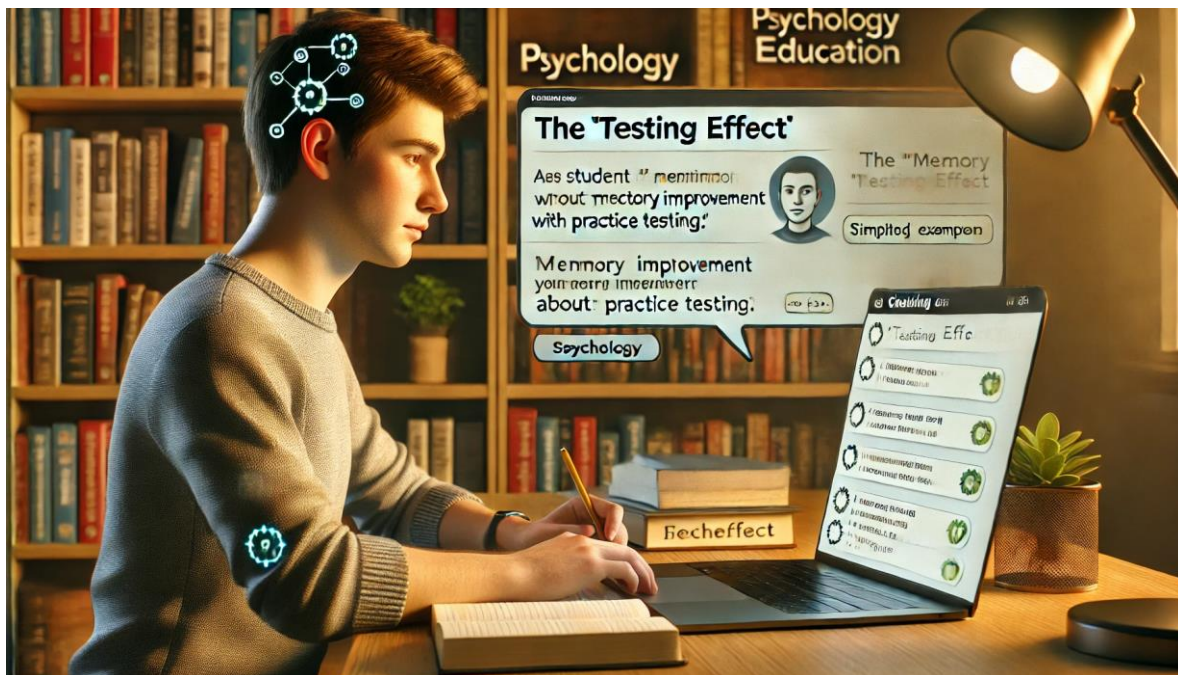
No relevant medical history

Inspection of the oral cavity



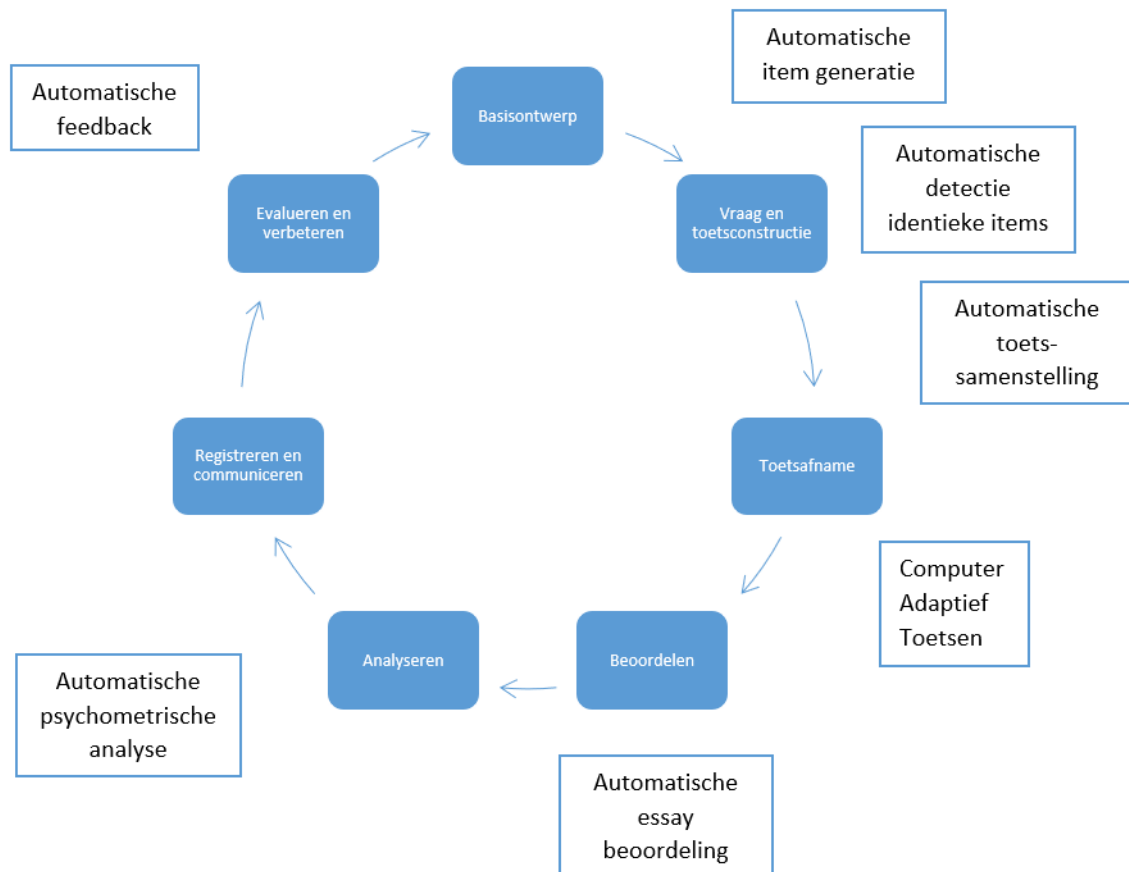
Figuur 9. Een schermafbeelding van een virtual patient.

En promovendus Tim Debets onderzoekt of een AI-gestuurde chatbot studenten kan helpen bij het selecteren en leren van geschikte leerstrategieën.



We kunnen niet meer om kunstmatige intelligentie (AI) heen. Generatieve AI zorgt ervoor dat we opnieuw naar ons onderwijs moeten kijken. Door de groeiende mogelijkheden van met name generatieve AI (GenAI) zijn er terechte zorgen over mogelijk misbruik. Je laten helpen door GenAI bij het schrijven van een werkstuk is verleidelijk en, bij goed gebruik, nauwelijks te onderscheiden van zelfgeschreven werk. De mogelijkheden van AI zorgen voor veranderingen in werkzaamheden op de arbeidsmarkt, met als gevolg dat ook de eindtermen van opleidingen om een herziening vragen, evenals leerdoelen en toetsing. De vraag die nu voorligt is: 'Welke aanpassingen zijn nodig om aan het einde van de opleiding een diploma af te kunnen geven waarbij we het vertrouwen hebben dat dit diploma aantoont dat de studenten beschikken over de kennis en vaardigheden die nodig zijn om als professional aan de slag te gaan?'.

