

SEQ MAGAZINE

Risbo
Research-Training-Consultancy
Erasmus



SENIOR EXAMINATION QUALIFICATION
COHORT 2024/2025

June 2025

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	Micheile Henderson, via Unsplash Daniel Agudelo, via Unsplash
Magazine photos	Alexander Santos Lima

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 second edition of the Risbo SEQ Magazine. In this magazine, you will be introduced to the participants of cohort 2024-2025 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 2024-2025, 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. For the first time we also had one participant from the University of Amsterdam and one participant 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 dilemma's and different approaches was found 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 inclusive assessment, to tackling the challenges posed by generative AI and from organizing quality assurance, to implementing assessment as learning.

We hope you will find the articles insightful and inspiring!

Marit Nieuwenhuys, Hanneke Bahlman & Celine van de Lienden
Risbo



Marit Nieuwenhuys
nieuwenuhuys@risbo.eur.nl



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



Celine van der Lienden
vanderlienden@risbo.eur.nl



TABLE OF CONTENTS

03 ABOUT THE PARTICIPANTS
Get to know our participants of cohort 2024/2025

06 ALEX KORTEKAAS
Beter zorgen zonder borgen

10 KOEN MIGCHELBRINK
An integrated and coordinated approach to assessing GenAI knowledge and skills

16 RIANNE KOK
Integratie van onderwijs over methoden en opzetten van wetenschappelijk onderzoek in het inhoudelijk curriculum: *kansen en uitdagingen*

24 EVA BOOMSLUITER
Assessment as an instrument for feedback and self-evaluation

28 ERIK ROS
Samenwerken tussen studenten op de uni is key! Maar hoe beoordeel je dat?

32 VICKI ERASMUS
Assessing student growth: assessment rounds in a programmatic curriculum

38 WILLEKE SCHIPPERS
De waarde van de SKE opleiding voor een docent van de politie academie

42 MARJOLEIN TIMMERS
Authenticiteit van de take home toets in het AI-tijdperk

54 SAMANTHA BREUER
The (in)visibility of assessment

58 JAY LEE
Invasion of the student- learning snatchers: the impact of generative AI on assessment and student performance

68 MARIJN FALING
Inclusive assessment or assessment for inclusion?

74 GEORGINA M. GÓMEZ
Navigating the digital frontier

78 PHILIPPE VERSIJP
May you live in interesting times – assessment of the thesis in the AI-age

ABOUT THE SEQ

The Senior Examiner Qualification (SEQ) course is offered by Risbo and the Community for Learning and Innovation (CLI) to stimulate the professionalisation of lecturers. It is intended for teachers that are (prospective) members of examination boards and/or program coordinators/directors at an Erasmus 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/or degree programs.

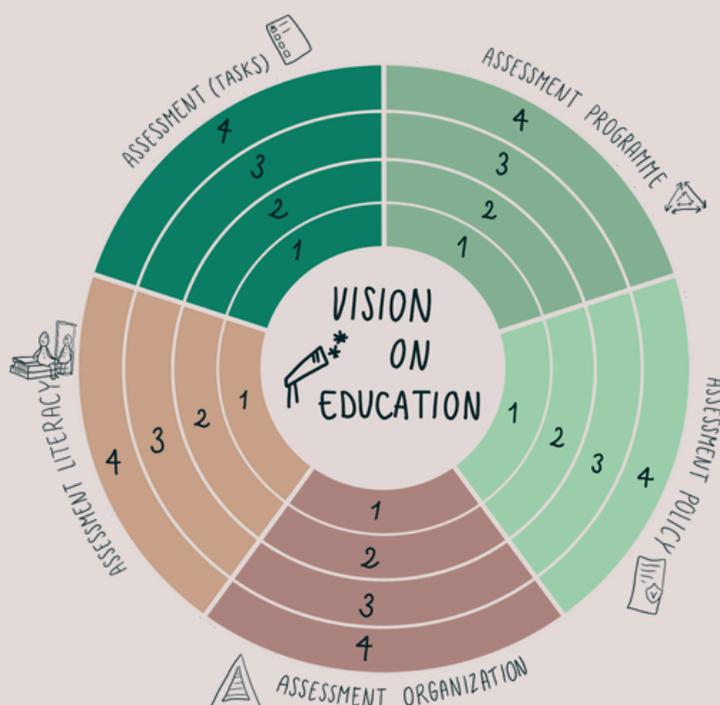
In the SEQ each participant is expected to bring in an ambition or challenge regarding the quality of assessment within their school, supported by their examination board and management team - to ensure knowledge dissemination within the school.

During the trajectory the participants learn about their vision on the role of assessment in higher education and the implications that this vision will have for the assessment (programma and/or policy) within their own programma 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 one's own organization is examined and participants work on an improvement plan.

Several topics covered in the SEQ are:

- the quality of components and processes in the assessment organization;
- the different responsibilities, roles and tasks related to assessment within a faculty/study programme;
- the competence enhancement and training of examiners in the field of assessment;
- the implementation process and evaluation of the assessment policy and (the framework of) the assessment organization.



ABOUT THE PARTICIPANTS



TLS

Alex kortekaas
Toetsdeskundige
Tilburg University



ESSB

dr. Koen Migchelbrink
Assistent professor
Department of Public Administration and
Sociology



ESSB

Dr. Rianne Kok
Associate Professor
Youth & Family



ESL

Mr. dr. Eva Boomsluiters
Universitair docent
Department Law & Tax



ESL

Mr. dr. Erik Ros
Universitair docent
Department Law & Tax



EMC

dr. Vicki Erasmus
Assistant professor
Public Health

ABOUT THE PARTICIPANTS



POLITIE
ACADEMIE

Willeke Schippers MSc
Docent jeugd/recht/criminologie,
sector Hoger Onderwijs



ESHPM

dr. mr. Marjolein Timmers
Assistent professor
Health Law & Ethics



EUC, ESSB

**Samantha Maria Serene Breuër,
MSc.**
Chair Examination Board Erasmus
University College, Lecturer
Psychology
Social Behavioural Sciences



ESHCC

Dr. Ju-Sung (Jay) Lee
Assistant Professor
Department: Media &
Communication



ISS

Marijn Faling, PhD
Assistant Professor



ISS

Dr. Georgina M. Gómez
Associate professor
Institutions and Local
Development



UVA

Dr. Philippe Versijp
Voorzitter examencommissie
Economie & Bedrijfskunde,
Docent
Finance Amsterdam Business
School

BETER BORGEN ZONDER ZORGEN

ALEX KORTEKAAS

Erasmus School of
History, Culture and
Communication



Een van de wettelijke taken van de examencommissie is “het borgen van de kwaliteit van tentamens en examens”[1]. De wet, noch de memorie van toelichting, geeft verdere context bij deze wettelijke bevoegdheid, waardoor examencommissies op verschillende manieren invulling geven aan deze wettelijke bevoegdheid. De examencommissie van ESHCC borgt de kwaliteit van toetsing – onder andere - door jaarlijks verschillende documenten van een aantal vakken per opleiding op te vragen en aan de hand daarvan een oordeel te vellen over de kwaliteit van de toetsing van het betreffende vak. De examinerator ontvangt een schriftelijk advies van dit oordeel. Dit advies komt vaak maanden nadat het vak is afgelopen en het is onduidelijk wat examineratoren met deze adviezen doen. Het doel van dit SKE-project is om te inventariseren hoe examencommissies van andere faculteiten en onderwijsinstellingen de kwaliteit van toetsing borgen en ‘best practices’ te formuleren.

Theoretisch kader

De wet geeft weinig houvast om het borgen van kwaliteit van tentamens en examens in te vullen. Volgens het Toetskader van de Open Universiteit borgt de examencommissie de kwaliteit van toetsen en examens door het uitvoeren van haar verschillende andere wettelijke taken, zoals het verlenen van vrijstellingen, het opstellen van een jaarverslag.[2] Volgens Martens en Moerkerke (2012) moet de examencommissie om deze wettelijke taak uit te voeren in ieder geval controleren of examineratoren deskundig en bekwaam zijn, of alle leerdoelen worden afgetoetst in de verschillende vakken en of zowel de toetsen als de beoordelingen voldoen aan kwaliteitseisen. Daarnaast kan de examencommissie op basis van art. 7.12b lid 1 onder b “richtlijnen en aanwijzingen vastleggen om de uitslag van tentamens en examens te beoordelen en vast te stellen” wat ook

bijdraagt aan de borging van toetskwaliteit.[3] Deze kwaliteitseisen aan toetsing zijn door de EUR in 2019 verder uitgewerkt, namelijk dat toetsen moeten voldoen aan de criteria van transparantie, betrouwbaarheid en validiteit[4]. Hier wordt ook geopperd dat de examencommissie gerichte steekproeven van vakken trekt, om zo te bepalen of de kwaliteit van toetsing op orde is. Verondersteld wordt dat de examencommissie specifiek de instrumenten opvraagt die door examineratoren worden gebruikt, zoals de grading rubric, het antwoordmodel en de toetsmatrijs.[5] Instrumenten zoals KIT-Plus geven examencommissies handvatten om te beoordelen of toetsen voldoen aan de verschillende kwaliteitseisen.[6]

Methode

Om inzicht te krijgen in de werkwijze van verschillende examencommissies is contact gezocht met verschillende secretarissen en/of leden examencommissies. Voor dit onderzoek is gesproken met medewerkers van de Erasmus Universiteit, de Hogeschool Utrecht, de Marnix Academie, de Universiteit van Amsterdam en de Radboud Universiteit. Alle deelnemers zijn geïnterviewd op basis van een aantal vooraf gedefinieerde vragen. Op basis van de antwoorden.

Resultaten

Toetscommissie

Sommige examencommissies hebben deze wettelijke taak gemandateerd aan een toetscommissie. Ook de rol van de toetscommissie staat niet (wettelijk) vast en wordt op allerlei plekken anders ingevuld[7]. Soms is de toetscommissie een subcommissie van de examencommissie en bestaat zij slechts uit leden van de examencommissie, in andere gevallen bestaat de toetscommissie juist niet of slechts voor een deel uit examencommissieleden.

Toetscarrousel

De meeste examencommissies borgen de kwaliteit van toetsing door steekproefsgewijs een aantal vakken op jaarbasis te controleren. Bij deze zogeheten 'toetscarrousel' worden alle (verplichte) vakken systematisch en periodiek door de examencommissie bekeken en beoordeeld. De meeste examencommissies kunnen elke 3 of 4 jaar alle vakken van het curriculum beoordelen, terwijl een enkele examencommissie de accreditatiecyclus volgt en er voor zorgt dat elke zes jaar de vakken uit het curriculum zijn nagelopen. De vorige evaluatie, die een aantal jaar eerder door de examencommissie is gedaan, dient dan als startpunt voor de nieuwe beoordeling om te bezien dat de resultaten zijn overgenomen. Daarnaast gaven verschillende examencommissies soms ook vakken op te nemen op basis van ontvangen signalen, zoals een klacht van een student of een hoog of laag slagingspercentage.

Vooraf borgen?

De meeste examencommissies beschouwen hun borgende taak als iets dat 'achteraf' plaatsvindt: eerst wordt de toets afgelegd, waarna de examencommissie controleert of de kwaliteit in orde is. Voordat het tentamen wordt afgelegd is het aan de opleiding om ervoor te zorgen dat examinatoren voldoende toegerust zijn om een toets te construeren en zien daar geen rol voor de examencommissie. Hierop waren twee uitzonderingen: een lid van een examencommissie van de Universiteit van Amsterdam fungeerde ook als 'toetscoördinator' waarbij zij ook de taak had om voordat tentamens werden afgelegd te controleren of de toetsvragen duidelijk gesteld waren, voldeden aan de richtlijnen van het toetsbeleid etc.

Daarnaast gaf een examencommissie van de EUR aan dat zij soms – op verzoek van de betreffende examinator – juist vooraf meekijken in plaats van achteraf. Dit wordt bijvoorbeeld gedaan bij een nieuw vak of als een examinator voornemens is om de toetsing ingrijpend wil veranderen en graag vooraf het advies van de examencommissie wil ontvangen.

Terugkoppeling resultaten

De meeste examencommissies koppelen de resultaten van de steekproef schriftelijk terug naar de examinatoren, zonder dat er bijvoorbeeld een gesprek plaatsvindt met de desbetreffende examinator. Zij herkennen het 'probleem' dat de brief met aanbevelingen laat naar de examinator wordt verstuurd. Eén examencommissie van de EUR nodigt wel altijd de examinator uit voor een gesprek om de bevindingen te bespreken voordat er een brief met aanbevelingen wordt verstuurd. Andere examencommissies nodigen examinatoren slechts uit als de evaluatie grote zorgen baarde, maar was dit niet de standaardprocedure.

Ook gaan examencommissie heel verschillend om met de terugkoppeling van de resultaten naar andere stakeholders, zoals de opleidingsdirecteur en de Vice Decaan Onderwijs. Sommige examencommissies doen dit standaard, terwijl andere het aan de examinator laten om dit zelf te delen. Examencommissies hechten veel waarde aan de dialoog met het opleidingsmanagement.

Conclusie en aanbevelingen

Uit de inventarisatie is gebleken dat niet één examencommissie eenzelfde werkwijze hanteert en dat iedere examencommissie op eigen manier invulling geeft aan haar wettelijke taak om de kwaliteit van toetsing te kunnen borgen welke past bij haar organisatie en middelen. Wel is gebleken dat de meeste examencommissies een vorm van een toetscarrousel hanteren, waarbij elk (verplicht) vak van de opleiding periodiek door de examencommissie wordt bekeken. Dit is in lijn met het 'course rotation schedule' wat de examencommissie van ESHCC sinds 2022-2023 hanteert.

De procedure zou nog wel kunnen verbeteren door de volgende zaken toe te voegen aan de procedure:

- Het opvragen/opzoeken van slagingspercentages, rendementen en cijfergemiddeldes die inzicht geven of een toets heel goed of heel slecht gemaakt is.

- Tentamens waar de examencommissie klachten over ontvangt opnemen in de steekproef van dat jaar.
- Docenten zo min mogelijk belasten door alleen documentatie op te vragen die het secretariaat niet zelf kan downloaden van Canvas of Ans (zoals een toetsmatrijs).

Ik zie geen toegevoegde waarde van een toetscommissie, als deze als subcommissie van de examencommissie fungeert en alleen uit examencommissieleden bestaat. Daarmee verschuif je slechts taken van de examencommissie naar de toetscommissie. Als de toetscommissie uit andere leden dan examencommissieleden bestaat kan het een duidelijk toegevoegde waarde hebben als:

- De leden van de toetscommissie voldoende toetsdeskundig zijn en de examencommissie kunnen adviseren.
- De leden van de toetscommissie voldoende tijd en middelen toebedeeld krijgen.

De vraag of de examencommissie ook aan borging 'vooraf' moet gaan doen vind ik erg interessant. Het lijkt me niet wenselijk als de examencommissie vooraf toetsvragen gaat controleren en examinatoren daarover gaat adviseren. Hiermee gaat de examencommissie teveel op de stoel van het opleidingsmanagement zitten en pakt zij de rol van het 'zorgen' te veel op. De opleiding zou er goed aan doen om een onderwijskundige of een toetsdeskundige in dienst te hebben die dit kan uitvoeren.

De grootste uitdaging blijft de terugkoppeling van de resultaten. Veel examencommissies kiezen ervoor om de resultaten schriftelijk terug te koppelen. Ondanks dat dit niet expliciet is uitgesproken, is een gebrek aan tijd hiervoor de voornaamste reden. Dit terwijl geschreven feedback waarschijnlijk minder effectief is dan een mondelinge toelichting/een dialoog tussen examencommissie en examinator. Omdat de examencommissie op dit moment ongeveer 30 vakken per jaar evalueert is het niet realistisch om te verwachten

dat een gesprek met iedere examinator kan worden toegevoegd aan de procedure om de kwaliteit van toetsing te borgen.

Wel kan de examencommissie, in samenwerking met het opleidingsmanagement, de volgende zaken overwegen:

- Het aantal FTE van de examencommissie wordt uitgebreid of er wordt een toetscommissie geïnstalleerd en getraind om de procedure uit te breiden met een dialoog met elke examinator.
- Het aantal vakken in de jaarlijkse steekproef wordt verminderd, zodat het haalbaar is om een gesprek met de examinator toe te voegen aan de procedure.
- Niet elke examinator wordt uitgenodigd voor een gesprek, maar alleen die examinatoren wiens evaluatie daar aanleiding toegeeft.
- Er vindt een gesprek plaats met meerdere examinatoren tegelijkertijd. Een voordeel is dat dit minder tijd kost voor de examencommissie. Een nadeel is dat dit mogelijk gevoelig ligt, zeker als één van de examinatoren een minder goede evaluatie heeft gehad.
- Er vindt een gesprek plaats met meerdere examinatoren waar vergelijkbare zorgen waren over de kwaliteit van de toetsing (bijvoorbeeld het ontbreken van een toetsmatrijs).
- Indien gesprekken met de examinatoren niet haalbaar blijken te zijn, moet het wel mogelijk zijn om jaarlijks de resultaten van de steekproef van zowel de toetskwaliteit van vakken als de kwaliteit van de scripties met de opleidingsdirecteur (en Vice Decaan Onderwijs) te bespreken.

Referenties

[1] Art. 7.12b lid 1 WHW

[2] Toetskader Open Universiteit, 9 juni 2020, [Microsoft Word - 140623_Toetskader OpenUniversiteit vastgesteldCvB.docx](#)

[3] Martens, R. en Moerkerke G. (2012) 'Ontwikkelen van Toetsbeleid' in 'Toetsen in het Hoger Onderwijs, p.38

[4] Handreiking examencommissies, Kerntaak 2 en 3 – Borging van de kwaliteit van tentamens en examens, Erasmus Universiteit Rotterdam, oktober 2019

[5] Ibid.

[6] Hogeschool Utrecht (2019) KIT-Plus Borgingsinstrument voor examencommissies, [Borgingsinstrument examencommissies KIT-plus | toetsing](#)

[7] Van de Veen, Pol en Moerkerke (2018), De toetscommissie: borgen of zorgen?, Hoger Onderwijs Management, p. 10-14

Alex over het SKE-traject: “ De SKE is een leerzaam traject, waarbij je echt uitzoomt en naar de kwaliteit van de hele toetsorganisatie kijkt. Het meest waardevol vond ik de verschillende perspectieven van de deelnemers. Hierdoor krijg je een goede discussie over toetskwaliteit en kun je van elkaars best practices leren.”



