

SEQ MAGAZINE

Risbo
Research-Training-Consultancy
Erasmus



**SENIOR EXAMINATION QUALIFICATION
COHORT 2025/2026**

May 2026

COLOPHONE

This magazine is compiled by

Risbo

Research - Training - Consultancy

Information about the SEQ/SKE

Name of the Qualification	Senior Examination Qualification
Number of hours	70 hours
Location	Rotterdam
Mode of Instruction	In person (offline)
Language of Delivery	English and Dutch
Department	Risbo, Erasmus University Rotterdam

Copyright

Cover photos

© Erasmus University Rotterdam -
Alexander Santos Lima, via Lytho

Magazine photos

© Erasmus University Rotterdam -
Alexander Santos Lima, via Lytho
Matt Hrusc, for Risbo

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright holder(s)

PREFACE

Dear reader,

In front of you is the third edition of the Risbo SEQ Magazine. In this magazine, you will be introduced to the participants of cohort 2025-2026 of our Senior Examiner Qualification and their projects. The SEQ focuses on strengthening assessment literacy and aims to contribute to quality of assessment in higher education, at faculty or program level. All SEQ participants work on a quality enhancement project related to quality of assessment in their own organization or program. This could, for instance, be a project related to a curriculum or policy change, an issue of quality assurance, or innovating the way in which we assess the intended learning outcomes of a program. In this magazine, you will get an overview of the different projects that the participants undertook in the academic year 2025-2026, what they resulted in, and what the participants learned from them.

This SEQ was offered by Risbo, in collaboration with the Community for Learning and Innovation for 12 participants from Erasmus University Rotterdam. In addition, we also had one participant from The New School for Information Services and three participants from the Police Academy, adding to the diversity of contexts and practices that we could learn from. During the SEQ, the exchange of views, shared dilemmas, and different approaches was found to be very valuable by the participants.

We are proud of what they have achieved and happy to share their experiences and lessons learned with a wider audience through this magazine. A diversity of topics is addressed in the articles, from exploring alternative graduation projects, to tackling the challenges posed by generative AI and from assessment literacy for authentic assessment, to developing and implementing new visions on assessment. We hope you will find the articles insightful and inspiring!

Marit Nieuwenhuys, Hanneke Bahlman, William Pitz & Celine van der Lienden



Marit Nieuwenhuys
nieuwenuhuys@risbo.eur.nl



Hanneke Bahlman - van Pel
bahlman@risbo.eur.nl



William Pitz
pitz@risbo.eur.nl



Celine van der Lienden





TABLE OF CONTENTS

- | | | | |
|-----------|---|-----------|--|
| 04 | ABOUT THE PARTICIPANTS
Get to know our participants of cohort 2025/2026 | 31 | KIM OUWEHAND
Criteria for Rubrics in Higher Education in the Age of AI |
| 06 | KJELL NOORDZIJ
“Geen eind-ding, maar een altijd-ding” | 36 | PLEUNI DE KOCK
Werken aan een toetsprogramma voor de toekomst |
| 10 | MARYSE KRUIHOF
Impactful Capstones: reimagining the graduation project for societal impact | 40 | NINA VAN SANTVOORT
Van papieren visie naar gedeelde praktijk |
| 14 | JULIA BAUR
Doorontwikkeling toetsprogramma in de master Geneeskunde EA2030: programmatische en competentiegerichte benadering | 44 | IRIS VAN DE GRAAF
Programmatisch toetsen van kennis in het bachelor Geneeskundecurriculum EA2030 |
| 21 | MANDY MEEUWSEN
Van praktijkexpert naar beoordelaar. Hoe versterken we de rol van examinatoren in beroepsauthentieke toetsing? | 51 | ERIK DE KLOE
Naar een toekomstbestendig scriptietraject in het AI-tijdperk |
| 26 | SANDRA KHOR MANICKAM
When, and when not, to use AI? Uneven values and expectations among History examiners | 56 | ANTHONY BROWN
Rethinking Assessment: Building Critical Thinking in 21st Century Management Education |
| | | 61 | FRANK PYNENBURG
Naar duurzame toetskwaliteit: samenwerken aan samenhang |

ABOUT THE SEQ

The Senior Examiner Qualification (SEQ) course is offered by Risbo in collaboration with the Community for Learning and Innovation (CLI) at Erasmus University to stimulate the professionalisation of lecturers. It is intended for teachers that are (prospective) members of examination boards and/or program coordinators/directors at a school or institute, or others with a role regarding assessment at curriculum level. The aims of this course are to improve knowledge and skills on quality of assessment in higher education, at the level of schools and programs.

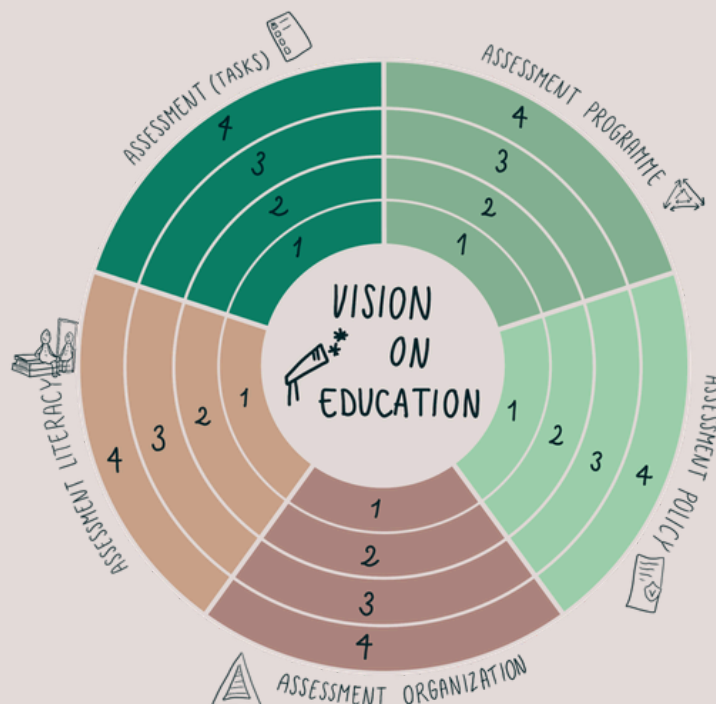
In the SEQ each participant is expected to bring in an ambition or challenge regarding the quality of assessment within their school, supported by their examination board and management team - to ensure knowledge dissemination within the school. The articles in this magazine showcase the outcomes of the projects participants conduct to tackle these challenges.

During the trajectory the participants also develop their personal vision on the role of assessment in higher education and reflect on the implications that this vision will have for the assessment (program and/or policy) within their own program and school. Various theoretical assessment frameworks are examined for the opportunities and possibilities, risks and policy choices to be made.

From this theoretical basis and the vision of assessment, the quality of assessment in their own organization is examined and participants develop recommendations for improvement.

Topics covered in the SEQ include:

- Visions and perspectives on assessment
- The quality of components and processes in the assessment organization;
- The continuous development of assessment literacy;
- The implementation process and evaluation of assessment policy
- The quality of assessment programs



ABOUT THE PARTICIPANTS



ESSB

Dr. Kjell Noordzij

Assistant Professor of Sociology
Erasmus University Rotterdam



EUC

Maryse Kruithof, Phd

Senior Lecturer Humanities, Phd.
Erasmus University Rotterdam



EMC

Julia Baur, Msc

Adviseur Onderwijs & toetsing
Erasmus University Rotterdam



POLITIE
ACADEMIE

Msc. Mandy Meeuwsen, Msc

Docent Gedragwetenschappen



ESHCC

Dr. Sandra Khor Manickam

Assistant Professor
Erasmus University Rotterdam



ESSB

Kim Ouwehand, Phd

Assistant professor
Erasmus University Rotterdam

ABOUT THE PARTICIPANTS



ESE

Pleuni de Kock, Msc
Beleidsmedewerker Onderwijs
& Toetsing
Erasmus University Rotterdam



POLITIE
ACADEMIE

Nina van Santvoort
Docent Gedragwetenschappen
Politieacademie



EMC

Iris van de Graaf- Weerd, Msc
Adviseur Onderwijs & toetsing
Erasmus University Rotterdam



ESL

Mr. Erik de Kloe
Universitair docent
Erasmus University Rotterdam



RSM

Dr. Anthony Brown
Senior Lecturer. Strategy &
Entrepreneurship
Erasmus University Rotterdam



TNS

Drs. Frank Pynenburg
Docent en lid examencommissie
The New School for Information
Services, Amsterdam

“GEEN EIND-DING, MAAR EEN ALTIJD-DING”

Perspectieven op de doorontwikkeling van het afstudeerproject van de masteropleiding Sociologie

KJELL NOORDZIJ

De master Sociologie aan de Erasmus Universiteit Rotterdam is de afgelopen jaren sterk vernieuwd. In die herziening is scherper geworden waar de opleiding, na de introductie van diverse nieuwe masterspecialisaties en nieuwe onderwijsvisies, voor staat. De ‘Rotterdamse sociologie’ is nauw verbonden met actuele maatschappelijke vraagstukken, met verschillende beroepspraktijken en met verschillende manieren om impact te maken. Toch bleef één belangrijk onderdeel nog buiten beeld: het afstudeerproject. Het afstudeerproject moet de breedte en ambities van de eindtermen van de herziene opleiding weerspiegelen, maar de vraag is of het huidige afstudeerproject daarin slaagt.

Door **Kjell Noordzij**¹

Het huidige afstudeerproject

Het huidige afstudeerproject van de masteropleiding Sociologie bestaat uit twee onderdelen: het traject en het product. Het traject draait om begeleiding in scriptiekringen, waarin studenten op doelgerichte wijze toewerken naar eerst een onderzoeksvoorstel en daarna de definitieve scriptie. Die opzet biedt structuur en zorgt voor vaste momenten van feedback.

Het product is een wetenschappelijk, empirisch onderzoeksartikel. Studenten formuleren een sociologisch relevante vraag, bouwen een theoretisch kader op, doen empirisch onderzoek en presenteren hun resultaten in de vorm van een wetenschappelijk artikel. Gezien de omvang en aard van het afstudeerproject weerspiegelen de leerdoelen van het project alle dertien eindtermen van de masteropleiding. Het project speelt daarmee een belangrijke rol in het afdekken van deze eindtermen, met name die gericht op het toepassen van kennis en inzicht en op oordeelsvorming.

Een professioneel-analytisch project

Zowel het afstudeertraject als het afstudeerproduct zijn erg klassiek. Dat betekent dat alle studenten hetzelfde product schrijven, namelijk een scriptie in de vorm van een onderzoeksartikel, en begeleid worden op een wijze die aansluit bij het doen van empirisch onderzoek. Die opzet is helder voor studenten en bekend bij begeleiders. Bovendien is de beoordeling zorgvuldig georganiseerd, met duidelijke handleidingen voor begeleiders, en volledig gedigitaliseerd. Begeleiders werken al enige jaren met deze kaders, waardoor de beoordeling betrouwbaar is.

Tegelijkertijd kent de vernieuwde masteropleiding verschillende specialisaties, met elk hun eigen sociologische benadering.² Die benadering geeft uitdrukking aan hoe deze specialisaties denken over de rol van de sociologie in de kennisproductie, en waar die kennis in de praktijk voor wordt benut. Zo zijn er specialisaties die zich sterk richten op het doen van sociologisch onderzoek om maatschappelijke vraagstukken op wetenschappelijke wijze te doorgronden, terwijl voor andere specialisaties onderzoek ten dienste staat van beleidsvorming. Andere willen met hun onderzoek een breder publiek in beweging brengen, terwijl weer andere specialisaties dat publiek in de eerste plaats een kritische spiegel willen voorhouden.

Het huidige afstudeerproject is in de kern een reflectie van slechts één sociologische benadering: die van de socioloog als professioneel-analytische onderzoeker. Hoewel die benadering nauw aansluit bij sommige specialisaties (zoals Polarisatie, Democratie en Samenleving), weerspiegelt het onderzoeksartikel als eindproduct, en de sociologische benadering die daarin besloten ligt, in mindere mate die van andere specialisaties.

“HET HUIDIGE AFSTUDEERPROJECT IS IN DE KERN EEN REFLECTIE VAN SLECHTS ÉÉN SOCIOLOGISCHE BENADERING.”

Hierdoor worden eindtermen van de opleiding rondom oordeelsvorming (zoals het kunnen positioneren in een krachtenveld van actoren), communicatie (richting diverse doelgroepen) en leervaardigheden (zoals het toepassen van inzichten en vaardigheden in nieuwe situaties of het bijhouden van ontwikkelingen die relevant zijn voor het beroepenveld) voornamelijk afgedekt binnen een professioneel-analytische, wetenschappelijke context.

Deze reflecties bieden aanleiding om kritisch na te denken over het huidige afstudeerproject van de opleiding. Wat gaat goed in het huidige project, en welke kansen liggen er voor de toekomst?

Bottom-up

Om die vraag te beantwoorden, is gesproken met mensen die op verschillende manieren betrokken zijn bij het afstudeerproject. Denk aan docenten, coördinatoren, studenten, alumni en mensen uit aanverwante opleidingen. In die gesprekken stonden twee vragen centraal: wat zijn de sterke en zwakke punten van het huidige afstudeerproject, en welke wensen leven er voor de toekomst?

Daarnaast is gekeken naar bredere onderwijsvisies binnen universiteit en faculteit, en naar de ambities en uitgangspunten van de recente masterherziening. Dit biedt een reflectie op het huidige afstudeerproject en op de vraag of het nog voldoende aansluit bij deze visies. Ook zijn voorbeelden uit andere opleidingen meegenomen, om te voorkomen dat het wiel opnieuw wordt uitgevonden.

Deze aanpak past bij een ontwerpgerichte, bottom-up manier van werken: niet meteen een oplossing opleggen, maar eerst goed begrijpen of er een probleem is, waar dat precies zit en hoe dat in de toekomst aangepakt kan worden.

Een dubbel beeld

Uit de gesprekken komt een dubbel beeld naar voren. Vrijwel niemand keert zich onomwonden tegen het huidige afstudeerproject.

Integendeel: het is duidelijk voor student en docent, zorgvuldig vormgegeven, voorzien van duidelijke beoordelingskaders en leerzaam voor studenten in relatie tot de eindtermen. Maar tegelijk leeft breed het gevoel dat het project smaller is dan de opleiding wil.

“HET GEVOEL LEEFT DAT HET PROJECT SMALLER IS DAN DE OPLEIDING WIL.”

Een belangrijk, terugkerend onderwerp van gesprek is daarbij de diversiteit in sociologische benaderingen, waar de gesproken stakeholders zich sterk van bewust zijn. Binnen de opleiding is de afgelopen jaren sterker zichtbaar geworden dat sociologie verschillende gezichten heeft. Veel gesprekspartners vinden dat het afstudeerproject daar meer ruimte voor zou mogen bieden, primair door het type afstudeerproduct te diversifiëren.

Dit betekent bovendien dat het huidige afstudeerproject nog onvoldoende aansluit bij de ‘authentieke’ wijze van toetsing die de faculteit en universiteit nastreven. Met andere woorden: het afstudeerproduct sluit niet voldoende aan bij de diversiteit aan professionele situaties waarin studenten na hun studie te recht komen. Dit terwijl de masteropleiding zich juist voorstaat op een nauwe verbinding met die praktijk.

Een tweede punt is de ervaring van studenten. De scriptie wordt vaak beleefd als intensief en stressvol. De huidige planning biedt structuur, maar wordt ook geregeld als krap ervaren, of als een “snelkookpan”. Daardoor blijft er soms weinig ruimte over voor verdieping of persoonlijke ontwikkeling. Bovendien wordt verwacht dat de studiemotivatie toeneemt wanneer studenten het gevoel hebben dat het afstudeerproject aansluit bij hun interesses en toekomstplannen.⁴ Ook de begeleiding is een belangrijk onderwerp. Docenten hechten sterk aan de persoonlijke begeleiding in het afstudeertraject, maar ervaren een hoge werkdruk. Studenten profiteren van deze kleinschaligheid, maar kunnen het traject ook als geïsoleerd ervaren: veel hangt af van de eigen begeleider.

“KAN EEN PRAKTIJKGERICHT PRODUCT ACADE MISCH ZIJN, ZOLANG HET STEVIG THEORETISCH EN EMPIRISCH IS ONDERBOUWD?”

Verder speelt de cruciale vraag hoe de opleiding haar wetenschappelijke identiteit ziet, en hoe zij die vormgeeft in het afstudeerproject. Moet dat altijd uit monden in een onderzoeksartikel op basis van empirisch onderzoek? Of kan ook een praktijkgericht product academisch zijn, zolang het stevig theoretisch en empirisch is onderbouwd? Juist daarover blijken verschillende opvattingen te bestaan. In dit kader wordt gesproken over ‘academische professionaliteit’: het vermogen om betrouwbare informatie te ontsluiten en verantwoord te gebruiken, conceptueel te denken, te werken vanuit een probleemanalyse die in diverse perspectieven is gegrond, en tot onderbouwde verklaringen en oplossingen te komen.

Tot slot komt de invloed van generatieve AI (GenAI) nadrukkelijk naar voren. Als studenten met GenAI teksten kunnen produceren, wordt het belangrijker om niet alleen een geschreven eindproduct van studenten te verlangen, maar ook aandacht te hebben voor bijvoorbeeld een presentatie van de bevindingen waarin studenten hun werk toelichten.

Daarnaast vraagt de ontwikkeling rondom GenAI, maar ook de nadruk op de zelfstandigheid van studenten op masterniveau, om meer aandacht voor de persoonlijke groei van studenten dan nu het geval is. Deze ontwikkeling maakt procesbegeleiding relevanter dan ooit en vraagt om innovatieve manieren van toetsing die reflexiviteit en zelfstandigheid bij studenten aanmoedigen.⁵

Een nieuw ambitieniveau

De perspectieven van de gesproken stakeholders hangen samen met diverse onderwijsvisies, waaruit een eenduidig beeld naar voren komt: de opleiding is de afgelopen jaren breder en diverser geworden, terwijl het afstudeerproject nog sterk leunt op één klassieke vorm. Daarmee ligt het ambitieniveau van de opleiding, faculteit en universiteit hoger dan wat het huidige afstudeerproject biedt.

“HET AMBITIENIVEAU VAN DE OPLEIDING, FACULTEIT EN UNIVERSITEIT LIGT HOGER DAN WAT HET HUIDIGE AFSTUDEERPROJECT BIJDT.”



Studenten van de masterspecialisatie Polarisatie, Democratie en Samenleving op werkbezoek bij het Ministerie van Binnenlandse Zaken (2025), een voorbeeld van de aansluiting van het onderwijs op de praktijk.

Dat betekent niet dat het huidige afstudeerproject volledig moet verdwijnen. Veel betrokkenen vinden juist dat de sterke kanten behouden moeten blijven: de heldere structuur, de validiteit en betrouwbaarheid van de beoordeling en de aandacht voor onderzoeksvaardigheden. Maar er is ook een duidelijke wens om te onderzoeken of het afstudeerproject beter kan aansluiten bij de verschillende specialisaties, de ambitie om maatschappelijke impact te maken, de verschillende beroepspraktijken waarmee de opleiding zich identificeert, en de veranderende realiteit door GenAI.

De noodzaak te beslissen

De probleemanalyse van het afstudeerproject wijst op een aantal cruciale beslispunten. Elk beslispunt werpt een kritisch licht op het huidige afstudeertraject en helpt te bepalen welke koers de doorontwikkeling ervan op gaat. Alvorens de opleiding het afstudeerproject kan doorontwikkelen, zou het goed zijn als zij hierover een formele positie inneemt.

De eerste vraag is hoe het afstudeerproject de diversiteit van de opleiding erkent. Op welke wijze krijgen de verschillende sociologische benaderingen een plaats in de opleiding, en moeten die benaderingen in het afstudeerproject zichtbaar zijn? Ook wordt aangevoerd dat de uniformiteit van het afstudeerproject bijdraagt aan duidelijkheid en een betrouwbare beoordeling. Als de diversiteit toeneemt, dan moeten de validiteit en betrouwbaarheid van de beoordeling onaangetast blijven. Daarnaast roept verdere diversificatie de vraag op of stafleden voldoende toegerust zijn om verschillende typen afstudeerproducten te begeleiden en te beoordelen.

“WAT VERSTAAT DE OPLEIDING PRECIES ONDER HET WETENSCHAPPELIJKE KARAKTER VAN HET AFSTUDEERPROJECT?”

De tweede vraag is wat de opleiding precies verstaat onder het wetenschappelijke karakter van het afstudeerproject. De opleiding leidt op tot ‘academische professionals’, maar wie zijn zij? En wat kunnen zij? Moeten zij in staat zijn om een empirische onderzoekscyclus te doorlopen? Of kunnen zij hun academische vaardigheden ook op andere manieren benutten voor vraagstukken uit de praktijk?

De derde vraag gaat over de duur en opzet van het traject. De huidige structuur biedt duidelijkheid, maar kan ook als te strak worden ervaren. Meer tijd kan ruimte geven voor verdieping, maar vraagt ook om organisatorische keuzes. De beslissing over de ideale duur van het afstudeertraject is dus geen triviaal besluit, maar ligt ten grondslag aan de mogelijkheden binnen het afstudeertraject en heeft directe implicaties voor studenten en hun leerproces.

De vierde vraag is of de huidige nadruk op individuele begeleiding op de lange termijn haalbaar blijft. Persoonlijke begeleiding is waardevol, maar ook arbeidsintensief. Zeker als het project diverser wordt, kan dat nieuwe eisen stellen aan docenten en aan de organisatie.

Denkrichtingen

Posities aan de hand van deze vier beslispunten leiden niet tot één eenduidige weg vooruit. Er is dus geen sprake van een enkele oplossing, maar wel van een aantal duidelijke denkrichtingen.

Een eerste richting is om het huidige model grotendeels te behouden, maar het traject te versterken. Denk aan meer ruimte voor oriëntatie, betere ondersteuning van op de benaderingen toegespitste onderzoeksvaardigheden, en meer aandacht voor reflectie op het leerproces achter het eindproduct.

Een tweede richting is om meer ruimte te geven aan verschillende afstudeerproducten. Studenten zouden dan bijvoorbeeld kunnen kiezen voor een onderzoeksartikel, maar mogelijk ook voor een beleidsgericht of maatschappelijk product, mits dat stevig wetenschappelijk onderbouwd is.

In deze richting is er ook meer aandacht voor de ontwikkeling van de academische professionaliteit van studenten. Dan gaat het niet alleen om onderzoek doen, maar ook om informatie kritisch gebruiken, verschillende perspectieven verbinden, maatschappelijke vraagstukken analyseren en onderbouwde keuzes maken voor de praktijk.

Welke richting het ook wordt, de belangrijkste opbrengst van deze verkenning is misschien wel dat het afstudeerproject niet langer alleen wordt gezien als het sluitstuk van een opleiding, maar ook als een uithangbord van de opleiding en een cruciaal onderdeel van de academisch-professionele ontwikkeling van studenten.

“WAT VOOR SOCIOLOGEN WIL DE OPLEIDING VORMEN?”