Naast de zorgen over het inzetten van AI in het onderwijs vind ik echter dat we ook heel nadrukkelijk de mogelijkheden van AI voor leren en toetsen moeten onderzoeken. AI kan ons helpen bij het personaliseren van het onderwijs (dat zien we al bij het adaptief toetsen), het over een langere periode volgen van de voortgang van studenten en suggesties geven voor vervolgstappen voor de studenten of aanpassingen aan het curriculum. Als we kijken naar het hele toetsproces zien we meerdere mogelijkheden om onze processen met AI efficiënter te maken (zie Figuur 10).



Figuur 10. Mogelijkheden om het toetsproces te optimaliseren met AI-technieken.

Met GenAI is het bijvoorbeeld mogelijk om toetsvragen efficiënter te construeren. Het schrijven van goede toetsvragen kost veel tijd, variërend van 1 tot 8 uur per toetsvraag. En dit vraagt om toetsexpertise. Met GenAI kan een veelvoud aan toetsvragen geformuleerd worden binnen enkele minuten, nadat er wel enige tijd besteed is aan het schrijven van een goede prompt. Deze toetsvragen voldoen mogelijk toetstechnisch en inhoudelijk nog niet aan alle eisen, maar kunnen op basis van de inhoudelijke deskundigheid en toetsdeskundigheid van docenten worden bijgesteld. Ook het ontwerpen van onderwijs, het brainstormen over didactische werkvormen en het uitwerken van werkbladen voor studenten kan efficiënter met GenAI. Ook hier staat expertise over goede werkvormen, kennis van de doelgroep en bijvoorbeeld klassenmanagement buiten kijf. Het automatiseren van bepaalde ontwerp- en toetstaken met AI kan docenten ontlasten, zodat meer tijd besteed kan worden aan persoonlijke begeleiding en instructie.