AN INTEGRATED AND COORDINATED APPROACH TO ASSESSING GENAI KNOWLEDGE AND SKILLS

KOEN MIGCHELBRINK

Erasmus School of
Social and
Behavioural Sciences



Introduction

Since the launch of ChatGPT in late December 2023, programme managers and exam boards have discussed the consequences of Generative Artificial Intelligence (GenAI) for the validity of assessment (Cotton et al., 2024; Dempere et al., 2023). Because of the ready availability of high quality GenAI tools such as ChatGPT, the unethical or plagiaristic use of GenAI can reduce the validity of non-proctored forms of written assessments as a measure of students' attainment of intended learning outcomes (ILOs) (Dawson et al., 2024; Goes & Van der Lienden, 2024; Herbold et al., 2023; Hutson, 2024). The unethical use of ChatGPT can serve to write high-quality texts, essays, reports, blogpost or articles that are near indistinguishable from students own work (Herbold et al., 2023), and in which the use of GenAI tools is nearly undetectable (Chaka, 2023). According to a 2023 Nature Scientific Report in which student-written reports were compared with GenAI written reports, ChatGPT written essays were rated higher regarding quality than human-written essays, though linguistic characteristics differed. The authors argued that's "educators must act immediately. We must re-invent homework and develop teaching concepts that utilize these AI models in the same way as math utilizes the calculator: teach the general concepts first and then use AI tools to free up time for other learning objectives" (Herbold et al., 2023, p. 1).

In an attempt to mitigate the negative consequences of GenAI tools on the validity of assessment, programme management called on course coordinators to 'ChatGPT-proof' their methods of assessment. This led to at least two unintended side effects (1) course-coordinators unilaterally started to 'ChatGPT-proof' their assessment programme without specific consideration for the consequences of students' attainment of programme ILOs. (2) Programme management and course coordinators inadvertently discouraged the use of GenAI in education altogether, preventing the ethical development of GenAI knowledge and skills among students.

Since then, programme management and course coordinators have also gone through a learning curve regarding GenAI in education and assessment. We have moved away from preventing the use of GenAI in education to doing education with GenAI. In this article, I instead argue that GenAI should receive a dedicated and integrated place within the program curricula. First, modern day professionals use GenAI in most facets of their work. Having the knowledge and skill to use GenAI ethically can be considered a crucial skill for successful professionals. Second, students use GenAI, whether we allow them to or not. Instead of trying to prevent the unpreventable, our curricula should teach students to use GenAI ethically and transparently instead.

The ethical use of GenAI means that students are and remain responsible and accountable for the wording and content of their texts and that students are aware of the broader societal and environmental consequences of GenAI use. Ownership is part of ethical GenAI use. Students are the owners of the texts submit and should check whether the information they use is truly from the sources they cite, and should correctly reference both sources and GenAI use (Cotton et al., 2024; Hutson, 2024).

Furthermore, ethical use of GenAI means that students are aware of the socio-environmental consequences of GenAI and can deal with them in a responsible manner. These issues include environmentally degrading practices (carbon-footprint of GenAI), exploitative labour practices in training and moderating of algorithms, biases in outcomes (racism, sexism, Anglophonic centrism), and privacy (what happens with personal data and copyrighted information?) (Baldassarre et al., 2023; Ferrara, 2024).

Increased coordination on a programme of assessment

An early consequence of the introduction of GenAI in higher education was the unilateral ChatGPT-proofing of course-level methods of assessment without specific consideration for students' ability to attain programme-level ILOs. A programmatic approach to assessment can help remedy this problem and help think about how to organize a more coordinated and integrated programme of assessment. Using elements of programmatic assessment allows for the increase of central coordination without the need to implement programmatic assessment as such. Instead of solely organizing assessment along the lines of individual courses and empowering individual course coordinators to develop assessment methods as they see fit, programmatic assessment offers the possibility to approach assessment programmatically, coordinating assessment across individual courses with the aim of optimally preparing students to attain programme ILOs (Govaerts et al., 2022; Van Der Vleuten et al., 2012). This does not mean that programme management should dictate what methods of assessment should take place in each course, but it does require programme management to coordinate and interactively design assessment programmes across courses with course coordinators.

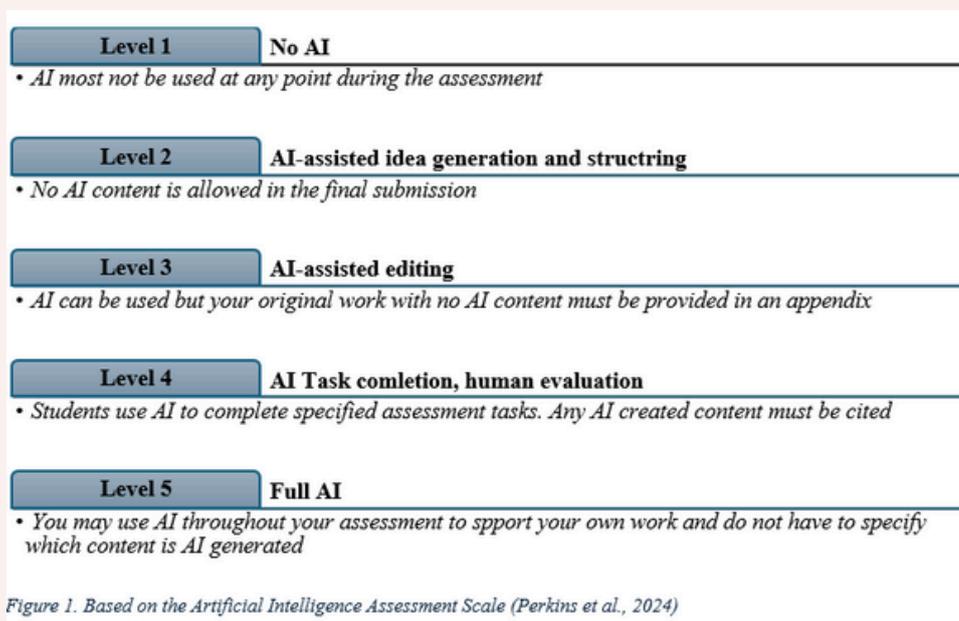
As discussed by Govaerts et al., (2022) in their article on successful implementation of programmatic assessment, *"the successful implementation of programmatic assessment requires strong leadership and central governance"* (p. 4). A strategic outlook on what the programme of assessment should look like is crucial and can be developed under the guidance of programme management, but in interaction with course coordinators. The successful implementation of a more integrated and coordinated programme of assessment requires top-down leadership and coordination, and bottom-up involvement and cocreation with all course-coordinators. This can provide essential buy-in and legitimacy among course-coordinators, and provides crucial experienced knowledge about what assessment is fit for purpose and applicable in specific educational and course settings (Govaerts et al., 2022; Van Der Vleuten et al., 2012).

Including GenAI in education and assessment

At the same time, increased coordination on assessment is not enough. Programme directors should actively include the teaching and assessment on GenAI knowledge and skills in their programme curricula in order to optimally prepare students for practice (Herbold et al., 2023). The Artificial Intelligence Assessment Scale (AIAS) (Perkins et al., 2024) provides a framework on how knowledge, skills and competencies regarding GenAI can be developed and assessed. The scale helps programme management and course coordinators to think about how to attain constructive alignment within and across courses, and helps programme management to design an integrated assessment programme on how to achieve relevant GenAI knowledge, skills and competencies across the educational programme (Hecker et al., 2024). Furthermore, from the perspective of assessment for learning (as well as from the perspective of assessment as learning), the AIAS can also help explain to students how and when GenAI can be used in line with principles of academic integrity (Perkins et al., 2024).

The Artificial Intelligence Assessment Scale (AIAS) (Perkins et al., 2024) is depicted in Figure 1. At level 1, students learn about GenAI, but are not allowed to use it in any form. This level is appropriate for assessment tasks in which students are expected to solely rely on their own understanding, knowledge or skills. This can include technology free exams or idea development, but also ad-hoc or planned oral exams and formative assessments. At level 2, students are allowed to use GenAI to generate ideas, brainstorm, and gain feedback. This level is appropriate for helping students develop ideas and forms of formative assessment, but in which the final product must be solely student authored. This level of GenAI involvement allows students to explore a wider range of ideas and to experiment with GenAI tools. At level 3, students are allowed to use GenAI to edit their texts.

At this level, students are allowed to use GenAI in grammar, punctuation and spelling, word choice, and structural or visual editing of their original documents in order to enhance the language or content of their original work. However, the argumentative content must still be students' own original work. Therefore, students should be obliged to hand in an un-edited version of their document alongside an AI-assisted version to ensure the authenticity of their own contribution. At level 4, students are encouraged and expected to use GenAI to complete (a specific part) of a learning assignment, while the emphasis remains on human evaluation and interpretation of the AI-generated content. This level of GenAI engagement encourages a deeper understanding of the capabilities and limitations of GenAI. At level 5, students use GenAI to complete their learning assignment at the own discretion. Students cocreate with GenAI tools in completing learning assignments. At this stage, students are expected to use GenAI at their own discretion (think about creating an artwork or writing a software) at which the attainment of skills or knowledge can be tested irrespective of the use of GenAI. (Perkins et al., 2024).



Using the AIAS, course coordinators can design, explain and organize whether and how GenAI can be used. The AIAS can be used to constructively align programme-level ILOs with learning activities and assessments across the three years of the bachelor's programme, as well as between course-level ILOs, activities and assessment (Biggs 1996; Hecker et al., 2024). The use and application of GenAI should be explicitly integrated into course and programme-level ILOs, it should be part of the learning activities and should be tested using appropriate methods of assessment. Preferably, ILOs, learning activities and assessment are aligned at the same AIAS level to ensure optimal alignment. Development across levels can be aligned across the years of the educational program, for example level 1 in year 1, levels 4 and 5 in year 3. Clear communication to students about whether and how GenAI can be used in the completion of formative and summative assessments is crucial to the application of the AIAS (Goes & Van der Lienden, 2024; Perkins et al., 2024).

Furthermore, from a programmatic assessment perspective, course coordinators are encouraged to think about assessment based on multiple assessment activities, focussing on low stakes assessment methods including GenAI and other tasks, which together form the basis for high-stakes assessment like passing a course. Course coordinators are encouraged to prevent single-data-point assessment bias and increase inclusivity of assessment by using multiple form of assessment to test the attainment of similar ILOs.

An integrated model

The approach of programmatic assessment allows programme directors and course coordinators to collaborate in a top-down and bottom-up process in developing an AIAS-inspired assessment programme at increasingly more advanced competency levels. One way to constructively align learning goals, activities and assessment at progressive levels of competency is using Miller’s (1990) pyramid of skills, competences and performance (Witheridge et al., 2019). Miller’s pyramid is based on the idea that not only theory-based knowledge, but also skills, competencies and performance should be tested. This type of learning and assessment is particularly valuable for GenAI use. Together, the AIAS (Perkins et al., 2024) and Miller’s Pyramid for assessment of skills, competences and performance (Miller, 1990) provide a framework on how student’s GenAI skills, competences and performance can be tested at progressively more advanced levels. This framework is presented in Figure 4 and provides programme directors with a tool to reassess GenAI directed programmatic assessment throughout the BA programme.

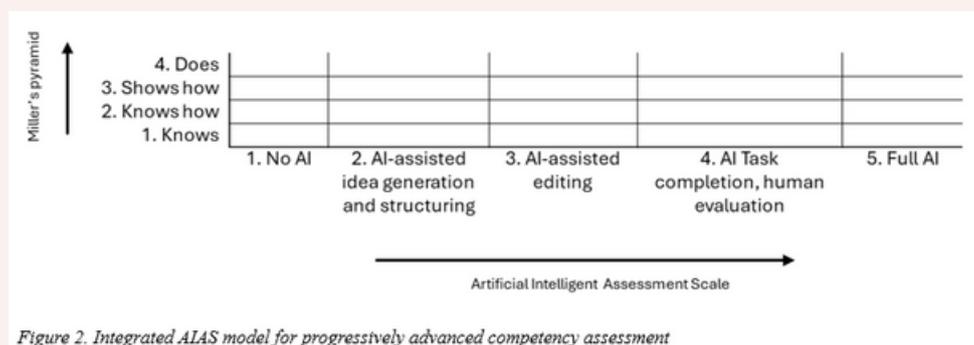


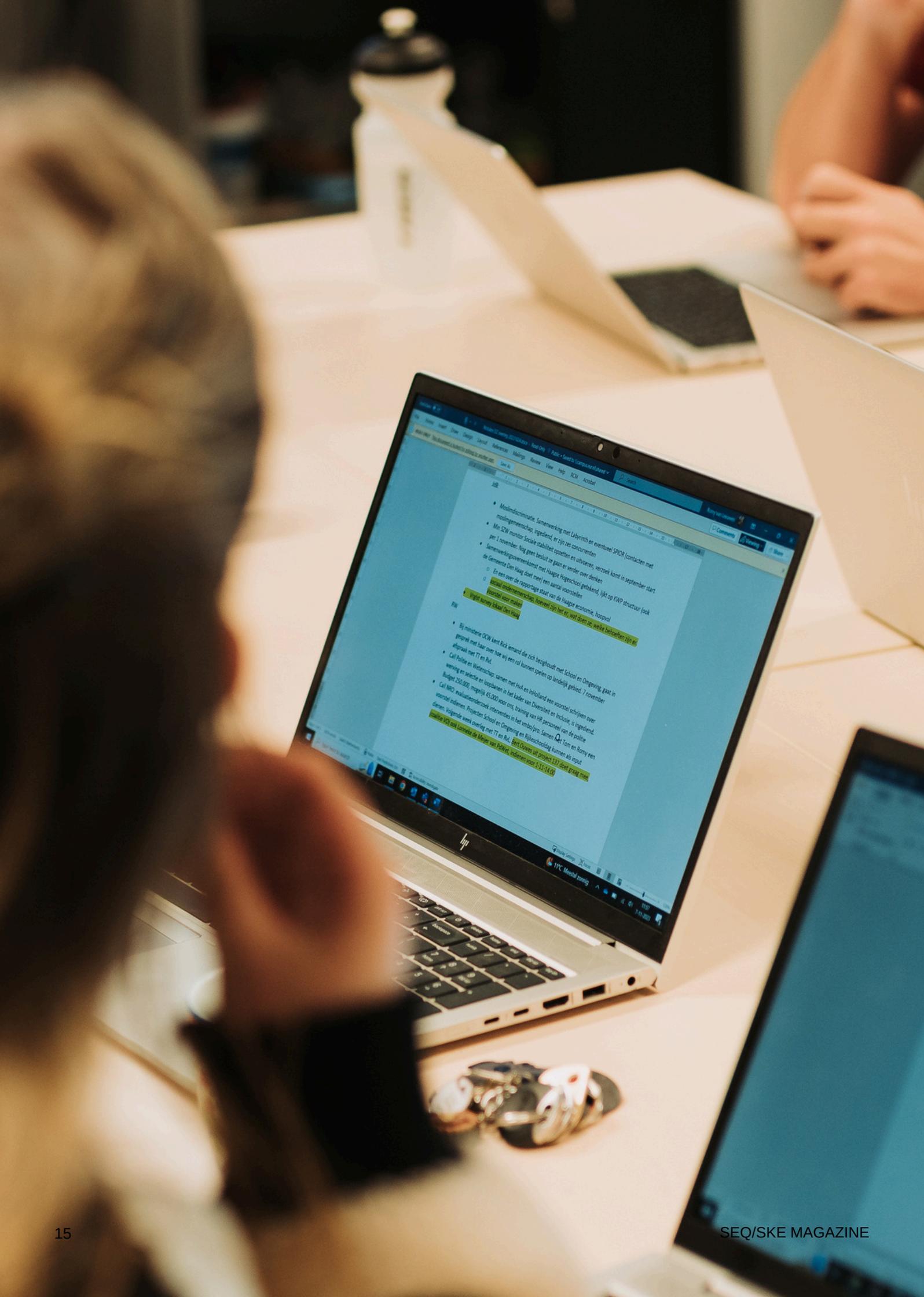
Figure 2. Integrated AIAS model for progressively advanced competency assessment

The model provides at least two clear benefits to the design of GenAI assessment for the bachelor's programme public administration. First, the model allows for different levels of the AIAS to be tested at different levels of skills, competences and performances depending on the ILOs. For example, AI-assisted idea generation and structuring can be tested at the level of knows (discuss what this is), at the level of knows how (discuss how this can be done), or at the level of shows how (show how this can be done) or at the level of does (make an assignment in which this is done). The same is true for the other levels of the AIAS scale. Second, the model allows for programme directors to develop a programmatic assessment programme in which progressively more advanced steps in the AIAS are tested in progressively higher levels of Miller’s pyramid across program-years and in courses. For example, a single level of the AIAS can be programmatically tested using different low-stake assessment methods for each of the levels of Miller’s pyramid. This way, the AIAS can be progressively integrated into a programme of assessment fit for purpose. Importantly, the benefits of this system rely in part on the ability of the programme director to co-design and manage this system of programmatic assessment with the separate course coordinators in the programme.

Conclusion

The rise of GenAI in higher education has presented programme managers and exam boards with challenges regarding the validity of assessments, in particular written, non-proctored forms of assessment. In response, programme managers have discouraged the use of GenAI in education called on course coordinators to ‘ChatGPT proof’ their courses. Since then, students, teachers and programme management have been able to get used to the use and practice of GenAI in education and a next step in preparing our students for practice is required. In this text, I argued for a more integrated and coordinated approach to assessment, in which the use and application of GenAI knowledge and skills is integrated throughout programme curricula. The Artificial Intelligence Assessment Scale, in combination with Miller’s Pyramid of skills, competences and performance provides a way to do this. Including GenAI in higher education and teaching students to use GenAI in an ethical and constructive way is crucial in preparing them for a successful career. Teaching and assessing GenAI knowledge and skills should therefore be an integrated and dedicated part of all undergraduate programs at our university.