De discussie over het afstudeerproject gaat uiteindelijk over de vraag: wat voor sociologen wil de opleiding vormen? Als de master Sociologie studenten wil voorbereiden op uiteenlopende maatschappelijke en professionele rollen, dan is het logisch om ook het afstudeerproject opnieuw tegen het licht te houden. ■

End notes

1. Dit project vormt het sluitstuk van de SUTQ en SEQ van dr. Kjell Noordzij (ESSB, EUR)
2. Deze benaderingen zijn gebaseerd op: Burawoy, M. (2005). For Public Sociology. *American Sociological Review*, 70(1), 4-28.
3. Lees hier meer over authentiek toetsen: Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67–86.
4. Lees hier meer over studiemotivatie in relatie tot authentiek toetsen: Jones, B. D. (2009). Motivating students to engage in learning: The MUSIC model of academic motivation. *International Journal of Teaching and Learning in Higher Education*, 21(2), 272–285; Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation*, 70, 101030.
5. Zie in dit kader, bijvoorbeeld, integrative assessment: Crisp, G. T. (2012). Integrative assessment: reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37(1), 33–43; of evaluative judgement: 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; of diverse visies op het leerproces: Schellekens, L. H., Bok, H. G., De Jong, L. H., Van der Schaaf, M. F., Kremer, W. D., & Van der Vleuten, C. P. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, 101094.

IMPACTFUL CAPSTONES: REIMAGINING THE GRADUATION PROJECT FOR SOCIETAL IMPACT

MARYSE KRUIHOF

Erasmus
University
College

Introduction

At Erasmus University College (EUC), our mission is to educate students to become impactful global citizens who can tackle complex societal challenges. This ambition runs through everything we do, from our teaching philosophy to our Intended Learning Outcomes (ILO's), which emphasize critical thinking, the ability to apply knowledge to real-world contexts and social responsibility. Yet when we look at our current graduation project, there's a gap between what we tell our students in Year 1 and what we ask of them in Year 3. We say we encourage students to engage with local communities and institutions and to pursue meaningful societal impact. Three years later, we ask them to demonstrate everything they have learned with a traditional academic thesis (the Capstone). Feedback from alumni and the Student Academic Advisory Council (SAAC) had been pointing in the same direction for some time as some students feel the Capstone simply does not reflect what they actually learned during the programme.

This creates a gap in EUC's educational ambitions. The standard thesis is excellent for evaluating academic research and writing skills, but the format can get in the way of other competencies we claim to care about: leadership, creativity, collaboration, and social engagement. And perhaps most tellingly, most of these theses are read by no more than two assessors before being archived. For a programme that holds societal impact as a central value, that's a tension worth acknowledging. The problem is not that our Academic Rules and Regulations (ARR) requires a thesis. It actually allows for a wide range of final project formats. The problem is that no structured assessment protocols or supervision models existed to support anything else. Without those, both students and supervisors default to what they know. This SEQ project, Impactful Capstones (ICaps), aims to close that gap by diversifying the ways students can demonstrate their growth throughout our BA program, and by giving them room to let their own interests, experiences and skills shape their graduation project.

This SEQ project, Impactful Capstones (ICaps), aims to close that gap by diversifying the ways students can demonstrate their growth throughout our BA program, and by giving them room to let their own interests, experiences and skills shape their graduation project.

Designing the ICaps track

With the ICaps track, students will be able to graduate through formats that are both academically rigorous and genuinely impactful. Whether that's a documentary, a policy paper, a podcast, or a community workshop, the format should serve the student's goals and reach a real audience rather than filling up a supervisor's drawer. The final product is not just evidence of learning, it is an intervention in its own right, designed to address a real problem and reach a real audience. Allowing students to choose their own format means they can tailor their experience to their own interests and skills, which tends to produce more motivated, more invested work. It also reflects something we know to be true but don't always act on: learning isn't a uniform process, and students show their knowledge and skills in very different ways.

Studying graduation projects in other programmes shows this is not uncharted territory. University College Roosevelt, for example, already graduates students on a portfolio rather than a thesis. The Applied History master at ESHCC gives students the option to submit a documentary or podcast alongside a shorter research paper. Students in the EsPhil program Societal Transitions tackle real-world social issues by designing a strategy and leading a project together with the people involved. These institutions have built rubrics that hold genuinely different products to the same academic criteria. EUC is not designing this from scratch, but is simply catching up with what peer institutions have already shown is possible.

The real challenge, and the focus of this project, is ensuring this shift happens without compromising our high academic standards. ICaps does this by shifting the focus of assessment from the final product to the academic process, documented in a portfolio built around four stages: Research, Design, Implementation, and Reflection. The structure follows the logic of action research, a tradition of inquiry that refuses to separate thinking from doing. Researchers identify a problem, design a response, act on it, and reflect on what they learned. ICaps follows the same cycle, with one key difference: the final product is not just evidence of learning. It is an intervention in its own right, designed to address a real problem and reach a real audience.

The Research stage asks students to establish genuine academic grounding: a minimum of fifteen sources, a stakeholder analysis, and a clear account of why the problem matters. The Design stage requires them to justify their chosen format rather than default to it. Justifying one's choices is essential, because a poorly designed product cannot be saved by good implementation. The third stage, Implementation, documents what actually happened, including what deviated from the plan. Reflection runs throughout the project rather than appearing as a conclusion written after the fact. Each stage makes the student's reasoning visible and assessable in ways a traditional thesis rarely does. By assessing the students' reasoning behind their choices and their ability to link theory to practice, we can hold every format to the same academic standard. A podcast built on fifteen peer-reviewed sources and a substantive stakeholder analysis can be held to the same standard as a 6,500-word thesis. In other words, the format changes; the rigor doesn't. In June, ICaps students present at the Capstone Conference just like every other graduating student, giving supervisors and second readers the opportunity to ask questions about the reasoning behind the work directly. The portfolio also meets the existing ILO's of the Capstone thesis, which means the ICaps track does not require any changes to the ARR. For the institution, nothing formally changes. For students, everything does.

The challenge of change

Launching the pilot was harder than I expected, and not for the reasons I anticipated. The first pilot had thirteen sign-ups, but in the end zero participants. This forced me to have more honest conversations with students and staff members. Talking to students confirmed that the demand is real, but that the word 'pilot' feels too uncertain for something as important as the final graduation project. They were missing clear examples of what is possible and what a portfolio could look like before committing to one. Without that, the standard thesis simply seemed the safer bet. Simultaneously, the words 'pilot' and 'innovation' are interpreted by staff members as an increase in an already heavy workload. And while our ARR allows for flexibility when it comes to the final project, the 'standard' thesis is still often considered the only way to graduate at a university. Moving toward more diverse assessment formats can be perceived by some as a threat to academic rigor.

I started this project thinking all we needed was a good rubric and some guidelines. Now that I'm almost finished with the SEQ trajectory, I am convinced that educational innovation is primarily a cultural problem, not a design one. Guidelines and rubrics certainly matter, but the right timing and a strong community of support matter more, and they are much difficult to design for. If we want to change how we assess our students, we have to take more time to talk to each other and learn more about each other's concerns. There's a difference between informing colleagues about a change and genuinely involving them in shaping it. That difference turned out to be the whole project.

Looking forward

Despite the failed initial launch, the year was not lost. The ICaps manual is complete, with portfolio guidelines, assessment rubric, and an ILO mapping. The framework has been discussed with staff from CLI and Risbo, giving it the necessary institutional backing. A promo video is in production with the Studio, built around concrete examples of what an ICaps project can look like.

One small but significant decision came out of my conversations with the students: we are no longer calling it a pilot. ICaps will simply be presented as one of the available Capstone tracks. That framing shift matters more than it might seem. It changes how both students and supervisors relate to it before anyone has read a single page of the manual.

For the 2026-2027 cycle, the plan is to run a first cohort of eight students, two from each department, with ICaps introduced during Year 2 orientation, well before Capstone proposals need to be written. The success of this first run will be measured not only by completion rates and grades, but by whether supervisors feel confident in the process, whether students feel genuinely supported, and whether the Exam Board responds positively to the transparency and reliability of the new rubric. The longer-term goal is to make the ICaps track a standard Capstone option and build a digital archive of completed projects, giving EUC student work a larger audience and giving future students the concrete examples they told me they needed.

Lessons for anyone trying something similar

The biggest lesson from this project is that educational innovation is not primarily a design issue. It is a trust issue. You can have a strong theoretical grounding and a detailed course manual, and still watch a launch fail because the conditions were not right. The right timing and visible institutional supports matters enormously. But most importantly is recognizing the difference between informing colleagues about a change and genuinely involving them in shaping it. So if I were starting this project again, I would spend the first few months doing almost nothing except having conversations. Not presenting a plan, but listening to concerns and building relationships across departments. Rubrics can wait, collegial support can not.

In the end, ICaps is about asking a question that every programme should ask regularly: are we assessing what we actually value? At EUC, the honest answer was 'not yet'. But with the framework in place and students eager for change, we are finally aligning our ending with our beginning.

DOORONTWIKKELING TOETSPROGRAMMA IN DE MASTER GENEESKUNDE EA2030: PROGRAMMATISCHE EN COMPETENTIEGERICHTE BENADERING



JULIA BAUR

Introductie:

Een student loopt coschappen (stage) en voert hierin onder andere een medisch consult uit. Dit consult is verplicht om 2 keer tijdens de coschappen op niveau te behalen. De begeleider observeert, geeft feedback en beoordeelt via een rubric of het consult van voldoende niveau was. Deze beoordeling wordt vastgelegd in het online dashboard van de student, dat inzichtelijk is voor de besliscommissie die later een oordeel velt over de algehele voortgang van deze student.

Binnen de master Geneeskunde van het Erasmus MC is de afgelopen jaren gewerkt aan een toetsprogramma waarin ontwikkeling centraal staat (Erasmus MC, 2026). Dit sluit aan bij een bredere verschuiving van assessment of learning naar assessment for learning, waarbij toetsing niet alleen prestaties vaststelt, maar vooral het leerproces ondersteunt en zichtbaar maakt (Schuwirth & Van der Vleuten, 2011). Met het gebruik van een studentgericht portfolio, loslaten van cijfers, focus op narratieve feedback en besluitvorming over een langere periode is een sterke basis gelegd voor programmatisch toetsen, waarin leren en beoordelen nauw met elkaar verbonden zijn. Daarnaast is de opleiding gebaseerd op competentiegericht opleiden volgens het Raamplan Artsopleiding 2020 (NFU, 2020), waarin het functioneren van de toekomstige arts centraal staat.

De herziening van het programma roept een fundamentele vraag op: wat zegt dit ene consult over de ontwikkeling van deze student als toekomstig arts? Hoe wordt deze informatie gebruikt binnen het bredere geheel van het toetsprogramma?

Competentiegericht opleiden vraagt om toetsing die zichtbaar maakt in hoeverre studenten kennis, vaardigheden en professionele attitude in samenhang kunnen toepassen in authentieke beroepssituaties (e.g. Frank et al., 2010). Een toetsprogramma moet daarom niet geïsoleerd meten wat studenten weten of doen, maar inzicht geven in hoe zij zich ontwikkelen over tijd en in verschillende contexten. Vanuit het perspectief van Competency Assessment Programmes (CAP) ligt de kwaliteit van toetsing daarbij in de samenhang en doelgerichtheid van het gehele programma (Baartman et al., 2007), waarbij de gehanteerde kwaliteitscriteria moeten aansluiten bij kernprincipes van competentiegericht onderwijs.

Programmatisch toetsen (PT) sluit hierop aan door niet te richten op afzonderlijke toetsmomenten, maar samenhangende informatie te aggregeren door een verzameling van datapunten die inzicht geven in ontwikkeling (van der Vleuten et al., 2012). Het concept veronderstelt dat toetsing geen optelsom is van losse momenten, maar berust op een samenhangende set uitgangspunten die richting geven aan hoe informatie wordt verzameld, geïnterpreteerd en gebruikt voor zowel leren als besluitvorming (Heeneman et al., 2021).

In de Nederlandse context zijn deze uitgangspunten recent geconcretiseerd in zeven herijkte ontwerpprincipes (Van Schilt-Mol et al., 2025; zie figuur 1). Deze principes laten ruimte voor verschillende ontwerpkeuzes in de praktijk (Agricola et al., 2024) maar veronderstellen wel dat zij in samenhang worden toegepast om te kunnen spreken van een programmatisch toetsprogramma.

In de praktijk blijkt echter dat de “reis” van een datapunt, zoals medische consultvoering bij een patiënt, niet altijd verloopt zoals bedoeld. Juist in de dagelijkse uitvoering ontstaan spanningen tussen de leerfunctie en de beslisfunctie van toetsing.



In dit project wordt onderzocht hoe het toetsprogramma van de master Geneeskunde van het Erasmus MC kan worden opgevat als een CAP dat programmatisch is ingericht, zodat de programmatische principes en de CAP-kwaliteitscriteria elkaar versterken in het ondersteunen van competentieontwikkeling. Tegen deze achtergrond wordt verkend in hoeverre het huidige toetsprogramma aansluit bij deze uitgangspunten en waar verdere ontwikkeling nodig is. De analyse in dit artikel is gebaseerd op een combinatie van literatuuronderzoek en praktijkervaring binnen de opleiding. Naast het bestuderen van relevante literatuur is de praktijkanalyse uitgevoerd op basis van participatie in en observatie van verschillende gremia en activiteiten binnen de opleiding, waaronder werkgroepen in de bachelor- en masterfase, vergaderingen van de besliscommissie en werksessies met clusterleiders. Daarnaast zijn gesprekken gevoerd met onderwijskundigen, toetsdeskundigen (waaronder het Amsterdam UMC), een kwalitatief onderzoeker en experts op het gebied van programmatisch toetsen (o.a. Lambert Schuwirth), en zijn bestaande documenten zoals het toetsplan geanalyseerd. Deze combinatie van bronnen biedt een rijk beeld van zowel het ontwerp als de uitvoering van het toetsprogramma.

Herijkte principes programmatisch toetsen*

De beoogde leerresultaten zijn de basis voor het verzamelen en aggregeren van informatie uit datapunten.

Informatie wordt verzameld door middel van een bewust gekozen en beargumenteerde mix van datapunten en omvat feedback die daarop wordt gegeven vanuit verschillende perspectieven.

Informatie staat ten dienste van (het sturen van) de ontwikkeling van de student op deze leerresultaten én het nemen van onderbouwde beslissingen over de beheersing van de beoogde leerresultaten: de leer- en beslisfunctie van toetsing zijn met elkaar verweven.

De student interpreteert en gebruikt, in georganiseerde dialoog, informatie uit datapunt(en) en geeft daarmee richting aan de verdere ontwikkeling van de beoogde leerresultaten.

De student en de opleiding monitoren doorlopend de beheersing van de beoogde leerresultaten aan de hand van geaggregeerde informatie uit datapunten.

Beslissingen worden niet genomen op basis van informatie afkomstig van één datapunt. Saturatie van informatie is voorwaardelijk voor het nemen van een onderbouwde beslissing.

Een commissie van examinatoren interpreteert de geaggregeerde informatie uit datapunten en neemt op basis daarvan gezamenlijk op een transparante en navolgbare wijze een onderbouwde beslissing over (het niveau van) de beheersing van de beoogde leerresultaten.

Tabel 1. Herijkte principes (Van Schilt-Mol et al., 2025).

*Alle 7 principes moeten worden toegepast. Anders is er geen sprake van programmatisch toetsen.

Drie spanningsvelden in de praktijk

De beschreven spanningsvelden in de praktijk zijn gebaseerd op de principes van programmatisch toetsen, zoals aangegeven in tabel 1.

1. Van beroepsactiviteit naar inzicht in competentieontwikkeling

In het voorbeeld wordt het medisch consult beoordeeld als een afzonderlijke activiteit. In de praktijk spelen kritische beroepsactiviteiten (KBA's), zoals een medisch consult, een belangrijke rol als herkenbare en authentieke taken (Ten Cate, 2005). Binnen een programmatisch en competentiegericht toetsprogramma ligt de kracht van dergelijke observaties voornamelijk in het feit dat deze KBA's inzicht geven in onderliggende competenties, zoals klinisch redeneren, communicatie of samenwerken. Dit vraagt om een benadering waarin het consult niet (alleen) wordt beoordeeld als taak, maar wordt geïnterpreteerd als informatief datapunt dat iets zegt over de competentieontwikkeling. De aandacht verschuift daarmee van de vraag "is deze activiteit voldoende uitgevoerd?" naar "wat laat deze activiteit zien over de ontwikkeling van de student?"

2. Van datapunt als norm, naar besluitvorming op basis van patronen

Het uitgevoerde consult wordt niet alleen vastgelegd, maar dient ook betekenis te krijgen in het besluitvormingsproces. Een belangrijk uitgangspunt hierbij is dat niet alle beschikbare informatie automatisch even relevant is. Binnen programmatisch toetsen wordt daarom gewerkt vanuit decision driven data-collection: de informatie die wordt verzameld, wordt bewust afgestemd op de beslissingen die genomen moeten worden (Williams & Harrison, 2019). Dit betekent dat vooraf moet worden bepaald welke informatie nodig is om een goed onderbouwd oordeel over de competentieontwikkeling te kunnen vormen, ook wel saturatie genoemd.

Wanneer observaties worden gekoppeld aan frequentienormen (e.g.: "je dient de sKBA medisch consult 2 keer op niveau te laten aftekenen tijdens coschap heelkunde"), kan de nadruk verschuiven naar afzonderlijke momenten, terwijl juist wordt beoogd om beslissingen te baseren op patronen in meerdere observaties, vanuit verschillende perspectieven en in uiteenlopende contexten, ook wel triangulatie (van der Vleuten et al., 2012). Het consult dient daarom betekenis in relatie tot andere observaties te krijgen: laat de student groei zien? Is er consistentie? Zijn er terugkerende ontwikkelpunten? Door op deze manier naar datapunten te kijken, ontstaat een rijker en beter onderbouwd beeld van de ontwikkeling van de student.

3. Van feedbackmoment naar doorlopend leerproces

Na het consult ontvangt de student feedback van de begeleider. Dit vormt een waardevol moment, waarin concrete aanknopingspunten worden gegeven voor verdere ontwikkeling. De vraag die hierbij centraal staat is: wat gebeurt er daarna met deze feedback? Binnen programmatisch toetsen heeft feedback niet alleen een beschrijvende functie, maar vooral een sturende rol in het leerproces (Schuwirth & Van der Vleuten, 2011). De waarde van het consult ligt daarmee niet alleen in de beoordeling of de feedback op dat moment, maar in de mate waarin de student deze informatie gebruikt in vervolgactiviteiten.

Dit vraagt om een benadering waarin feedback cyclisch wordt benut: eerdere observaties en feedback worden meegenomen in nieuwe situaties, waardoor ontwikkeling zichtbaar wordt over tijd. De aandacht verschuift daarmee van welke feedback is gegeven? naar hoe wordt deze feedback gebruikt in de vervolgactiviteit? (Winstone & Boud, 2022).



Kwaliteit van het geheel

Bovenstaande spanningsvelden raken direct aan de kwaliteit van het toetsprogramma als geheel. Zij maken duidelijk dat de uitdaging niet zozeer ligt in het verzamelen van voldoende informatie, maar in de manier waarop deze informatie wordt geïnterpreteerd en gebruikt voor besluitvorming en leren. Dit roept de vraag op aan welke criteria ons toetsprogramma moet voldoen om op basis van deze rijke, contextgebonden informatie tot betekenisvolle en verantwoorde beslissingen te komen. Kwaliteit van toetsing wordt traditioneel benaderd vanuit een perspectief waarin objectiviteit, betrouwbaarheid en standaardisatie centraal staan (Kickert & Langeveld, 2026). Deze benadering is waardevol bij toetsvormen waarbij prestaties relatief contextonafhankelijk kunnen worden beoordeeld. In de context van het coschap, waarin een student een consult uitvoert, spelen echter juist de specifieke situatie, de patiënt en het professioneel oordeel van de begeleider een belangrijke rol. Dit vraagt om aanvullende kwaliteitscriteria waarin de interpretatie van rijke, contextgebonden informatie centraal staat.

Binnen CAPs wordt toetskwaliteit daarom niet primair beoordeeld op het niveau van afzonderlijke observaties, maar op het niveau van het toetsprogramma als geheel (Baartman et al., 2007). De vraag is daarbij niet of dit ene consult valide of betrouwbaar is, maar of het geheel van observaties leidt tot een onderbouwd en betekenisvol oordeel over de achterliggende competentieontwikkeling van de student. Dit sluit aan bij inzichten uit kwalitatief onderzoek, waarin validiteit niet wordt opgevat als een eigenschap van afzonderlijke metingen, maar als de geloofwaardigheid van de interpretaties die op basis van rijke, contextgebonden informatie worden gemaakt (Creswell & Miller, 2000). Deze benadering impliceert dat de manier waarop kwaliteit wordt gedefinieerd en geborgd, in lijn moet zijn met de visie op toetsing en opleiden. Wanneer een opleiding kiest voor een programmatisch en competentiegericht toetsprogramma, vraagt dit om kwaliteitscriteria die recht doen aan de integratie van informatie, de rol van professioneel oordeel en de ontwikkeling van studenten over tijd. Met andere woorden: ontwerp van het toetsprogramma en de manier waarop kwaliteit wordt beoordeeld, moeten met elkaar in lijn zijn. Hoewel meerdere kwaliteitscriteria relevant zijn, sluiten drie criteria, verder toegelicht in figuur 1, in het bijzonder aan op de beschreven spanningsvelden.

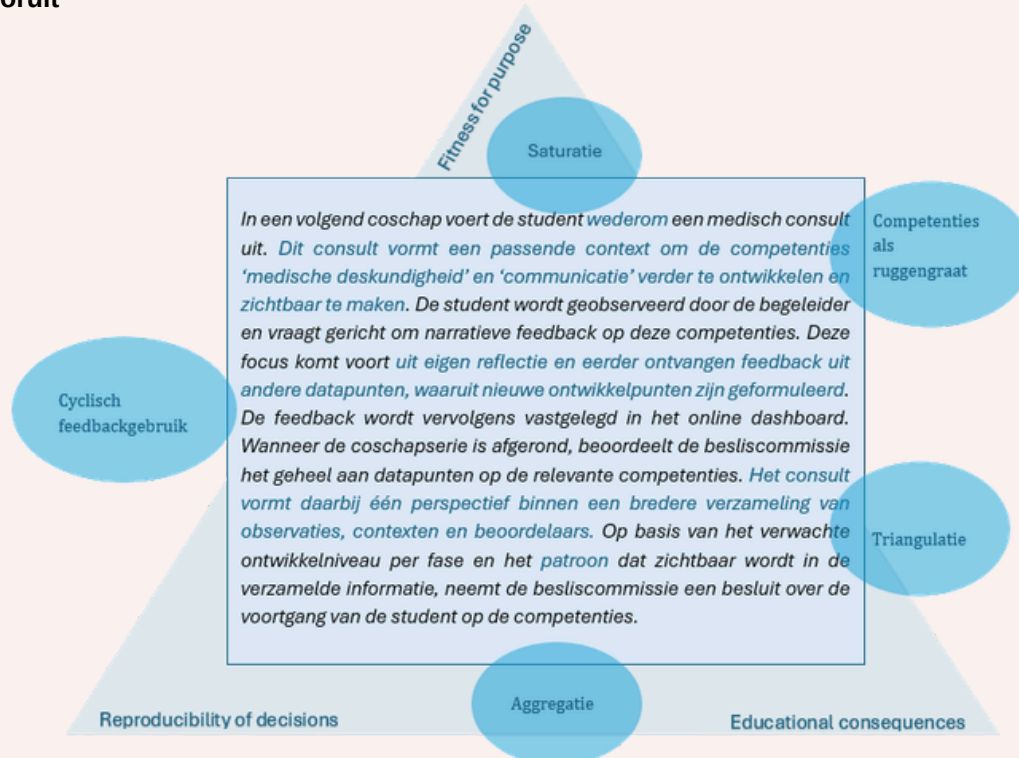
- **Fitness for purpose:** Sluit het toetsprogramma aan bij het doel van de opleiding?
- **Reproducibility of decisions:** Zijn beslissingen navolgbaar en consistent?
- **Educational consequences:** Stimuleert het toetsprogramma het gewenste leergedrag?