Een aanpak om toch vertrouwen te hebben dat de student zelf beschikt over de kennis in het werkstuk en om de authenticiteit te checken, is bijvoorbeeld het toevoegen van een mondeling examen. Maar bij opleidingen met honderden studenten is dit niet haalbaar. Collega's bij de faculteit Science en Engineering hebben een eerste prototype voor een automatisch mondeling tentamen (Automatic Conversational Assessments) gebouwd op basis van taaltechnologie. Samen onderzoeken we of het mogelijk is om met deze methode de kennis van studenten in kaart te brengen op hogere cognitieve niveaus, zoals toepassing en analyse, waarbij het systeem zich tijdens het tentamen aanpast aan de antwoorden van studenten. Daarbij willen we kijken naar de betrouwbaarheid en validiteit van de methode, de gevoeligheid voor verschillen in taalvaardigheid van studenten en zeer belangrijk naar de ethische vraagstukken die horen bij het inzetten van AI bij toetsing.



Ethiek bij AI bij leren en toetsen

Die ethische vraag bij toetsing is niet nieuw – in alle stappen van het toetsproces maken we keuzes waarvan we ons moeten afvragen doen we het goede en doen we het goed? (Joosten-ten Brinke & Arts, 2023) – maar met de mogelijkheden van AI is die ethische vraag nog crucialer. Er moet bewust aandacht zijn voor privacy, informatiebescherming en mogelijke vooroordelen in AI-beoordelingssystemen zodat eerlijke en onbevooroordeelde beoordelingen gewaarborgd worden. Deze ethische vraag komt ook terug in het onderzoek van promovendus Marco Dumont, waarbij hij kijkt of het mogelijk is om op basis van data die we produceren op de werkplek inzicht te krijgen in iemands competenties om op die manier te beschikken over informatie die samen met observaties van collega's en leidinggevende leiden tot informatie-rijkere jaargesprekken. Een uiteindelijk oordeel moet genomen worden door menselijke beoordelaars.

Een onderzoek van enkele collega's van de OU, laat zien dat er zes belangrijke ethische waarden zijn van AI-gebruik in het onderwijs (Agarwal et al., 2024): Menselijk toezicht, Non-discriminatie, verklaarbaarheid, educatieve geschiktheid, aandacht voor well-being van docenten en studenten, en een goed gegevensbeheer.

Ethische waarden voor AI-gebruik in het onderwijs (Agarwal et al., 2024)

1. Menselijk toezicht: Betrokken (toegewijde) mensen moeten in staat zijn om beslissingen van AI te begrijpen, in twijfel te trekken en terzijde te schuiven.
1. Non-discriminatie: AI-systemen in het onderwijs moeten eerlijk en onbevooroordeeld zijn, openstaan voor iedereen, inclusief en respectvol voor diversiteit.
2. Verklaarbaarheid: Beslissingen van een AI-systeem moeten verklaarbaar, transparant en controleerbaar zijn.
3. Educatieve geschiktheid: AI-systemen in het onderwijs moeten worden ontworpen met een duidelijk educatief doel en voordeel voor ogen.
4. Aandacht voor well-being van studenten en docenten: AI-systemen in het onderwijs moeten mensgericht zijn, met een focus op het welzijn van studenten en docenten.
5. Gegevensbeheer: Gegevens over studenten en docenten moeten privacy, veilige opslag, nauwkeurigheid en transparantie garanderen. De mogelijkheid om gegevens te verwijderen of te corrigeren is ook vereist.

Docenten moeten een centrale rol blijven spelen in het interpreteren van AI-gegenereerde informatie en bij het nemen van beslissingen over de voortgang van studenten. Deze menselijke laatste stap is cruciaal. Op basis van de resultaten van toetsen worden immers beslissingen genomen over studenten die grote consequenties kunnen hebben. Voor het ontwerpen van onderwijs is het nodig dat de risico's van AI en de voordelen van AI met elkaar in balans gebracht worden.