- Baldassarre, M. T., Caivano, D., Fernandez Nieto, B., Gigante, D., & Ragone, A. (2023). The Social Impact of Generative AI: An Analysis on ChatGPT. *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, 363–373. <https://doi.org/10.1145/3582515.3609555>
- Biggs J. (1996). Enhancing Teaching Through Constructive Alignment. *Higher Education*, 32(3), 347–364.
- Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning & Teaching*, 6(2). <https://doi.org/10.37074/jalt.2023.6.2.12>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016. <https://doi.org/10.1080/02602938.2024.2386662>
- Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education*, 8, 1206936. <https://doi.org/10.3389/educ.2023.1206936>
- Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7(1), 549–569. <https://doi.org/10.1007/s42001-024-00250-1>
- Goes, H., & Van der Lienden, K. C. (2024). *Betekenisvol gebruik maken van chatbots in toetsing in het hoger onderwijs*.
- Govaerts, M., Van Der Vleuten, C., & Schut, S. (2022). Implementation of Programmatic Assessment: Challenges and Lessons Learned. *Education Sciences*, 12(10), 717. <https://doi.org/10.3390/educsci12100717>
- Hecker, K., Danielson, J. A., Van der Vleuten, C., & Bok, H., G. J. (2024). Constructive Alignment and Programmatic Assessment. In J. L. Hodgson & J. M. Pelzer (Eds.), *Veterinary Medical Education: A Practical Guide* (2nd ed.). John Wiley & Sons Inc.
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(1), 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Hutson, J. (2024). Rethinking Plagiarism in the Era of Generative AI. *Journal of Intelligent Communication*, 4(1). <https://doi.org/10.54963/jic.v4i1.220>
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance: *Academic Medicine*, 65(9), S63-7. <https://doi.org/10.1097/00001888-199009000-00045>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>
- 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>
- Witheridge, A., Ferns, G., & Scott-Smith, W. (2019). Revisiting Miller's pyramid in medical education: The gap between traditional assessment and diagnostic reasoning. *International Journal of Medical Education*, 10, 191–192. <https://doi.org/10.5116/ijme.5d9b.0c37>



MSB

- Moelndiadministrie: Samenwerking met Labyrinth en eventueel SPOR (contacten met moelndiadministrie, ingediend, er zijn res concurrenten)
- Mo. SW monitor Sociale stabiliteit opzetten en uitvoeren, verzoek komt in september start per 1 november. Nog geen besluit te paas er vorder over denken
- Samenwerkingsovereenkomst met Haagse Hogeschool getekend, lijkt op KWP structuur (ook de Gemeente dien laag doel mee) een aantal voorstellen
 - o En een over de rapportage staat van de Haagse economie, hoopvol
 - o **aanval ondernemerschap, hoeveel zijn het er, wat doen ze welke behoeften zijn er**
 - o **aanval voor markt**

RW

- Bij ministerie OCW kent Rick iemand die zich bezighoudt met School en Omgeving, gaat in gesprek met haar over hoe wij een rol kunnen spelen op landelijk gebied. 7 november afspraak met TT en RA.
- Call Politie en Wetenschap: samen met HVA en Infoland een voorstel schrijven over weging en selectie en loonbanen in het kader van Diversiteit en Inclusie, is ingediend. Budget 250.000, mogelijk 45.000 voor ons, training van HR personeel van de politie
- Call NRC evaluatieonderzoek interventies in het veld/pros. Samen met Tom en Rommy een voorstel indienen. Projecten School en Omgeving en Eijkendoodslag kunnen als input dienen. Volgende week overleg met TT en RA. **Best Overtes uit project 137 doet graag twee**
- **realise 100 gis Lomada de Mijne van Politie, indienen voor 1-31-14-15**

INTEGRATIE VAN ONDERWIJS OVER METHODEN EN OPZETTEN VAN WETENSCHAPPELIJK ONDERZOEK IN HET INHOUDELIJKE CURRICULUM: KANSSEN EN UITDAGINGEN

Erasmus School of
Social and
Behavioural Sciences



RIANNE KOK

Binnen alle sociaalwetenschappelijke opleidingen is kennis over methoden en opzetten van wetenschappelijk onderzoek en de vaardigheid om wetenschappelijk onderzoek te verrichten cruciaal, zoals ook beschreven in de Dublin descriptor 1 en 2 (Kennis en Inzicht; Toepassen van Kennis en Inzicht). Dit vertaalt zich in een verankering hiervan in de eindtermen, bijvoorbeeld bij Pedagogische Wetenschappen:

“Afgestudeerde studenten hebben kennis over geavanceerde opzetten en methoden voor sociaalwetenschappelijk onderzoek”;

“Afgestudeerden zijn in staat om wetenschappelijk onderzoek uit te voeren binnen de gekozen specialisatie en op het raakvlak met andere gerelateerde sub disciplines”.

Maar ook bij andere sociaalwetenschappelijke opleidingen worden vergelijkbare eindtermen omschreven, zie Tabel 1.

In dit SKE-project stond de verankering van deze eindtermen in het inhoudelijk curriculum van Pedagogische Wetenschappen centraal, omdat daarmee een betere aansluiting mogelijk was tussen methodologische en statistische kennis enerzijds en de onderzoekspraktijk en beroepspraktijk anderzijds. In dit artikel wordt de term “methoden onderwijs” gebruikt om te verwijzen naar het onderwijs waarin studenten kennis opdoen over methoden en opzetten van wetenschappelijk onderzoek en het onderwijs waarin zij de vaardigheid aanleren om wetenschappelijk onderzoek te verrichten.

Het nut van integratie van methoden onderwijs in het curriculum.

Hoewel in een aantal masteropleidingen het methoden onderwijs in een apart studieonderdeel wordt verwerkt (bijvoorbeeld in een los vak dat gericht is op statistiek en onderzoeksmethoden), wordt bij andere opleidingen er bewust voor gekozen om het methoden onderwijs te integreren in het inhoudelijke curriculum. Er zijn verschillende redenen waarom een integratie van methoden onderwijs in het inhoudelijke curriculum van een masterprogramma relevant kan zijn.

De integratie van het methoden onderwijs in het inhoudelijke curriculum biedt de mogelijkheid voor studenten om hun kennis en vaardigheden op het gebied van wetenschappelijk onderzoek direct toe te passen in het vakgebied waarin zij worden opgeleid. Vanuit het idee van integrative education, integrated curriculum (Shoemaker, 1989) of integrated learning (Huber & Hutchings, 2004) is het belangrijk om het onderwijs zo te organiseren dat er betekenisvolle connecties gemaakt kunnen worden binnen een vak, in plaats van het aanbieden van aparte onderwerpen in aparte vakken. Door deze aanpak leggen wij de regie en verantwoordelijkheid niet bij studenten om de verschillende puzzelstukken met elkaar te combineren, maar kunnen docenten hen actief begeleiden in het integreren van de kennis en vaardigheden op gebied van methoden in het vakinhoudelijke werken. Deze integratie heeft verschillende voordelen (Huber & Hutchings, 2004). Ten eerste, biedt integratie de mogelijkheid om kennis en vaardigheden direct toe te passen in de wetenschappelijke dan wel beroepspraktijk.

Tabel 1

Overzicht van eindtermen gerelateerd aan kennis en vaardigheden van methoden en opzetten van wetenschappelijk onderzoek binnen de masteropleidingen van ESSB, EUR.

Pedagogische W.	Psychologie	Sociologie	Bestuurskunde
Afgestudeerden hebben kennis over geavanceerde opzetten en methoden voor [...] onderzoek.	Afgestudeerden hebben kennis over [...] wetenschappelijke onderzoeksmethoden en statistiek op een geavanceerd niveau	Het hebben van diepgaande kennis en inzicht in [...] onderzoeksmethoden.	n.v.t.
Afgestudeerden zijn in staat om [...] wetenschappelijk onderzoek uit te voeren binnen de gekozen specialisatie en op het raakvlak met andere gerelateerde subdisciplines van de pedagogische wetenschappen.	Afgestudeerden zijn in staat om met een hoge mate van zelfstandigheid en met inachtneming van de ethische en wettelijke gedragsregels wetenschappelijk onderzoek op te zetten, uit te voeren en de bevindingen te rapporteren volgens de gangbare conventies van het vakgebied. Psychologisch en wetenschappelijk werk uit te voeren [...]	Zelfstandig en op wetenschappelijk verantwoorde wijze een onderzoek kunnen verrichten [...].	Can design and conduct advanced qualitative, and/or mixed methods [...] research
n.v.t.	n.v.t.	In staat zijn een wetenschappelijk standpunt te betrekken, maar deze ook kritisch op waarde kunnen schatten.	Can critically and extensively assess public administration research.

Ten tweede, is de kans groter dat kennis en vaardigheden beklijven, als deze kennis en vaardigheden een duidelijkere betekenis krijgen voor studenten. Ten derde, kan de directe toepassing en connectie met het inhoudelijke vakgebied ook bijdragen aan de motivatie van studenten om in dit, voor sommigen meer complexere, studieonderdeel te investeren. Dit laatste punt wordt onder andere ondersteund door de expectancy-value theory (Wigfield & Eccles, 2000), waarin wordt gesteld dat de motivatie van iemand om een taak te doen groter is op het moment dat iemand deze taak nuttiger acht voor zijn of haar toekomstige doelen.

Het belang van een eenduidige operationalisatie van eindtermen.

In dit SKE-project is een analyse gemaakt van het inhoudelijke curriculum van Pedagogische Wetenschappen, waarbij een overzicht is gemaakt van de inhoud en werkvormen in het onderwijs, de toetsing en de leerdoelen per studieonderdeel en de daaraan gekoppelde eindtermen. Uit deze analyse bleek onder andere dat er onduidelijkheid of inconsistenties waren in de interpretatie van de betreffende eindtermen. Deze verschillen in interpretatie hebben implicaties voor zowel de validiteit als de betrouwbaarheid (Baartman et al., 2014). Ten eerste, leiden de verschillen in interpretatie ertoe dat de beoordeling welke eindtermen door een vak worden afgedekt wisselt tussen examinatoren, wat een probleem is voor de betrouwbaarheid van deze beoordeling. Ten tweede, zijn de verschillen tussen de examinatoren een gevaar voor de validiteit van de eindterm en daarmee het onderwijs zelf: leiden wij studenten op voor hetgeen we ze willen opleiden? Ten derde, is de betrouwbaarheid van het Toetsplan in het geding, met name het overzicht waar de eindtermen in het curriculum worden getoetst (de zogenaamde 'kruisjestabel'). Een duidelijkere operationalisatie is dus cruciaal.

Afgestudeerde studenten hebben kennis over geavanceerde opzetten en methoden voor sociaalwetenschappelijk onderzoek.

In de formulering van bovenstaande eindterm ligt de focus op het verhogen en versterken van de inhoudelijke kennis van studenten over opzetten en methoden van wetenschappelijk onderzoek (gelinkt aan Dublin descriptor 1). Cruciaal voor deze eindterm lijkt enige mate van kennisoverdracht of kennisverrijking bij studenten. Dit kan verschillende vormen aannemen. In het traditionele (niet-geïntegreerde) methoden en statistiekonderwijs wordt de kennis aangereikt in colleges en werkgroepen en via het bestuderen van literatuur, maar meer in abstracte vorm en los van de wetenschappelijke praktijk. In het inhoudelijk curriculum zijn er kansen om deze kennis te vergroten door bij het bestuderen van de wetenschappelijke literatuur bij een thema aandacht te besteden aan de aard van de methoden en de implicaties hiervan voor de interpretatie van de resultaten. Dit vraagt ofwel een actieve rol van een docent die studenten hierin kan begeleiden en verantwoordelijk is voor de kennisoverdracht in het onderwijs, ofwel gerichte literatuur die aansluit bij de methoden die in de literatuur terugkomen. Bijvoorbeeld, er kunnen specifieke hoofdstukken of paragrafen uit een methodenboek (bv. Passer, 2014) worden gekoppeld aan de literatuur die wordt voorgeschreven die studenten helpt deze methoden te begrijpen en te duiden. De operationalisatie van 'opzetten en methoden' is daarnaast relevant. Onder opzetten kunnen we de verschillende onderzoeksdesigns verstaan die relevant zijn voor het sociaalwetenschappelijk domein. Mogelijke methoden en opzetten zijn (gebaseerd op Passer, 2014): kwantitatief versus kwalitatief onderzoek; experimenteel versus beschrijvend onderzoek; laboratorium versus veldonderzoek; longitudinaal versus cross-sectioneel onderzoek, literatuurstudie (meta-analyse, systematische review), N=1 onderzoek of casestudies en interventie onderzoek (bv. *randomized-controlled trials*).

De term methoden kan betrekking hebben op zowel de verschillende methoden van dataverzameling als de verschillende (statistische) methoden van dataverwerking. Dataverzameling kan gebeuren via vragenlijsten, vrije of (semi-) gestructureerde interviews, focusgroepen, vrije of systematische gedragsobservatie, of bijvoorbeeld experimenten. De methoden van dataverwerking kunnen vervolgens kwalitatief of kwantitatief van aard zijn. Binnen deze brede stromingen zijn er specifieke vormen van methoden, zoals thematische analyse of een grounded theory benadering voor kwalitatief onderzoek, of structural equation modelling of multilevel analyse bij kwantitatief onderzoek.

Het lijkt nuttig om binnen elke wetenschappelijke discipline (bv. Pedagogische Wetenschappen), maar ook over disciplines heen binnen de faculteit, een overzicht te maken van de opzetten en methoden van wetenschappelijk onderzoek die op dit moment als relevant en/of essentieel worden beschouwd voor de studenten. Dit biedt niet alleen houvast voor het bepalen van de focus in onderwijs en toetsing, maar draagt daarnaast bij aan een meer uniforme interpretatie van de reikwijdte van deze eindterm.

Afgestudeerden zijn in staat om wetenschappelijk onderzoek uit te voeren binnen de gekozen specialisatie en op het raakvlak met andere gerelateerde sub disciplines.

Deze eindterm is gericht op het uitvoeren van wetenschappelijk onderzoek, maar ook hier is de vraag hoe breed of hoe smal we deze eindterm willen interpreteren. De afstudeerscriptie is de proeve van bekwaamheid voor deze eindterm, waarin zoveel mogelijk de hele empirische cyclus wordt doorlopen. Echter, ook als delen van de empirische cyclus worden doorlopen, kan deze eindterm relevant zijn. Immers, om een betrouwbaar oordeel te kunnen vellen over de competentie van de studenten is het van belang om voor elke eindterm meerdere toetsmomenten te hebben (Baartman et al., 2014). In de breedte bestaat het uitvoeren van wetenschappelijk onderzoek uit veel verschillende aspecten: bv. de formulering van een onderzoeksvraag, de opzet van het onderzoek, de verzameling van de onderzoeksgegevens, de analyse van de onderzoeksgegevens, de interpretatie van de resultaten en de formulering van hypothesen voor toekomstig onderzoek. Door een overzicht te maken van de verschillende onderdelen waaruit wetenschappelijk onderzoek in de betreffende discipline bestaat (bv. met als leidraad de empirische cyclus) wordt het mogelijk om deze eindterm ook meer fragmentarisch te kunnen toetsen, naast het sluitstuk, de afstudeerscriptie, waar alle onderdelen samenkomen.

Diversiteit in toetsingsmogelijkheden.

Om de eindtermen op een passende manier te toetsen, is het van groot belang om de reikwijdte die hierboven is beschreven ook terug te laten komen in een diversiteit aan toetsingsmogelijkheden (Baartman et al., 2014). Een belangrijk aandachtspunt hierbij is dat er zorgvuldig moet worden bekeken of er voldoende aandacht is voor de diverse aspecten die hierboven zijn beschreven, bv. voor zowel kwalitatieve als kwantitatieve onderzoeksliteratuur. Vervolgens moeten deze aspecten voldoende worden verankerd in toetsing, zodat de kennis van geavanceerde methoden en opzetten (eindterm 5) en de toepassing ervan in wetenschappelijk onderzoek (eindterm 8) op voldoende momenten wordt getoetst. Er zijn verschillende manieren van toetsing mogelijk. Het is van groot belang om hierin niet alleen uit te gaan van summatieve toetsing, maar ook de mogelijkheden en voordelen van formatieve toetsing te benutten (Schellekens et al., 2021). Onderdeel van de analyse voor dit SKE-project van het inhoudelijke curriculum van Pedagogische Wetenschappen, was een overzicht maken van de verschillende manieren van toetsing van de eindtermen over methoden van onderzoek. Deze voorbeelden zijn aangevuld met voorbeelden uit de curriculumherziening van het Masterprogramma Orthopedagogiek (in ontwikkeling, start per 2025/2026).

Tabel 2

Mogelijkheden om eindtermen te toetsen o.b.v. curriculum Pedagogische Wetenschappen.

Vorm	Voorbeeld	Eindterm*
Kennis reproduceren	Multiple choice vragen of open vragen over specifieke methoden	5
Kennis toepassen	Uitvoeren van en rapporteren over analyses in een losse opdracht. Wetenschappelijk gefundeerde tekst schrijven met gebruik van bronnen.	5, 8
Participeren	Bijdrage van een student aan dataverzameling van een onderzoeksproject.	8
Kritisch reflecteren	Reflectie op kwaliteit van wetenschappelijke bronnen in een journal club. Presenteren van een onderzoeksvoorstel.	5, 8
Ontwerpen	Schrijven van een onderzoeksvoorstel bij een onderzoeksvraag.	8

*Eindterm 5: Afgestudeerde studenten hebben kennis over geavanceerde opzetten en methoden voor sociaalwetenschappelijk onderzoek; Eindterm 8: Afgestudeerden zijn in staat om wetenschappelijk onderzoek uit te voeren binnen de gekozen specialisatie en op het raakvlak met andere gerelateerde sub disciplines.

Door bijvoorbeeld in toetsing via multiple choice of open vragen directe vragen te stellen die betrekking hebben op de specifieke methoden die zijn aangeleerd, kan kennis over deze methoden worden getoetst. Dit lijkt minder passend voor het inhoudelijke curriculum en een nadeel hiervan kan zijn dat de kennis wordt getoetst zonder directe link met de wetenschappelijke praktijk.

Ook de bijdrage van een student aan de dataverzameling van een onderzoeksproject kan getoetst worden, ook wel 'observatie' genoemd in het overzicht van instrumenten van beoordeling in Van der Klink en Boon (2014), gebaseerd op de toetsingspiramide van Miller (1990). Door studenten een actieve rol te geven in dataverzameling dan wel dataverwerking in een lopend wetenschappelijk onderzoek, tonen zij aan dat zij in staat zijn ook dit deel van de empirische cyclus te doorlopen. Kennisoverdracht is in dit geval vrij specifiek gericht op de taak die zij zullen uitvoeren: bv. het interviewen van proefpersonen, het begeleiden van lab experimenten of het observeren van schoolklassen. Beoordeling en/of feedback kan gebaseerd zijn op documentatie, maar zal vaak ook gebaseerd zijn op het direct observeren van de student in een bepaalde setting.