Lessons learned

Dit project heeft laten zien dat de ontwikkeling van het toetsprogramma voor de master EA2030 vraagt om een samenhangende benadering van zowel programmatisch toetsen als competentiegericht opleiden. Het gaat niet alleen om hoe informatie verzameld en gebruikt wordt, maar ook om wat deze informatie zichtbaar maakt. De kern ligt daarmee in het verschuiven van losse observaties naar het herkennen en interpreteren van patronen in de ontwikkeling van studenten over tijd, binnen betekenisvolle beroepscontexten. Programmatisch toetsen biedt het kader om deze informatie te verzamelen en te wegen, terwijl competentiegericht opleiden richting geeft aan de inhoud en betekenis ervan.

Deze beweging raakt niet alleen het ontwerp van het toetsprogramma, maar vraagt voornamelijk om een bredere cultuurverandering binnen de opleiding. Het betekent dat betrokkenen op alle niveaus anders moeten willen leren kijken naar toetsing: minder gericht op afvinken en normeren, en meer op begrijpen, duiden en begeleiden van ontwikkeling. Daarbij vraagt het expliciet om het loslaten van vertrouwde referentiekaders. De manier waarop wijzelf zijn opgeleid, vormt niet vanzelfsprekend de beste basis voor toekomstgericht opleiden. Juist in het openstaan voor nieuwe benaderingen en het vertrouwen op professioneel oordeel, ondersteund door rijke informatie uit meerdere perspectieven, ligt de sleutel tot een toetsprogramma dat recht doet aan een waardevolle opleiding voor de arts van de toekomst.

Een blik vooruit



Literatuur

- Agricola, B.T., De Vos, M., Baartman, L., & Van Schilt-Mol, T. (2024). Naar een nieuwe balans in toetsfuncties: van formatief en summatief naar een continuüm van beslissingen. [NRO Overzichtsstudies Onderzoeksrapport]. Hogeschool Utrecht, Utrecht. <https://www.hu.nl/onderzoek/publicaties/naar-een-nieuwe-balans-in-toetsfuncties-3>
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, 2(2), 114-129. <https://doi.org/10.1016/j.edurev.2007.06.001>
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, 39(3), 124-130.
- Erasmus MC. (2026). Ontwikkeling master Erasmusarts 2030. <https://www.erasmusmc.nl/nl-nl/erasmusarts2030/ontwikkeling-master-erasmusarts-2030>
- Frank, J. R., Snell, L. S., Cate, O. T., Holmboe, E. S., Carraccio, C., Swing, S. R., Harris, P., Glasgow, N. J., Campbell, C., Dath, D., Harden, R. M., Lobst, W., Long, D. M., Mungroo, R., Richardson, D. L., Sherbino, J., Silver, I., Taber, S., Talbot, M., & Harris, K. A. (2010). Competency-based medical education: Theory to practice. *Medical Teacher*, 32(8), 638-645.
- Heeneman, S., de Jong, L. H., Dawson, L. J., Wilkinson, T. J., Ryan, A., Tait, G. R., ... & van der Vleuten, C. P. (2021). Ottawa 2020 consensus statement for programmatic assessment–1. Agreement on the principles. *Medical Teacher*, 43(10), 1139-1148.
- Kickert, R., & Langeveld, E. (2026). Toetsen is onderzoeken. In A. van Berkel & D. Sluijsmans (Red.), *Toetsen in verbinding: Perspectieven en praktijkvoorbeelden voor het hoger onderwijs* (blz. 284-303). Telos Uitgevers.
- Kickert, R., Meeuwisse, M., Stegers-Jager, K. M., Prinzie, P., & Arends, L. R. (2022). Curricular fit perspective on motivation in higher education. *Higher Education*, 83(4), 729-745.
- Nederlandse Federatie van Universitair Medische Centra (NFU). (2020). Raamplan Artsopleiding 2020. NFU. <https://www.nfu.nl>
- Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, 33(6), 478–485. <https://doi.org/10.3109/0142159X.2011.565828>
- Ten Cate, O. (2005). Entrustability of professional activities and competency-bases training. *Medical Education*, 39, 1176-1177. <https://doi.org/10.1111/j.1365-2929.2005.02341.x>
- Van Schilt-Mol, T., Baartman, L., & De Jong, L. (2025). Herijking principes programmatisch toetsen [Presentatie]. Week van Toetsbekwaamheid. https://toetsbekwaamheid.nl/wp-content/uploads/2025/11/Tamara-van-Schilt-Mol_Herijking-principes-programmatisch-toetsen-work-in-progress-week-van-toetsbekwaamheid.pdf
- Van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Govaerts, M. J. B., & Heeneman, S. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205-214.
- William, D., & Harrison, C. (2019). Planning for learning. Decision-driven data collection. <https://www.youtube.com/watch?v=1MMDLZxJz0k>
- Winstone, N. E., & Boud, D. (2022). The need to disentangle assessment and feedback in higher education. *Studies in higher education*, 47(3), 656-667.



Testimonials, quotes & more:

Kjell Noordzij- Assistant Professor of Sociology (ESSB) (p.6)

Kjell Noordzij is Assistant Professor of Sociology in the Department of Public Administration and Sociology at Erasmus University Rotterdam. He studies the social aspect of representative democracy. Much of his research concerns how different groups of citizens perceive democracy and their representation in it, and how political conflicts play out inside and outside electoral politics. For more information, see www.kjellnoordzij.nl

Julia Baur - Adviseur Onderwijs & Toetsing (p.16)

“As an assessment advisor, I work on assessment every day, but the SEQ program gave me the opportunity to reflect on this at a more fundamental level and engage in dialogue with colleagues outside my own assessment bubble.”

VAN PRAKTIJKEXPERT NAAR BEOORDELAAR

Hoe versterken we de rol van examinatoren in beroepsauthenticke toetsing?

MANDY MEEUWSEN

Inleiding: beoordelen in de werkelijkheid

De afgelopen jaren heeft de Politieacademie een duidelijke verschuiving doorgemaakt richting beroepsauthentic onderwijs. Studenten worden niet langer uitsluitend beoordeeld op losse kennis of vaardigheden, maar op hun vermogen om in realistische situaties professioneel te handelen. Dit vraagt om een andere manier van kijken naar toetsing én naar de rol van de examinator.

Binnen de leerlijn Integrale Politiële Vaardigheden (IPV) wordt deze ontwikkeling concreet zichtbaar. Studenten combineren juridische kennis, fysieke vaardigheden en communicatieve competenties in complexe praktijksituaties. De beoordeling van deze prestaties vraagt om deskundige examinatoren die verder kijken dan checklists alleen.



“Beoordelen in een realistische context vraagt meer dan het toepassen van criteria; het vraagt professioneel oordeelsvermogen.”

Het vraagstuk: wat vraagt beroepsauthentic toetsen van een examinator?

In de praktijk bleek dat examinatoren zich niet altijd voldoende toegerust voelen voor deze taak. Verwachtingen zijn niet altijd expliciet, beoordelingscriteria worden verschillend geïnterpreteerd en structurele afstemming ontbreekt. Dit leidt tot risico's voor de betrouwbaarheid en validiteit van beoordelingen.

Achter deze vraag gaat een fundamenteel spanningsveld schuil. Enerzijds vraagt beroepsauthenticke toetsing om ruimte voor professionele oordeelsvorming: situaties zijn complex, contextafhankelijk en niet volledig te standaardiseren. Anderzijds stelt kwaliteitsborging eisen aan consistentie, transparantie en onderbouwing van beoordelingen.

De centrale vraag van dit project luidde daarom: hoe kan de professionalisering van examinatoren zodanig worden ingericht dat beroepsauthenticke toetsen op een consistente en onderbouwde manier worden beoordeeld?

Om deze vraag te beantwoorden is een iteratieve aanpak gehanteerd waarin analyse, literatuurstudie en ontwerp met elkaar verbonden waren. Gesprekken met stakeholders vormden de basis voor inzicht in de huidige praktijk. Vervolgens is literatuur geraadpleegd over assessment literacy en beoordelaarsdeskundigheid. De inzichten uit praktijk en theorie zijn vertaald naar een concreet praktijkproduct: een examinatorprofiel en een professionaliseringsaanpak.

In de praktijk bleek dat deze spanning niet altijd expliciet werd gemaakt. Er werd vaak impliciet uitgegaan van het idee dat een examiner die inhoudelijk expert is, ook in staat is om goed te beoordelen. Literatuur laat echter zien dat toetsbekwaamheid meer omvat dan inhoudelijke expertise alleen. Het gaat ook om het vermogen om gedrag systematisch te observeren, te interpreteren en te relateren aan vooraf vastgestelde criteria.

Daarnaast blijkt uit onderzoek dat authenticiteit geen vaststaand gegeven is, maar afhankelijk van perspectief. Wat voor een docent authentiek lijkt, hoeft dat voor een student of professional uit het werkveld niet te zijn. Dit onderstreept het belang van afstemming: tussen examinatoren onderling, en tussen onderwijs en beroepspraktijk.

Het project richtte zich daarom niet alleen op de vraag wat examinatoren moeten kunnen, maar ook op hoe die bekwaamheid ontwikkeld en geborgd kan worden.

Resultaten: van impliciet naar expliciet

Inzicht in de praktijk

De eerste stap bestond uit het in kaart brengen van de huidige situatie. Door gesprekken met examinatoren, kwaliteitscoördinatoren en onderwijskundigen werd duidelijk waar knelpunten lagen. Examinatoren gaven onder meer aan dat:

- het niet altijd duidelijk was wat precies van hen verwacht werd;
- beoordelingscriteria ruimte lieten voor verschillende interpretaties;
- afstemming tussen examinatoren beperkt was;
- professionalisering voornamelijk bestond uit e-learning, zonder sterke koppeling met de praktijk.

Gesteld moet worden dan met de ontwikkeling van nieuwe bacheloropleidingen en de vorming van de sector Hoger politieonderwijs al heel mooie stappen gemaakt zijn. Er is geïnvesteerd in een onderwijsvisie, er wordt nagedacht over een visie op examinering en beleid krijgt langzaam vorm. Ook op het vlak van deskundigheidsbevordering en kwaliteit liggen er verschillende documenten waarin processen beschreven worden. In de praktijk echter blijkt dit vooral een papierenwerkelijkheid die niet goed op elkaar aansluiten, en wordt een visie en werkwijze nog nauwelijks gedragen in het team.

Deze bevindingen maakten zichtbaar dat het probleem niet zozeer lag in de motivatie of inzet van examinatoren en anderen betrokkenen, maar in het ontbreken van een helder en gedeeld kader dat aansluit bij de huidige onderwijs- en toetspraktijk.



Verkenning van de gewenste situatie

Vervolgens is literatuur geraadpleegd over assessment literacy en beroepsauthenticke toetsing (o.a. Meijer et al. 2020; Sluijsman et al, 2015, Derksen et al., 2020, Chan, 2022), . Daaruit komt een consistent beeld naar voren: goede beoordeling vraagt om een combinatie van kennis, vaardigheden en gezamenlijke betekenisgeving.

Tabel 1. Definities en componenten van toetsbekwaamheid uit (vrij naar Pastore & Andrade, 2019 verkregen uit Meijer en collega's, 2020).

Stiggins (1999)	Toetsbekwame docenten: 1. Verbinden beoordelingen met duidelijke doelen. 2. Verduidelijken de verwachtingen van de prestaties. 3. Passen de juiste beoordelingsmethoden toe. 4. Ontwikkelen beoordelingsmodellen, scorecriteria en steekproeven op de juiste manier. 5. Vermijden bias. 6. Communiceren effectief over de prestaties van studenten. 7. Gebruiken toetsen als een educatieve interventie.
Fullan & Waton (2000)	De bekwaamheid van docenten - alleen of samen - (a) om werk- en prestatiegegevens van studenten te toetsen en nauwkeurig te begrijpen, en, dienovereenkomstig (b) klas en schoolplannen te ontwikkelen om betere resultaten te bereiken.
Shepard (2000)	Toetsbekwame docenten moeten in staat zijn om te evalueren welke verbeteringen op de scholen moeten worden gerealiseerd, naast het kunnen ontwerpen en afnemen van summatieve eindtoetsen.
McMillan (2001)	Docenten met een solide achtergrond in beoordelen zijn goed in staat om beoordelingen te integreren met onderwijs zodat ze toetsen kunnen gebruiken voor geschikte vormen van lesgeven.
Paterno (2001)	Toetsbekwaamheid is het bezit van kennis over de basisprincipes van een degelijke beoordelingspraktijk, inclusief terminologie, de ontwikkeling en gebruik van beoordelingsmethoden en -technieken, bekendheid met kwaliteitsnormen bij de beoordeling en vertrouwdheid met alternatieven voor traditionele toetsen.
Webb (2002)	Toetsbekwaamheid is het hebben van kennis van middelen om te beoordelen wat studenten weten en kunnen, hoe de resultaten van deze beoordelingen moeten worden geïnterpreteerd, en hoe deze resultaten toe te passen om het leren van studenten en de effectiviteit van het onderwijs te verbeteren.

Mertler (2003)	Toetsbekwame docenten herkennen degelijke toetsen, evaluaties, en communicatiepraktijken; zij begrijpen beoordelingsmethoden om betrouwbare informatie over prestaties van studenten te verzamelen en communiceren beoordelingsresultaten effectief, of het nu gaat om het gebruik van rapportcijfers, testcores, portfolio's of presentaties. Toetsbekwame docenten kunnen beoordelingen gebruiken om de motivatie en het leren van studenten te maximaliseren door studenten als volwaardige partners te betrekken bij de beoordeling, de administratie en de communicatie.
DeLuca & Klinger (2010)	Toetsbekwame docenten begrijpen en gebruiken de juiste beoordelingsmethoden en hebben kennis over de theoretische en filosofische onderbouwing bij het meten van het leren van student.
Willis, Adie, & Klenowski (2013)	Toetsbekwaamheid is een dynamische, contextafhankelijke sociale praktijk wat impliceert dat docenten met elkaar en met studenten spreken en onderhandelen over praktische en algemene kennis in de aanvang, de ontwikkeling en de praktijk van toetsing om de leerdoelen van studenten te bereiken.

Examinatoren moeten niet alleen weten wat ze beoordelen, maar ook hoe ze dat doen en waarom. Dat betekent dat zij een scherp beeld moeten hebben van de competenties die centraal staan, en dat zij deze op vergelijkbare wijze interpreteren. Ook wordt in de literatuur benadrukt dat kalibratie, het gezamenlijk bespreken en afstemmen van beoordelingen, een cruciale rol speelt in het vergroten van betrouwbaarheid.

Opvallend is bovendien dat er in het kader van authentiek toetsen, veelal algemeen wordt gesteld dat een examiner inhoudelijk expert moet zijn op het thema dat hij beoordeeld. En laat daar nu een deel van het probleem zitten: het docententeam binnen HPO bestaat voor het grootste gedeelte uit collega's wie geen achtergrond hebben in het politiewerk op straat. Zij komen, zoals we dat vaak zeggen, niet 'uit het blauw', dus hoe kunnen zij inhoudelijk expert worden?

De opbrengsten: examiner profiel en advies voor professionalisering

Een belangrijke opbrengst van het project is het expliciteren van verwachtingen. Waar voorheen impliciete aannames domineerden, is nu uitgewerkt wat examinatoren daadwerkelijk nodig hebben om hun rol goed te vervullen.

“Kwaliteit van toetsing zit niet alleen in instrumenten, maar vooral in mensen.”

Het ontwikkelde profiel beschrijft drie kerncomponenten: inhoudelijke expertise, toetsbekwaamheid en contextbegrip. Deze combinatie maakt het mogelijk om complexe prestaties op een onderbouwde manier te beoordelen.

Het ontwikkelde profiel beschrijft drie kerncomponenten: inhoudelijke expertise, toetsbekwaamheid en contextbegrip. Deze combinatie maakt het mogelijk om complexe prestaties op een onderbouwde manier te beoordelen.

Daarnaast is het belang van kalibratie benadrukt. In complexe toetsvormen is volledige standaardisatie niet mogelijk. Betrouwbaarheid wordt daarom vooral bereikt door gezamenlijke interpretatie en kalibratie. Door structureel casuïstiek te bespreken en beoordelingen te vergelijken, ontstaat een gedeelde interpretatie van criteria en neemt de betrouwbaarheid toe. Professionalisering wordt niet langer gezien als een eenmalige training, maar als een doorlopend proces. Deskundigheid ontwikkelt zich in de praktijk, door herhaling, reflectie en uitwisseling. Dit vraagt om structurele inbedding in de organisatie en voortdurende aandacht voor ontwikkeling.

Implicaties voor de praktijk

De resultaten van dit project laten zien dat beoordeling een gezamenlijke verantwoordelijkheid is. Goede toetsing vraagt om samenwerking, afstemming en een gedeeld begrip van kwaliteit. Beroepsauthenticke toetsing brengt onderwijs dicht bij de praktijk, maar stelt tegelijkertijd hogere eisen aan de beoordeling. Dit project laat zien dat het versterken van examinatoren een sleutelrol speelt in het waarborgen van kwaliteit. Daarnaast wordt duidelijk dat de relatie met de beroepspraktijk essentieel is om de authenticiteit van toetsing te waarborgen. Niet door alles op papier helemaal vaststaand te regelen, maar door helderheid te creëren, samenwerking te stimuleren en ruimte te bieden voor professionele ontwikkeling maak je als organisatie stappen in toetskwaliteit.

De volgende stap is implementatie binnen pilots, gevolgd door evaluatie en opschaling. Hierbij wordt gekeken naar effecten op beoordelingsconsistentie en ervaren bekwaamheid van examinatoren.

Literatuur

- Chan, C. K. Y. & Luk, L. Y. Y. (2022) A four-dimensional framework for teacher assessment literacy in holistic competencies, *Assessment & Evaluation in Higher Education*, 47:5, 755-769, DOI: 10.1080/02602938.2021.196280
- Derksen, K., van Middelkoop, D., van den Berg, L., Bay, N., Horsseleberg, P., & Kaarsemaker, A. (2020). Samensterk of nog meer werk? Onderwijsteams in het HBO. Zestor. <https://www.zestor.nl/nieuws/samen-sterk-nog-meer-werk>
- Meijer, K., van Schilt-Mol, T., Dobbelaer, M., Baartman, L., van der Linden, J., & Munneke, L. (2020). Voortdurende toetsbekwaamheid in het hoger beroepsonderwijs, een eerste verkenning. *Examens*, 2, 24-34
- Sluijsman, D., Joosten-ten-Brinke, D., & Van Schilt-Mol, T. (Red., 2015) *Kwaliteit van toetsing onder de loep*.

WHEN, AND WHEN NOT, TO USE AI? UNEVEN VALUES AND EXPECTATIONS AMONG HISTORY EXAMINERS

Erasmus School of
History, Culture and
Communication



SANDRA KHOR MANICKAM

In ESHCC and the History department, there are three options available to lecturers for how AI can be used by students in the class – yes, yes with qualification, and not at all. In previous SEQ research conducted by ESHCC colleague Jay Lee in April 2025, only 67% of History courses in 2024 updated their AI policy page in Canvas - the lowest % among the three departments. When it comes to individual assignments, however, there is often not an accounting of what ways AI can be used in assignments, and how using AI instead of students completing certain tasks themselves might interfere with learning goals. An assessment policy that specifically takes into consideration the myriad ways students can use AI in history assignments has not yet been formulated. Leaving aside for the moment the question of AI detection, the question of what, if any, AI tools are acceptable at different stages of completing assignments has yet to be answered among our staff or as a programme. This leaves open the possibility that there are differing unspoken expectations from class to class about what students are actually being tested on.

My project aimed to:

1. Create and circulate a questionnaire among History examiners about what they find to be acceptable AI use in graded assessments.
2. Collate and analyze the results to come up with better instructions for assessment, and as material to formulate a larger AI policy within the department.

The survey questions were created by me after researching similar kinds of questionnaires and specifications for GenAI use already in circulation in higher education. A main consideration was the length of the questionnaire. Though I could ask about many other components of assignments and possible ways GenAI could be used, I prioritized keeping the questionnaire to 10-15 minutes so as to ensure that colleagues would complete it. Part 1 consisted of 19 questions with five possible answers to teachers' preference for GenAI use. These questions addressed various tasks in completing assignments ranging from copy-editing to summarizing texts and coming up with research questions. The fifth option, "Other", gave respondents the opportunity to fill in their own unique answers.

After this series, three questions were asked to see if their answers would change if there were specific GenAI tools used for academic purposes. To round off the section, an AI-index table created by Tilburg University was shown and respondents asked which levels would be appropriate for different years of the bachelor programme.

The next section gave respondents statements and asked them to respond the extent of their agreement on a five-point scale. Nine statements were given and the results are collated into "Strongly agree/agree", "Neutral" and "Disagree/Strongly Disagree". Finally, several questions about the respondents' age, gender, role in education and AI use were asked. Respondents were asked further if they would be willing to be contacted further for more discussion on this issue.

Respondent composition

An email was sent out to History department colleagues on 12 January 2026. After approximately two weeks, I received 18 responses and closed the questionnaire. Half of respondents were course coordinators, and about three-quarters were tutorial lecturers. There was an even sample of colleagues who were permanent staff and temporary staff which usually comprise PhD candidates and external lecturers. I found it important to understand how permanent staff thought about these issues as they hold a leading role in setting standards within the department. PhD candidates' views were also essential because this generation of scholars are working in an environment where GenAI is available and I expect them to be more familiar with GenAI tools and their current usage in academia.