Met de komst van technologie en veranderende onderwijs- en beoordelingsaanpakken verandert ook de rol van de docent. De ontwikkeling van docenten en trainingsprogramma's voor docenten zijn essentieel om docenten uit te rusten met de nodige vaardigheden om door dit veranderende landschap te navigeren.

Docenten moeten worden getraind in het gebruik van genAI als leermiddel en bij beoordelingen, zodat ze effectief en ethisch verantwoord kunnen begeleiden. Om te zorgen voor gelijke mogelijkheden voor alle studenten en docenten is het toegankelijkheid tot AI-gebaseerde hulpmiddelen en materialen nodig voor zowel studenten als docenten. Om meer zicht te krijgen op de impact van AI op het onderwijs en zicht te krijgen op hoe we met AI het onderwijs effectiever, efficiënter en leuker kunnen maken is het nodig dat docenten, studenten en onderzoekers experimenteren. Alleen door te experimenteren met de mogelijkheden, leren we waar de grenzen liggen. Dit vraagt om een onderwijsbeleid dat flexibel genoeg is om de voordelen van generatieve AI te benutten, terwijl het tegelijkertijd de nodige waarborgen biedt om de integriteit, eerlijkheid en ethiek in het onderwijs te handhaven.

Samen met AI moeten we komen tot succes!



Samenvatting

In deze oratie ben ik ingegaan op de verbinding tussen leren, toetsen en technologische vooruitgang die in mijn leerstoel centraal staat. Ik begon met sport als metafoor voor academische ontwikkeling en benadrukte daarbij dat toewijding, feedback en het doorzettingsvermogen nodig zijn voor succes, maar ook dat leren en toetsen samen de ontwikkeling in kaart brengen en zorgen voor verdere ontwikkeling. Deze verdere ontwikkeling is ook nodig voor opleidingen die nu nog teveel focussen op het vastleggen of er geleerd is. Onderzoek ten aanzien van Assessment for learning, Assessment as learning of een volledige integratie volgens programmatisch toetsen biedt hiervoor handreikingen. Tot slot ben ik ingegaan op de rol van technologie, met name kunstmatige intelligentie (AI), als ondersteuning voor zowel leren als toetsen. Ethische overwegingen, waaronder menselijk toezicht, zijn hierbij cruciaal. Het doel van ons onderzoek is om de voordelen van technologie op verantwoorde wijze te omarmen en zo een boeiende en effectieve leeromgeving te creëren.

Tot besluit

Ik ben hiermee gekomen aan het einde van mijn oratie, maar voordat ik ga afsluiten wil ik graag de mensen bedanken die het mogelijk maken dat ik hier sta en me kan blijven ontwikkelen.

Ten eerste dank ik het College van bestuur van Maastricht University, de decaan en de begeleidingscommissie van de Faculty of Health, Medicine and Life Sciences voor het vertrouwen dat zij in mij hebben gesteld door mij aan te stellen als hoogleraar Leren en Toetsen.

Dank aan de mensen die meestal als laatste genoemd worden in een dankwoord, maar voor mij zodanig belangrijk zijn, dat ze hier op de voorste rij mogen zitten en dus ook vooraan zitten in mijn dankwoord. Mijn man, Marcel, al 30 jaar mijn steun en toeverlaat. Bij elke stap in mijn carrière heb je me gesteund en geholpen bij het nemen van belangrijke beslissingen. Mijn kinderen en schoonkinderen Wesley, Rianne, Joëlle, Amanda, Somay en Len, jullie zijn zes totaal verschillende personen, maar alle zes toppers! Mijn moeder, die me geleerd heeft hoe belangrijk het is om goed voor mezelf te zorgen. Allen, dankjewel! Ook dank aan mijn broers en zussen. Ik vind het enorm fijn dat jullie er ook weer bij zijn.

En ook onze vriendengroep vanuit mijn studententijd aan de Witbreuksweg in Enschede: Ton en Hester, Erik en Franneke, Erik en Sylvia, Gehard, Frank en Alice en de 'mossel'vriendinnen Sandra, Ilona en Wil ik ben heel blij dat jullie er zijn! Jullie zorgen naast een Leven Lang Leren ook voor een Leven Lang Lol.