Studenten kunnen ook oefenen in een opdracht met het uitvoeren en rapporteren van analyses. In traditioneel methoden en statistiek curriculum is deze vorm van toetsing een standaard onderdeel, maar er zijn wellicht ook mogelijkheden om dit binnen een inhoudelijk curriculum te integreren, als studenten bijvoorbeeld de kans krijgen om met bestaande data een inhoudelijke onderzoeksvraag te beantwoorden.

Kennis en vaardigheden kunnen ook getoetst worden in de vorm van kritische reflectie op kwaliteit van wetenschappelijke bronnen. In een opdracht dan wel een toets kan deze kritische reflectie van studenten worden gevraagd, waarbij indirect wordt getoetst of studenten de methoden kennen/begrijpen, of ze de mogelijkheden en beperkingen kunnen beschrijven en of ze deze kunnen gebruiken om de resultaten op waarde te schatten.

Naast het eindwerk, kunnen studenten ook al eerder in het curriculum oefenen met het schrijven van een wetenschappelijk gefundeerd stuk met gebruik van bronnen. In een opdracht dan wel een toets zou deze wetenschappelijke onderbouwing gebruikt kunnen worden om een beeld te krijgen van de mate waarin een student in staat is om de aard van het design en de methoden mee te nemen in de interpretatie van de resultaten. Bijvoorbeeld, een meta-analyse heeft meer zeggingskracht over een bepaalde associatie dan een los empirisch paper. Dit begrip zou moeten blijken uit de kwaliteit van de integratie van deze bronnen. Dit is een meer indirecte manier van toetsing. Afhankelijk van de vorm van het geschreven stuk, zou deze vorm van toetsing ook relevant kunnen zijn voor het toetsen van de vaardigheid om wetenschappelijk onderzoek uit te voeren, mits het stuk voldoende te vergelijken is met een formele literatuur review, bv. een narratieve review of systematische review.

Ook het ontwerpen van een onderzoeksopzet en/of het schrijven van een onderzoeksvoorstel is een manier van toetsing die passend kan zijn. Deze vorm van toetsing, op een hoger niveau van Bloom, is geschikt om te zien of studenten de kennis over geavanceerde methoden kunnen gebruiken om een passende keuze te maken voor een opzet/methode die aansluit bij hun onderzoeksvraag. Tegelijkertijd is het schrijven van een onderzoeksopzet onderdeel van de empirische cyclus en kan het dus gezien worden als een vorm van uitvoeren van wetenschappelijk onderzoek.

Ook een verdediging of bevraging van onderzoeksopzet of onderzoek kan een vorm van toetsing zijn, waarbij met name de mate van kritische reflectie en begrip kan worden getoetst.

Kansen voor opleidingen om samen te werken.

Er zijn verschillende voorwaarden om een samenwerking in methoden onderwijs over programma's heen op te zetten en succesvol te laten zijn. Een eerste belangrijke voorwaarde is een gedeelde interpretatie en operationalisatie van de eindtermen in de opleidingen met betrekking tot methoden onderwijs. Een tweede voorwaarde is een overzicht van welke kennis over methoden en opzetten van onderzoek en welke vaardigheden van wetenschappelijk onderzoek als cruciaal worden gezien binnen de verschillende opleidingen (opleidingsbreed overleg) en over de verschillende opleidingen heen (faculteitsbreed overleg) voor de vorming van studenten. Hoewel bijvoorbeeld kwalitatief onderzoek een methode kan zijn die voor alle vakgebieden binnen de faculteit relevant is, betekent dit niet automatisch dat elke opleiding deze methode van onderzoek cruciaal genoeg acht om te onderwijzen en toetsen in het onderwijsprogramma. Een dergelijk overzicht is nodig om te zien waar de overlap is en waar men kan en wil samenwerken in het vormgeven van het onderwijs dan wel de toetsing. In een dergelijk overzicht zou ook ruimte moeten zijn om te noteren welke expertise aanwezig is in de verschillende onderwijs-/onderzoeksteams, bij welke individuen.

Vervolgens kan er op facultair niveau gesproken worden over de kansen en (on)mogelijkheden om in het curriculum nauwer samen te werken in onderwijs en toetsing van kennis en vaardigheden op het gebied van wetenschappelijk onderzoek. Er kan bijvoorbeeld een methoden-werkgroep opgericht worden die het overzicht in kaart kan brengen en een advies kan formuleren hoe en waar de opleidingen onderling kunnen samenwerken.

Een belangrijke kanttekening is dat het methoden onderwijs zich zal moeten blijven ontwikkelen en vernieuwen, ook na een initiële analyse. Voor sociale wetenschappen is het bijvoorbeeld wenselijk om ook principes van Open en Responsible Science op te nemen in het methoden onderwijs, gekoppeld aan de internationale beweging richting open science praktijken, zoals preregistratie en open data, en de ethische uitdagingen die nieuwe vormen van onderzoek oproepen, bijvoorbeeld de ethische kanten van citizen science en engaged research. Er moet in het curriculum en in de interpretatie van de eindtermen voldoende ruimte zijn om in te kunnen blijven inspelen op deze nieuwe ontwikkelingen in het wetenschappelijk vakgebied.

Referenties

- Baartman, L., Kloppenburg, R., & Prins, F. (2014). Kwaliteit van toetsprogramma's. In H. van Berkel, A. Bax, & JD. Joosten-ten Brinke (Eds.), *Toetsen in het hoger onderwijs* (pp. 5262). Bohn Stafleu van Loghum, Houten.
- Huber, M.T., & Hutchings, P. (2004). Integrative learning: Mapping the terrain. *The Academy in Transition*. The Association of American Colleges and Universities, Washington.
- Miller, G.E. (1990). The assessment of clinical skills/competence/performance. *Academic medicine*, 65, 63–67.
- Passer, M.W. (2014). *Research methods: Concepts and Connections*. Worth Publishers, New York.
- Schellekens, L.H., Bok, H.G.J., De Jong, L.H., Van der Schaaf, M.F., Kremer, W.D.J., & Van der Vleuten, C.P.M. (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, <https://doi.org/10.1016/j.stueduc.2021.101094>
- Shoemaker, B. (1989). Integrative education: A curriculum for the twenty-first century. *Oregon School Study Council Bulletin*, 33(2). (ERIC No.: ED311602).
- Van der Klink, M., & Boon, J. (2014). Toetsen van werkplekleren. In H. van Berkel, A. Bax, & JD. Joosten-ten Brinke (Eds.), *Toetsen in het hoger onderwijs* (pp. 230-238). Bohn Stafleu van Loghum, Houten.
- Wigfield, A., & Eccles, J.S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81. <https://doi.org/10.1006/ceps.1999.1015>

Pair up & Pitch



Feedback

ASSESSMENT AS AN INSTRUMENT FOR FEEDBACK AND SELF-EVALUATION

EVA BOOMSLUITER

Erasmus
School of
Law



Introduction

Assessments play a central role in higher education. Three main purposes of assessment can be distinguished:

1. Assessment of learning: to evaluate students' performance and ensure they meet the program's final qualifications and the learning objectives of the specific course;
2. Assessment for learning: to provide students with practice tests and assignments as part of the learning process;
3. Assessment as learning: to encourage students to take responsibility for their own learning.

Students place great value on assessments and invest significant time and effort in preparing for them. This strong focus on the test and its outcome makes it a suitable instrument for effectively using feedback and self-evaluation to support students' learning processes (assessment for learning and assessment as learning).

The project "Assessment as instrument for feedback and self-evaluation" provides an overview of the opportunities for utilizing assessments in the context of students' learning development. Two aspects are explored in more detail:

- a. Feedback from the teacher on the assessment, in dialogue with the student.
- b. Self-evaluation and reflection by the student based on the assessment.

a. Feedback in dialogue

Effective feedback must answer three questions (Hattie & Timperley 2007):

1. Where am I going?
2. How am I doing? and
3. What is the next step?

And (4.) effective feedback must be aimed at developing a growth mindset. The feedback should be a two-way dialogue rather than a one-sided process. It must be reflective and encourage students to engage actively. Dialogic feedback supports student autonomy and self-regulated learning.

Effective feedback based on the assessment should therefore meet the following criteria:

Ad 1. Where am I going?

It is essential to formulate and communicate clear learning objectives. Assessments should be aligned with these objectives to help students understand what they have mastered and what still needs development. Feedback must directly relate to the learning goals to be meaningful. Rubrics are a valuable tool in this process. They clarify assessment criteria and describe different levels of performance, helping students understand both the goals and how to achieve them. This approach not only increases the transparency of assessment but also promotes self-regulated learning.



Ad 2. How am I doing?

The feedback must be individual, specifically aimed on the student's performance on the test. It is helpful for students to receive a detailed analysis of their answers, including explanations of mistakes and suggestions for improvement. The feedback should be timely, specific, and constructive, helping students learn from their mistakes and improve their performance.

Ad 3. What is the next step?

Feedback should not only relate to the completed assessment at hand, but it can also serve as a tool for future learning. Feedback on assessments is more effective when it is a cyclical process in which students learn how to improve their performance in the future. Anonymous grading hinders the possibility of organizing the feedback process in a cyclical manner (Winstone & Boud 2020). An important requirement for implementing a cyclical feedback process is the presence of a well-designed curriculum, both at the course and program level. In addition, the feedback provided within this cyclical process must meet the criteria as outlined under 2: it should be timely, specific, and constructive. This ensures that students have the opportunity to actually apply previously received feedback in a focused and meaningful way throughout the curriculum.

Ad 4. Growth mindset

Regarding feedback on the assessment aimed at a growth mindset, the following aspects can be taken into account. First, to refer not only to the mistakes, but also to the success-points. Further, to make the feedback personal. Examples are one-on-one meetings, voice recordings and video feedback. Second, to make it dialogic, engaging and focused on autonomy, giving room for the students to take initiatives and ask questions. Where possible, separate the feedback from the grade, so that the students keep an open mindset. The feedback should furthermore be supportive and development-oriented, in other words safe, positive, non-judgmental, and allowing room for mistakes (Nicol 2010).

b. Self-evaluation and reflection

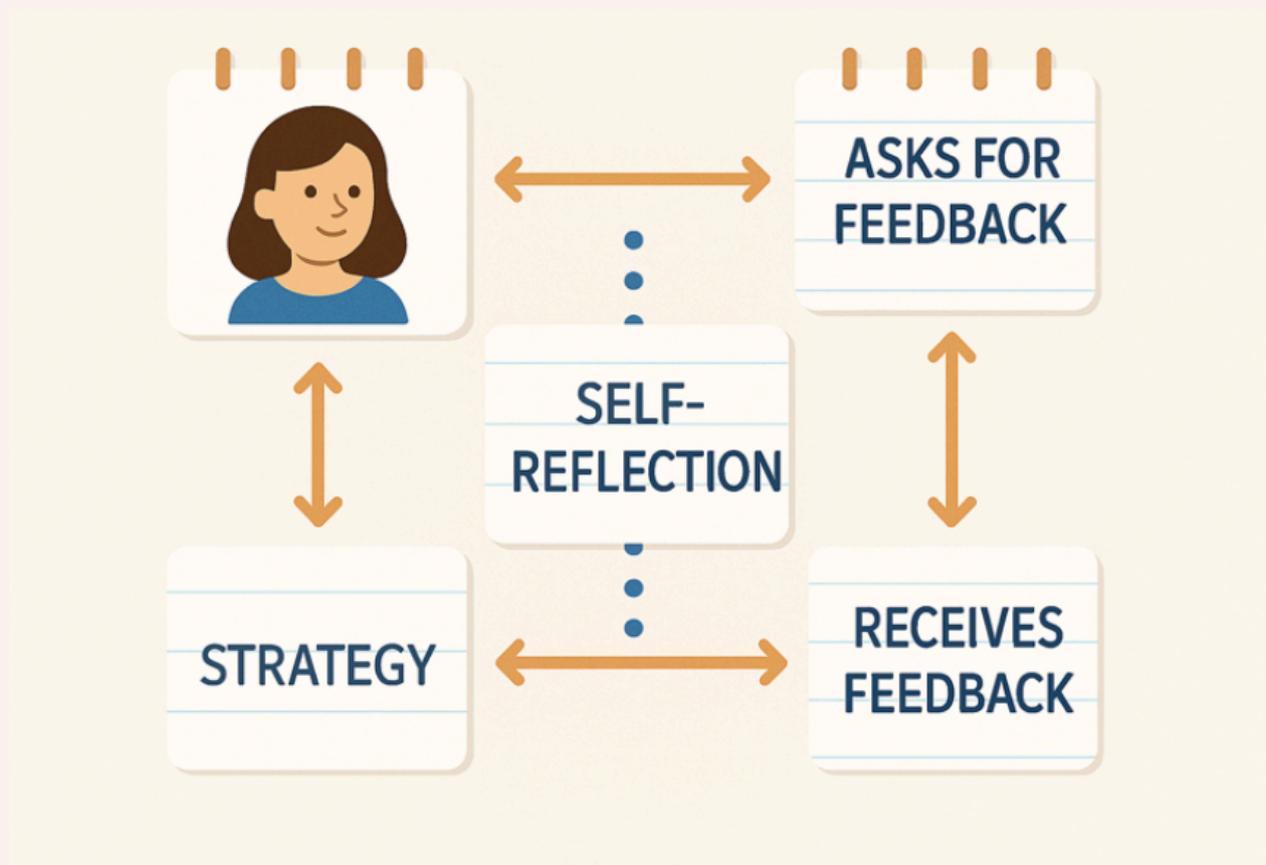
Effective feedback not only informs students about their performance but also helps them develop self-regulation skills. By actively reflecting on the feedback received and linking it to their learning strategies, students can adjust their approach and thereby promote their learning development (Nicol & Macfarlane-Dick 2006).

Self-evaluation and reflection play a crucial role in the learning development of students. By systematically reflecting on their learning outcomes, students gain insight into their strengths and weaknesses, allowing them to take targeted steps to improve their learning process and preparing them for lifelong learning.

An effective way to stimulate self-evaluation is through written reflections, such as learning journals. Reflection helps to organize thoughts and gain deeper insight into the learning process. For example, when students write a reflection report after a test, they can not only analyze their mistakes but also develop strategies to improve their performance.

Applied to the assessment process, two aspects can be distinguished:

- a. Self-evaluation and reflection relating to the assessment; and
- b. (Circular) self-evaluation and reflection throughout the curriculum.



Ad a. Self-revaluation and reflection relating to the assessment

Opportunities for self-evaluation and reflection following the test include:

- Keeping learning journals during the preparation for the test;
- After taking the test (possibly before the results are published), using the assessment model and the rubric to indicate what went well, where mistakes were made, and what the correct answers should have been;
- Writing a reflection report after the test, analyzing both the successful aspects and the mistakes made, and identifying strategies for improvement;
- Writing an analysis of the combination of the learning process in preparation for the test and the result as reflected in the test outcome.

Ad b. (Circular) self-evaluation and reflection throughout the curriculum

Self-evaluation and reflection following the test can be conducted in a cyclical manner by comparing the outcomes of different tests and examining patterns in development. Questions that may be asked in this process include:

- Is there a recurrence of mistakes? What type of errors are most common? This may include, for example: errors in knowledge, skill-based errors, misreading the question, recurring issues in answering (e.g., certain topics, calculation skills, writing skills, skipping steps, failure to mention legal articles, not applying knowledge to the case);
- What is the relationship between test preparation and test outcomes? Consider factors such as: number of study hours, study schedule, timing of self-study sessions, attendance of lectures (in-person or online), whether assignments were completed prior to tutorials, time spent on reading literature, whether or not questions were asked during learning sessions;
- Is there visible improvement over the years? If so, in what areas? What strategies have been employed in achieving this?

Conclusion

Assessments are suitable tools for using feedback effectively to support students' learning processes. Both teacher feedback and student self-evaluation and reflection play an important role in students' learning development. In order for feedback to be effective, several conditions must be met. The timing of the feedback is important, as is a dialogical approach, a safe learning environment, a clear connection to learning objectives, a forward-looking perspective, a cyclical approach, a supportive attitude from the teacher and an active attitude from the student. Effective feedback helps students to develop self-regulation skills. By actively reflecting on the feedback received and linking it to their learning strategies, students can adjust their approach and promote their learning development. These are effective stepping stones for lifelong learning.

Literature

David J. Nicol & Debra Macfarlane-Dick (2006), Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.

John Hattie & Helen Timperley (2007), The power of feedback. *Review of Educational Research*, 77(1), 81-112.

David J. Nicol (2010), From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501-517.

Naomi E. Winstone & David Boud, The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*, 47(3), 656-667.

SAMENWERKEN TUSSEN STUDENTEN OP DE UNI IS KEY! MAAR HOE BEOORDEEL JE DAT?

ERIK ROS

Erasmus
School of
Law



Inleiding

Binnen het hedendaagse onderwijs en de professionele praktijk is samenwerking een essentiële vaardigheid geworden. De toenemende complexiteit van maatschappelijke vraagstukken vereist dat professionals effectief kunnen functioneren in multidisciplinaire teams. Dit geldt in het bijzonder voor de fiscale praktijk, waar juristen, waaronder fiscalisten, intensief samenwerken met collega's uit andere disciplines, zoals accountants, economen en andere juristen. In het onderwijs, en met name binnen academische masteropleidingen, is het dan ook van groot belang dat studenten deze vaardigheid ontwikkelen. Tegelijkertijd is het noodzakelijk dat deze ontwikkeling op een gestructureerde manier wordt beoordeeld. Binnen de masteropleiding Fiscaal Recht aan de Erasmus Universiteit Rotterdam werd geconstateerd dat bij een aantal vakken een formeel beoordelingskader voor samenwerking ontbrak. Ondanks het toenemende belang van samenwerking als kerncompetentie voor fiscaal juristen, werd deze vaardigheid binnen de opleiding onvoldoende gestimuleerd, gewaardeerd en benut als leerervaring. Deze lacune leidde tot de behoefte aan een onderbouwd en praktisch toepasbaar beoordelingsmodel. Het doel hiervan is om samenwerking op een valide, betrouwbare en transparante wijze te kunnen evalueren, zodat studenten gericht kunnen werken aan hun samenwerkingsvaardigheden en deze ook daadwerkelijk worden erkend binnen hun academische ontwikkeling.

Erik over het SKE-traject:

Deelname aan SKE was een game-changer: het versterkte mijn zelfvertrouwen, verdiepte mijn kennis van het onderwerp en zorgde ervoor dat ik oprecht enthousiast was om de klas binnen te stappen tijdens de teachinglabs.