Introduction to the questionnaire

The introductory text asked respondents to think about the following assessments when answering questions:

- Open answer questions (in Canvas, exam, or written in class)
- Short (<1000 words) opinion/position pieces based on research
- Short research papers (<1000 words)
- Long research papers/BA thesis
- Podcasts, oral recordings
- Presentations
- Poster presentations

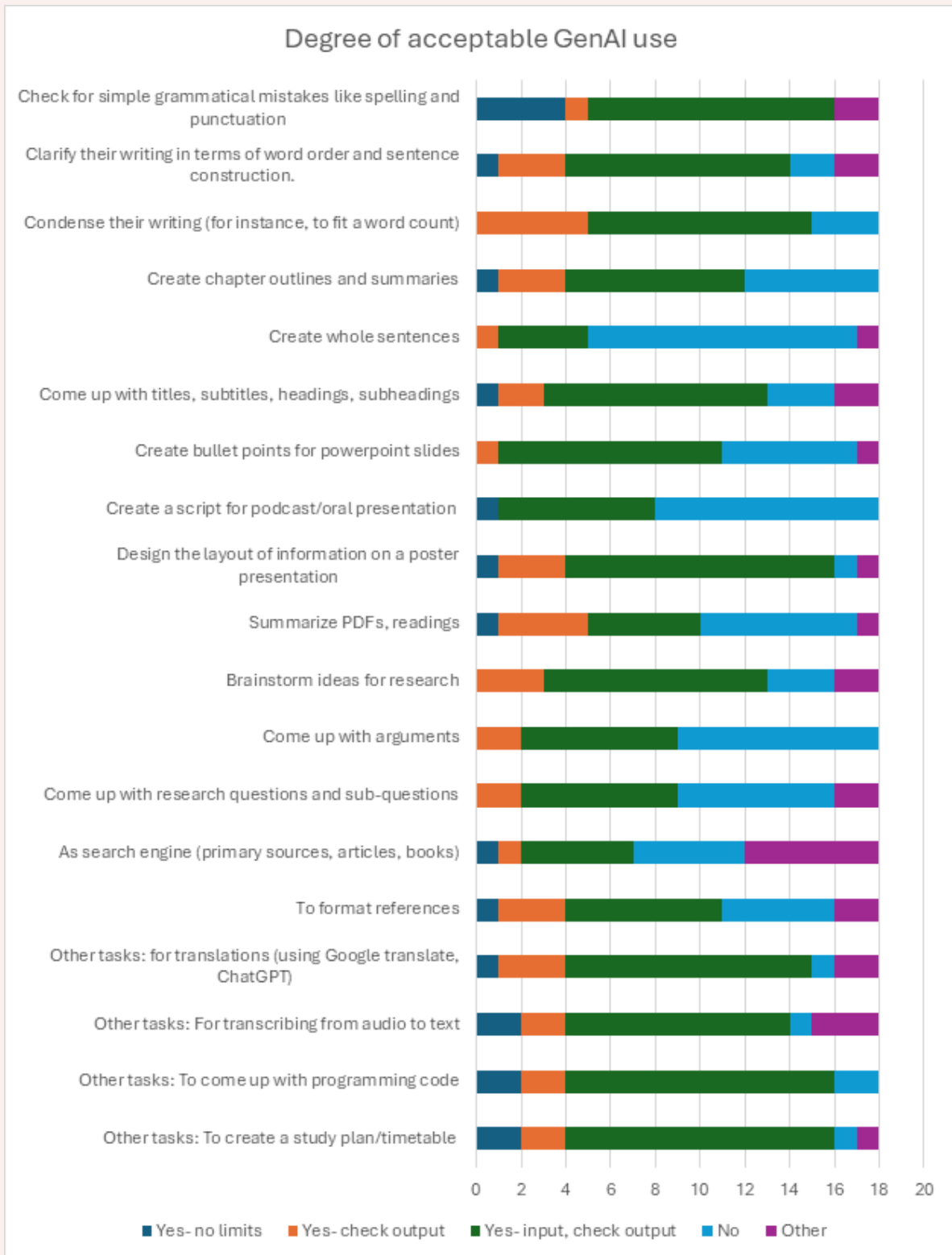
Respondents were asked if there were other types of assessments that should be included, and the only additional format was creating images and videos, indicating that the list of assessments were largely representative of the forms we use. Respondents were asked to leave aside the question of AI detection and to focus on what they think students should be able to do as part of their history study for graded assessments.

There were 5 possible answers per task:

- Option 1: Yes (GenAI can be used, no limitations)
- Option 2: Yes (GenAI can be used provided there is input from student, and output is checked by student)
- Option 3: Yes (GenAI can be used provided output is checked by student)
- Option 4: No (GenAI should not be used at all)
- Other. Please provide further clarification, for instance, if it depends on a specific kind of class or assignment, on which year a student is in, on which kind of GenAI tool is used, etc.

Options 1 (no limitations) and 4 (a complete ban on GenAI use) represent two extremes of views on whether students should use GenAI. Options 2 and 3 allowed varying degrees of input from the student with most respondents choosing Option 2 for most assessment components as will be shown below. It is interesting to see what GenAI use respondents considered to be 'out-of-bounds' (Option 4) and how this differed among teachers as this indicates where discussion is needed within the department about what skills should be assessed without students using GenAI.

Results: Degree of acceptable GenAI Use



The analysis of this initial exploration of lecturers' views demonstrates where lecturers are divided in their views of the importance of specific tasks and skills in graded assessments. Where there is a high level of agreement about the extent to which some tasks and skills can be completed with some form of AI use, for others, there is high level of divergence. Staff were split about whether the use of AI is allowed for important components of historical learning and scholarship such as "Summarize PDFs, readings"; "Come up with arguments"; and "Come up with research questions and sub-questions". Of note, there were a wide variety of responses for whether AI is allowed "As search engine". Searching for material is a very basic skill on which other historical competencies are built. It is also, if anecdotal evidence from students is to be believed, the one task that students most often use GenAI for. That lecturers themselves do not have a similar idea of how students should go about this in assessments is indicative that further discussion is needed within the department about how we wish to teach students some of our most basic disciplinary skills.

Results: Level of agreement with statements on GenAI1

The aim of asking respondents' level of agreement on the following statements was to gauge respondents' views on the role of GenAI more generally in student learning, whether they think that it improves student work, and how they see student engagement with GenAI as a tool.

The first thing to note is that most feel that it is not desirable (or practical) that GenAI be banned completely from use as students are already using it. That being said, respondents have a low opinion of how good students are at using GenAI in their assignments. There is also the issue of how students see output from GenAI with most respondents feeling that students see it as "truthful". Such results say nothing about how students actually think about GenAI and how well they think they are using the tool. Further research would be needed among the student population on these points.



Conclusion

The attempt to break down assessments to their component pieces and ask lecturers about acceptable AI use shows that the move from assessing the final product to assessing the process will entail agreement about what components should be completed without GenAI. Most agree to the use of GenAI in checking for simple grammatical mistakes like spelling and punctuation. However, there is disagreement about how much GenAI should play in other important components such as reading documents and searching for sources. A move towards grading the process would mean such tasks could be assessed and graded separately and not just as part of a larger project like writing a research paper. However, to do this, more internal agreement is needed about what whether GenAI is allowed in these processes, and how to assess such that use of GenAI can be controlled.

Other results from the questionnaire support the need to address “academic integrity, privacy, and sustainability in relation to AI use” as is stated in the faculty assessment policy.

According to teachers, students are not proficient at using GenAI and they misunderstand the very basis of its operation and what the output is since they consider it in terms of truthfulness. They are also not aware of basic ethical issues surrounding the use of GenAI in historical work which is perhaps not surprising since this is not given consistent attention in the curriculum. Assessments also do not clearly detail when using GenAI is allowed as is evidenced by the Canvas page AI-policy not being consistently updated. The results of the questionnaire will be shared with the Education Programme Director, the department head, as well as the Vice-Dean for Education in ESHCC. Disseminating the questionnaire has already sparked discussion among colleagues about the use of GenAI in assessments with colleagues discussing ways to implement “no Gen-AI use” assessments where appropriate. The last section of the questionnaire on “GenAI use and the history curriculum” elicited many responses which called for more discussion internally, a more consistent vision and approach throughout all years, and a change in the kinds of assessments used throughout the programme. It is hope that this material will aid in coming discussions within the department and faculty.



CRITERIA FOR RUBRICS IN HIGHER EDUCATION IN THE AGE OF AI

KIM OUWEHAND

Introduction

The rapid rise of generative AI tools (such as ChatGPT, Copilot, and DALL·E) has profoundly affected assessment in higher education. Authenticity, validity, and reliability are under pressure because traditional written assignments can increasingly be completed by AI. Research shows that institutions worldwide struggle with authorship, assessment design, and policy formation; fewer than 40% of institutions have developed a formal AI governance framework (Alvira et al., 2025; Ncube et al., 2025).

AI undermines validity when assignments focus primarily on reproductive knowledge that AI can easily generate. Studies show that AI-susceptible tasks often contain low complexity or limited requirements for originality. Recent literature also argues for a shift from product-based to process-based assessment to prevent AI from replacing the learning process (Alvira et al., 2025; Ncube et al. 2026).

AI-supported or AI-generated texts make consistent scoring more difficult, especially because detection tools are often inaccurate and unreliable. Weber-Wulff (2023) highlights that AI detectors have structural limitations: they regularly produce false positives and false negatives. This increases the need for clear assessment criteria that help assessors make consistent judgments.

Authentic assessment is defined as assessment in which students perform tasks resembling those in their future professional practice. Recent studies argue that authenticity cannot be enforced through surveillance or detection; instead, it must be redesigned in contexts where AI is expected, declared, and assessed. Process transparency, reflection, and justification of choices are essential to maintain authenticity (Kickbush et al., 2025).

Programs focusing on diagnostics (Psychology, Pedagogy, Medicine) face an additional challenge as diagnostic assignments often involve (reused) cases that are AI-susceptible. Case-assignments might be relatively authentic, however repetitive use over the years facilitates misuse of AI.

It is therefore necessary to formulate assessment criteria and design rubrics to address both traditional quality requirements and AI-related risks.

AUTHENTIC ASSESSMENT IS DEFINED AS ASSESSMENT IN WHICH STUDENTS PERFORM TASKS RESEMBLING THOSE IN THEIR FUTURE PROFESSIONAL PRACTICE

Design of assignments

Authenticity, critical thinking, originality

Authentic assessment is defined as assessment in which students perform tasks resembling those in their future professional practice. Recent studies argue that authenticity cannot be enforced through surveillance or detection; instead, it must be redesigned in contexts where AI is expected, declared, and assessed. Process transparency, reflection, and justification of choices are essential to maintain authenticity (Kickbush et al., 2025)

Rubric Criteria by Assignment Type

As assignments can take multiple forms in different programs, I provide some criteria/points of attention for common types of assignments in the program of Pedagogical Sciences, of the Erasmus University Rotterdam.

Theses

Core criteria:

Research design & methodological justification

AI may support literature exploration, but students must critically evaluate sources and methods. Overreliance on AI limits students' critical thinking, autonomous authorship and problem-solving qualities (for a review see Ncube et al., 2026). Students must explicitly justify their choices, including whether and how AI tools were used.

Originality and contribution to the field

Instead of judging AI-use, educators must teach students how to use Gen-AI critically and responsibly (Kickbush et al., 2025; Van Broekhoven et al., in press) proposed that. They point out that AI reproduces information from sources rather than (re)conceptualizes to a novel and or original synthesis. Rubrics should explicitly assess originality, focusing on interpretation, conceptualization, and theoretical integration—domains where human nuance is central.

Process documentation

Literature identifies visible work processes as a strong fraud-reducing strategy (Kickbush et al., 2025). In the review of Kickbush et al. (2025) it is also addressed that educators should reconceptualize assessment or how they frame it "evidence of learning" is. Authors propose to shift focus main from assessment of end products to assessments of processes and products together. Examples mentioned are keeping logs, drafts, feedback cycles, and progress notes. Because Gen-AI can generate smooth end products without effort and higher order thinking and learning skills of students, the process pathway from the problem state to the desired end state should be emphasis of assessment (Corbin et al., 2025). This requires also that assessment should be approached more as an interactive process between learner and educator instead of a one-moment evaluation (Connolly, 2023).

Other types of written assignments

Core criteria:

Quality of argumentation

Studies show essays are increasingly vulnerable to AI-generated content and call for assessment of thinking processes rather than end products (Ncube et al., 2026).

Therefore, it is important to assess logical reasoning, depth, and originality; AI texts often contain generic reasoning. Students should be able to evaluate the value of the sources used in relation to the goal of the assignment (e.g. how do specific empirical findings support your argumentation, or how do they challenge your argumentation). Similarly, the problem arises with re-used cases for diagnostic reports in which the AI can be asked to decide on a diagnosis (e.g. based on the DSM V) and to paraphrase a report of another student to prevent plagiarism. In this case it is advisable to add a reflection and process report as well. In these reports students should describe the steps made in the diagnostic process and underscore the aspects they used to make decisions. For practice reports that describe role play and assessment situations, students should also reflect on the procedure, professional behavior, and personal learning goals (strengths and weaknesses).

Use of sources & critical thinking

AI frequently fabricates or misuses references.

Students must use verifiable, relevant sources and justify their selections.

Digital literacy is essential for academic integrity (Ncube et al., 2026).

Reflection on AI use

Rubrics should encourage students to report on how AI supported their work.

Reflective artefacts are shown to serve as evidence of learning and integrity (Kickbush et al., 2025).

Design Assignments

Core criteria:

Design process & iterative development

Logs, sketches, prototypes, and decision documents help verify human creativity.

Multistage evaluation is recommended to limit AI misuse (Ncube et al., 2026). Short (online) meetings might suffice. Two specific moments are crucial to monitor 1) the brainstorming phase to get insight in the way the students thinks, which ideas they generate, and 2) decision making (which idea is selected for further processing).

Creativity & problemsolving

AI often produces standard solutions; rubrics should reward original problem analysis.

Research on authentic assessment highlights creativity as an AI-resistant competence (Alvira et al. 2025). A good idea here might be to collaborate with field partners (organizations that are in need of a product or a solution that falls in the criteria of the assignment). Often problems by organizations are more ill-defined than examples teachers might design, which is harder for AI to tackle.

In addition, I suggest that real-life problems develop/change over time and the problem exploration and definition of a human problem solver guides the interpretation to a unique perspective on the problem, rather than a collection of generic facts on topics related to the problem. This development can only be monitored by the designer, and thus for a sharp tailor-made solution, AI will not surpass a human problem solver/designer.

Verifiable design rationale

Students must justify decisions based on user needs, context analysis, or data.

Presentations and Oral Exams

Core criteria:

Oral defense of choices

AI can generate products, but authentic oral defense remains hard to fake.

Oral exams are widely recognized as effective against AI-based cheating (Ncube et al., 2026).

Depth and flexibility in responses

Rubrics should assess realtime reasoning and adaptability.

Use of multimodal materials

AI may assist with visualizations, but assessment focuses on interpretation and justification.

Criteria & Review Tips for Identifying AI-Related Fraud

Evidence-Based AI Fraud Indicators

Overly general or superficial reasoning

AI output is often generic and lacks context-specific nuance (Ncube et al., 2026).

Incorrect or fabricated references

Generative AI frequently hallucinated citations or combines names into nonexistent references (Ncube et al., 2026).

Stylistic homogeneity without variation

AI texts tend to be overly polished, consistent, and free of natural imperfections. Human writing contains variation and idiosyncrasies.

Conceptual inconsistencies

AI sometimes misinterprets theoretical concepts or contradicts itself (Kickbush et al., 2025).

Lack of personal or contextual detail

AI cannot reproduce practice-based insights or personal experiences expected in authentic assignments.

Review Tips for Instructors

Tip 1 — Request process materials

Logs, drafts, prompt histories, etc., help verify authorship (Kickbush et al., 2025).

Tip 2 — Use brief oral checks

Short oral defenses validate authorship effectively (Ncube et al., 2026).

Tip 3 — Verify sources

Check whether cited sources exist and were used correctly (Ncube et al., 2026).

Tip 4 — Inspect reasoning quality

Look for depth, causal logic, and contextual understanding; lack thereof can suggest AI involvement.

Tip 5 — Use AI detection tools only as supportive evidence

Detection tools are unreliable and should never be used as standalone proof.

Summary Guidelines for Policy Implementation

- Use rubrics that give equal weight to process and product.
- Multiple studies highlight the importance of visible thinking and iteration (Kickbush et al., 2025).
- Embed complexity, creativity, and critical thinking in every assignment.
- These reduce susceptibility to AI-generated responses (Alvira et al., 2025).
- Require transparency regarding AI use.
- Students must describe which tools they used, how, and why.
- Invest in staff development on AI ethics and digital literacy.
- Institutions should adopt AI policies and literacy training (Ncube et al., 2026).
- Include oral or multimodal checkpoints for verification.
- Widely recognized as effective antifraud measures (Ncube et al., 2026).
- Redesign authenticity by integrating AI into assessment.
- Authenticity arises through responsible AI integration rather than prohibition (Kickbush et al., 2025).

Conclusion

Rubrics in higher education must be reimagined for an era in which generative AI reshapes how students learn, write, design, and present. The literature makes clear that detection alone is ineffective; instead, assessment must prioritize process, reflection, authenticity, ethical AI use, and complex learning outcomes.

Assessment should move from reproduction to reasoning, from product to process, and from passive to active knowledge construction.

This policy provides guidance for institutions seeking to build AI-resilient, valid, reliable, and ethically grounded rubrics—supporting academic integrity and educational innovation.

References

- Brufau Alvira, N., Bannister, P., & Santamaría Urbieta, A. (2025). Validating the PANDORA GenAI Susceptibility Rubric for Higher Education Assessment: A field test of all Translation and Interpreting BA assignments. *Higher Education Quarterly*, 79(4), e70056. <https://doi.org/10.1111/hequ.70056>
- Kickbusch, S., Ashford-Rowe, K., Kemp, A., Boreland, J., & Huijser, H. (2025). Beyond Detection: Redesigning Authentic Assessment in an AI-Mediated World. *Education Sciences*, 15(11), 1537. <https://doi.org/10.3390/educsci15111537>
- Ncube, P.D.N., Dzvapatsva, G.P., Matobobo, C. & Ranga, M. M. (2026). Redefining student assessment in AI-infused learning environments: a systematic review of challenges and strategies for academic integrity. *AI Ethics* 6, 68. <https://doi.org/10.1007/s43681-025-00871-w>
- van Broekhoven, K., Ouwehand, K., de Koning, B. B., Nederhand, M. L., & Bocanegra, B. R. (in press). Human-AI Interaction in Higher Education: A Multidisciplinary Cognitive Perspective on LLM-Support in Structured and Exploratory Learning. In D. Robinson (Ed) *The Handbook of Generative AI in Education: Integrating Research into Practice*.



WERKEN AAN EEN TOETSPROGRAMMA VOOR DE TOEKOMST

Erasmus
School of
Economics



PLEUNI DE KOCK

Visie op het onderwijs

De Erasmus School of Economics (ESE) bevindt zich midden in een herziening van haar bacheloropleidingen. Het doel is om de inhoudelijke en onderwijskundige samenhang binnen de curricula te versterken, waarmee de opbouw sterker en transparanter wordt voor studenten. Een onderdeel van de curriculumherziening is de introductie van een nieuwe toetsvisie. Deze visie markeert een verschuiving van traditioneel, vakspecifiek en summatief toetsen naar een programmatische benadering waarin toetsing het leren van studenten ondersteunt. De kernprincipes van de nieuwe toetsvisie zijn:

- Een evenwichtige verhouding tussen assessment of learning en assessment for learning
- Variatie in toetsvormen, passend bij de leerdoelen
- Structurele integratie van feedback
- Afstemming van toetsing op leerlijnen in plaats van alleen op afzonderlijke vakken

Daarnaast vormt de opkomst van generatieve AI een extra reden om toetsing te heroverwegen. AI brengt risico's met zich mee voor betrouwbaarheid en validiteit van toetsen. Het toekomstige toetsprogramma vraagt daarom om een aanpak waarin helder wordt waar AI een bedreiging is in toetsing en waar AI op een verantwoorde manier kan worden ingezet zonder de validiteit van toetsing of het behalen van eindtermen te ondermijnen. Deze uitdagingen vragen om programmabrede ontwerpkeuzes, waaronder een aanpak zoals de Two-lane approach, die in het ontwerp verder wordt toegelicht.

Ontwerp van programmatische samenhang met leerlijnen als ruggengraat

Het curriculumherontwerp wordt ontwikkeld binnen een project op basis van drie fasen: de Orientation Phase, de Design Phase en de Development Phase. Op dit moment bevindt het project zich in de tweede fase, de Design Phase.

In de Orientation Phase zijn de conceptuele fundamenten gelegd voor de nieuwe bacheloropleidingen:

- Eindtermen van de opleidingen zijn geactualiseerd
- Kernconcepten van het curriculum zijn herijkt
- Leerlijnen zijn gedefinieerd en in kaart gebracht

Deze leerlijnen vormen nu de kapstok voor het gehele programma. Ze beschrijven de ontwikkeling van vaardigheden en kennis over de jaren heen en maken de leerprogressie transparant. Hierdoor ontstaat betere constructive alignment tussen leerdoelen, leeractiviteiten en toetsing op programmaniveau (Biggs, 1996).

In de huidige Design Phase werkt academische staf gezamenlijk aan de invulling van de leerlijnen en de bredere didactische keuzes van de nieuwe bacheloropleidingen, waaronder het toetsprogramma. Om het toetsprogramma aan te laten sluiten bij de leerlijnen van de opleidingen komt er ruimte voor meer variatie, spreiding en opbouw in de toetsing. Binnen leerlijnen zal kennis en vaardigheden namelijk over een langere periode opgebouwd worden, waardoor toetsing van deze kennis en vaardigheden meer gespreid en doelgericht kan worden ingericht. Dit vraagt wel van de opleidingen om naar bredere onderliggende constructen te kijken, zodat de leerlijn logisch opbouwt binnen de vakken. Naast het beter plannen van de toetsen, stelt het centraal stellen van onderliggende constructen studenten in staat om betekenisvol te leren, wanneer zij nieuwe concepten te koppelen aan hun bestaande kennis, oftewel deep learning (Biggs & Tang, 2007).

Voor het toetsprogramma betekent dit dat er een balans gevonden moet worden tussen assessment of learning van de aangeleerde kennis binnen het vak en assessment for learning, waarbij studenten kunnen oefenen en hun kennis over de breedte van leerlijnen kunnen opbouwen (Black & William, 2009). Hiervoor is het belangrijk dat feedback een structurele rol krijgt binnen de opbouw van de leerlijnen. Door studenten te voorzien van feedback op de onderliggende constructen binnen leerlijnen, weten studenten beter waar zij staan en is transfer naar nieuwe kennis en vaardigheden mogelijk.

Een nieuw toetsprogramma op basis van leerlijnen ontwikkelen, kan ook helpen de bedreigingen van AI op toetsing te adresseren. Hierbij kan een Two-lane approach gebruikt worden (University of Sydney, Bridgeman et al. 2026), om docenten realistische handvatten te geven hoe om te gaan met AI in toetsing. Deze benadering gaat ervan uit dat er onderscheid gemaakt wordt tussen toetsing die in een open lane plaats kunnen vinden, omdat deze toetsing gericht is op assessment for learning. Hierbij wordt de invloed van AI niet uitgesloten, en deze toetsen dienen om studenten informatie te geven waar zij staan in hun leerproces. Dit staat tegenover de toetsvormen die worden afgenomen in de secure lane, waarin wordt getoetst zonder invloed van AI waarbij studenten hun kennis en kunde demonstreren voor een beoordeling (assessment of learning).

Aanpak om tot verandering te komen

In de eerdere fase, de Orientation Phase, waren al een aantal uitgangspunten geformuleerd waarop de nieuwe curricula worden gebaseerd. Tijdens de Design Phase werd echter duidelijk dat het ontwikkelen van een toekomstbestendig toetsprogramma niet alleen om het adresseren van uitgangspunten gaat, maar om een gedragen beeld te vormen over wat dat betekent voor het onderwijs.

Binnen het project hebben we gewerkt in leerlijnteams en in thematische groepen om de didactische concepten door middel van co-creatie met de academische staf vorm te geven. Voor toetsing is ook een themagroep ingesteld. De verschillende themagroepen delen hun opbrengsten met elkaar, waarin gezamenlijk naar een gedeeld begrip wordt gewerkt.

In de themagroep Toetsing is gewerkt aan een vertaling van de toetsvisie in specifieke onderwerpen, die aan bod moeten komen in de nieuwe curricula. De onderwerpen zijn in gezamenlijkheid bepaald en zijn: Assessment on Learning Trajectory level, Role of Feedback and Assessment for Learning, Mix of Assessment Methods en Two-lane approach.