Dan een dankwoord voor alle collega's van FHML, de vakgroep Onderwijs en Onderzoek en SHE, School for Health Professions Education. Vanaf dag 1 heb ik me hier welkom gevoeld. De collega's van de taakgroep Toetsing en taakgroep Instructional Design: Sanne, Sanne, Jeroen, Saskia, Maryam, Joyce, Nynke, Ilse, Peter, het is fijn om met jullie samen te werken.

Dank aan mijn promovendi: Agnes, Tim, Marco, Claudia, Sander, Anne, Marjolein, Janneke, Debby, Jannie, Joost, Linda, Anne, Mohamed, Zhien,

En hun begeleiders: Gerry, Gino, Tanja, Gideon, Kazem, Roland, Xander, Jos, Ruud, Kim, Karin, Kelly, Tamara Schleepen, Cees, Tamara van Schilt-Mol, Bas, Herma, Piet, Diana, Maryam. Ik leer zo ontzettend veel van jullie.

Dank aan de vele fijne 'toets'-collega's in het land. Hier kan ik zeker niet iedereen opnoemen, dus ik beperk me tot Kelly, Tamara, Liesbeth, Judith, Kim, Harry, Annie, Jan, Jackelien, Henk, Anneke en Dorien. Het is altijd weer fijn om elkaar te spreken en om samen te publiceren over leren en toetsen.

Dank aan mijn oud-collega's van de Open Universiteit, Fontys lerarenopleiding en de masteropleiding Toetsdeskundige. Zo fijn dat velen van jullie ook vandaag hier zijn.

Ik wil ook graag alle presentatoren van vandaag bedanken voor hun lezing of workshop tijdens het symposium Leren en toetsen, voorafgaand aan deze oratie.

En tot slot, heel erg bedankt aan de collega's die me bij de voorbereiding van het symposium en deze oratie hebben geholpen: Lisa, Hennie, Monique, Willine, Nicky, en Sandy. Dankjewel!

Ik heb gezegd.

Dank aan allen op mijn leerpad



Referenties

- Agarwal, B., Urlings, C., Van Lankveld, G., & Klemke, R. (2024). Identifying the Ethical Values and Norms of Artificial Intelligence in Education: a Systematic Literature Review. <https://doi.org/10.35542/osf.io/e7t3f>
- Ama, J. U., & Emetarom, U. G. (2020). Equipping higher education students with the 21st century skills. *European Journal of Education Studies*. <https://doi.org/10.46827/ejes.v0i0.3136>
- AssessmentReformGroup. (1999). *Assessment for Learning: Beyond the black box*. University of Cambridge School of Education. https://www.nuffieldfoundation.org/sites/default/files/files/beyond_blackbox.pdf
- Baartman, L. K. J., Van Schilt-Mol, T., & Van der Vleuten, C. P. M. (2021). *Programmatisch toetsen in de praktijk. Voorbeelden en geleerde lessen*. Boom.
- Beekman, K., Joosten-ten Brinke, D., & Boshuizen, E. (2021). Sustainability of Developed Self-Regulation by Means of Formative Assessment Among Young Adolescents: A Longitudinal Study. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.746819>
- Bjornavold, J. (2001). Making learning visible: Identification, assessment and recognition of non-formal learning. *European Journal*, 22, 24-32. https://www.cedefop.europa.eu/files/etv/Upload/Information_resources/Bookshop/117/22_en_bjornavold.pdf
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). Assessment for learning: 10 principles. <https://dspace.stir.ac.uk/bitstream/1893/32458/1/Assessment%20for%20learning%2010%20principles%202002.pdf>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Day, I. N. Z., van Blankenstein, F. M., Westenberg, M., & Admiraal, W. (2018). A review of the characteristics of intermediate assessment and their relationship with student grades. *Assessment & Evaluation in Higher Education*, 43(6), 908-929. <https://doi.org/10.1080/02602938.2017.1417974>
- De Kleijn, R. A. M. (2023). Supporting student and teacher feedback literacy: an instructional model for student feedback processes. *Assessment & Evaluation in Higher Education*, 48(2), 186-200. <https://doi.org/10.1080/02602938.2021.1967283>
- Dirkx, K. J. H., Joosten-ten Brinke, D., & Camp, G. (2019). *Ontwerprichtlijnen voor formatief toetsen vanuit de geheugenpsychologie. 1 + 1 = 3? Eindrapport NRO-PPO overzichtsstudie dossiernummer 405-17-711*. https://www.nro.nl/sites/nro/files/migrate/Eindrapport-405-17-711_Definitief.pdf
- Dirkx, K. J. H., Joosten-ten Brinke, D., Schleepen, T. M. J., & Camp, G. (2024). *Designing Intermediate Tests: Guidelines Based on the Memorial Effects of Testing*.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84(5), 795-805. <https://doi.org/10.1037/0021-9010.84.5.795>
- Driessen, E. W., Overeem, K., van Tartwijk, J., van der Vleuten, C. P., & Muijtjens, A. M. (2006). Validity of portfolio assessment: which qualities determine ratings? *Med Educ*, 40(9), 862-866. <https://doi.org/10.1111/j.1365-2929.2006.02550.x>