Theoretisch kader

Bij het opstellen van het beoordelingsmodel is gebruik gemaakt van diverse onderwijskundige theorieën en benaderingen die samenwerking in onderwijscontexten onderbouwen:

Sociaal-constructivisme

Volgens het sociaal-constructivisme leren studenten het meest effectief in interactie met anderen. Kennis wordt geconstrueerd in een sociale context, waarbij samenwerking essentieel is. Dit betekent dat samenwerkend leren een krachtige leerstrategie is, mits goed begeleid en beoordeeld.[1]

Coöperatief leren

Coöperatief leren is een didactische werkvorm waarbij studenten in kleine groepen gezamenlijk werken aan een leerdoel. Belangrijke principes zijn positieve wederzijdse afhankelijkheid, individuele verantwoordelijkheid, directe interactie, sociale vaardigheden en groepsreflectie. Deze principes vormen een belangrijke basis voor het beoordelingsmodel.[2]

Teameffectiviteitmodel (Hackman)

Het Teameffectiviteitmodel van Richard Hackman is een bekend model binnen de organisatiepsychologie dat beschrijft welke factoren bijdragen aan het effectief functioneren van teams. Hackman richtte zich vooral op wat teams nodig hebben om goed te presteren, in plaats van alleen te kijken naar de uitkomsten. De belangrijkste elementen van zijn model zijn heldere en uitdagende doelen, gestructureerde groepsprocessen en een ondersteunende omgeving. Het succes van een team wordt niet alleen bepaald door het eindresultaat, maar ook door de mate waarin leden van het team nieuwe vaardigheden leren en plezier hebben in de samenwerking.[3]

Bloom's taxonomy

Bloom's taxonomie biedt een raamwerk om leerdoelen te classificeren van eenvoudige naar complexe niveaus: onthouden, begrijpen, toepassen, analyseren, evalueren en creëren. Voor samenwerking betekent dit dat studenten niet alleen samen informatie verwerken, maar ook leren reflecteren op hun samenwerking en deze verbeteren.[4]

GIPS-model

Het GIPS-model (Groupwork, Individual accountability, Peer assessment, Self-assessment) biedt een concreet beoordelingskader voor samenwerkend leren. Door zowel groepsprestaties als individuele bijdragen te beoordelen, en ruimte te geven voor peer- en self-assessment, ontstaat een eerlijk en motiverend systeem. [5]

Ontwikkeling van het beoordelingsmodel

Op basis van bovenstaande theorieën is een beoordelingsmodel ontwikkeld dat samenwerking inzichtelijk maakt op verschillende niveaus. De samenwerking tussen studenten is beoordeeld op verschillende niveaus om een genuanceerd en volledig beeld te krijgen van hun gezamenlijke functioneren binnen het project/team. Elk niveau richt zich op een ander aspect van samenwerking en draagt met een specifieke weging bij aan het eindoordeel. De beoordeling is opgebouwd uit de volgende niveaus en weging:

1. Observatie docent (40%)
2. Productevaluatie (20%)
3. Peer assessment (25%)
4. Self-assessment (15%)

De wegingspercentages zijn vastgesteld ter illustratie en kunnen naar eigen inzicht van de opleiding of opleidingseenheid worden aangepast, afhankelijk van de context, leerdoelen of het onderwijstype. Elk van deze niveaus is uitgewerkt in concrete beoordelingscriteria die aansluiten bij het niveau van een academische masteropleiding en de beroepspraktijk van fiscalisten.

De criteria zijn onderverdeeld in de volgende hoofdthema's en uitgewerkt in beoordelingsschalen (1-5):

- Communicatie en overleg
- Taakverdeling
- Probleemoplossend vermogen binnen de groep
- Reflectie en leervermogen
- Betrokkenheid

Voor de implementatie van het model zijn drie formulieren ontwikkeld:

1. **Beoordelingsformulier observaties docent.** Dit formulier wordt ingevuld door de docent en richt zich op de beoordeling van de samenwerking en de bijdrage van iedere student aan het gezamenlijke eindproduct.
2. **Peer assessment formulier.** Ieder elk groepslid vult dit formulier in over de andere groepsleden.
3. **Zelfreflectieformulier.** Dit formulier nodigt studenten uit om stil te staan bij hun eigen functioneren en leerproces binnen de groep.

Implementatie en gebruik

Het ontwikkelde beoordelingsmodel voor samenwerking tussen studenten is inmiddels besproken met een aantal vakdocenten binnen de vakken van de master Fiscaal recht waar dit thema actueel is. Over het algemeen waren de docenten positief over de opzet, onderbouwing en toepasbaarheid van het model. Zij gaven aan dat het model inzichtelijk en helder is, en dat het bijdraagt aan het bewust beoordelen van samenwerkingsvaardigheden. Tegelijkertijd werden er enkele praktische aandachtspunten gedeeld. Met name bij het onderdeel observaties door de docent werden kanttekeningen geplaatst. Docenten gaven aan dat hiervoor in de praktijk wellicht onvoldoende tijd en capaciteit beschikbaar is, waardoor dit onderdeel lastig uitvoerbaar kan zijn in de huidige onderwijssetting. Daarnaast werd de suggestie gedaan om de beoordelingsformulieren – en de bijbehorende berekening van het eindcijfer – onder te brengen in een digitale omgeving. Dit zou volgens de docenten het beoordelingsproces aanzienlijk vergemakkelijken en bijdragen aan de efficiëntie van de beoordeling.

Conclusie

Samenwerking is een onmisbare vaardigheid voor studenten Fiscaal Recht, zowel binnen de opleiding als in de beroepspraktijk. Door het ontbreken van een beoordelingskader werd deze competentie tot nu toe onvoldoende ontwikkeld en gewaardeerd. Het nieuwe beoordelingsmodel, gebaseerd op onderwijskundige theorieën biedt een solide basis voor het stimuleren en beoordelen van samenwerking. Door het gebruik van concrete beoordelingsformulieren wordt de beoordeling transparant, eerlijk en leerzaam voor studenten. Het model draagt bij aan de kwaliteitsverbetering van het onderwijs binnen de masteropleiding Fiscaal Recht en sluit aan bij de eisen van de moderne beroepspraktijk waarin samenwerking centraal staat. Het beoordelingsmodel is besproken met vakdocenten binnen de Master Fiscaal recht voor wie de beoordeling van samenwerking actueel is. De intentie is de input van vakdocenten mee te nemen in de verdere verfijning en implementatie van het beoordelingsmodel.

Literatuur

- [1] Zie <https://nivoz.nl/nl/lev-vygotsky-en-de-zone-van-naaste-ontwikkeling-door-samen-te-werken-en-in-gesprek-te-zijn-zijn-mensen-in-staat-tot-dingen-te-komen-die-nieuw-zijn>. Laatst geraadpleegd 11 januari 2025.
- [2] Zie <https://wij-leren.nl/samenwerkend-leren-positief-effect-leerprestaties.php>.
- [3] Zie <https://www.bitesizelearning.co.uk/resources/hackman-model-team-effectiveness>.
- [4] Zie https://quincycollege.edu/wp-content/uploads/Anderson-and-Krathwohl_Revised-Blooms-Taxonomy.pdf.
- [5] Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350 en Lejk, M., & Wyvill, M. (2001).
- [6] Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, 26(1), 61–72.



ASSESSING STUDENT GROWTH: ASSESSMENT ROUNDS IN A PROGRAMMATIC CURRICULUM

VICKI ERASMUS



*“Tussen droom and daad staan wetten in de weg en praktische bezwaren”
Willem Elschot*

In programmatic assessment students study and grow, and (mainly formative) assessment enables them to do so. Feedback on their progress is collected through datapoints, that are the result of students showing their progress in for example projects or practical test skills [1, 2]. This calls for thorough alignment of the entire assessment programme, with a clear planning of which datapoints will be collected during which activities, as well as the necessity for valid and reliable datapoints and feedback [1]. By then assessing students' overall growth at set time points (e.g. twice a year), students' the combination of all these datapoints and feedback can provide a robust view of how a student is progressing through the curriculum. But the step of doing this overall, holistic assessment over all these varying datapoints, is less easy than it might seem, and calls for a different approach compared to a traditional curriculum. An assessment committee (AC) then assessment students' progress twice a year. But how exactly will they reach their evaluation? How easy it is to use the collected datapoints, and how can this process be described in a practice guideline? And last but not least: are the datapoints collected useful enough for the AC to be able to assess growth?

The new medical curriculum at Erasmus MC (EA2030) is based around programmatic assessment. In the first year student take 5 tests on knowledge domains. These tests at the moment have a summative nature (i.e. they receive a pass/fail score for their exam result, per exam), but will in time also become formative (i.e. students will collect datapoints over all exams for specific knowledge domains. Per knowledge domain, over several exams student will assess their knowledge growth, and no longer per exam). Furthermore student collect datapoints on competencies. Throughout the bachelor curriculum students have a VIP (shaping identity and professionalism) coach who they periodically see (9 group lessons through year 1) and who also assesses them on their personal growth (twice a year, individually).

The curriculum is based on students reflecting their growth every 6 months, after which an assessment committee (AC) reviews their entire performance over the previous period: knowledge, competencies (skills) and also personal reflection and growth. The AC then take a decision that can be either: 1) you are doing well; 2) you are not doing that well yet, so work on... or 3) things are really not going well. Please seek help. Prior to the assessment round students have reflected on their growth and formulated a POP (a personal development plan). The result of these assessment rounds is caught in a decision regarding progress: these are considered high stake assessments.

A programmatic curriculum requires different working methods, compared to the traditional programme. In my project I will work on the decisions necessary for the assessment committee and examination committee (EC) to be able to work together ensuring the best support for student learning journeys but also ensuring the quality needed to train good doctors for the future. I supported in formulating a practice guideline for the AC, as well as the form they can use to document their findings. I advised on how to train the AC and ensured connection with the EC.

My last step, which is described in this article in more detail, was the exploration of first experiences by the members of the AC after the first assessment round, Feb 2025. To this end I conducted semi-structured interviews with 4 members of the assessment committee, and one member from the support staff. The study goals were to gain insight into:

- The perceptions on the role of the advisory board and possible influence this might have had on the manner of assessment
- To which extent the previously developed practice guidelines for the assessment committee was feasible and where possible areas for improvement were
- The role of the VIP coach and POP (personal development plan) in the assessment rounds
- What important steps might need to be taken to prepare the assessment committee for the second round of assessment in the summer, as this round would include the binding study advise, and therefore a high stake for students.

The interviews were conducted using a semi-structured interview guide and transcribed verbatim.

The main outcomes are described along a number of themes, which I will link to recommendations for further exploration.

1. The role of the assessment committee

Although all members felt that calibration of the work processes had been performed well, **they did not share the same ideas regarding the function of the assessment committee itself.** Some members felt the assessment committee's most important task was to sift students likely to succeed from student unlikely to succeed (based on their perception). Others felt that the most important task was identify the 'bad apples' from the rest. In both cases the idea was that past results could be an indication of future performance. A different take was that the assessment committee should look only at the results and not make any assumptions as to future performance. The main task of the assessment committee in this case is to provide an unbiased review of past performance.

Interestingly, the members who felt the AC should only provide an unbiased assessment, also indicated the greatest need to more clarity on normative roles and cut-off points. Members who felt the AC should predict or steer future performance indicated this less. Members indicated feeling surprised how some of their peers in the AC conducted assessment, and it was clear that there was no shared method of view on purpose of assessment. **Therefore, calibration on the function of the AC seems paramount.**

It appears at the moment that most calibration had been performed on outcome: so the different members of the assessment committee agree very much on the outcomes of their assessment, but in discussion it turns out that the reasoning they apply to reach these outcomes, can be very different. **Therefore calibration on the process would also be advisable.** One way to reach this goal, would be for each AC duo to explain in detail their process and reasoning for a random student they assessment. This would be a very successful form of post-assessment calibration, but has as yet not been applied. Prior to commencement of the first round, there was no detailed description of the role and function of the AC, other than providing an holistic assessment of students' performance.

2. The practice guidelines, and role of POP and VIP coach

The one aspect all members agree on in unison, is that time had been too tight, both in the period leading up to the first assessment round i.e. preparation phase, as during the assessment round itself. Some also felt that the amount of time for preparation did not align with the importance of this assessment moment.

The members indicated that they generally applied the practice guidelines, although some use the POP earlier to 'get a feel for the student'. Other members only read the POP as the final phase of assessment. Some members felt reading the POP could create bias: it could be difficult to remain neutral towards the student after reading the POP. Additionally, the POP could be very different from student to student. Some members questioned how VIP coaches prepared student for the assessment round. Were student well informed what to include in their POP? Interestingly, although some members of the AC placed much importance on the content and quality of the POP, there is no assessment rubric in place. This could explain the variation in how the POP is used. I would suggest re-investigating the role of the POP, and if important, in some phase an examiner must assess it, before it reaches the AC. In addition, a rubric for evaluating the POP should be developed and shared with students. At the moment, all datapoints are assessed before review by the AC, except the POP, which is only assessed by the AC, but without publishing a result. There is an additional area for concern regarding the use of the POP. As one participant described it: "the assessment is based on knowledge, competencies and personality" with the latter being found in the POP. Some members indicate "looking for signs in the POP, like being able to plan, then I know things are likely to work out". This cannot be in alignment with the faculty assessment plan. Lastly, one member mentioned possible cultural bias when using the POP in the assessment round.

Calibration of what is required from VIP coaches regarding preparing students for the assessment rounds could be necessary. Members now indicate feeling students were prepared in a very different manner.

When it came to the practical application of the practice guidelines, one frequently mentioned problem was what to do with **missing data**. Members reported very different ideas of what missing data meant, and how they dealt with it. Some members tried to look for indication why data might be missing, through other forms of feedback. Is it likely a student missed the data-point collection? Or could it be that the examiner made a mistake or omitted entering data?

Missing data was more likely to be interpreted in a less favourable light. **Regarding quality of assessment, it seems paramount that it is unequivocal for the AC why data is missing.** A no show by a student must be clear. Additionally, it is not possible to enter that data could not be collected or is incomplete. Lastly, narrative data (feedback) must be of a certain quality to make it useful for programmatic assessment. [3] Members mentioned cases where a teacher used the same text for all students, or that feedback is so general as to almost lose meaning. **Alignment between the purpose of feedback, and training of those who must provide it is essential. [4]**

3. Feedback

Members of the AC also shared different views on how they wish to formulate their decision to the student. Some members wish to include detailed feedback on what the student should tackle before the next assessment round, i.e. a coaching role (and did so in the last round). Other members feel this is not part of the function of the assessment round, and that their focus should be on assessment only. It is evident that the role of the assessment rounds needs to be clear cut, both for the members, and for the students. This might help to better understand the role and function of the VIP coach. The role of the VIP coach has been defined coach (in the VIP coach manual) has been described as predominantly coaching, and not assessing by nature. A specific definition such as this is lacking for members of the AC, **it is advisable to define the role of AC and VIP clearly and check for overlap, as this is likely to also raise questions among students.** It is advisable to explore this among students, as well as their views on the feedback they received from the AC. It is perhaps good to note that this feedback is sent to the student only, not to the VIP coach, whose role it is to further coach the student.

Some anecdotal evidence from students indicates that the feedback from the AC does not necessarily add value to the assessment decision. **Further exploration of the function of feedback, and clear guidelines for the AC members on how to formulate their feedback, and importantly what types of this to give feedback on, seem important. Importantly, the choice must be made whether the AC is providing an explanation/justification for their decision, or actual feedback/feedforward for the student. At the moment members of the AC apply this in different manners.**

4. Some challenges ahead

Some members mentioned some concerns regarding the coming assessment round, particularly around the Binded Study Advice (BSA), deciding whether or not a student can pass to year 2 (and where applicable under which conditions). Who is responsible for which part? Members indicate these aspects are not defined well enough in the practice guidelines, and need more attention and calibration for the July assessment round. Members feel “all data must be complete, nothing can be missing, as it’s a much more important round.” Achieving this is very important and places an important role for the programme directors and chair of the AC. Some members indicate that it is unclear how the signals picked up by the AC (e.g. regarding quality issues) are brought under the attention of the programme directors, and what is done with this information.

Furthermore, they also indicate being a bit unclear on how the AC, EC and programme directors work together “who is responsible for what”. Possibly the answers to some of these questions are only being formulated through practice. From a teamwork point of view, **that importance of the leadership view role of the chair of the AC (really being able to keep a complete picture and overview) might be under threat by the necessity to take on a hands-on role in most steps of the assessment rounds.** This could be a consideration as the AC further professionalises in the future.

I have formulated some recommendations to help the AC forward.

1. Define the function of the AC and select members who agree with this. Ensure the chair has time for the important leadership role.

A clear description of what the AC is, and is not, is essential in order for its member to function as one. Just as each new member of the EC must learn what the EC is and isn't for, the same applies for the AC. Learning to also leave your other professional opinions at the door is also important. Every member of the AC also has other teaching and or assessment ideas. There is an important role for the chair and secretary, but also for the EC [1]. Ensure the chair has enough time for the overall decisions/activities, as these form the pillar that the day to day activities

2. Label each assessment round by its stake value

At the moment the assessment rounds are presented as equal to students, i.e. twice a year there is a review by the AC. But in effect the rounds might not be so equal after all. For example, the first review round in year 1, seems like a medium stake assessment at best. There are relatively few data points, with some competencies only having been measure once, and there aren't really any consequences for the student, regardless of their assessment. Better describing the value of each assessment round could help students better prepare, and understand, their holistic assessment, decision and feedback. In addition, it could help the AC as their goal is clearer.

3. Agree on what to do with missing or inferior data and invest in/train teachers to give better data/feedback

Since the feedback students receive is the cornerstone of programmatic assessment, ensuring this feedback is complete and of a quality that it is useful for student and AC is essential. At the moment, feedback is often very general (show yourself a bit more in the group), or highly specific (you disinfect your hands for too short a period of time). The general feedback is probably not very useful at all, and teachers need to be trained what type of feedback is useful. Feedback literacy is required from both teachers and students [1]. Specific feedback can be useful for the student, but not for the AC. The AC ideally needs feedback at a slightly higher level, regarding growth on the competency in general. It would be wise to investigate how this can be achieved.