Alle themagroepsleden hebben zich verdiept in deze onderwerpen. In meerdere bijeenkomsten is gekeken wat dit betekent voor de nieuwe curricula en is dit gedeeld met de andere groepen uit het project. Hierbij is veel input opgehaald wat de academische staf belangrijk vindt en kan helpen bij het uitwerken van deze onderwerpen binnen de toetsing van hun eigen vak. Dit zijn bijvoorbeeld achtergrondinformatie wat een onderwerp precies betekent, waarom en wat noodzakelijk is om te integreren in vakken, wat best practices zijn en waar extra informatie te vinden is.

Het is cruciaal gebleken dat de academische staf zelf actief heeft bijgedragen aan het formuleren van de didactische uitgangspunten en de onderwijshoud, om het draagvlak en eigenaarschap te vergroten. Het gevolg is dat de academische staf elkaar zelf kan informeren over gemaakte keuzes.

Door deze aanpak met de themagroepen heeft dit ertoe geleid dat in deze fase belangrijke procesuitkomsten zijn bereikt:

- er is een gedeelde taal ontstaan rond toetsing en leerlijnen
- kaders voor ontwerpprincipes zijn aangescherpt en breed besproken
- docenten denken meer programmatisch dan voorheen
- de basis is gelegd voor verdere implementatie in volgende fasen

Deze procesuitkomsten zijn cruciaal, want zonder dit fundament kan toetsvernieuwing niet duurzaam worden doorgezet.

Handvatten voor concrete ondersteuning voor docenten

In de volgende fase van de curriculumherziening, Development Phase, zullen docenten concreet aan de slag gaan met het ontwerp van hun vak. Om docenten te helpen bij het toepassen van de toetsvisie zal de themagroep een reeks one-pagers gaan schrijven. Hierin kunnen we kort en toegepast informatie geven over de onderwerpen, de formele kaders en uitgangspunten, praktische suggesties voor de onderwijspraktijk en best practices. De opzet van deze korte documenten biedt docenten op een overzichtelijke, laagdrempelige en direct toepasbare manier inzicht in wat van hen verwacht wordt bij het ontwerpen van vakken. Dit zal bijdragen aan consistentie in het ontwerp van de vakken.

Deze one-pagers zullen binnen de thematische groepen worden gemaakt, en dus direct afgestemd worden met de academische staf. Dit verstevigt het draagvlak verder en verankert daarmee het gedeelde begrip voor de ontwikkeling van de programmatische benadering van de curricula en toetsing. Daarmee one-pagers vormen dus een belangrijk instrument binnen de bredere implementatieaanpak van de didactische concepten binnen het curriculum.

Hoewel de curriculumherziening nog volop gaande is, verwachten we met deze ontwikkelingen een belangrijke stap richting toekomstbestendig onderwijs te zetten. Toetsing krijgt een positie waarin het niet alleen beoordeelt, maar structureel bijdraagt aan het leren en de ontwikkeling van studenten.

Literatuur

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364. <https://doi.org/10.1007/BF00138871>

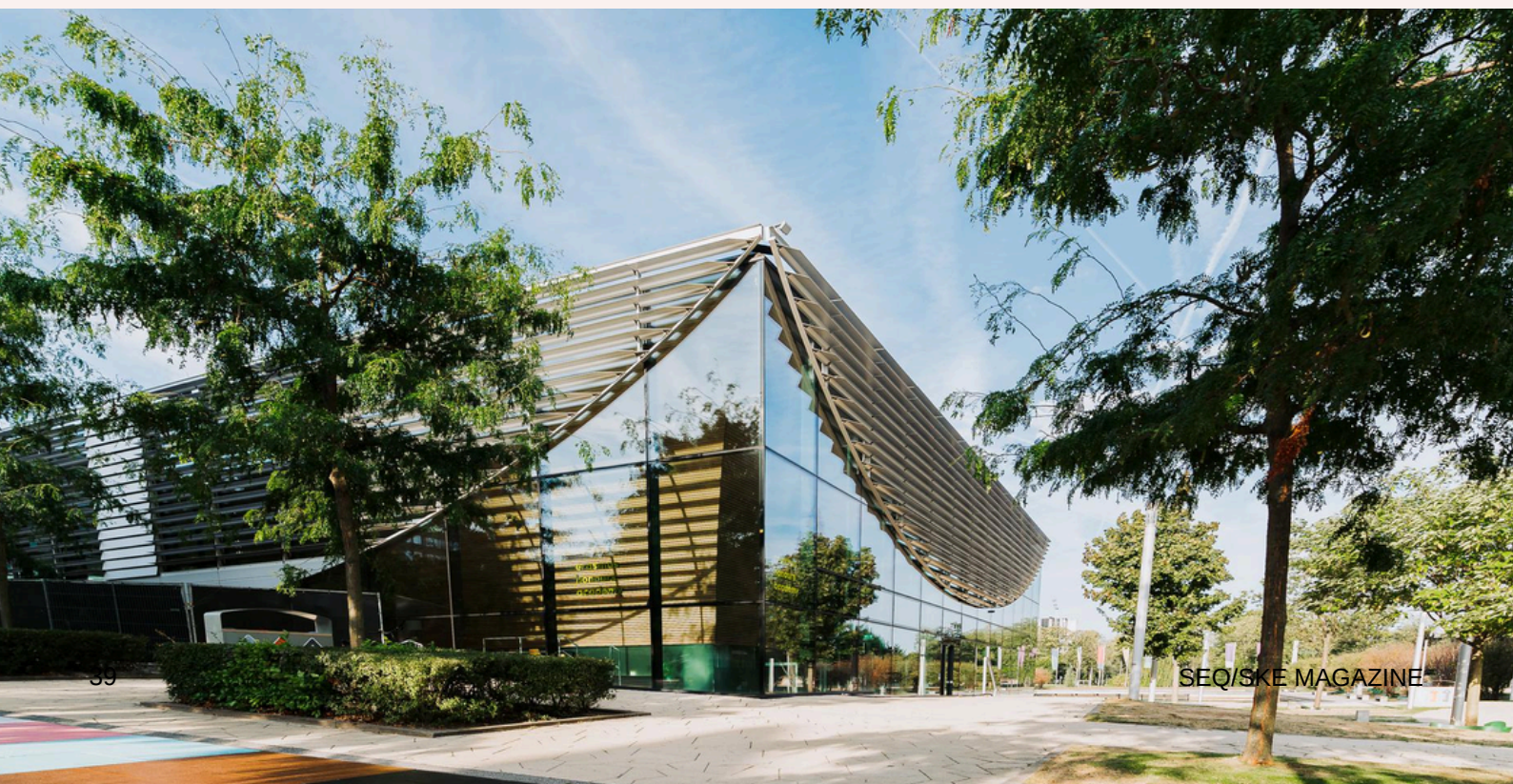
Biggs, J., & Tang, C. (2007). *Using Constructive Alignment in Outcomes-Based Teaching and Learning Teaching for Quality Learning at University*, pp. 50-63. Maidenhead: Open University Press.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31. <https://doi.org/10.1007/s11092-008-9068-5>

Bridgeman, A., Weeks, R., & Liu, D. (2024, 26 november). *Aligning our assessment to the age of generative AI*. The University of Sydney. Geraadpleegd op 16 februari 2026, van [Aligning our assessments to the age of generative AI – Teaching@Sydney](#).

Testimonial Pleuni de Kock:

“Een grote meerwaarde van de SKE is om te leren van anderen. Het is heel zinvol om inzicht te krijgen hoe er met toetsing wordt omgegaan bij andere instellingen en faculteiten. Dit heeft mijn blik op mijn eigen praktijk zeker verrijkt.”



VAN PAPIEREN VISIE NAAR GEDEELDE PRAKTIJK



Hoe programmabrede dialoog de kwaliteit van toetsing kan versterken

NINA VAN SANTVOORT

Hoe programmabrede dialoog de kwaliteit van toetsing kan versterken

In veel onderwijsprogramma's is de onderwijsvisie expliciet uitgewerkt in beleidsdocumenten. Deze visie omvat doorgaans heldere leeruitkomsten, aandacht voor curriculumcoherentie en een doordachte positionering van toetsing binnen het leerproces. Ook binnen het hoger politieonderwijs is een dergelijke visie aanwezig. De bacheloropleiding voor politie- en wijkagenten is gebaseerd op principes als integratief leren, beroepsauthentieke examinering en reflectie.

Desondanks blijkt dat deze visie niet vanzelfsprekend doorwerkt in het toetskundig handelen van docenten en examinatoren. Gesprekken met betrokkenen laten zien dat docenten beperkte samenhang ervaren tussen modules, weinig inzicht hebben in de longitudinale ontwikkeling van leeruitkomsten en onvoldoende mogelijkheden zien voor onderlinge afstemming.

Dit leidt tot de centrale vraag hoe een op papier coherente onderwijsvisie daadwerkelijk kan worden verankerd in de onderwijs- en toetspraktijk.

In dit quality enhancement project is onderzocht hoe de discrepantie tussen beleid en praktijk kan worden verkleind.

De onderzoeksvraag luidde:

Welke interventies of strategieën zijn effectief om de onderwijsvisie van de Politieacademie programmabreed te verankeren in het toetskundig handelen van docenten, en hoe draagt dit bij aan hun assessment literacy?

Projectaanpak

Het onderzoek is uitgevoerd op basis van een combinatie van documentanalyse, semigestructureerde interviews en literatuurstudie.

Allereerst zijn beleidsdocumenten, waaronder het examenbeleid, de onderwijsvisie en de studiegids geanalyseerd om inzicht te verkrijgen in de formele uitgangspunten van toetsing. Hieruit blijkt dat het programma nadrukkelijk inzet op samenhang, authenticiteit en programmabreed toetsen. Instrumenten zoals de dekkingsmatrix maken inzichtelijk waar en hoe leeruitkomsten binnen het curriculum worden getoetst.

Vervolgens zijn interviews afgenomen met docenten en examinatoren. Deze boden inzicht in de feitelijke uitvoeringspraktijk en de ervaren knelpunten.

Daarnaast is relevante literatuur bestudeerd met betrekking tot assessment literacy en het belang van gedeelde visievorming. Deze literatuur benadrukt dat assessment literacy niet uitsluitend betrekking heeft op kennis van toetsinstrumenten, maar vooral op het vermogen om toetsing doelgericht en onderbouwd in te zetten als onderdeel van het leerproces. Dit vermogen ontwikkelt zich in sociale interactie en binnen een specifieke organisatorische context.

De combinatie van deze methoden resulteerde in een geïntegreerd beeld van zowel de beoogde als de gerealiseerde toetspraktijk.

Wat gaat er mis in de praktijk

De analyse laat een duidelijke spanning zien tussen formele uitgangspunten en feitelijke uitvoering. Hoewel het toetsprogramma op papier zorgvuldig is opgebouwd, met herhaalde en toenemende complexiteit van leeruitkomsten, ervaren docenten deze samenhang in de praktijk beperkt. Zij geven aan weinig zicht te hebben op de positionering van hun eigen module binnen het bredere programma. De dekkingsmatrix functioneert daarbij voornamelijk als ontwerpinstrument en minder als een operationeel hulpmiddel in de uitvoeringsfase. Daarnaast blijkt dat toetsbeslissingen vaak worden gebaseerd op bestaande formats en rubrics, zonder expliciete koppeling aan de onderliggende onderwijsvisie. Dit wijst op een procedurele benadering van toetsing, waarin het instrumentele aspect centraal staat en de pedagogische onderbouwing minder expliciet is. Contextuele factoren, zoals werkdruk, organisatorische complexiteit en beperkte tijd voor overleg, versterken deze tendens. Hierdoor krijgt praktische uitvoerbaarheid prioriteit boven inhoudelijke afstemming. Het gevolg is dat toetsing grotendeels een individuele activiteit blijft, terwijl de literatuur juist wijst op het belang van collectieve afstemming voor de borging van kwaliteit.

De rol van assessment literacy

De bevindingen sluiten aan bij recente theoretische inzichten waarin assessment literacy wordt opgevat als een sociaal en contextgebonden fenomeen. De ontwikkeling van toetskundige expertise vindt niet uitsluitend individueel plaats, maar in interactie met collega's en binnen een gedeeld referentiekader. Bij het ontbreken van een dergelijke gedeelde basis blijft assessment literacy beperkt tot het correct toepassen van instrumenten. Toetsing krijgt dan het karakter van een technische handeling, in plaats van een doelbewuste bijdrage aan leren. Een gedeelde visie fungeert in dit proces als verbindend kader. Zij maakt het mogelijk om individuele praktijken te relateren aan programmadoelen en bevordert gezamenlijke oordeelsvorming. De literatuur onderstreept dat de combinatie van gedeelde visie en professionele dialoog essentieel is voor duurzame kwaliteitsontwikkeling.

Interventies

Op basis van de analyse zijn twee interventies ontworpen die gericht zijn op het versterken van programmabrede afstemming en professionele dialoog.

1. Programma-visualisatiesessies

Deze interventie bestaat uit interactieve werksessies waarin docenten gezamenlijk het toetsprogramma expliciteren. Daarbij wordt inzichtelijk gemaakt:

- welke leeruitkomsten waar in het curriculum aan bod komen,
- hoe deze zich in de tijd ontwikkelen,
- en op welke wijze zij worden getoetst.

Deze gezamenlijke visualisatie bevordert inzicht in samenhang, overlap en opbouw, en maakt het mogelijk om de eigen bijdrage te positioneren binnen het geheel.

Interventies

2. Verankering in bestaande overlegstructuren

De tweede interventie richt zich op het doelgericht benutten van bestaande overlegmomenten, zoals onderwijsdagen en professionaliseringsbijeenkomsten. Deze worden ingezet voor concrete afstemming, waaronder:

- het bespreken van beoordelingscriteria,
- het vergelijken van interpretaties,
- en het expliciteren van onderliggende aannames.

Hierdoor verschuift de focus van individuele naar collectieve beoordeling en wordt gezamenlijke betekenisgeving gestimuleerd.

Verwachte opbrengsten

Hoewel de interventies zich nog in de ontwerpfase bevinden, wijzen de analyses op meerdere potentiële opbrengsten:

- versterkte curriculumcoherentie door beter inzicht in de opbouw van leeruitkomsten;
- ontwikkeling van assessment literacy via gedeelde reflectie en referentiekaders;
- verbetering van feedbackkwaliteit door meer programmabreed perspectief;
- versterking van de kwaliteitscultuur, waarbij toetsing als gezamenlijke verantwoordelijkheid wordt opgevat.

Implicaties

De resultaten zijn niet uitsluitend relevant voor het politieonderwijs, maar raken aan bredere vraagstukken binnen het hoger onderwijs. Veel programma's staan voor vergelijkbare uitdagingen met betrekking tot de implementatie van beleid, het realiseren van samenhang en het ontwikkelen van assessment literacy.

De bevindingen suggereren dat verbetering niet primair ligt in het ontwikkelen van nieuw beleid of aanvullende instrumenten, maar in het herontwerpen van bestaande interactiemomenten. Het faciliteren van professionele dialoog en gezamenlijke analyse vormt daarbij een cruciale voorwaarde voor het realiseren van een gedeelde praktijk.

Lessons learned

Het project levert tevens enkele generieke inzichten op:

1. Een expliciete visie is noodzakelijk, maar onvoldoende voor implementatie.
2. Professionele dialoog vereist structurele inbedding en facilitering.
3. Contextuele factoren beïnvloeden in sterke mate de uitvoerbaarheid van interventies.
4. Duurzame verandering vraagt om actieve betrokkenheid en eigenaarschap van docenten.

PROGRAMMATISCH TOETSEN VAN KENNIS IN HET BACHELOR GENEESKUNDECURRICULUM EA2030

IRIS VAN DE GRAAF



Vertrekpunt - achtergrond en opzet EA2030

De geneeskundeopleiding van het Erasmus MC wil studenten opleiden tot breed academisch gevormde artsen die voorbereid zijn op het beargumenteerd inzetten van technologische en maatschappelijke ontwikkelingen en die maatschappelijk betrokken zijn. Hiervoor is het profiel van de Erasmusarts 2030 (EA2030) ontwikkeld.

Toetsing binnen EA2030 ondersteunt de ontwikkeling tot deze nieuwe generatie zorgprofessionals en is gericht op het bevorderen en faciliteren van leren, zoals vastgelegd in de onderwijsvisie en het toetsplan. Het toetsconcept is gebaseerd op programmatisch toetsen (Van der Vleuten et al., 2012). Uit de principes van programmatisch toetsen zijn vier kernprincipes gedestilleerd die de basis vormen voor het toetsprogramma van de bachelor:

1. Toetsen binnen EA2030 nodigen studenten uit tot leren.
2. Studenten werken aan groei met behulp van ontwikkelingsgerichte feedback en reflectie.
3. Studenten monitoren zelf hun kennisgroei en competentie-ontwikkeling en sturen deze bij.
4. Robuuste voortgangsbesluiten worden genomen op basis van een steeds rijkere mix aan datapunten.

Het curriculum EA2030 is gestart in studiejaar 2024/2025. Bij het opleveren van de OER (maart, 2024) was nog onvoldoende inzicht in kennisdatapunten over de jaren heen. Hierom is besloten de kennistoetsing nog niet programmatisch in te richten.

De huidige vorm van toetsing is beschreven in het SKE-project (2023) 'Bloктоets van ontwikkeling naar borging'. Binnen EA2030 wordt gewerkt met bloктоetsen die kennisdomeinen toetsen met behulp van casussen. Dit betekent dat:

- Elke toetsvraag is gekoppeld aan een kennisdomein.
- Studenten hun beheersingspercentage zien per bloктоets (afbeelding 1).
- Studenten ook hun beheersingspercentage zien per kennisdomein (afbeelding 2).
- Studenten, coaches en de besliscommissie deze gegevens gebruiken voor inzicht en advies; besluiten worden uitsluitend gebaseerd op bloктоetsresultaten.
- Er daarnaast een voortgangstoets is die (vier keer per jaar) inzicht geeft in ontwikkeling binnen kennisdomeinen, maar dat deze niet wordt gebruikt voor besluitvorming.

Dit betekent dat:

- Elke toetsvraag is gekoppeld aan een kennisdomein.
- Studenten hun beheersingspercentage zien per bloктоets (afbeelding 1).
- Studenten ook hun beheersingspercentage zien per kennisdomein (afbeelding 2).
- Studenten, coaches en de besliscommissie deze gegevens gebruiken voor inzicht en advies; besluiten worden uitsluitend gebaseerd op bloктоetsresultaten.
- Er daarnaast een voortgangstoets is die (vier keer per jaar) inzicht geeft in ontwikkeling binnen kennisdomeinen, maar dat deze niet wordt gebruikt voor besluitvorming.



Afbeelding 1



Afbeelding 2

Op de route - programmatisch toetsen, de theorie

Het gekozen onderwijsconcept is programmatisch toetsen. Het doel van dit concept is om toetsinformatie te gebruiken om zowel de betrouwbaarheid van beslissingen over studenten te optimaliseren, als om het leerproces van studenten te stimuleren (Baartman et al., 2020; Van der Vleuten et al., 2012). Bij programmatisch toetsen vormen de leeruitkomsten de ruggengraat van het toetsprogramma. De datapunten, waarin informatierijke feedback een belangrijke rol speelt, leveren informatie over waar een student zich in diens ontwikkeling bevindt. Het wel of niet behalen van een losse toets of datapunten binnen die toets leidt niet tot een zak-slaag beslissing (Van Schilt-Mol, 2021). Een voortgangsbesluit wordt genomen op basis van datapunten die de student tot het beslismoment heeft verzameld per leeruitkomst (aggregatie). Door informatie uit verschillende datapunten te combineren (triangulatie) en te zorgen voor voldoende informatie om het besluit te rechtvaardigen (saturatie), kan een betrouwbare en robuuste beslissing worden genomen. Deze beslissing berust niet op één enkele evaluatie, maar op meerdere datapunten die gezamenlijk leiden tot een gedeelde beslissing (intersubjectiviteit) (Baartman et al., 2020, zoals beschreven in Toetsplan Bachelor Geneeskunde, 2025). Programmatisch toetsen is theoretisch onderbouwd en internationaal uitgewerkt in het Ottawa Statement, waarin twaalf principes voor toetsing zijn geformuleerd (Heeneman et al., 2021). In Nederland zijn deze uitgangspunten vertaald naar zeven ontwerpprincipes (Tabel 1, linkerkolom) voor de praktijk (van Schilt-Mol, Baartman, & de Jong, 2025). Deze principes geven richting aan de inrichting van toetsprogramma's, zonder een vaste vorm voor te schrijven. Opleidingen kunnen hierin ontwerpkeuzes maken, zolang de onderliggende programmatische principes worden gevolgd (Agricola et al., 2024).



Zijpad - kennis en competenties

Een van de elf ontwerpkeuzes is dat “kennis wel of niet als datapunt wordt geprogrammeerd” (Agricola et al., 2024). In het huidige curriculum van EA2030 worden kennis en competenties grotendeels gescheiden aangeboden en gewaardeerd. Hoewel kennis een geïntegreerd onderdeel vormt van competenties, wordt de verwerving ervan vaak afzonderlijk getoetst. Momenteel gebeurt het toetsen van kennis met behulp van bloктоetsen en een voortgangstoets. In een curriculum dat is gebaseerd op het Raamplan artsopleiding 2020, waarin competenties centraal staan, is het relevant om te overwegen of deze scheiding bijdraagt aan het beoogde leerproces. Het doel is immers om artsen op te leiden die competenties en medische kennis integraal kunnen toepassen en die duurzaam en breed inzetbaar zijn (NFU, 2020).

Van Eenbergen en van Ierland (2025) hebben onderzocht hoe kennis beter kan worden geïntegreerd in een curriculum waarin programmatisch wordt getoetst. Datapunten rondom kennis zouden niet alleen moeten dienen om voortgang te monitoren, maar vooral als leermomenten waarbij studenten gerichte feedback krijgen om hun leerproces te verdiepen. Daarnaast kan kennis sterker worden geïntegreerd in praktijkgerichte situaties, zodat de relevantie en toepasbaarheid duidelijker worden. Dit wordt samengevat in drie adviezen.

- Zorg voor voldoende herhaling en integratie van kennis door regelmatig terug te komen op eerdere onderwerpen, zodat studenten hun begrip kunnen verdiepen en versterken.
- Versterk de koppeling tussen theorie en praktijk door opdrachten en projecten waarin studenten theoretische concepten toepassen in realistische situaties.
- Evalueer het toetsprogramma regelmatig en pas het waar nodig aan op basis van input van zowel studenten als docenten.

Hieruit volgt de eerste aanbeveling:

Analyseer hoe kennis en competenties meer geïntegreerd en betekenisvol kunnen worden getoetst, zodat de toetsing aansluit bij de leerresultaten, praktijk en de ontwikkeling van competente en duurzame artsen.

Overweeg daarbij of een aparte eindterm voor kennis nodig blijft.



Terug op de route - kennis programmatisch toetsen

Om te onderzoeken in welke mate de opleiding momenteel voldoet aan de principes van programmatisch toetsen en welke knelpunten nog bestaan, is een analyse uitgevoerd op basis van de zeven ontwerpprincipes (van Schilt-Mol, Baartman, & de Jong, 2025). Voor elk principe worden de huidige stand van zaken en de bijbehorende knelpunten beschreven. De input voor deze analyse is verzameld tijdens een overleg met de longitudinale kennisdomeincoördinatoren in december 2025 en een toetsgroepvergadering in 2026. Aan deze bijeenkomsten namen de opleidingsdirectie, leden van de examencommissie, docenten en onderwijskundigen deel.