- Govaerts, M., & Joosten – ten Brinke, D. (submitted). Assessment as Learning. In J. Dent, R. M. Harden, & D. Hunt (Eds.), *A Practical Guide for Medical and Health Professions Teachers* (7th ed.). Elsevier health sciences.
- Gulikers, J., & Baartman, L. (2017). *Doelgericht professionaliseren. Formatief toetsen met effect! Wat DOET de docent in de klas?: Eindrapport NRO-PPO overzichtsstudie dossiernummer 405-15-722*.
- Harris, J. (2006). Questions of knowledge and curriculum in the recognition of prior learning. In P. Andersson & J. Harris (Eds.), *Re-theorising the Recognition of Prior Learning* (pp. 51-76). National Institute of Adult Continuing Education.
- Hawe, E., & Parr, J. (2014). Assessment for Learning in the writing classroom: an incomplete realisation. *The Curriculum Journal*, 25(2), 210-237.
<https://doi.org/10.1080/09585176.2013.862172>
- Joosten-ten Brinke, D., & Arts, J. (2023). Ethiek bij toetsen. In H. v. Berkel, A. Bax, D. J.-t. Brinke, K. Beekman, & T. v. Schilt-Mol (Eds.), *Toetsen in het hoger onderwijs* (pp. 409-415). Boom Uitgevers.
- Joosten – ten Brinke, D. (2008). *Assessment of Prior Learning* [Proefschrift]. Open Universiteit.
https://research.ou.nl/ws/portalfiles/portal/934651/Thesis_Desiree_2008.pdf
- Langbroek, S., Duchatelet, D., Joosten-ten Brinke, D., & Camp, G. (2024). Enhancing teachers' instruction on how to study: an exploration of the effectiveness of learning strategies for particular secondary school subjects. *Frontiers in Education*, 9.
<https://doi.org/10.3389/feduc.2024.1340120>
- Moonen-van Loon, J. M. W., Govaerts, M., Donkers, J., & Van Rosmalen, P. (2022). Toward Automatic Interpretation of Narrative Feedback in Competency-Based Portfolios. *IEEE Transactions on Learning Technologies*, 15(2), 179-189. <https://doi.org/10.1109/TLT.2022.3159334>
- Naidoo, O., Tai, J., & Penman, M. (2021). Preparing students for the future through developing evaluative judgement. *The Clinical Teacher*, 18(2), 115-120.
<https://doi.org/10.1111/tct.13268>
- Nair, B. R., Moonen - van Loon, J. M. W., van Lierop, M., & Govaerts, M. (2024). Leveraging Narrative Feedback in Programmatic Assessment: The Potential of Automated Text Analysis to Support Coaching and Decision-Making in Programmatic Assessment. *Advances in Medical Education and Practice*, 15, 671-683. <https://doi.org/10.2147/AMEP.S465259>
- Poortman, C., & Visser, K. (2009). *Leren door werk: De match tussen deelnemer en werkplek*. Expertisecentrum Beroepsonderwijs. <https://ecbo.nl/wp-content/uploads/sites/3/2009-11-Leren-door-werk-de-match-tussen-deelnemer-en-werkplek.pdf>
- Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1(3), 181-210.
<https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Sadler, D. R. (2009). Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 1-19). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8905-3_4
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467-481. <https://doi.org/10.1007/s10734-017-0220-3>
- Tigelaar, D. E. H., Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., & van der Vleuten, C. P. M. (2005). Quality issues in judging portfolios: implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education*, 30(5), 595-610.
<https://doi.org/10.1080/03075070500249302>
- Torre, D., Rice, N. E., Ryan, A., Bok, H., Dawson, L. J., Bierer, B., Wilkinson, T. J., Tait, G. R., Laughlin, T., Veerapen, K., Heeneman, S., Freeman, A., & van der Vleuten, C. (2021). Ottawa 2020 consensus statements for programmatic assessment – 2. Implementation and practice. *Medical Teacher*, 43(10), 1149-1160. <https://doi.org/10.1080/0142159X.2021.1956681>