4. Define clear criteria for the POP, and how to use it in the AC

The guidelines students receive for the POP are quite general, and expect from a 500 word limit (very specific) and some suggestion what they should include in the POP, they are free to write it as they see fit. The POP itself is not assessed by the VIP coach but is included in the review of the AC. The AC have no rubric how to assess the POP. A rubric for the POP must be included and students need to receive feedback on the POP, if it is to be included in the review by the AC. In addition, calibration is needed to ensure all members apply the POP in the same manner

Wrap up

These recommendations could help the AC move forward to the next round, which will be a most higher stake round. A thorough evaluation, looking at not only calibration of decision but especially of process and reasoning, is essential. And although it is a very general recommendation, it is perhaps the most important one: begin on time. Decisions regarding vision and purpose take time and this cannot be rushed as a deadline approaches. Currently the assessment rounds are being included in an overall assessment plan. But even until these ideas are put into practice, each round should receive an evaluation. Even more challenging are the incorporation of the assessment rounds for year 2 next year, and the year after also for year 3. Right now is the time that a total plan must be made for all AC assessment rounds. It is encouraging that the members of the AC feel confident of their assessments, and that they agree so much on outcome. Some extra formalisation of the function and processes, as well as fixing the technical difficulties making work extra difficult at the moment, would be most valuable to ensure a working AC for the years to come.

To return to the line of poetry on the first page: between the dreams of programmatic assessment and actually achieving it, law and practicalities can get in the way. These recommendations can hopefully help alleviate some practical issues, and ensure the examination board can uphold the higher education laws, even though they were not necessarily designed to programmatic curricula.

References:

[1] Van Berkel et al. Toetsen in het hoger onderwijs. Boom, 2023.

[2]https://cris.maastrichtuniversity.nl/ws/portalfiles/portal/78063049/Van_der_Vleuten_2018_Programmatic_assessment_the_process_rationale.pdf

[3] Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

[4] Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478-85. doi: 10.3109/0142159X.2011.565828. PMID: 21609177.

DE WAARDE EN IMPACT VAN DE SKE-OPLEIDING VOOR EEN DOCENT VAN DE POLITIEACADEMIE

WILLEKE SCHIPPERS



Inleiding

Examinering is een essentieel onderdeel van het onderwijsproces. Het biedt niet alleen een min of meer objectieve maatstaf voor de prestaties van studenten (zoals bij summatief toetsen), maar draagt ook bij aan hun persoonlijke en professionele ontwikkeling (wat ook gestimuleerd kan worden door formatief toetsen). Toen ik met de SKE-opleiding startte, werd ik tegelijkertijd geplaatst in een ontwikkelteam dat een geheel nieuwe bacheloropleiding, de Bachelor Politieleider, moest gaan ontwikkelen. Mijn opdracht binnen dit team was een grofontwerp te maken voor het onderwijs en de examinering van de leerlijn Professionele Beroepsidentiteit. Dit is een leerlijn waarbij de student zichzelf en het politievak goed moet leren kennen en ontwikkelen. Daarnaast moest ik voor deze leerlijn een examendocument ontwikkelen. Dankzij de SKE-opleiding lukte het mij om een focus te creëren en te houden op vier belangrijke punten: het belang van een duidelijke visie, de integratie van opleidingsdoelen in tussentijdse examens, de keuze voor formatief danwel summatief toetsen, en de integratie van diverse kennisgebieden: Recht, Onderzoekend Vermogen en Gedrag en Maatschappij. Daarnaast geef ik aan wat de impact van deze opleiding is geweest op mijn eigen professionele ontwikkeling en hoe ik het geleerde heb toegepast in de praktijk.

4 belangrijke punten:

- *Belang van een duidelijke visie*
- *De integratie van opleidingsdoelen in tussentijdse examens*
- *De keuze voor formatieve danwel summatief toetsen*
- *Integratie van diverse kennisgebieden*

Het Belang van een Duidelijke Visie

Een duidelijke visie over wat je met de opleiding wilt bereiken is cruciaal voor het succes van zowel de student als de opleiding zelf. Een visie biedt richting en focus, en helpt docenten en studenten om hun doelen te stellen en te bereiken. Zonder een heldere visie kan het onderwijsproces fragmentarisch en ongericht worden, wat de effectiviteit van de opleiding kan verminderen (Glasbeek, 2015).

Het helpt om een goed gedefinieerde visie te hebben die een realistisch en helder beeld schetst van wat er gedurende de opleiding bereikt moet worden. Dit ondersteunt docenten en studenten om de voortgang te monitoren en aan te passen daar waar nodig. Ik merkte dat voor de leerlijn Professionele Beroepsidentiteit deze visie goed hielp om voor ogen te houden hoe het eindproduct eruit moest zien. Wat wilden we dat studenten zouden leren binnen deze leerlijn? Een duidelijke visie biedt een helder kader voor wat er getoetst moet worden. Vooral bij het definiëren van specifieke leerdoelen en competenties die studenten moeten bereiken bleek de visie een uitstekend richtpunt. Bovendien moet de visie aansluiten bij de bredere doelen van het opleidingsinstituut (in dit geval de Politieacademie), de vraag vanuit het werkveld (de Nationale Politie) en de verwachtingen van de maatschappij. Een duidelijke visie kan helpen docenten en studenten gemotiveerd te houden en zorgt dat ze zich bewust zijn van de waarde van hun opleiding. In dit specifieke geval bleek een heldere visie ook onontbeerlijk om de staf van de Politieacademie uit te leggen wat het doel was van de nieuwe bacheloropleiding en hoe de verschillende elementen van de visie op de nieuwe Politieleider naar voren kwamen gedurende de opleiding.

Integratie van Opleidingsdoelen in Tussentijdse Examens

De integratie van opleidingsdoelen in tussentijdse examens is een andere belangrijke factor die bijdraagt aan de waarde van een opleiding examinering. Tussentijdse examens bieden een regelmatige feedbackmoment dat docenten en studenten helpt om hun voortgang te volgen en hun leerstrategieën aan te passen. Door de opleidingsdoelen expliciet te integreren in deze examens, worden studenten continu herinnerd aan de belangrijkste leerdoelen en kunnen ze hun inspanningen richten op het bereiken van deze doelen (Sheppard, 2021).

Bovendien helpt de integratie van opleidingsdoelen in tussentijdse examens om de consistentie en continuïteit van de opleiding te waarborgen. Dit zorgt ervoor dat studenten een coherent en samenhangend leertraject volgen, wat hun begrip en toepassing van de leerstof verbetert.

Het was een puzzel om de leerlijn Professionele beroepsidentiteit te integreren in de verschillende modules. Door aan de verschillende modules verschillende leerdoelen te koppelen, die samenkwamen in de aangeboden stof en in de leerlijn is de integratie uiteindelijk goed geslaagd.

Bovendien helpt de integratie van opleidingsdoelen in tussentijdse examens om de consistentie en continuïteit van de opleiding te waarborgen.

De Rol van Formatief Toetsen

Formatief toetsen speelt een cruciale rol in het onderwijsproces van de nieuwe Bachelor Politieleider. In tegenstelling tot summatief toetsen, dat voornamelijk gericht zijn op het meten van de eindresultaten, is formatief toetsen bedoeld om de leerprocessen van studenten te ondersteunen en te verbeteren (Koenka et al, 2021). Ze bieden studenten de mogelijkheid om hun vaardigheden te oefenen, fouten te maken en feedback te ontvangen, zodat ze zich kunnen ontwikkelen gedurende de opleiding.

Het formatief toetsen kreeg in de opleiding Politieleider een belangrijke rol binnen de leerlijn Professionele beroepsidentiteit. Ik wilde dat studenten zich zouden focussen op hun ontwikkeling als politieprofessional. Niet op het examen dat behaald moest worden. Dit is hopelijk gelukt door de voorwaardelijke opdrachten die horen bij de leerlijn Professionele Beroepsidentiteit in verschillende modules door de jaren heen zonder cijfer aan hun portfolio toe te voegen.

Het halen van examens wordt hierdoor niet langer gezien als het uiteindelijke doel, maar als een mijlpaal in het leerproces. De focus verschuift naar de continue ontwikkeling van de student, wat leidt tot een diepgaander begrip en zelfkennis. Formatief toetsen helpen studenten om hun sterke en zwakke punten te identificeren en waar nodig te verbeteren, wat hun zelfvertrouwen en motivatie verhoogt.

Integratie van Diverse Kennisgebieden

De integratie van diverse kennisgebieden door de verschillende modules heen is een ander belangrijk aspect van de SKE. In de moderne wereld is interdisciplinariteit essentieel, omdat veel problemen en uitdagingen niet binnen de grenzen van één enkel vakgebied liggen (Gulikers & Oonk, 2019). De Nederlandse politieagent wordt aan de ene kant gezien als een creatieve duizendpoot, die voor elk probleem een oplossing moet kunnen bedenken. Aan de andere kant is de maatschappij zo ingewikkeld geworden, dat specialisatie onontkoombaar is. De Politieleider wordt geacht de verbanden tussen alle specialisaties te kunnen zien en het grote plaatje in beeld te houden.

Bij het ontwerpen van de toets voor de leerlijn Professionele beroepsidentiteit werden diverse kennisgebieden geïntegreerd. Er is gekozen voor een holistische benadering bij het beoordelen van studenten. Dit betekent dat niet alleen individuele kennisgebieden worden beoordeeld, maar ook hoe goed studenten deze gebieden kunnen combineren om complexe problemen op te lossen.

Deze interdisciplinaire benadering verrijkt het leerproces en maakt het relevanter voor de politiepraktijk. Studenten leren om complexe problemen vanuit meerdere perspectieven te benaderen, wat hun probleemoplossingsvaardigheden en creativiteit verbetert. Bovendien bereidt het hen beter voor op de wereld buiten de opleiding, waar interdisciplinaire samenwerking vaak noodzakelijk voor is.

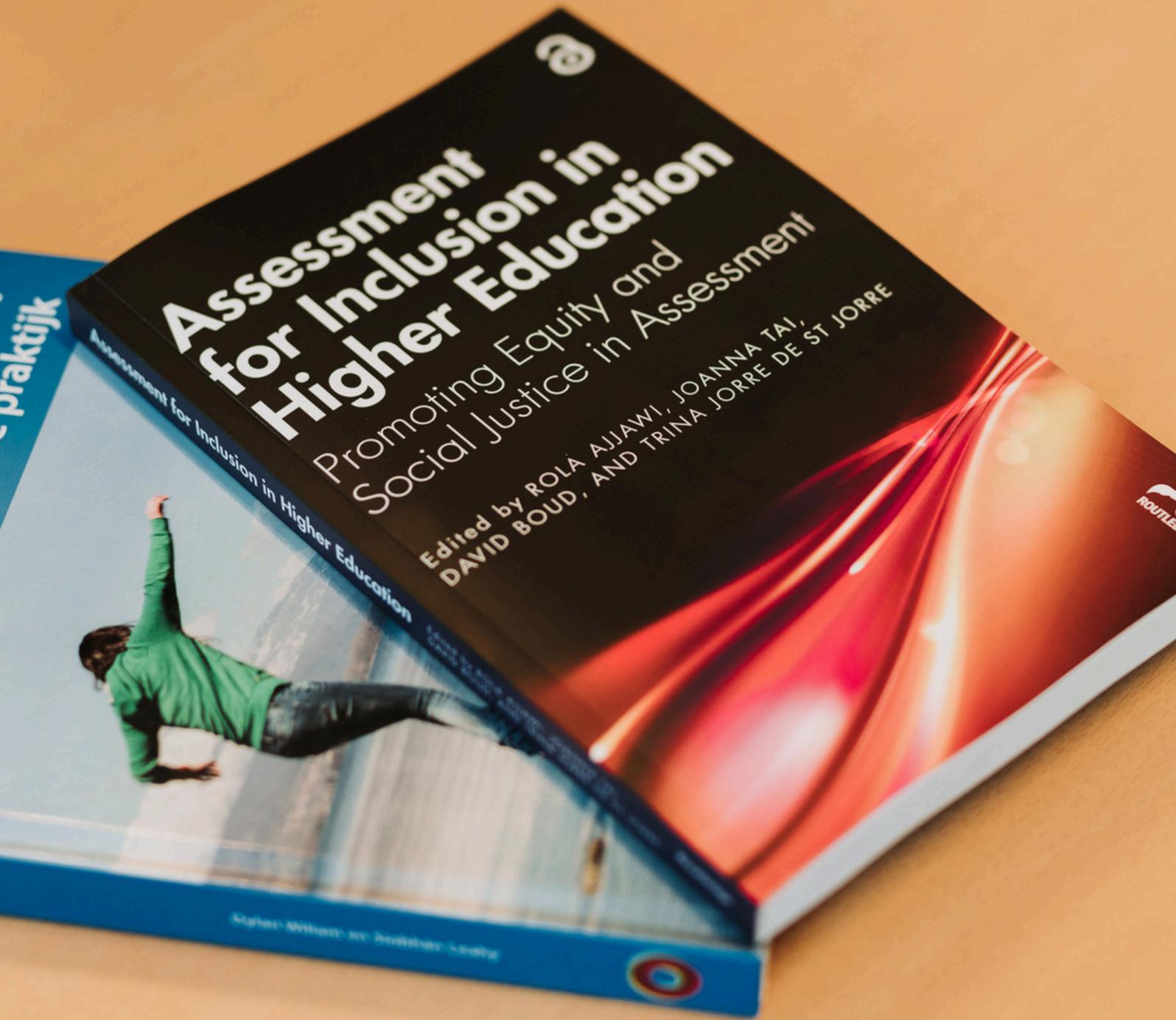
Conclusie

De waarde van de SKE- opleiding gaat verder dan alleen het meten van kennis en vaardigheden. Het biedt een structuur en richting voor het leerproces en de ontwikkeling van nieuwe opleidingen. Het helpt docenten en studenten om hun doelen te bereiken, en maakt van een opleiding meer dan een serie hoepeltjes waar doorheen gesprongen moet worden om het diploma te behalen. Een duidelijke visie, het inzichtelijk houden van opleidingsdoelen in tussentijdse examens, de rol van formatief toetsen, en de integratie van diverse kennisgebieden zijn allemaal essentiële elementen die ruimschoots aan bod zijn gekomen tijdens de SKE-opleiding.

Hierdoor is deze opleiding niet alleen een krachtige motor voor de persoonlijke en professionele groei van docenten die de opleiding volgen, maar ook voor de studenten die het ontwikkelde onderwijs gaan volgen, dat hen moet voorbereiden op de uitdagingen van de moderne wereld.

Referenties

- Glasbeek, H (2015) Onderwijscentrum, V. U. Wat werkt in hoger onderwijs? Aanbevelingen voor het ontwikkelen van een leerzaam en studeerbaar curriculum.
- Gulikers, J., & Oonk, C. (2019). Towards a Rubric for Stimulating and Evaluating Sustainable Learning. *Sustainability*, 11(4), Article 969. <https://doi.org/10.3390/su11040969>
- Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology*, 41(7), 922-947.
- Shepard, L. A. (2021). Ambitious Teaching and Equitable Assessment: A Vision for Prioritizing Learning, Not Testing. *American Educator*, 45(3), 28.



AUTHENTICITEIT VAN DE TAKE-HOME TOETS IN HET AI-TIJDPERK

MARJOLEIN TIMMERS

1 Introductie

Binnen de bacheloropleiding van ESHPM vormt de take-home toets – in de vorm van schriftelijke integratieopdrachten – een cruciaal instrument om de opgedane kennis en vaardigheden binnen een onderwijsblok te toetsen in een uitdagende en wetenschappelijk of praktisch relevante opdracht. De opkomst van Generatieve AI-tools (GenAI), zoals ChatGPT, Perplexity en Copilot, roept vragen op over de validiteit en betrouwbaarheid van deze toetsvorm.[1] Deze technologieën kunnen studenten ondersteunen bij het uitvoeren van onderzoeks- en schrijfopdrachten, waardoor de vraag ontstaat in hoeverre de opleiding nog kan vertrouwen op de autonome prestaties van studenten in overeenstemming met de vastgestelde eindtermen van de bacheloropleiding.

De examencommissie heeft de wettelijke taak om zowel de kwaliteit van het diploma als de toetsing binnen de opleiding te borgen. In het licht van deze opdracht richt het kwaliteitsverbeteringsproject zich op de waarborging van het bachelor diploma bij het gebruik van integratieopdrachten als take-home toets, met name in de context van de toenemende AI-mogelijkheden.[2] De centrale onderzoeksvraag binnen het project is dan ook: “Kan met het afnemen van schriftelijk take-home integratieopdrachten de kwaliteit van het diploma worden gewaarborgd?”

2 Aanpak van het project

Eerst is in samenspraak met verschillende stakeholders – examencommissie, opleidingsmanagement, afdeling onderwijsbeleid – het beoordelingsprobleem geïnventariseerd en afgebakend. Hierbij werd duidelijk dat het kwaliteitsverbeteringsproject zich specifiek richt op de integratieopdrachten van blok 1 tot en met 9 van de bacheloropleiding. Andere schriftelijke opdrachten, zoals het bachelor eindproject en de scripties, vallen onder de verantwoordelijkheid van de beleidsmedewerkers onderwijs en het opleidingsmanagement.

Na de afbakening van het project is gestart met het verzamelen van relevante informatie en documenten met betrekking tot de integratieopdrachten. Het Onderwijs Service Centrum heeft toegang geboden tot alle negen Canvaspagina's van de huidige onderwijsblokken in de bacheloropleiding. Daarnaast werden oude (beleids)documenten beschikbaar gesteld via de senior beleidsmedewerker onderwijs en verschillende docenten die nauw betrokken waren bij de opzet van de integratieopdrachten.

De volgende stap was de analyse. Om de hoofdvraag van het project te kunnen beantwoorden, werden de volgende zeven sub-vragen geformuleerd, die de basis vormen voor het verdere onderzoek:

- Wat is het ESHPM-beleid betreffende het gebruik van AI in toetsing?
- Hoe kan de inzet van AI worden gestimuleerd en voorkomen in onderwijs en toetsing?
- Wat zijn de eindkwalificaties van de bacheloropleiding en kunnen deze eindkwalificaties geborgd worden in het AI-tijdperk?



- Wat was het oorspronkelijke doel van de invoering van de integratieopdrachten?
- Hoe zijn de huidige integratieprojecten vormgegeven?
- Zijn de leerdoelen accuraat en passen de schriftelijke (integratie)opdrachten bij de leerdoelen?
- Wat zijn de gewenste vaardigheden die studenten van de bacheloropleiding bereiken met het afleggen van een integratieopdracht?