Op basis van deze gesprekken, aangevuld met inzichten uit de literatuur, relevante documenten zoals het toetsplan, input vanuit het leernetwerk programmatisch toetsen en een gesprek met de visitatiecommissie, zijn knelpunten geïdentificeerd die aangepakt moeten worden om aan elk principe te voldoen. Onderstaande tabel geeft per principe de huidige situatie en de bijbehorende knelpunten weer. De tabel richt zich volledig op de rol van kennis binnen de bachelor EA2030.

Tabel 1.

Principe	Huidige situatie (EA2030)	Knelpunt
1) Leerresultaten zijn de basis voor het verzamelen en aggregeren van informatie uit datapunten.	Kennis is samengevat in één eindterm. Datapunten/bloktoetsvragen zijn wel gekoppeld aan een kennisdomein, maar het 'bloktoetsbeheersingsniveau' wordt gebruikt voor besluitvorming. Voortgangstoetsresultaten worden niet meegenomen.	De koppeling tussen de kennisdatapunten en de leerresultaten ontbreekt.
2) Informatie wordt verzameld door middel van een bewust gekozen en beargumenteerde mix van datapunten en omvat feedback die daarop wordt gegeven vanuit verschillende perspectieven.	Datapunten worden zorgvuldig door de blokcoördinator gekozen. Feedback bestaat uit de beheersingspercentages per toets en per kennisdomein. Feedback vindt achteraf plaats en is generiek.	De mix van datapunten is niet systematisch op opleidingsniveau geborgd. Feedback is weinig persoonlijk en/of weinig ontwikkelingsgericht.
3) Informatie staat ten dienste van (het sturen van) de ontwikkeling van de student op deze leerresultaten én het nemen van een onderbouwde beslissing over de beheersing van de beoogde leerresultaten.	Het dashboard en met name de percentages per kennisdomein geven inzicht in de voortgang, maar beslissingen worden vooral op 'bloktoetsbeheersingsniveau' genomen.	Toetsinformatie wordt onvoldoende ingezet om ontwikkeling te sturen of onderbouwde beslissingen te nemen. De leeruitkomsten moeten concreter worden gemaakt.
4) De student interpreteert en gebruikt, in georganiseerde dialoog, informatie uit datapunt(en) en geeft daarmee richting aan de verdere ontwikkeling van de beoogde leerresultaten.	*Idem. De georganiseerde dialoog vindt plaats, maar is vrij beperkt.	Er is onvoldoende zicht op wanneer kennisdomeinen opnieuw aan bod komen. De dialoog is beperkt waardoor de ontwikkeling van studenten niet continu wordt gestimuleerd.

Principe	Huidige situatie (EA2030)	Knelpunt
5) Student en opleiding monitoren doorlopend de beheersing van de beoogde leerresultaten aan de hand van geaggregeerde informatie uit datapunten.	Het dashboard toont de beheersingspercentages per toets en per kennisdomein. De koppeling met leerresultaten is niet expliciet (weergegeven).	De aggregatie en koppeling met leerresultaten ontbreekt, waardoor doorlopende monitoring van de leerresultaten niet mogelijk is.
6) Beslissingen worden niet genomen op basis van informatie afkomstig van één datapunt; saturatie van informatie is voorwaardelijk voor het nemen van een onderbouwde beslissing.	Beslissingen zijn vaak gebaseerd op individuele bloktoetsen. Er bestaan geen afspraken over het aantal benodigde datapunten per leerresultaat.	Beslissingen kunnen afhankelijk zijn van losse toetsmomenten. Er is sprake van onvoldoende saturatie voor betrouwbare besluiten.
7) Een commissie van examinatoren interpreteert de geaggregeerde informatie uit datapunten en neemt op basis daarvan gezamenlijk op een transparante en navolgbare wijze een onderbouwde beslissing over het niveau van de beheersing van de beoogde leerresultaten.	Beslissingen worden genomen door minimaal twee examinatoren en zijn navolgbaar. Beslissingen zijn gebaseerd op bloktoetspercentages in plaats van op de beheersing van de beoogde leerresultaten.	Het is onzeker of de beslissingen het werkelijke beheersingsniveau reflecteren. De expliciete koppeling tussen kennisdata en leerresultaten ontbreekt.

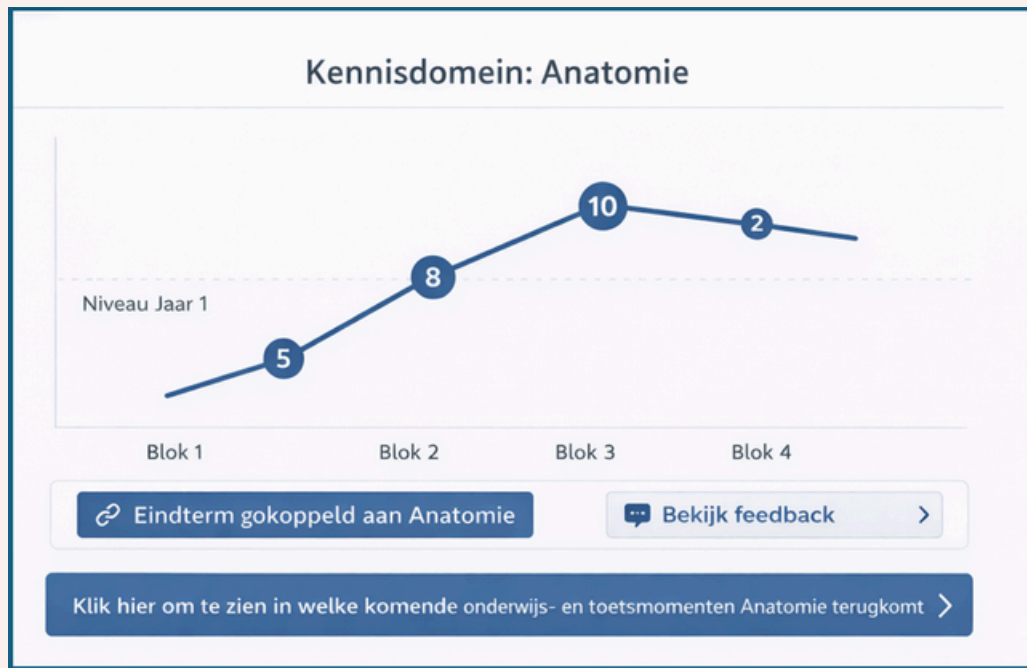


Eindpunt - aanbevelingen

Op basis van de analyse van de huidige toetspraktijk kunnen vier aanbevelingen worden geformuleerd. Ten eerste volgt uit de analyse van de verschillende principes en bijbehorende knelpunten dat moet worden vastgesteld bij welke leerresultaten (of welk leerresultaat) de kennisdatapunten horen en welke rol de al geformuleerde kennisdomeinen daarbij krijgen. Alle informatie over kennis, inclusief bijvoorbeeld de voortgangstoets, moet systematisch worden gekoppeld aan deze leerresultaten. Deze koppeling moet bovendien zichtbaar worden gemaakt in het dashboard, zodat studenten, VIP-coaches en de besliscommissie inzicht hebben in voortgang, beheersing en ontwikkeling over de tijd. Ten tweede verdient het aanbeveling om een systematische en geborgde mix van datapunten te ontwikkelen. Toetsvormen en datapunten moeten op opleidingsniveau gestructureerd worden geselecteerd, bijvoorbeeld met behulp van een toets- of datapuntenmatrijs, zodat keuzes consistent zijn en niet afhankelijk van individuele blokcoördinatoren. Ten derde is het essentieel om kennisgerichte feedback te ontwerpen en te integreren in het leerproces. Ontwikkelingsgerichte feedback over beheersing van bijvoorbeeld de kennisdomeinen moet terugkomen in de georganiseerde dialoog met de student. De feedback moet niet alleen na afloop van toetsen beschikbaar zijn, maar ook gedurende het leerproces, zodat studenten hun leren actief kunnen bijsturen. Tot slot is het van belang om ontwikkeling over tijd inzichtelijk te maken. Niet alleen het gemiddelde, maar de groei van studenten over de tijd moet zichtbaar zijn (voorbeeld afbeelding 3). Daarnaast moet duidelijk zijn welke leerresultaten in toekomstig onderwijs en toetsing aan bod komen. Zo kan de besliscommissie onderbouwde beslissingen nemen over zowel beheersing als ontwikkeling van de studenten.

Afbeelding 3

Voorbeeld van ontwikkeling op het kennisdomein anatomie over de tijd heen.



Geleerde lessen

Tijdens dit project is duidelijk geworden dat de kloof tussen theorie en praktijk van programmatisch toetsen groter is dan vooraf verwacht. Deze kloof wordt mede beïnvloed door de uiteenlopende visies en achtergronden van betrokkenen, wat leidt tot spanningen in de vertaling van het theoretisch kader naar de onderwijspraktijk. Torre et al. (2021) benadrukken dat programmatisch toetsen een cultuurverandering vraagt, waarbij traditioneel summatief 'toetsdenken' plaatsmaakt voor een meer constructivistische benadering. Succesvolle implementatie vereist volgens hen een veranderstrategie op drie niveaus: macro, meso en micro. Dit betekent dat niet alleen toetsmethoden moeten worden aangepast, maar ook de onderliggende toetscultuur. Een verschuiving naar een cultuur waarin leren en feedback centraal staan in plaats van cijfers is daarbij cruciaal. Dit vraagt om gezamenlijke inspanning van docenten, studenten, opleidingsmanagers, examencommissies en beleidsmakers om toetsing als ontwikkelinstrument te benutten. Ter afsluiting kan met de woorden van Barend Last (2023) worden gesteld dat programmatisch toetsen een complex onderwijsconcept is dat niet onderschat moet worden bij de implementatie. Het vereist een andere manier om te denken over de waarde en functie van toetsing.

Literatuur

- Agricola, B. T., De Vos, M., Baartman, L., & van Schilt-Mol, T. (2024). Naar een nieuwe balans in toetsfuncties: Van formatief en summatief naar een continuüm van beslissingen [NRO-overzichtsstudies onderzoeksrapport]. Hogeschool Utrecht. <https://www.hu.nl/onderzoek/publicaties/naar-een-nieuwe-balans-in-toetsfuncties-3>
- Baartman, L. K. J., van Schilt-Mol, T., & van der Vleuten, C. P. M. (2020). Programmatisch toetsen: Voorbeelden en ervaringen uit de praktijk. Hogeschool Utrecht.
- Erasmus MC. (z.d.). Onze visies – Erasmusarts2030. <https://www.erasmusmc.nl/nl-nl/erasmusarts2030/onze-visies>
- Heeneman, S., de Jong, L. H., Dawson, L. J., Wilkinson, T. J., Ryan, A., Tait, G. R., ... van der Vleuten, C. P. M. (2021). Ottawa 2020 consensus statement for programmatic assessment – 1: Agreement on the principles. *Medical Teacher*, 43(10), 1139–1148. <https://doi.org/10.1080/0142159X.2021.1957088>
- Nederlandse Federatie van Universitair Medische Centra. (2020). Raamplan artsopleiding 2020. https://www.nfu.nl/sites/default/files/2020-08/20.1577_Raamplan_Artsenopleiding_-_maart_2020.pdf
- Torre, D., Rice, N. E., Ryan, A., Bok, H., Dawson, L. J., Bierer, B., ... van der Vleuten, C. P. M. (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 Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–214. <https://doi.org/10.3109/0142159x.2012.652239>
- van Eenbergen, M., & van Ierland, C. (2025). De rol van kennis bij programmatisch toetsen binnen het hoger onderwijs. *Examens: Tijdschrift voor de Toetspraktijk*, 2025(1).
- van Schilt-Mol, T. (2021). Eigentijds beoordelen en beslissen: Intreerede bij de aanvaarding van het ambt van lector [Installatierede]. HAN University of Applied Sciences Press.
- van Schilt-Mol, T., Baartman, L., & De Jong, L. (2025). Herijking principes programmatisch toetsen [Presentatie]. Week van Toetsbekwaamheid. https://toetsbekwaamheid.nl/wp-content/uploads/2025/11/Tamara-van-Schilt-Mol_Herijking-principes-programmatisch-toetsen-work-in-progress-week-van-toetsbekwaamheid.pdf
- Bij de totstandkoming van dit document is AI gebruikt om teksten te optimaliseren, reflectieve vragen te formuleren en afbeeldingen te creëren.

Testimonial Iris van de Graaf- Weerd:

Het SKE-traject gaf mij de kans om een laag dieper te gaan en op een bredere, meer doordachte manier na te denken over toetsing binnen het gehele curriculum. Het traject samen volgen met een directe collega maakte de ervaring bovendien extra waardevol.



NAAR EEN TOEKOMSTBESTENDIG SCRIPTIETRAJECT IN HET AI-TIJDPERK

ERIK DE KLOE

1. Inleiding

Generatieve AI heeft de afgelopen jaren een enorme vlucht genomen. Uit een onderzoek naar 1.000 college students uit de Verenigde Staten in maart 2023 bleek dat een op de vijf studenten een AI-tool gebruikte voor het maken van opdrachten of tentamens (Welding 2023). In augustus 2025 was dat 60% (Leckrone 2025). Aan de Erasmus School of Law moeten studenten hun AI-gebruik verantwoorden door een scriptielogboek in te leveren. Het overgrote deel van de studenten laat hierin zien AI te gebruiken bij het schrijven van de scriptie. Onder docenten en in de media wordt de waarde van de scriptie als eindwerk om die reden betwijfeld (Klute, De Bruijn & Van der Wees 2026).

Vanwege de opkomst van AI is het belangrijk de scriptie tegen het licht te houden. Is het schrijven van een scriptie als eindwerk in het AI-tijdperk nog een zinvolle manier van toetsing? En als dat zo is, welke wijzigingen zijn dan nodig om het scriptietraject toekomstbestendig te maken? Dit artikel biedt een probleemanalyse (par. 2), beargumenteert waarom de scriptie als eindwerk behouden moet blijven (par. 3) en formuleert uitgangspunten voor herijking (par. 4).

2. Probleemanalyse

Generatieve AI-systemen zoals ChatGPT en Gemini (hierna: AI) veranderen de manier waarop academisch schrijven wordt vormgegeven. Dat heeft ook gevolgen voor schrijfp opdrachten die in een academische opleiding worden gegeven aan studenten, zoals het schrijven van een scriptie als eindwerk.

Studenten gebruiken AI op verschillende manieren: om ideeën te genereren, om hun gedachten om te zetten in een gestructureerd en goed geschreven stuk, om geschreven werk op structuur en taal te controleren of om output van AI (eventueel met beperkte aanpassingen) als opdracht in te leveren (Walton et al. 2026).

Docenten zijn onvoldoende in staat om onderscheid te maken tussen AI-gegenereerde tekst en tekst die door mensen is gegenereerd (Kumar & Mindzak 2024; Hostetter et al. 2025). Dit zorgt, afhankelijk van de leerdoelen, voor gebrekkige validiteit van opdrachten. Onder een valide opdracht wordt verstaan dat de uitwerking van de opdracht door een student de daadwerkelijke bekwaamheid van deze student reflecteert (Dawson et al. 2024).

Het gebruik van AI bij academisch schrijven heeft verschillende voor- en nadelen. Als voordeel kan erop worden gewezen dat het gebruik van AI de kwaliteit van de tekst kan verbeteren. Een AI kan ook op ieder moment feedback geven, wat leidt tot meer motivatie, plezier en schrijfsucces. Iedereen die de kosten van een abonnement op een goede AI-tool kan betalen, heeft toegang tot schrijfondersteuning op hoog niveau. Als belangrijk nadeel van het gebruik van AI wordt genoemd dat te veel afhankelijkheid van AI kritisch denken, creativiteit en originaliteit beperkt. Een beperking bij het gebruik van AI is dat de *AI literacy* zowel bij docenten als studenten nog vaak ontbreekt. Hierdoor worden AI-tools niet efficiënt ingezet door studenten en kunnen docenten geen feedback geven op het AI-gebruik van studenten (zie voor deze voor- en nadelen, inclusief verdere verwijzingen, Chanpradit 2025).

In het vervolg van dit artikel wordt antwoord gegeven op de vraag of het mogelijk is de validiteit van een scriptie te borgen, gegeven het feit dat studenten de mogelijkheid hebben om AI te gebruiken en wat hiervoor nodig is. In par. 3 wordt eerst de vraag beantwoord wat de toegevoegde waarde is van een scriptie als eindwerk: waarom wordt de scriptie niet volledig afgeschaft?

3. Afschaffen scriptie als eindwerk?

Zowel in de wetenschap als in de juridische praktijk is schrijfvaardigheid van cruciaal belang. Om te promoveren moet een proefschrift worden geschreven en kennisuitbreiding vindt grotendeels plaats door het publiceren van artikelen in wetenschappelijke tijdschriften. In de praktijk worden procedures voor een groot deel schriftelijk gevoerd en wordt een advies veelal (mede) vastgelegd in een schriftelijk adviesrapport. Deze schriftelijke stukken staan zelden op zichzelf. Ze vereisen ook mondelinge vaardigheden, zoals de verdediging van een proefschrift, de mondelinge behandeling van een procedure en de mondelinge toelichting van een advies. Om studenten in de studie voor te bereiden op een vervolg in de wetenschap en/of praktijk, moeten zij schriftelijke stukken op het vereiste niveau kunnen produceren en deze mondeling kunnen verdedigen.

Sommigen stellen voor de scriptie te vervangen door korte schrijfp opdrachten in een gecontroleerde, AI-vrije omgeving. Dit is echter geen goede weg vooruit. In de eerste plaats kunnen studenten na hun studie ook gebruikmaken van AI-tools als zij schriftelijke stukken moeten presenteren. Door studenten ook binnen de studie de mogelijkheid te geven gebruik te maken van AI, wordt de authenticiteit van toetsing vergroot, wat positieve effecten heeft op het leerproces (Villarreal et al. 2018). In de tweede plaats kunnen de eerdergenoemde voordelen van AI-gebruik alleen worden benut als AI is toegestaan. Ten derde ontwikkelt een mens gedachten door te schrijven (Corbin et al. 2026). Dat is een creatief proces dat niet elk moment afgedwongen kan worden en voor de ene persoon meer tijd in beslag neemt dan voor de andere persoon. Gedachten moeten ook 'rijpen'. Die mogelijkheid is er niet als uitsluitend binnen een bepaald tijdsbestek in een gecontroleerde omgeving wordt getoetst.

De uitdaging is dus om studenten de voordelen van AI te laten benutten, terwijl de nadelen worden beperkt. Eventuele wijzigingen van een scriptietraject moeten bovendien financieel haalbaar zijn en zijn daarom bij voorkeur budgetneutraal.

4. Uitgangspunten

In deze paragraaf worden op basis van de probleemanalyse enkele uitgangspunten genoemd om de validiteit van de scriptie in het huidige AI-tijdperk te borgen en de uitvoering van de begeleiding van scripties haalbaar te laten zijn/blijven.

a. Integreer mogelijkheid gebruik AI in leerdoelen

In het kader van constructive alignment moeten leerdoelen, onderwijsactiviteiten en toetsing op elkaar aansluiten (Biggs 2003). Hoewel studenten bij het schrijven van de scriptie gebruik kunnen maken van AI, wordt hier in (veel) juridische opleidingen aan de Erasmus School of Law geen rekening mee gehouden in de leerdoelen en in de rubric die wordt gebruikt bij de beoordeling van de scriptie. Een leerdoel van het scriptietraject van de master ondernemingsrecht is bijvoorbeeld dat een student na afloop van het traject in staat is zelfstandig een juridische vraagstelling te formuleren, bijbehorend onderzoek te verrichten en daarvan verslag te doen. Wordt met zelfstandig bedoeld zonder hulp van AI, dan staat de validiteit van een dergelijk leerdoel onder druk. Het begrip 'zelfstandig' kan uit het leerdoel worden verwijderd, maar het is beter om een apart leerdoel over het gebruik van AI op te nemen, bijvoorbeeld:

NA VOLTOOIING VAN HET SCRIPTIETRAJECT KAN DE STUDENT UITLEGGEN OF EN IN WELKE MATE AI IS GEBRUIKT, DE VOOR- EN NADELEN VAN DIT GEBRUIK BENOEMEN EN REFLECTEREN OP HET EIGEN AI-GEBRUIK.

b. Zorg voor AI literacy onder studenten en docenten

Om AI op een efficiënte en geschikte manier te gebruiken, is het van belang dat zowel docenten als studenten AI-geletterd zijn. Tijdens een verplichte workshop aan het begin van een scriptietraject kan worden ingegaan op de voor- en nadelen van het gebruik van AI bij het doen van juridisch onderzoek, de kunst van het prompten en de vraag welke tools waarvoor geschikt zijn (vgl. Rinnooy Kan & Salomon 2026). Docenten moeten in staat worden gesteld dergelijke workshops te verzorgen of in ieder geval in staat zijn met een student mee te denken over het AI-gebruik en de reflectie van een student hierop te beoordelen. Voor docenten moeten daarom mogelijkheden worden aangeboden om zich verder te bekwamen in het gebruik van AI in de context van academisch schrijven.

c. Stimuleer kritisch denken, creativiteit en originaliteit op een andere manier

De kern van academisch onderwijs is kritische reflectie, zowel op de eigen vooronderstellingen als op uitgangspunten en regels die gelden in de samenleving en in het recht (vgl. de Educational Vision van de EUR (<https://www.eur.nl/en/media/2023-09-education-vision-2023>)). Studenten moeten in staat zijn op basis van onder meer maatschappelijke tendensen en juridische beginselen te reflecteren op het bestaande recht en/of na te denken over de ontwikkeling van nieuw recht. Voor dat laatste zijn ook creativiteit en originaliteit van belang. Voor de komst van AI was het mogelijk op basis van schriftelijke stukken te beoordelen of een student deze vaardigheden in huis heeft. Te veel afhankelijkheid van AI doet juist afbreuk aan deze vaardigheden. Door deze vaardigheden te toetsen in begeleidingsgesprekken en tijdens een mondelinge verdediging, worden studenten gedwongen de stof te internaliseren.

d. Laat studenten reflecteren op hun AI-gebruik

Door het gebruik van AI worden evaluatieve judgements belangrijk. Studenten moeten niet alleen in staat zijn de output van AI te beoordelen, maar moeten ook hun gebruik van AI kunnen evalueren. Waarom is AI voor bepaalde taken wel of niet geschikt, ook vanuit ethisch perspectief, wat zijn good practices, wat maakt een goede prompt (Bearman et al. 2024)? Voorwaarde voor een dergelijke reflectie en beoordeling daarvan is dat docenten en studenten AI-geletterd zijn (zie onder b).

e. AI-gebruik als methode

AI wordt regelmatig nog gezien als bron. Zo bevatten syllabi van scriptietrajecten aan de rechtenfaculteit een passage waarin is opgenomen op welke wijze naar een AI-tool moet worden verwezen in voetnoten. Hoewel het mogelijk is AI als bron te gebruiken, bijvoorbeeld om de output van AI te laten zien en daar commentaar op te geven, wordt AI in de regel echter niet als bron gebruikt. AI wordt meer gebruikt als sparringpartner of als coach (zowel inhoudelijk als op het gebied van taal). Het is daarom belangrijk om de nadruk te leggen op de methodologische verantwoording van het AI-gebruik en de verplichting om te verwijzen naar AI als bron te beperken tot uitzonderingsgevallen (Engelfriet 2025).

f. Verplichte begeleidingsgesprekken en mondelinge verdediging.