- Van der Schaaf, M., Donkers, J., Slof, B., Moonen-van Loon, J., van Tartwijk, J., Driessen, E., Badii, A., Serban, O., & Ten Cate, O. (2017). Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. *Educational Technology Research and Development*, 65(2), 359-380. <https://doi.org/10.1007/s11423-016-9496-8>
- Van der Steen, J., Van Schilt-Mol, T., Van der Vleuten, C., & Joosten- ten Brinke, D. (2022). Supporting Teachers in Improving Formative Decision-Making: Design Principles for Formative Assessment Plans. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.925352>
- Van der Steen, J., Van Schilt-Mol, T., Van der Vleuten, C., & Joosten-ten Brinke, D. (2023). Designing Formative Assessment That Improves Teaching and Learning: What Can Be Learned from the Design Stories of Experienced Teachers? *Journal of Formative Design in Learning*, 7(2), 182-194. <https://doi.org/10.1007/s41686-023-00080-w>
- Van der Vleuten, C., Sluijsmans, D., & Joosten-ten Brinke, D. (2017). Competence Assessment as Learner Support in Education. In M. Mulder (Ed.), *Competence-based Vocational and Professional Education: Bridging the Worlds of Work and Education* (pp. 607-630). Springer International Publishing. https://doi.org/10.1007/978-3-319-41713-4_28
- Van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Med Educ*, 39(3), 309-317. <https://doi.org/10.1111/j.1365-2929.2005.02094.x>
- Van Merriënboer, J. G., Kirschner, P. A., & Frèrejean, J. (2025). *Ten Steps to Complex Learning. A Systematic Approach to Four-Component Instructional Design* (4th ed.). Routledge.
- Van Schilt-Mol, T., & Baartman, L. (2024). *Programmatisch toetsen: een verdieping*. Boom.
- Veldhuijzen, N. H., Goldebeld, P., & Sanders, P. J. (1993). Klassieke testtheorie en generaliseerbaarheidstheorie. In T. Eggen & P. F. Sanders (Eds.), *Psychometrie in de praktijk* (pp. 33-82). Cito. <https://cito.nl/media/ywmcz3kk/klassieke-testtheorie-en-generaliseerbaarheidstheorie-psychometrie-in-de-praktijk-boek.pdf>
- Veugen, M. J., Gulikers, J. T. M., & den Brok, P. (2022). Secondary school teachers' use of online formative assessment during COVID-19 lockdown: Experiences and lessons learned. *Journal of Computer Assisted Learning*, 38(5), 1465-1481. <https://doi.org/10.1111/jcal.12699>
- Yan, Z., & Boud, D. (2021). conceptualising assessment-as-learning In Z. Yan & L. Yan (Eds.), *Assessment as Learning. Maximising Opportunities for Student Learning and Achievement*. Routledge. <https://doi.org/10.4324/9781003052081>

Verantwoording afbeeldingen

Voorblad afbeelding en afbeeldingen pagina 16 en 18: gemaakt door Desirée Joosten-ten Brinke met generatieve AI (DALI-E)

Afbeeldingen pagina 2, 6, 13, 19 verkregen via Stock Images voor Microsoft 365 subscriber.

Figuur 2: Competentiedomeinen van artsen – raamplan artsopleiding 2020 en Bekwaamheidseisen leraren – Rijksoverheid 2017

Figuur 3: Van Schilt-Mol and Baartman (2024, p. 25)

Figuur 4: Van Schilt-Mol and Baartman (2024, p. 26)

Afbeelding pagina 21: Stock Images, persoonlijke aanschaf