Gedurende de looptijd van het project, en aansluitend op de analyse, zijn drie belangrijke interventies uitgevoerd: (1) het betrekken van stakeholders, (2) de oprichting van de werkgroep AI-integratieprojecten, gericht op de inzet van acties op de korte termijn, en (3) een brainstormsessie voor het formuleren van acties op de langere termijn. De inzet van deze verschillende interventies is mede het gevolg van de vaststelling van het AI-beleid van ESHPM.

Ter afsluiting van het project werd een antwoord geformuleerd op de hoofdvraag en gereflecteerd op het kwaliteitsverbeteringsproject en de resultaten van de analyse en de bevindingen uit de interventies.

3. Resultaten en bevindingen

3.1 De analyse

Om het onderwerp van het kwaliteitsverbeteringsproject in perspectief te zetten en te komen tot een antwoord op de hoofdvraag zijn verschillende sub-vragen geformuleerd voor nader onderzocht. De belangrijkste resultaten en bevindingen worden hieronder per sub-vraag weergegeven.

Wat is het ESHPM-beleid betreffende het gebruik van AI in toetsing?

In het AI-beleid van ESHPM wordt vastgesteld dat de toepassing van AI door studenten is geoorloofd, tenzij dit een negatieve impact heeft op de toetsing. Onder een negatieve impact wordt in het beleid verstaan dat AI het leerproces van de student verstoort of de toetsing belemmert, waardoor de docent de kennis en vaardigheden van de student niet kan beoordelen. Het is aan de opleiding om te monitoren waar AI een negatieve impact heeft op de toetsing en hier een alternatief voor aan te bieden. In het beleid wordt extra aandacht gevraagd voor schrijfoopdrachten die worden ingezet als summatieve toetsing. Uit het beleid blijkt dat een thuis geschreven product dat als op zichzelf staande toets wordt beoordeeld, zonder aanvullende toetsing (bijvoorbeeld een mondeling) geen valide meetinstrument is.

Hoe kan de inzet van AI worden gestimuleerd en voorkomen in onderwijs en toetsing?

AI kan studenten op veel verschillende manier ondersteunen bijvoorbeeld door informatie te zoeken, onderzoeksvragen te formuleren en data te analyseren. Daarnaast kan GenAI studenten helpen bij het schrijven van opdrachten en het maken van reflecties Om het leereffect te vergroten is het belangrijk dat studenten AI als hulpmiddel gebruiken en niet als producent van het eindproduct.[1] Daarbij is actieve betrokkenheid, samenwerking en reflectie op het leerproces noodzakelijk om dit te bewerkstelligen. Voor het voorkomen van AI-gebruik bij toetsen, zijn mondelinge toetsen en praktijkopdrachten effectief, omdat ze persoonlijke input vereisen. Portfolio-assessments helpen de authenticiteit van het leerproces te waarborgen.

Wat zijn de eindkwalificaties van de bacheloropleiding en kunnen deze eindkwalificaties geborgd worden in het AI-tijdperk?

De bacheloropleiding hanteert twaalf eindtermen (zie Tabel 1). Uit de analyse blijkt dat het bij eindtermen 7 tot en met 10 onzeker is of studenten hier zelf(standig) aan voldoen wanneer zij gebruikmaken van GenAI als hulpbron. Om de borging van deze eindtermen te garanderen, zijn er twee mogelijke benaderingen. Ten eerste kan de huidige toetsvorm – voornamelijk bestaande uit schriftelijke onderzoeks- en schrijfoopdrachten – worden uitgebreid met toetsmethoden die minder vatbaar zijn voor beïnvloeding door AI-tools.

Ten tweede kan de interpretatie van de eindtermen zelf worden aangepast. Onder de formulering 'de student kan' zou expliciet vermeld kunnen worden dat dit geldt 'met inachtneming van de stand van de techniek', waarmee het gebruik van AI binnen zekere grenzen wordt geaccepteerd. Dit vraagt wel om een aanpassing in de Onderwijs- en Examenregeling (OER), en opent tegelijkertijd een bredere discussie over de kwaliteitsborging van toetsing en de discrepantie tussen de vaardigheden van studenten.

Tabel 1 Eindtermen zoals opgenomen in de OER 2024-2025

Kennislijn	
1	(DD 1) De student kan beschrijven hoe de (Nederlandse) gezondheidszorg functioneert op het niveau van het stelsel, de organisatie en de rollen van stakeholders, en hoe dit functioneren mede is bepaald door ontwikkelingen in ziekte en gezondheid.
2	(DD 1) De student kan beschrijven welke transities in de zorg plaatsvinden, deze historisch duiden, in internationaal perspectief plaatsen en de implicaties voor de verschillende stakeholders beargumenteren.
3	(DD 1) De student kan laten zien dat hij de belangrijkste basisbeginselen en concepten uit de disciplines die gezamenlijk de gezondheidswetenschappen vormen kan toepassen. Te weten: economie van de gezondheidszorg; gezondheidsrecht; management- en organisatiewetenschappen in de context van de gezondheidszorg; beleid- en bestuurswetenschappen in de context van de gezondheidszorg; sociaal-medische wetenschappen; filosofie en ethiek in relatie tot wetenschap, technologie en zorg.
4	(DD 1,2) De student kan uitleggen en analyseren hoe de beginselen en concepten van de verschillende disciplines zich tot elkaar verhouden.
5	(DD 2) De student kan met behulp van de aangereikte beginselen en concepten uit de disciplines van de gezondheidswetenschappen actuele vraagstukken in de gezondheidszorg analyseren: a. de student kan de oorsprong van deze vraagstukken verklaren; b. de student kan het vraagstuk vanuit verschillende disciplines en (internationale) contexten analyseren door het stellen van goede vragen en het toepassen van adequate concepten; c. de student kan het vraagstuk omzetten naar een afgebakende onderzoeksvraag; d. de student kan verschillende oplossingsrichtingen aandragen en verkent daarbij ook nieuw terrein; e. de student kan deze oplossingsrichtingen onderling wegen, rekening houdend met rollen en belangen van betrokken stakeholders en relevante sociaal-medische, economische, beleidsmatige, organisatorische, juridische en ethische aspecten.
Academische Vorming & Vaardigheden	
6	(DD 1) De student kent de uitgangspunten en kwaliteitseisen van wetenschappelijk onderzoek, binnen de context van gezondheidszorg.
7	(DD 2,3,4,5) De student kan kwantitatieve en kwalitatieve onderzoeksmethoden gebruiken en onder begeleiding, op bachelorniveau, een onderzoek in een organisatie binnen de gezondheidszorg ontwerpen, uitvoeren en daarover rapporteren.
8	(DD 2,3) De student kan op een basisniveau interdisciplinair denken en handelen, door kritisch te denken, samen te werken en te reflecteren.
9	(DD 4) De student kan overtuigend schriftelijk en mondeling communiceren en daarbij een vorm en stijl kiezen die past bij de doelgroep en context.
10	(DD 2,4) De student kan projectmatig werken, individueel en in groepsverband.
Attitude (Persoonlijke & Professionele Ontwikkeling)	
11	(DD 2,3) De student neemt eigen verantwoordelijkheid voor het leerproces en toont daarin zelfstandigheid.
12	(DD 2,3) De student kan een professionele houding aannemen: a. de student stelt zich open en onderzoekend op; b. de student is in samenwerkingsrelaties bewust van de belangen van anderen.

Wat was het oorspronkelijke doel van de invoering van de integratieopdrachten en hoe zijn de huidige integratieprojecten vormgegeven?

De integratieprojecten zijn oorspronkelijk bedoeld om onderdelen van zowel de inhoudelijke module als de AVV-module (Academische Vorming en Vaardigheden) met elkaar te verbinden. In het eerste jaar van de bacheloropleiding ligt de focus op het integreren van verworven kennis en vaardigheden. In de latere studie jaren verschuift de focus naar verdieping binnen de afzonderlijke disciplines. Studenten worden daarbij uitgedaagd om vanuit diverse perspectieven – individueel en /of in groepsverband – tot een oplossing of onderbouwd advies te komen. Tabel 2 geeft per onderwijsblok een overzicht van de huidige invulling en beoordelingswijze van de integratieopdrachten.

Tabel 2 Huidige vormgeving integratieopdrachten

Blok	Opdracht	Wijze van beoordeling
1	<p>Informatiepagina: De studenten maken in groepsverband een informatiepagina over een bepaalde deelsector van de gezondheidszorg. Beoogd wordt om aan de hand van vragen de kennis uit de verschillende disciplines (recht, economie en beleid & bestuur) samen met de aangeleerde vaardigheden te integreren.</p>	<p>De informatiepagina wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld. Daarbij wordt verwacht dat de bronvermelding in orde is en dat de tekst vrij is van schrijf-, spel-, en taalfouten.</p>
2	<p>Onderzoeksverslag: De studenten maken in groepsverband een onderzoeksverslag naar aanleiding van het onderzoek dat zij gedaan hebben naar een omstreden aandoening die zij als groep gekozen hebben. De studenten stellen een onderzoeksvraag op, houden elk een interview met iemand die lijdt aan de gekozen aandoening, werken dit interview uit en maken hier samen een onderzoeksverslag van. In dit onderzoek wordt het onderwerp besproken vanuit de verschillende disciplines (recht, wetenschapsfilosofie en sociaal-medische-wetenschappen).</p>	<p>De studenten worden in het onderzoek begeleid tijdens een aantal begeleid werken bijeenkomsten. Het uiteindelijke onderzoeksverslag wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld.</p>
3	<p>Managementadvies: De studenten schrijven in groepsverband aan de hand van een casus en een bijbehorende dataset een managementadvies. Om tot een advies te komen moeten de studenten de data analyseren in Excel. Daarnaast gebruiken ze de kennis uit de functionele managementgebieden en theoretische perspectieven om het probleem te analyseren.</p>	<p>Het managementadvies wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld. Daarnaast houdt elke groep een presentatie en discussie waarin zij ondervraagd worden door het managementteam.</p>
4	<p>Beleidsadvies: De studenten maken in groepsverband een beleidsadvies aan het Zorginstituut (ZIN) over hoe de transparantie van zorg verbeterd kan worden. De studenten analyseren gezamenlijk de huidige stand van zaken. Per student beschrijven ze een stakeholder perspectief om vervolgens gezamenlijk te bekijken welke perspectieven overeenkomen of conflicteren. De studenten geven elkaar twee keer peer-feedback en zij voeren een rondetafel-discussie waarin elke student het perspectief van de onderzochte stakeholder onder woorden brengt.</p>	<p>Het beleidsadvies en de stakeholderanalyse worden thuis gemaakt en de ingeleverde schriftelijke opdrachten wordt beoordeeld. De peer-feedback en het rondetafel-discussie worden beoordeeld met een voldaan of niet voldaan.</p>

5	<p>Innovatieadvies: De studenten maken in groepsverband een innovatieadvies voor een unieke opdrachtgevers uit de praktijk. De studenten analyseren het praktijkprobleem, ontwikkelen een mogelijke oplossing (prototype) en onderzoeken de haalbaarheid daarvan in termen van kosten en opbrengsten voor de organisatie. Daarnaast onderzoeken de studenten hoe de innovatie kan worden geïmplementeerd.</p>	<p>Alle onderdelen van het innovatieadvies worden thuis gemaakt en de ingeleverde schriftelijke opdrachten worden beoordeeld. Daarnaast worden twee presentaties gehouden en pitchen de studenten het gemaakte prototype via een video. Ook deze opdrachten worden beoordeeld. Als laatste leveren de studenten nog een projectmanagementverslag in wat beoordeeld wordt met een voldaan of niet voldaan.</p>
6	<p>Advies: De studenten maken in groepsverband een advies voor een buitenlandse zorgverzekeraar met winsttoegmerk. Het gaat om een basiszorgverzekering voor een doelgroep met een specifieke zorgbehoefte die de studenten zelf mogen kiezen. Het uiteindelijke eindadvies wordt gepresenteerd aan de RvB van de buitenlandse zorgverzekeraar.</p>	<p>Het advies wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld. De eindpresentatie telt in de beoordeling enkel mee als afrekpunten indien dit niet goed genoeg is.</p>
7	<p>Kwantitatief leeronderzoek: De studenten voeren in groepsverband een kwantitatief leeronderzoek uit. De studenten passen de methodologische en statistische kennis die zij tijdens de studie hebben opgedaan toe op een gezondheidswetenschappelijk probleem rondom de late levensfase. De studenten doorlopen daarbij zoveel mogelijk de gehele onderzoekscyclus. De studenten ontwikkelen een onderzoeksvraag, werken een onderzoeksplan uit, voeren een kwantitatieve analyse uit van empirische data met behulp van SPSS, interpreteren de resultaten en schrijven een rapportage over het onderzoek en de bevindingen.</p>	<p>Tijdens een begeleid werken bijeenkomst worden studenten geholpen bij het analyseren van de data. Het Short Report wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld.</p>
8	<p>Kwalitatief leeronderzoek: De studenten voeren in groepsverband een kwalitatief leeronderzoek uit naar een vraagstuk rondom zorg en welzijn in het sociaal domein. De studenten kiezen daarvoor zelf een casus en formuleren daarbij een passende probleemanalyse. Studenten gebruiken 2 concepten uit de kennislijn (beleidsparadigma's, padafhankelijkheid, juridische en ethische dilemma's enz.) en 2 methoden uit de AVV lijn (discoursanalyse, interviews, observaties enz.).</p>	<p>Het verslag wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld. Daarnaast maken de studenten een infographic over de inhoud van het onderzoek. Deze infographic wordt tijdens de laatste bijeenkomst gepresenteerd. Als laatste leveren studenten ook een individueel reflectierapport en wordt het onderzoek voorzien van peer-feedback.</p>
9	<p>Essay: De studenten schrijven individueel een essay over een morele vraag naar keuze. De morele vraag moet recent zijn en gerelateerd aan een nieuwsbericht van na januari 2022. De vraag moet specifiek gemaakt worden voor het gezondheidsbeleid of gezondheidsmanagement. Er moet empirisch materiaal over de morele vraag worden opgenomen. Bij het vormen van een moreel argument moeten minimaal twee van de morele theorieën die centraal staan in het vak worden toegepast (deontologie, consequentialisme, rechtsethiek, enz.) en gebruik worden gemaakt van de voorgeschreven en aanvullende literatuur.</p>	<p>Tussentijds wordt de student voorzien van feedback. Het essay wordt thuis gemaakt en de ingeleverde schriftelijke opdracht wordt beoordeeld, daarnaast vindt een aanvullende mondelinge toets plaats.</p>

Zijn de leerdoelen accuraat en passen de schriftelijke (integratie)opdrachten bij de leerdoelen?

Uit de analyse blijkt dat de huidige schriftelijke opdrachten qua vorm goed aansluiten bij de beoogde leerdoelen. De opkomst van GenAI-tools en het vastgestelde AI-beleid van ESHPM maken het echter niet langer mogelijk om uitsluitend via schriftelijke opdrachten te garanderen dat studenten daadwerkelijk aan alle leerdoelen voldoen. Om deze leerdoelen wél te kunnen borgen, zijn aanvullende formatieve of summatieve toetsvormen noodzakelijk. In vrijwel alle onderwijsblokken worden al aanvullende toetsvormen toegepast, zoals presentaties, discussies en (peer)feedback. Of deze vormen voldoende zijn om aan te tonen dat studenten de leerdoelen behalen, hangt sterk af van de inhoud en de wijze van uitvoering. Twee voorwaarden zijn hierbij cruciaal. Ten eerste moet de aanvullende toetsing voldoende diepgang bieden. Dit kan bijvoorbeeld door verdiepende vragen of vervolgdiscussies naar aanleiding van een presentatie of schriftelijke opdracht, waarmee studenten hun inzicht en vaardigheden daadwerkelijk kunnen tonen. Ten tweede moet de kwaliteit van deze toetsing aantoonbaar geborgd worden. Dit vereist dat een bevoegde examinerator betrokken is bij de beoordeling en kan garanderen dat de toetsvorm voldoet aan de eisen van betrouwbaarheid en validiteit.

Wat zijn de gewenste vaardigheden die studenten van de bacheloropleiding bereiken met het afleggen van een integratieopdracht?

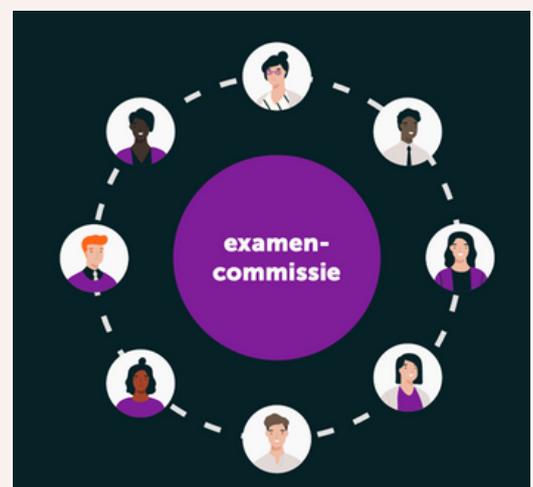
Binnen de integratieopdrachten wordt van studenten verwacht dat zij – individueel, in duo's of in groepsverband – een beroepsgericht product kunnen ontwikkelen en opleveren. Tijdens dit proces werken zij actief aan het versterken van diverse academische vaardigheden, waaronder het zoeken van wetenschappelijke literatuur, het uitvoeren van onderzoek, academisch schrijven, presenteren en kritisch redeneren. Daarnaast ligt de nadruk op het toepassen en verdiepen van opgedane kennis. Studenten worden uitgedaagd om gezondheidswetenschappelijke vraagstukken vanuit verschillende invalshoeken te benaderen en gezamenlijk tot een onderbouwde oplossing te komen. Naast de inhoudelijke integratie, wordt verwacht dat de studenten samenwerking en het interdisciplinair denken.

3.2 De interventies

De opkomst van GenAI-tools en het vastgestelde AI-beleid van ESHPM hebben directe implicaties voor het waarborgen van de diploma-kwaliteit, met name bij schriftelijke take-home integratieopdrachten. Ter ondersteuning van deze kwaliteitsborging zijn drie interventies geïnitieerd. Allereerst wordt ingezet op tijdige betrokkenheid van relevante stakeholders, waarbij de examencommissie een centrale rol speelt. Daarnaast is de werkgroep AI-integratieprojecten, opgericht met als doel de inzet van acties op de korte termijn. Tot slot heeft een brainstormsessie plaatsgevonden voor het formuleren van acties op de langere termijn. De belangrijkste resultaten en bevindingen worden hieronder per interventie weergegeven.