Om kritisch denken te toetsen, het proces te monitoren en te beoordelen of een student de stof heeft geïnternaliseerd, moeten begeleidingsgesprekken en een verdediging onderdeel uitmaken van de toetsing (Rinnooy Kan & Salomon 2026). Tijdens deze gesprekken en de verdediging kan worden doorgevraagd op het onderzoeksproces, de reflectie op het AI-gebruik, maar ook op de diepgang en eigen inbreng. Omdat kritisch denken wezenlijk is voor een wetenschappelijke opleiding, moeten deze mondelinge onderdelen voor een relatief groot deel meewegen bij de beoordeling van het scriptietraject, bijvoorbeeld de helft. Het gewicht van deze onderdelen kan ook tot uitdrukking komen door als regel te stellen dat voor deze onderdelen een voldoende moet worden behaald. Een heldere rubric moet hierbij de focus leggen op inhoudelijke diepgang en niet op presentatievaardigheden.

g. Less is more: minder woorden voor de scriptie

De toevoeging van mondelinge elementen verhoogt de werkdruk voor docenten.

Om de werklast haalbaar te houden, moet de vereiste omvang van de scriptie omlaag. Het woordenaantal bij veel opleidingen aan de Erasmus School of Law is momenteel 12.000 tot 14.000 woorden. Dat kan ingekort worden tot bijvoorbeeld 8.000-10.000 woorden. Dit dwingt studenten tot de kern te komen en stelt docenten in staat stukken sneller te beoordelen, waardoor de extra tijd die ze kwijt zijn aan de mondelinge elementen wordt gecompenseerd.

5. *Conclusie*

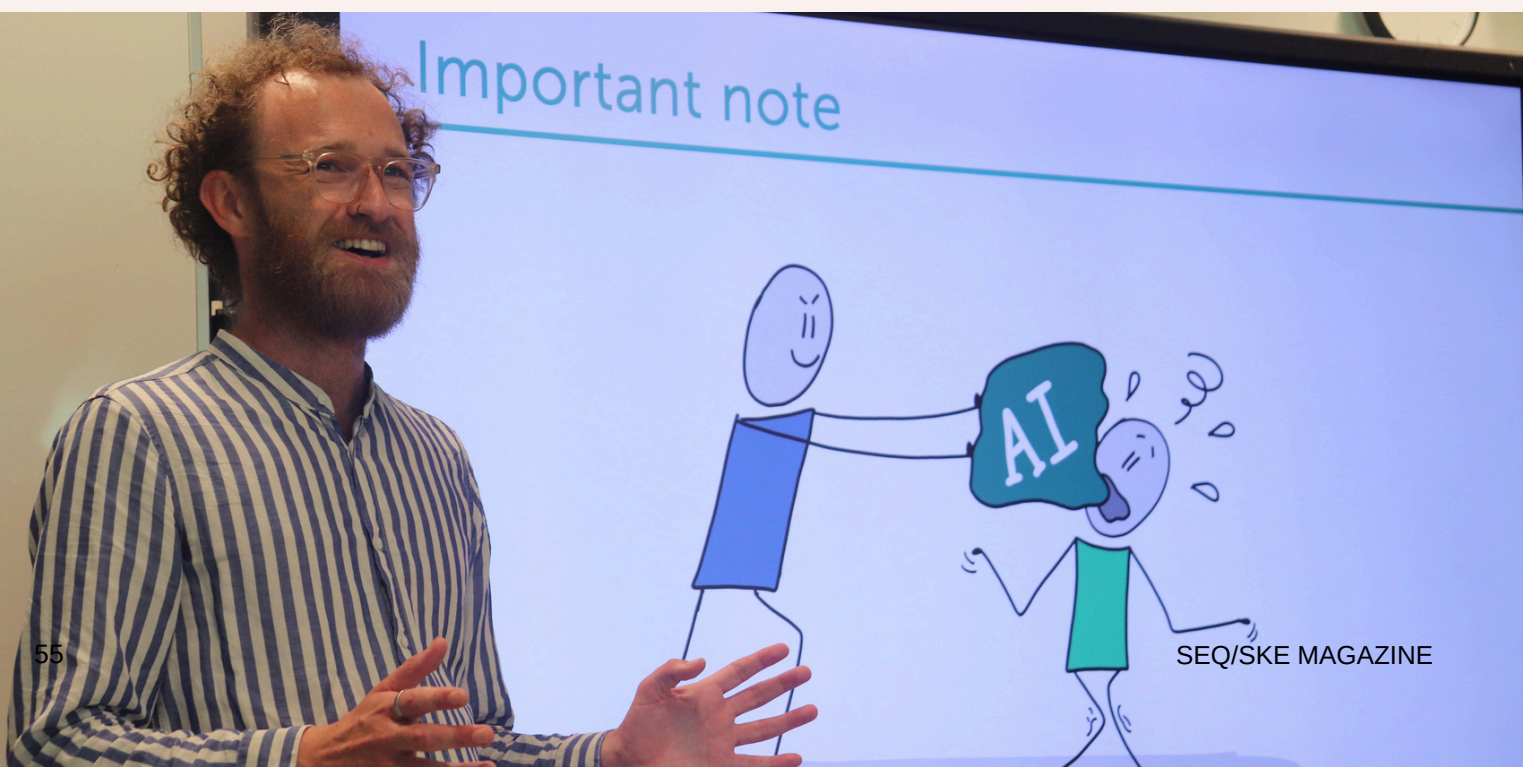
De opkomst van AI dwingt niet tot afschaffing van de scriptie, maar wel tot herijking van het traject. Door AI te integreren in de leerdoelen en de focus te verschuiven naar mondeling verantwoording en methodologische transparantie, blijft de scriptie een eindwerk waarvan de validiteit is geborgd. De ontwikkelingen gaan snel, dus het is belangrijk dat het wetenschappelijk onderwijs hierin meebeweegt. In die zin is dit artikel geen eindstation, maar onderdeel van een reis die vooralsnog geen concrete bestemming heeft.



(Infographic, gemaakt met behulp van Gemini op basis van de tekst van dit artikel)

Literatuur

- Bearman, M. et al. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893–905. <https://doi.org/10.1080/02602938.2024.2335321>
- Biggs, J. (2003). *Aligning Teaching and Assessment to Curriculum Objectives*, (Imaginative Curriculum Project, LTSN Generic Centre).
- Chanpradit, T. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology*, 9(4), 889-906.
- Corbin, T. et al. (2026). On the essay in a time of GenAI. *Educational Philosophy and Theory*, 58(3), 198–210. <https://doi.org/10.1080/00131857.2025.2572802>
- Dawson, P. et al. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016. <https://doi.org/10.1080/02602938.2024.2386662>.
- Engelfriet, A.P. (2025). ChatGPT is geen bron maar een methodologie, en daarmee basta. *NJB* 2025/2454.
- Hostetter, A.B. et al. (2025). Student and Faculty Perceptions of Generative Artificial Intelligence in Student Writing. *Teaching of Psychology*, 52(3), 319-329. <https://doi.org/10.1177/00986283241279401>.
- Klute, E., De Bruijn, M. en Van der Wees, J. (2026). Wat zijn scripties waard, nu AI meeschrijft? 'Je weet niet meer wat je aan het beoordelen bent, de student of de chatbot'. *NRC* 11 januari 2026.
- Kumar, R. & Mindzak, M. (2024). Who wrote this? Detecting artificial intelligence-generated text from human-written text, *Canadian Perspectives on Academic Integrity*, vol. 7, no. 1.
- Leckrone, B. (2025). 6 in 10 Online Students Report Using AI, Survey Finds, 16 augustus 2025, <https://www.bestcolleges.com/research/college-student-attitudes-on-ai/>.
- Rinnooy Kan, A. en Salomon, A (2026). *Opinie: AI hoeft geen bedreiging te zijn voor het universitaire onderwijs*. *De Volkskrant* 26 februari 2026.
- Villarroel, V., et al (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43 (5), 840-854.
- Walton, J. et al. (2026). How university students work on assessment tasks with generative artificial intelligence: matters of judgement. *Assessment & Evaluation in Higher Education*, 1-17.
- Welding, L. (2023). Half of College Students Say Using AI on Schoolwork is Cheating or Plagiarism, 17 maart 2023, <https://www.bestcolleges.com/research/college-students-ai-tools-survey/>.



RETHINKING ASSESSMENT: BUILDING CRITICAL THINKING IN 21ST CENTURY MANAGEMENT EDUCATION



ANTHONY BROWN

Introduction

Many educational systems today suffer from what I will describe as a 'legacy of tiredness'. By this, I mean that while educational systems face completely new challenges in order to keep preparing students to be fit-for-purpose in the 21st century labor market, both teaching and in particular assessments are firmly rooted in the 20th century models and thinking. To address these issues, creative, integrated and courageous solutions may be required, especially in the area of assessment. In parallel there have been concerns across the academic and practice worlds about the quality and type of learning standards as research as well as student quality has come under increasing scrutiny. As early as 1996 two reports published in the UK – the Harris and Dearing reports, cited the increased pressure on student numbers, teaching standards, class sizes and student staff ratios highlighted the importance of assurance and the need of innovation in assessment techniques primarily in post-graduate education (Pathirage et al., 2004).

Other studies have highlighted the challenges faced by higher education institutions both on the quality of instruction but also to their graduate's employability (Bisouz, 2008). Both policy makers and some employers have criticized degree conferring institutions on perceived deficiencies in both skills and learning (Casner-Lotto, 2006) and business schools have had their curriculums questioned on their relevance in delivering workforce-ready students (Rhew et al, 2019). Employers (and the availability of advanced digital data analytical and end-product generation technologies) require us, as academics, to rethink and refocus and how we develop and measure student's ability to sort through increasing amounts of information, solve complex issues and have the ability to critically analyze and think.

Higher education is increasingly confronted with a structural tension: while the demands placed on graduates have evolved rapidly, the methods used to assess their learning have often remained anchored in earlier pedagogical traditions. In management education in particular, this misalignment is evident in the continued reliance on assessment formats that privilege knowledge recall over the development of critical, analytical, and collaborative competencies.

This project emerges from the recognition that assessment should not merely serve as a mechanism for grading, but as a central driver of learning. By repositioning assessment at the core of course design, it seeks to create a more coherent, engaging, and skills-oriented educational experience that better reflects the realities of contemporary professional environments.

"In management education in particular, this misalignment is evident in the continued reliance on assessment formats that privilege knowledge recall over the development of critical, analytical, and collaborative competencies."

Background

The more traditional vision of assessment is that it is there to improve learning by linking it to the objectives and teaching strategies (Sewagegn, 2020; Shute, 2007), and that it should be used for 3 main reasons: as a form of feedback on the quality of teaching; to encourage students to learn and to give stakeholders a better view of knowledge absorption and use thus the idea of encouraging more formative than summative assessment forms (for learning rather than of learning) (Shute, 2009). As mentioned in the introduction, current approaches to assessment not only carry the 20th century legacy of being if not fully, at least, partially divorced from learning. Many courses have and still rely for the most part on one form of assessment: exams, although there has been significant movement to integrate other forms such as cases study analysis, written papers, oral defenses and presentations and partnerships with local businesses, this area is still ripe for further experimentation and application. All of these are needed to meet the ILOs that the program requires from the course; the perceived need to remain current -usually with reference to research in the field and lastly to meet the governance requirements by the examination board and or the dean of education and the program directors. Recently a new element to the mix has been added, that of cost control.

Project Approach: Assessment as the Starting Point

In this article I propose that assessment is at the center of course design and delivery (see figure 1). In this model and in line with the program ILOs and the agreed course thematic, the learning outputs and best forms to encourage and evaluate it are outlined. Once this rough outline of intended outcomes is completed it is evaluated and the rest of the course materials developed to deliver these and ensure a clear student journey and evaluation (see figure 2).

Figure 1 – Assessment development

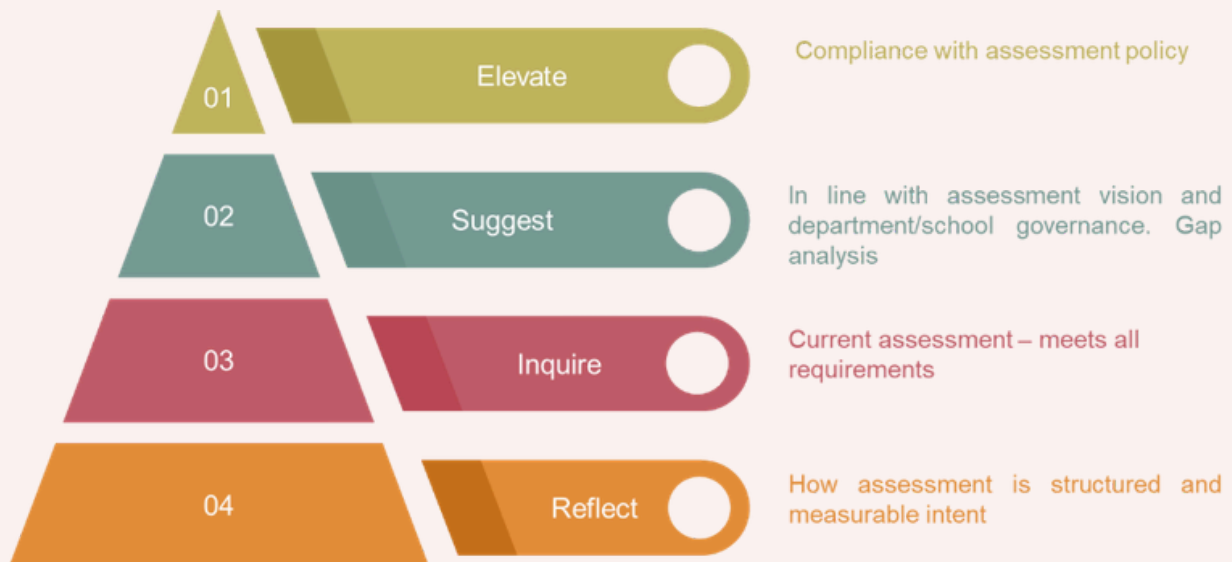
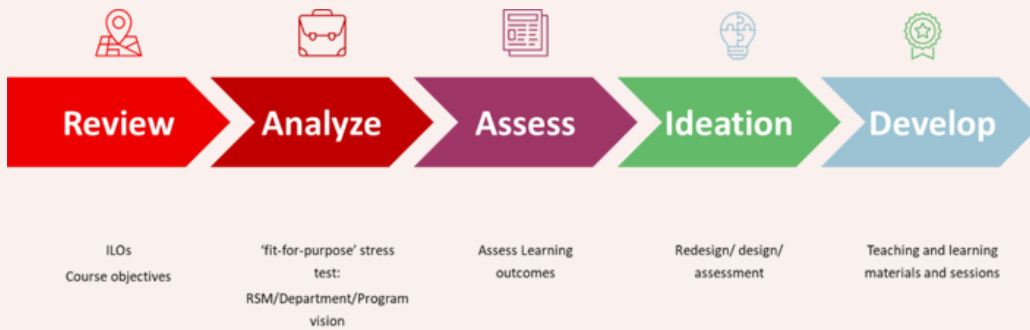


Figure 2 - Course design



Once the learning objectives and measurement parameters have been clearly established, the form of assessment is created. They need to ensure that all the ILOs are forefront rather than an afterthought, the clear skills development that will be communicated to the students and program competencies.

The base idea is to design and develop an intervention for the course that will better assess the overall learning and application of the same while delivering in a more transparent manner on the course materials, objectives, ILOs, skills and competencies of the program.

The central premise of this project is that assessment should shape, rather than follow, course design. Instead of aligning assessments to pre-existing teaching activities, the process begins by identifying the intended learning outcomes and the competencies students are expected to develop. These outcomes then inform the design of assessment tasks, which in turn structure the learning journey.

To operationalize this approach, the project introduces a two-part assessment model built around a simulation of real-world decision-making. The intention is to move beyond traditional examination formats and create an environment in which students must actively engage with complexity, ambiguity, and competing perspectives.

The first component of the model is a structured simulation inspired by international trade negotiations. Students are assigned roles as either firms or countries and are tasked with developing strategies for entering new markets. Over the course of the simulation, they participate in plenary sessions, working groups, and negotiations that require them to apply theoretical frameworks, interpret contextual information, and respond dynamically to the actions of others. In doing so, they are compelled to integrate knowledge with practice, while simultaneously developing communication and collaboration skills.

The second component builds on this experiential foundation through reflective and analytical work. Students produce a short group video in which they present and justify their strategic decisions, followed by an individual reflection essay that critically examines their experience in light of relevant theories and concepts. This combination of collective and individual outputs is designed to reinforce both shared learning and personal insight.

Key Deliverables and Educational Value

The assessment model generates a set of deliverables that are closely aligned with the development of critical thinking skills. The simulation itself functions as a live environment in which students must interpret information, evaluate alternatives, and make strategic decisions under conditions of uncertainty. The subsequent video presentation requires them to articulate and defend their reasoning, while the individual reflection essay encourages deeper analysis and self-evaluation.

Taken together, these elements contribute to a more holistic form of assessment. Rather than isolating knowledge from application, the model integrates cognitive, interpersonal, and reflective dimensions of learning. Students are not only asked to demonstrate what they know, but also how they use that knowledge, how they communicate it, and how they evaluate their own performance.

Implications for Teaching and Learning

Positioning assessment at the center of course design has broader implications for pedagogical practice. It necessitates a shift from viewing assessment as a summative endpoint to understanding it as an ongoing, formative process that shapes the entire learning experience.

This approach enhances transparency by making explicit the connection between learning objectives, teaching activities, and assessment tasks. It also encourages greater student engagement, as learners are able to see the relevance of their work and the skills they are developing. Moreover, it supports the cultivation of higher-order competencies, including critical thinking, problem-solving, and collaboration, which are increasingly valued in professional contexts.

For educators, this model requires a reconsideration of course structure and sequencing. Teaching activities must be deliberately designed to prepare students for the assessment tasks, while feedback mechanisms need to be integrated throughout the process to support continuous improvement.

Next Steps

Although the model is conceptually robust, it remains in a developmental phase. The next stage of the project will focus on refinement through iterative feedback and evaluation. Input from students will be particularly useful in assessing the clarity, fairness, and perceived value of the assessment tasks. Engagement with program directors and academic colleagues will also be essential to ensure alignment with broader curricular objectives and institutional constraints.

Further work will be required to develop detailed assessment criteria and rubrics, as well as to evaluate the practical implications of the model in terms of workload and resource allocation. These considerations will be critical in determining the scalability and sustainability of the approach.

Conclusion

Reconceptualizing assessment as the foundation of course design represents a meaningful shift in management education. By integrating simulation-based learning with reflective analysis, this project demonstrates how assessment can be leveraged to support the development of critical thinking and other essential competencies.

While further refinement is required, the approach offers a promising pathway toward more engaging, relevant, and effective educational practices. It invites educators to reconsider not only how they assess learning, but how assessment itself can become a catalyst for deeper and more meaningful student engagement.

References

Bisoux, T. (2008). The socially responsible curriculum. *BizEd*, 7(4), 22.

Casner-Lotto, J., & Barrington, L. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce. Partnership for 21st Century Skills. 1 Massachusetts Avenue NW Suite 700, Washington, DC 20001.

Pathirage, C. P., Haigh, R., Amaratunga, D., Baldry, D., & Green, C. M. (2004, September). Improving dissertation assessment. In *Education in a Changing Environment: Scholarship, Educational Research and Development at the University of Salford: Proceedings of the University of Salford's 2nd Learning and Teaching Research Conference*. University of Salford.

Rhew, N. D., Black, J. A., & Keels, J. K. (2019). Are we teaching what employers want? Identifying and remedying gaps between employer needs, accreditor prescriptions, and undergraduate curricular priorities. *Industry and Higher Education*, 33(6), 362-369.

Shute, V. J. (2007). Focus on formative feedback. *ETS Research Report Series*, 2007(1), i-47.

Shute, V. J. (2009). Simply assessment. *International Journal of Learning and Media*, 1(2), 1-11.

Testimonial Anthony Brown:

“The SEQ and the process within it has helped me contextualize and clarify my ideas around assessment for one of my postgraduate courses. The course allowed me to clearly and succinctly articulate my vision and thinking around why I felt a need for a second form of assessment, the requirement of completing the products has helped me both frame and articulate these better. This has included looking at the course from a different set of lenses.”

NAAR DUURZAME TOETSKWALITEIT: SAMENWERKEN AAN SAMENHANG

FRANK PYNENBURG

The New School

Communication, Creativity & Marketing

De opgave van the new school

The New School for Information Services (TNS) is een particuliere hogeschool in Amsterdam, die één hbo-opleiding op het gebied van marketing, communicatie, creativiteit en ondernemen aanbiedt. De vierjarige bacheloropleiding bestaat uit een divers curriculum met een brede waaier aan kennis en vaardigheden. Sinds haar oprichting, zo'n 40 jaar geleden, heeft TNS zich gepositioneerd als een opleidingsinstituut dat niet alleen beroepsvaardigheden aanbiedt, maar ook sterk inzet op de persoonlijke ontwikkeling van haar studenten. De kleinschaligheid van de school creëert een omgeving waarin persoonlijke aandacht en maatwerk centraal staan. De school neemt hierdoor een unieke plek in het Nederlandse onderwijssysteem in.

The New School steunde de afgelopen jaren, naast een vast kernteam, op een pool van freelance docenten. Hoewel deze freelance docenten betrokken en vakinhoudelijk deskundig waren, maakte hun flexibele inzet het een uitdaging om continuïteit en duurzame kwaliteitsontwikkeling te waarborgen. Mede hierdoor is de afstemming van toetsing tussen de verschillende onderwijseenheden tot twee jaar geleden beperkt geweest. Dit heeft geleid tot verschillen in interpretatie en toepassing van beoordelingscriteria. Dit heeft de consistentie en transparantie van toetsing beïnvloed. Sinds september 2023 is er een nieuwe directeur en sinds 2024 is er ook een nieuwe opleidingscoördinator actief. Er zijn verschillende processen rondom integratie, standaardisering en kwaliteitsborging in gang gezet. Er wordt door de invoering van de DBA minder gewerkt met freelancers, waardoor meer ruimte ontstaat voor structurele verbetering en afstemming tussen docenten.

The New School had van oorsprong een profiel dat sterk aansloot bij de liberal arts-traditie, maar is de afgelopen jaren geëvolueerd naar een business- en marketinggeoriënteerde opleiding met een focus op ondernemerschap.

Deze koerswijziging vraagt om een herijking van het toetsbeleid, inclusief de eindbeoordeling, zodat deze beter aansluiten bij de huidige identiteit en ambities van de opleiding. TNS wil het toetsbeleid versterken door de invoering van assessments op drie niveaus, zodat de zes competenties per niveau beter kunnen worden geborgd. De herziene TNS-competentiematrix wordt daarbij als uitgangspunt genomen en op drie niveaus zichtbaar gemaakt. Ook wordt het eindassessment losgetrokken van het afstudeertraject, zodat dit meer nadruk krijgt als afzonderlijk meetmoment van de eindcompetenties.

DEZE KOERSWIJZIGING VRAAGT OM EEN HERIJKING VAN HET TOETSBELEID, INCLUSIEF DE EINDBEOORDELING, ZODAT DEZE BETER AANSLUITEN BIJ DE HUIDIGE IDENTITEIT EN AMBITIES VAN DE OPLEIDING.

De toetsrealiteit bij TNS

In de praktijk worden toetsen vaak per onderwijseenheid en per docent ontwikkeld. Dat geeft ruimte voor eigen invulling, maar het maakt ook zichtbaar dat de onderlinge afstemming beperkt is. De borging van de zes TNS-competenties is daardoor niet altijd even helder. In opdrachten, toetsen en beoordelingen komt competentiegericht werken terug, maar niet op een consequente en voor iedereen herkenbare manier. Dat knelt juist op een moment waarop TNS studenten wil opleiden voor een beroepspraktijk die snel verandert. In het toetsbeleid van 2024-2025 staat dat TNS-toetsing ziet als een integraal en richtinggevend onderdeel van het leerproces, als middel om ontwikkeling zichtbaar te maken en als manier om studenten gericht voor te bereiden op de beroepspraktijk. Het uitgangspunt van constructieve alignment dat genoemd wordt in het toetsbeleid is sterk. Toch laat de dagelijkse praktijk zien dat dit niet altijd het geval is

Het huidige eindassessment in jaar 4 is een aandachtspunt. Dat assessment is nu het enige moment waarop studenten aan het einde van de opleiding de zes competenties op eindniveau moeten laten zien. Dit gebeurt in tijdens een korte eenmalige presentatie zonder portfolio's of structurele tussentijdse toetsing, waardoor de beoordeling van competentieontwikkeling kwetsbaar wordt. De kern van het vraagstuk is daarom niet of TNS al een toetsbeleid heeft, maar of dat beleid voldoende doorwerkt in het hele programma. De centrale vraag van het SKE-traject luidt dan ook hoe de afstemming en integratie van assessments binnen het vierjarige TNS-programma kan worden verbeterd, zodat de zes TNS-competenties tussentijds beter worden geborgd.

Hoe is dit probleem onderzocht?

Het onderzoek voor de SKE-project had als doel een onderbouwd advies te geven voor versterking van het toetsbeleid en de competentieontwikkeling en de invoering van assessments op elk niveau. Voor de interne inventarisatie werd het curriculum geanalyseerd met Toetsweb, aangevuld met interviews en vragenlijsten voor docenten. De theoretische basis kwam uit literatuurstudie en desk- /expertinterviews over competentiegericht toetsen en best practices. Extern werd gekeken naar andere hogescholen middels interviews met drie onderwijscoördinatoren bij andere hogescholen. Ervaringen van studenten met competenties en toetsing werden opgehaald via vragenlijsten. Het onderzoek onder docenten combineerde kwalitatieve interviews met zeven docenten en staf en een vragenlijst ingevuld door uiteindelijk negen docenten. Ondanks de beperkte omvang van de enquête, bieden de resultaten een goede basis voor een concreet advies over programma-brede assessments bij TNS en de versterking van het toetsprogramma.

Analyse huidige toetsbeleid: ambitie op papier

Het huidige toetsbeleid van TNS bevat op papier al een aantal stevige uitgangspunten. Het momenteel geldend document Toetsbeleid- en plan noemt validiteit, betrouwbaarheid, objectiviteit, transparantie en constructieve alignment als inhoudelijke kwaliteitseisen voor toetsing

Het beleidsplan beschrijft ook dat toetsen moeten voortkomen uit leerdoelen, dat de examencommissie kwaliteitsbewaking uitvoert en dat toetsing onderdeel is van een groter geheel, namelijk het toetsprogramma. Ook staat in het beleid dat het curriculum is gericht op competentieontwikkeling, dat beroepsproducten een belangrijke rol spelen en dat een digitaal portfolio in principe een geschikte vorm kan zijn om ontwikkeling zichtbaar te maken. Op papier is er al meer aanwezig dan alleen losse procesafspraken. Het beleid noemt een onderwijsvisie, naar intensieve didactiek en naar de relatie met beroepsproducten. Er wordt ook gesproken over een digitaal portfolio. De realiteit laat namelijk zien dat veel van deze uitgangspunten nog niet consequent in praktijk worden gebracht. Het verschil tussen papier en praktijk vormt daarmee de kern van het probleem.

Docenten: samen werken aan samenhang

De interviews met zeven collega's schetsen een opleiding met sterke praktijkgerichtheid, veel docentvrijheid en een krachtige leerlijn voor onderwijseenheid Persoonlijke en Professionele Ontwikkeling (PPO) maar ook een gefragmenteerde toetspraktijk waarin de zes TNS-competenties nog niet overal en doorlopend zijn ingebed. Er zijn witte vlekken bij enkele competenties, onbewuste overlap tussen vakken, weinig structurele kalibratie en onduidelijke kaders rond doorstroming tussen verschillende jaren. Dit is ook mede ingegeven door het ontbreken van een curriculumcommissie (cuco). Deze cuco wordt op dit moment opgetuigd, maar zal op korte termijn nog niet volledig operationeel zijn. Tegelijk is er brede bereidheid van het docententeam om te werken aan een samenhangende toetslijn met ontwikkelassessments.

Als richting voor TNS adviseren docenten om te starten met een volledige inventarisatie van alle toetsen in een "toetswerkplaats", drie integrale ontwikkelassessments over de jaren heen op te bouwen, de competentiematrix en niveau-indicaties gezamenlijk te concretiseren en kalibratie structureel te organiseren.

Intern is er ook gekeken naar de manier waarop docenten het huidige toetsbeleid kennen en gebruiken. Daarbij kwam naar voren dat de bekendheid met het toetsbeleid beperkt is en dat niet iedereen het document daadwerkelijk heeft gelezen. Dat is relevant, want een beleid kan nog zo zorgvuldig zijn opgesteld, maar het krijgt pas betekenis als teams het kennen, bespreken én vooral toepassen. Daarnaast is er ook een enquête uitgezet onder docenten. De resultaten van de anonieme vragenlijst – ingevuld door negen TNS-docenten - toont dat de meerderheid weliswaar bewust is van TNS competenties in eigen onderwijs, maar deze voornamelijk koppelt aan eigen leerdoelen en opdrachten. Beoordeling binnen eigen vak/project voelt zeker, maar actieve inzet van de TNS competentiematrix bij toetsontwerp is beperkt. Intercollegiale afstemming scoort laag, met minimale samenwerking bij een aantal vakken. Er is onder respondenten veel onduidelijkheid over de competentieniveaus per studiejaar en de verwachtingen op afstudeerniveau. Docenten reageerden voornamelijk neutraal of negatief op stellingen hierover en ook de kennis van de competenties en leerdoelen van andere docenten wordt als beperkt ervaren. Open antwoorden benadrukken de behoefte aan meer duidelijkheid over niveaus, en betere teamkalibratie. Toch is er weinig animo voor extra scholing of overleg. Dit benadrukt de noodzaak voor gestructureerde afstemmingsmomenten en regelmatige besprekingen over de competentiematrix.

Inspiratie van buitenaf: leren van andere hogescholen

De externe oriëntatie bestond uit gesprekken met specialisten van de Hogeschool van Amsterdam, Codarts en de Politieacademie. Dit zijn voorbeelden van opleidingen waar assessments programmatisch zijn ingebed en waar gewerkt wordt met opbouw over meerdere jaren en waar beoordeling niet alleen op een eindmoment rust. Uit de gesprekken bleek dat integrale, doorlopende assessments op TNS kansrijk zijn, mits ze ingebed worden in een duidelijke leerlijn met oplopende complexiteit en echte beroepsproducten als uitgangspunt.

Een centrale ruggengraat in de vorm van een dekkingsmatrix over vier jaar is daarbij cruciaal, zodat basisvaardigheden in jaar 1 doorontwikkelen naar geïntegreerde opdrachten en beroepsproducten in latere jaren.

Literatuur en experts: onderbouwing van de koers

Ook literatuuronderzoek en expertinterviews zijn meegenomen voor dit onderzoek. Uit de gesprekken Ard Zonneveld, Lisette Munnike en Judith Gulikers blijkt dat TNS veel kan leren van de manier waarop andere hogescholen hun assessments hebben ingericht. De centrale rode draad in de gesprekken met de experts is dat toetsing niet bestaat uit losstaande momenten, maar onderdeel moet zijn van een doorlopende leerlijn waarin studenten zich stapsgewijs ontwikkelen in complexiteit en zelfstandigheid. Daarbij is het belangrijk dat assessments aansluiten op authentieke beroepsproducten, zodat studenten worden beoordeeld op werkzaamheden die passen bij de praktijk van het vak. De expertinterviews laten bovendien zien dat een goede opbouw in het curriculum noodzakelijk is, waarbij eerdere opdrachten steeds worden verdiept en uitgebreid.

Een tweede belangrijk inzicht is dat betrouwbare beoordeling vraagt om duidelijke structuur en zorgvuldige afstemming. De geïnterviewde specialisten benadrukken allemaal het belang van meerdere beoordelaars, kalibratie en heldere beoordelingscriteria. Tegelijkertijd zeggen zij dat dit in de praktijk niet vanzelf goed gaat. Vage rubrics, beoordelaarsbias en te veel flexibiliteit kunnen leiden tot fragmentatie en inconsistentie in de beoordeling. Juist daarom is het nodig om afspraken over toetsing stevig vast te leggen en beoordelaars goed te trainen. Op basis van deze interviews kan worden geconcludeerd dat integrale en geïntegreerde assessmentpraktijken bij TNS kans van slagen hebben, mits er een sterke inhoudelijke en organisatorische basis aanwezig is. Een complete dekkingsmatrix is een belangrijke voorwaarde. Met een dekkingsmatrix wordt in dit kader een overzichtstabel waarin zichtbaar is waar, wanneer en op welk niveau de competenties in het curriculum én in de toetsing terugkomen.

Een andere voorwaarde is doorlopende portfolio-opbouw, betrokkenheid van het werkveld en structurele kalibratie zijn daarbij essentiële voorwaarden. De expertinterviews maken duidelijk dat de kwaliteit van assessment niet alleen afhangt van wat er wordt getoetst, maar vooral ook van hoe de toetsing is georganiseerd en door wie zij wordt uitgevoerd.

De blik van studenten

De resultaten van een enquête onder TNS-studenten toont dat het begrip 'competentie' breed bekend is. Het aantal respondenten was met elf weliswaar beperkt, maar de meest respondenten kozen 'bekend', 6 'redelijk' en 1 'zeer bekend' te zijn met de competenties. Respondenten beschrijven competenties consistent als geïntegreerd geheel van kennis, vaardigheden en houding voor praktijktoepassing, met de zes TNS-kerncompetenties (zoals omgevingsbewustzijn en onderzoeken/analyseren) veel genoemd. Een meerderheid (72%) bevestigt heldere koppeling aan opdrachten en leerdoelen, en toetsbelasting en moeilijkheidsgraad worden als optimaal ervaren (89% 'precies goed'). Toch blijft progressie in competentieniveaus (jaar 1-4) onduidelijk voor de helft van de respondenten, en feedback is positief, maar mist vaak coachende feedforward op ontwikkeling. Open antwoorden roepen op tot meer zichtbaarheid: competenties voelen nu vaak administratief aan, en er is behoefte aan structurele begeleiding en groeitracking. Dit wijst op kansen voor explicietere integratie in curriculum en feedback.

De volgende stap voor TNS

TNS moet de komende jaren bewust inzetten op een stevige versterking van het toetsbeleid, waarbij toetsing niet langer wordt gezien als een verzameling losse handelingen, maar als een samenhangend en gedragen geheel binnen de opleiding. De ambitie moet zijn om toe te groeien naar een procesgeoriënteerde aanpak waarin toetsactiviteiten niet afhankelijk zijn van individuele voorkeuren of werkwijzen, maar worden gestuurd vanuit duidelijke afspraken, vaste kwaliteitscriteria en een gezamenlijke visie binnen het docententeam.

Bovenstaande betekent dat toetsbeleid niet alleen op papier aanwezig moet zijn, maar ook daadwerkelijk richtinggevend moet worden voor de dagelijkse praktijk van ontwerp, afname, beoordeling en evaluatie. Een belangrijk vertrekpunt daarbij is dat TNS het toetsbeleid actueel en concreet maakt door doelen, verantwoordelijkheden en kwaliteitsafspraken helder vast te leggen.

In de huidige situatie is er al een competentiematrix aanwezig, maar deze is nog niet volledig geïntegreerd in het curriculum en de competentiedekking is nog niet overal consequent zichtbaar in toetsen, opdrachten en beoordelingen. Juist daarom is het van belang dat TNS deze matrix niet als los instrument gebruikt, maar als inhoudelijke basis voor de hele toetsstructuur. Daarmee wordt duidelijk welke competenties wanneer, op welk niveau en met welk type bewijs worden beoordeeld, zodat er meer samenhang ontstaat tussen leeruitkomsten, toetsvormen en het gewenste eindniveau.

Daarnaast is het wenselijk dat TNS de stap zet naar integrale assessments. Dat houdt in dat toetsen niet alleen afzonderlijke momenten blijven, maar samen een doorlopende beoordelingslijn vormen waarin studenten hun ontwikkeling stap voor stap laten zien. Hiervoor is een helder toetsprogramma nodig waarin per leerjaar zichtbaar is welke toetsen worden afgenomen, welke competenties daarmee worden gedekt, hoe de weging is opgebouwd en hoe de opbouw richting het eindniveau verloopt. Door dit systematisch in kaart te brengen, ontstaat meer inzicht in hiaten, overlap en toetsdruk, en kan de opleiding gericht werken aan een evenwichtige en betekenisvolle toetsstructuur.

Een essentieel onderdeel van deze ontwikkeling is de invoering van portfolio's, formatieve- en summatieve toetsmomenten. Studenten moeten gedurende hun opleiding structureel bewijsstukken en beroepsproducten verzamelen, zodat hun groei en ontwikkeling beter zichtbaar wordt. Die portfolio opbouw maakt het mogelijk om niet alleen eindproducten te beoordelen, maar ook tussentijdse ontwikkeling, reflectie en groei in competentieontwikkeling.

Tegelijkertijd biedt dit TNS de mogelijkheid om formatieve toetsing sterker te verankeren in het onderwijs, waardoor studenten vaker feedback krijgen op hun voortgang en gerichter kunnen werken aan hun ontwikkeling naar het eindniveau. Om deze verandering betrouwbaar en uitvoerbaar te maken, moet ook de toetsorganisatie worden versterkt. Alle toetsgerelateerde processen, zoals ontwerp, afname, beoordeling, herkansing, inzage, bezwaar en cijferregistratie, moeten opnieuw helder worden vastgelegd en bewaakt.

Daarbij is het belangrijk dat voor iedereen duidelijk is wie eigenaar is van welk proces, wie uitvoert, wie controleert en wie wordt geïnformeerd. Een dergelijke heldere verantwoordelijkheidsverdeling voorkomt onduidelijkheid en zorgt voor meer voorspelbaarheid en kwaliteit in de uitvoering. Ook is het zinvol om toetsing regelmatig expliciet op de agenda van teamoverleggen en examencommissies te zetten, zodat knelpunten tijdig besproken en aangepakt kunnen worden. Minstens zo belangrijk is de verdere ontwikkeling van toetsbekwaamheid binnen het docententeam. Als TNS wil werken met integrale assessments, dan moeten docenten en examinatoren beschikken over voldoende kennis en vaardigheden om toetsen goed te ontwerpen, uit te voeren en te beoordelen. Daarom is het verstandig om toetsbekwaamheid structureel onderdeel te maken van jaargesprekken, scholing en intervisie. Jaarlijkse interne scholingsmomenten kunnen helpen om gezamenlijke afspraken te verdiepen en om met elkaar te spreken over casuïstiek uit de TNS-praktijk.

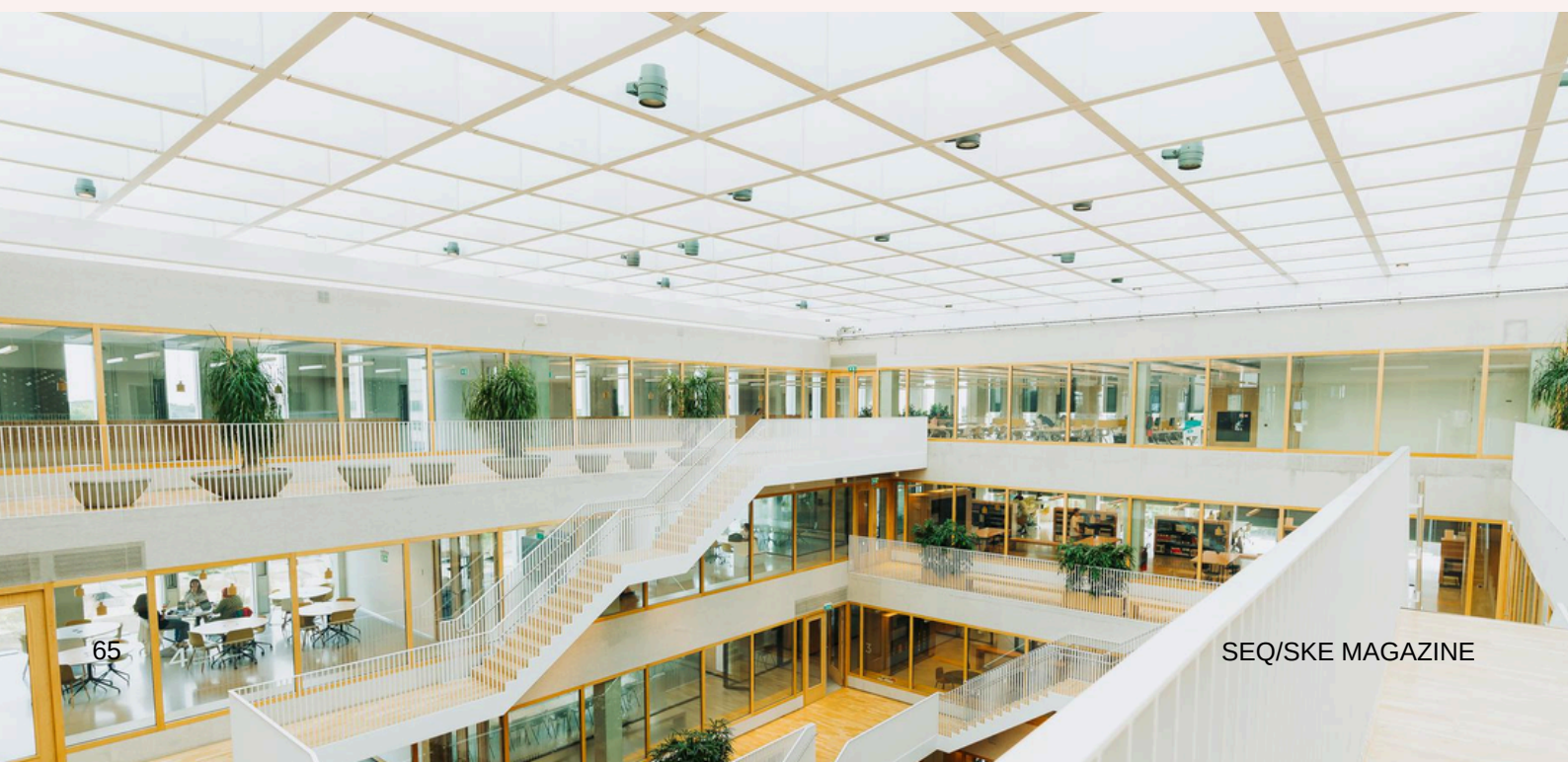
DAAROM IS HET VERSTANDIG OM TOETSBEKWAAMHEID STRUCTUREEL ONDERDEEL TE MAKEN VAN JAARGESPREKKEN, SCHOLING EN INTERVISIE.

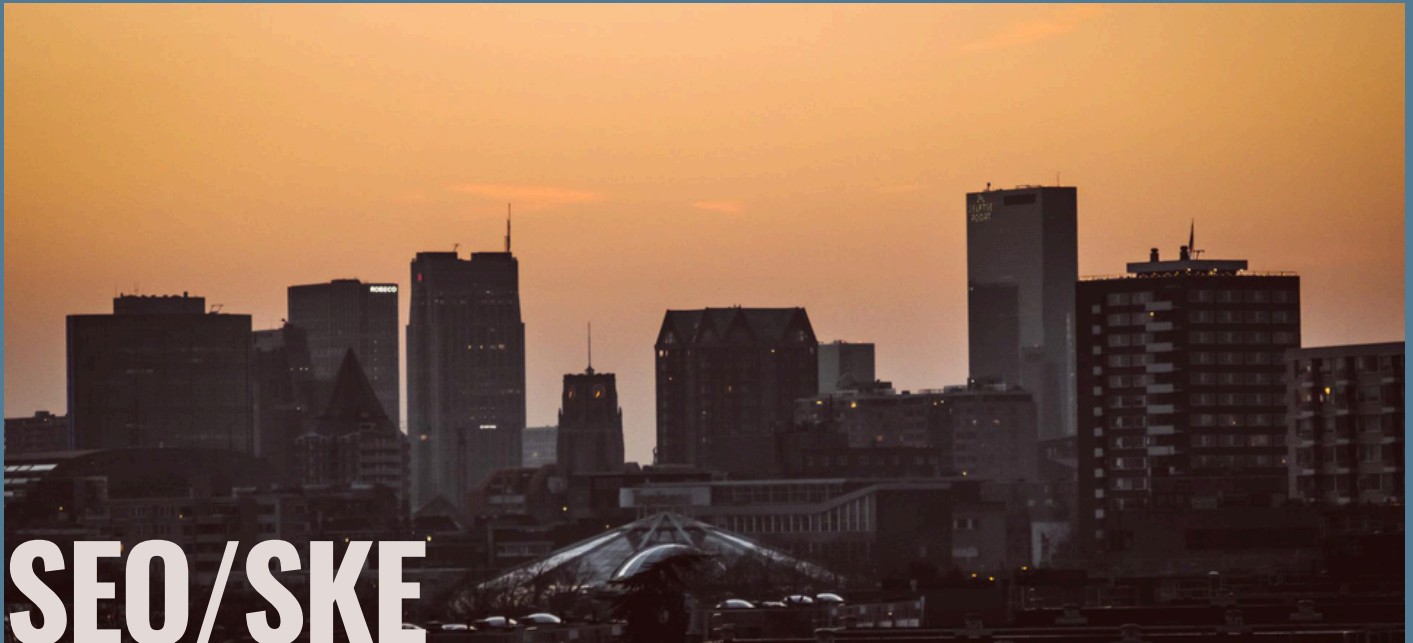
. Tot slot is het van belang dat TNS het toetsbeleid niet ziet als een eenmalige verandering, maar als een proces van continue ontwikkeling. Door jaarlijks te evalueren hoe het toetsbeleid werkt, waar verbeteringen nodig zijn en hoe de toetsorganisatie functioneert, kan de opleiding gericht blijven werken aan kwaliteitsverbetering. De kern van het advies is daarom dat TNS moet groeien naar een toetscultuur waarin integrale assessments, portfolio-opbouw, gezamenlijke kalibratie en goed getrainde examinatoren samen zorgen voor een robuuste en toekomstbestendige toetspraktijk.

Een stevige basis voor later

De kernboodschap van dit betoog is dat toetsing pas echt waarde krijgt wanneer zij onderdeel wordt van een bredere onderwijskundige visie en gedragen wordt de hele organisatie. Voor TNS ligt er daarmee niet alleen een opgave, maar vooral ook een kans: door toetsbeleid, assessment en competentieontwikkeling met elkaar te verbinden, kan de opleiding bouwen aan een stevig fundament voor de komende jaren. Een toekomst die TNS qua toetsbeleid met vertrouwen tegemoet kan zien

...





SEQ/SKE MAGAZINE

Visit our sites for more info!



Senior Examiner Qualification

NEVER MISS AN ISSUE!

- Enjoy huge savings
- Free home delivery
- Get your copy before everyone else

VISIT US ONLINE

- Catch the freshest features
- Updated daily
- Read anytime, anywhere