Stakeholders betrekken

Het betrekken van stakeholders – zijnde het opleidingsmanagement, de afdeling onderwijsbeleid en onderwijscoördinatoren en examinatoren die betrokken zijn bij de uitvoering en ontwikkeling van de integratieopdrachten – vormt een essentieel onderdeel van het kwaliteitsverbeteringsproject en is nauw verweven met de opzet en uitvoering ervan. Gedurende de looptijd van het project werd het AI-beleid van ESHPM vastgesteld, hetgeen direct gevolgen had voor de uitvoering ervan.



Het nieuwe beleid onderstreept de urgentie van het onderwerp en vereist de inzet van een twee sporen aanpak: enerzijds het voortzetten van de analyse vanuit de borgende taak van de examencommissie, anderzijds de inzet van korte termijn acties vanuit de zorgende taak van het opleidingsmanagement.

Deze eerste interventie werd uitgevoerd in meerdere fasen. De eerste twee fasen markeerden de start van het project. In deze fasen lag de focus op het betrekken van stakeholders bij het vaststellen en afbakenen van het onderwerp, én bij het delen van de eerste bevindingen om het belang van het project te duiden. In de afsluitende fase werden de resultaten van de analyse teruggekoppeld aan de betrokken stakeholders. Dit leverde waardevolle inzichten op in wat voor hen van belang is. Denk hierbij aan de interpretatie van formuleringen als 'de student kan', en de borging van de kwaliteit van formatieve toetsing bij de inzet van examinatoren die werken onder de verlengde arm van een aangewezen zelfstandig examinator. Daarnaast is duidelijk geworden dat het opleidingsmanagement behoefte heeft aan verdere advisering, zoals de vraag of aanpassingen in het aantal en de omvang van de integratieopdrachten wenselijk zijn om de eindkwalificaties ook in de toekomst voldoende te borgen.

Werkgroep AI-integratie

Het opzetten van de werkgroep AI-integratie is onderdeel van de twee sporen aanpak. Het opleidingsmanagement neemt hierin het voortouw, terwijl de examencommissie als actief lid bijdraagt aan de borging van de toetskwaliteit.

Deze tweede interventie is net als de eerste interventie gefaseerd uitgevoerd. In de eerste twee fasen werd de werkgroep officieel gevormd en werd systematisch in kaart gebracht waar het gebruik van GenAI mogelijk knelpunten oplevert binnen de constructieve alignment van het curriculum. Daarbij werd ook onderzocht waar alternatieve vormen van toetsing noodzakelijk zijn. In de vervolgfase werkt de werkgroep aan het uitzetten van de grote lijnen en het geven van adviezen aan coördinatoren en examinatoren. Gezien ESHPM een voorloper is in het toestaan van AI-gebruik binnen zowel onderwijs als toetsing, is het van belang dat diverse externe deskundigen – zoals onderwijskundigen en AI-experts – worden betrokken bij de vormgeving van deze transitie. De uitvoering van de benodigde aanpassingen ligt uiteindelijk de coördinatoren en examinatoren van de onderwijsblokken. Specifiek zijn de coördinatoren van zogenoemde hoog-risicoblokken op de hoogte gebracht van de bevindingen van de werkgroep. Met ondersteuning van externe experts werken zij momenteel aan het AI-proof maken van de integratieopdrachten op korte termijn. In de toekomst richt de werkgroep zich op de langer termijnvisie aangaande de inzet van integratieopdrachten in de bacheloropleiding.

Brainstormsessie

Deze laatste interventie kijkt op een explorerende wijze naar de inzet van de schriftelijk take-home integratieopdrachten op de langer termijn. In deze opdrachten staan enerzijds de leerdoelen samenwerken, integreren en verdiepen centraal. Anderzijds is er nadrukkelijk aandacht voor de ontwikkeling tot academicus en bijbehorende academische vaardigheden. Ideaal gezien ligt de verantwoordelijkheid voor het aantonen van deze academische vaardigheden bij de student zelf. In dat kader wordt het gebruik van een portfolio als toetsvorm geopperd.[4] Hiermee verschuift de nadruk van enkel het (eind)product naar het leerproces, waarbij ontwikkeling en reflectie centraal staan.

Daarnaast werden binnen deze interventie twee alternatieve scenario's verkend voor de toekomstige inrichting van de integratieopdrachten:

1. *Verschuiving van de focus*: Hierbij worden de leerdoelen op het gebied van kennis en AVV expliciet geborgd binnen de bijbehorende modules zelf. De integratieopdrachten kunnen zich dan volledig richten op de leerdoelen samenwerken, integreren en verdiepen.

2. *Drie omvangrijke integratieopdrachten:* In dit scenario wordt ingezet op drie grote integratieopdrachten waarbij de studenten werken aan actuele opdrachten met verschillende deelopdrachten. Naast het innovatieadvies van blok 5 zijn naar verwachting de integratie opdrachten van blok 7 kwantitatief leeronderzoek en blok 8 kwalitatief leeronderzoek het meest geschikt om uit te bouwen tot opdrachten waar de studenten in staat zijn samen te werken, te verdiepen en academische vaardigheden te ontwikkelen.

4. Discussie en Conclusie

4.1 Discussie

De examencommissie heeft de wettelijke taak om zowel de kwaliteit van het diploma en de toetsing binnen de opleiding te borgen. De opkomst van GenAI-tools en het toestaan van deze tools in onderwijs en toetsing, trekt de authenticiteit van de take-home toets in twijfel. In het licht van de borgende taak van de examencommissie staat dit beoordelingsprobleem centraal in het kwaliteitsverbeteringsproject, wat een antwoord geeft op de vraag “Kan met het afnemen van schriftelijk take-home integratieopdrachten de kwaliteit van het diploma worden gewaarborgd?”

Eindkwalificaties

De borging van kwaliteit van het diploma is vastgelegd in artikel 7.12 lid 2 Wet op het hoger onderwijs en wetenschappelijk onderzoek (WHW). Daarin staat dat de examencommissie het orgaan is dat op objectieve en deskundige wijze vaststelt of een student voldoet aan de voorwaarden die de onderwijs- en examenregeling stelt ten aanzien van kennis, inzichten en vaardigheden die nodig zijn voor het verkrijgen van een graad.

Voor de bacheloropleiding van ESHPM moet de examencommissie vaststellen dat de student voldoet aan de twaalf eindkwalificaties (zie Tabel 1). Deze eindkwalificaties zijn overwegend geformuleerd met de aanzet: ‘de student kan’. Een ogenschijnlijk eenvoudige zinsnede, die echter in het huidige AI-tijdperk aan complexiteit wint. Deze formulering impliceert traditioneel dat de student zelf(standig), zonder externe hulp, beschikt over de betreffende kennis, inzichten of vaardigheden. Echter, de opkomst van GenAI en het vaststellen van het AI-beleid van ESHPM maken deze interpretatie minder vanzelfsprekend. Dit AI-beleid van ESHPM staat het gebruik van GenAI door studenten toe, tenzij dit een negatieve invloed heeft op de toetsing. Deze voorwaardelijke ruimte stelt de toetspraktijk voor een fundamentele uitdaging: wat betekent het daadwerkelijk om vast te stellen dat een student kan voldoen aan de eindkwalificaties? Betreft dit uitsluitend individuele beheersing, of valt het adequaat inzetten van AI binnen deze competentie? Indien de formulering ‘de student kan’ strikt wordt geïnterpreteerd als het zelf(standig) beheersen van de eindtermen zonder inzet van GenAI, dan vereist de toetsing methoden die ondubbelzinnig de individuele prestaties van studenten aantonen. Dit geldt in het bijzonder voor eindkwalificaties 7 tot en met 10. Om de validiteit van deze eindtermen te waarborgen, is een verscheidenheid aan toetsvormen nodig die gericht zijn op authenticiteit en autonomie. Tegelijkertijd valt te betogen dat ‘de student kan’ óók mag betekenen: *de student kan met behulp van AI-tools*. In dat geval verandert de focus: dan moet de opleiding niet alleen toetsen op kennis en vaardigheden, maar ook op het bewust, effectief en ethisch gebruik van GenAI. De vraag wordt dan niet langer óf AI wordt ingezet, maar hóe – en of de student in staat is de tool kritisch en verantwoord in te zetten als verlengstuk van het eigen denken. Het is aan de opleiding om hier een keus in te maken. Vanuit het borgingsperspectief is het essentieel dat de keus expliciet wordt gemaakt in de Onderwijs- en Examenregeling (OER), zodat toetsing, begeleiding en beoordeling daarop afgestemd kunnen worden. Wetende dat de techniek in ontwikkeling blijft kan, gelijk aan terminologie in wetgeving, gebruik worden gemaakt van een zinsnede als ‘*de student kan met inachtneming van de stand van de techniek*’.

Kwaliteit examens

De borging van kwaliteit van de examens is vastgelegd in artikel 7.12b lid I sub a WHW. Daarin staat dat de examencommissie de taak en bevoegdheid heeft tot het borgen van de kwaliteit van de tentamens en examens. Vastgesteld dient te worden of de schriftelijk take-home integratieopdrachten in het huidige AI-tijdperk nog voldoende geschikt zijn om de beoogde leerdoelen te realiseren.

De integratieprojecten zijn initieel ontworpen met als doel een brug te slaan tussen de inhoudelijke modules en de module AVV. In het eerste jaar van de bacheloropleiding ligt de nadruk voornamelijk op het toepassen van opgedane kennis en vaardigheden in een geïntegreerde context. Naarmate de opleiding vordert, verschuift het accent naar een verdere verdieping binnen de afzonderlijke disciplines. Naast samenwerken ontwikkelen de studenten uiteenlopende academische vaardigheden, waaronder literatuur zoeken, onderzoeksvaardigheden, academische schrijfvaardigheden, kritisch denken en presentatievaardigheden.

Hoewel de huidige opdrachten qua vorm aansluiten bij de leerdoelen, werpt de opkomst van generatieve AI-tools, in combinatie met het AI-beleid van ESHPM, vragen op over de borging van het individuele kunnen behalen van leerdoelen. Het is niet langer vanzelfsprekend dat studenten via uitsluitend schriftelijke opdrachten zelf(standig) aan alle leerdoelen voldoen. Aanvullende formatieve of summatieve toetsingsvormen zijn daarom noodzakelijk. Uit de resultaten van de analyse blijkt dat vrijwel alle onderwijsblokken reeds gebruik maken van dergelijke aanvullende toetsvormen, zoals presentaties, discussies en (peer)feedback. Of deze vormen voldoende zijn om aan te tonen dat studenten de leerdoelen behalen, hangt sterk af van de inhoud en de wijze van uitvoering. Zoals uiteengezet in de analyse zijn daarbij twee voorwaarden cruciaal. Ten eerste moet de aanvullende toetsing voldoende diepgang bieden. Ten tweede moet de kwaliteit van deze toetsing aantoonbaar geborgd worden. In de huidige opzet blijkt dat deze twee aspecten – diepgang en toetskwaliteit – binnen sommige integratieprojecten nog onvoldoende gerealiseerd worden.

Interventies

De verschillende interventies richten zich op het in kaart brengen van de directe implicaties van het AI-tijdperk voor de waarborging van de kwaliteit van het bachelorsdiploma, met specifieke aandacht voor schriftelijke take-home integratieopdrachten. Daarbij wordt onderzocht welke mogelijke acties geïmplementeerd kunnen worden om deze kwaliteit te behouden.

Een belangrijke eerste stap was de tijdige betrokkenheid van relevante stakeholders, waarbij de examencommissie een cruciale rol vervult. Het betrekken van stakeholders is essentieel voor het succes en de borging van het project, vooral wanneer het gaat om het vaststellen van het onderwerp en het delen van bevindingen. Dit bevordert niet alleen de relevantie van de acties, maar stimuleert ook concrete maatregelen, zoals het verduidelijken van de eindkwalificaties en het verbeteren van de kwaliteit van formatieve toetsing. Stakeholders kunnen bovendien waardevolle feedback leveren, bijvoorbeeld met betrekking tot de gewenste omvang en inzet van integratieopdrachten in de toekomst.

Daarnaast is er de oprichting van de werkgroep AI-integratieprojecten, met als doel het ontwikkelen van concrete acties op korte termijn. Een twee sporen traject werd ingezet om zowel de impact van AI op onderwijs als toetsing te onderzoeken. De werkgroep richt zich op het uitzetten van de grote lijnen en de risicoanalyse van de leerdoelen. Coördinatoren van risicovolle onderwijsblokken werken met ondersteuning van externe deskundigen intensief aan de noodzakelijke aanpassingen om de kwaliteit van de toetsing snel te waarborgen.

Tegelijkertijd heeft een brainstormsessie plaatsgevonden om acties voor de lange termijn te formuleren. Naast de kortetermijnmaatregelen is het essentieel om na te denken over de toekomst van schriftelijke take-home integratieopdrachten in het AI-tijdperk.

De oorspronkelijke leerdoelen van de integratieopdrachten – samenwerking, integratie en verdieping – blijven belangrijk, maar de nadruk moet verschuiven van het (eind)product naar het proces. Een mogelijke benadering is het gebruik van academische portfolio's waarin studenten hun ontwikkeling aantonen. Het is bovendien niet noodzakelijk om in elk onderwijsblok een integratieopdracht aan te bieden. Het zou effectiever kunnen zijn om drie grote integratieopdrachten te organiseren – het innovatieadvies (blok 5), het kwantitatief leeronderzoek (blok 7) en het kwalitatief leeronderzoek (blok 8) –, met meer nadruk op het proces in plaats van op het uiteindelijke resultaat.



4.2 Conclusie

Kan met het afnemen van schriftelijk take-home integratieopdrachten de kwaliteit van het diploma worden gewaarborgd?

De komst van GenAI tools en het ESHPM-beleid aangaande het gebruik van AI heeft tot gevolg dat de kwaliteit van het diploma met het afnemen van schriftelijk take-home integratieopdrachten niet volledig kan worden gewaarborgd. Enerzijds zal vastgesteld moeten worden wat verstaan wordt onder de formulering 'de student kan'. Anderzijds moeten aanvullende toetsvormen worden ingezet om de kwaliteit van de huidige schriftelijk take-home integratieopdrachten te borgen. De werkgroep heeft hiertoe een voorzet gedaan en de coördinatoren van risicovolle onderwijsblokken zijn nu aan zet voor aanpassing op de korte termijn. Hierbij is het van belang in acht te nemen dat de kennis en vaardigheden van de examinatoren cruciaal zijn om de kwaliteit van de (formatieve) toetsing te borgen. Indien nodig zullen de examinatoren de juiste scholingsmogelijkheden moeten worden aangeboden en bij examinatoren die onder verlengde arm werken, zal gekeken moeten worden hoe de examinator de kwaliteit van de toetsing kan verifiëren. Om de kwaliteit van het diploma te kunnen borgen in het AI-tijdperk is het belangrijk om op de langer termijn in te zetten op het terugbrengen van het aantal integratieopdrachten. Tegelijkertijd is het verstandig de integratieopdrachten die behouden blijven in omvang te laten toenemen, waarin verschillende deelproducten moeten worden aangeleverd. Hierbij is het belangrijk de focus op de (eind)producten te verschuiven naar het proces, bijvoorbeeld door de inzet van een academisch portfolio waarin de student verantwoordelijk is voor zijn ontwikkeling tot academicus.

5. Opedane ervaringen en vervolgstappen

5.1 Ervaringen

De urgentie van de uitvoering van dit kwaliteitsverbeteringsproject laat zien dat het speelveld dynamisch is. Het betrekken van de stakeholders is essentieel gebleken. Door elkaar te informeren kan je inspelen op de dynamiek en actualiteit. Zo bleek de inzet van een twee sporen traject onvermijdelijk. Enerzijds het voortzetten van de analyse en interventies vanuit de borgende taak van de examencommissie, anderzijds de inzet van korte termijn acties vanuit de zorgende taak van het opleidingsmanagement. Met de vaststelling van het AI-beleid werd de druk om te innoveren in het onderwijs en de toetsing opgevoerd. Vooralsnog wordt ingezet op haalbare aanpassingen op de korte termijn, met als doel straks vanuit een rustiger vaarwater in te zetten op mogelijke grotere veranderingen op de langer termijn.

5.2 Vervolgstappen

De ontwikkeling is nog in volle gang. De examencommissie blijft actief betrokken in de werkgroep AI-integratieopdrachten. Daarbij wordt ingezet op de borging van de kwaliteit van het diploma met het afnemen van schriftelijk take-home integratieopdrachten op de langer termijn.

Uitgezocht moet worden of de het opleidingsmanagement af wil wijken van de huidige opzet van de bacheloropleiding met daarin in elk onderwijsblok een integratieopdracht.

Daarnaast moet gewaarborgd worden dat – indien overgegaan wordt tot het uitvoeren van minder integratieopdrachten – de studenten in het onderwijs voldoende ruimte krijgen om zich te ontwikkelen tot een academicus.

Ook zal de inzet van een academisch portfolio als toetsvorm nader onderzocht moeten worden. Hiervoor kunnen andere opleidingen die al gebruik maken van een dergelijk portfolio als voorbeeld en inspiratiebron dienen. Te denken valt aan het leiderschapsdossier van de opleiding Master of Health Business Administration (MHBA) van het Erasmus Centrum voor Zorgbestuur.

Literatuurlijst

[1] Voor meer informatie over de inzet van AI bij het maken van essay en het detecteren daarvan, zie: M. Sharples, 'Automated Essay Writing: An AIED Opinion', *International Journal of Artificial Intelligence in Education*, (2022) 32:1119-1126, doi.org/10.1007/s40593-022-00300-7.

[2] Voor meer voorbeelden uit de praktijk over de borgende taak van de examencommissie en de opkomst van AI, zie: M. Luth-Morgan, C. Elion-Valter, L. Poort, 'ChatGTP als katalysator voor meer authentieke toetsing. Praktijkvoorbeelden uit het ethiekonderwijs', AA20240790.

[3] Voor meer informatie over ChatGTP in het onderwijs, zie: J. Rudolph, S. Tan S. Tan, 'ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?', *Journal of Applied Learning & Teaching*, (2023) 6:342-263, doi.org/10.37074/jalt.2023.6.1.9

[4] Voor meer informatie over het gebruik van een portfolio, zie: R. van der Gulden, et. al., 'How does portfolio use affect self-regulated learning in clinical workplace learning: What works, for whom, and in what contexts?' *Perspect Med Educ* (2022) 11:247-257, doi.org/10.1007/s40037-022-00727-7



THE (IN)VISIBILITY OF ASSESSMENT

SAMANTHA BREUER

Erasmus
University
College

Introduction

According to the 'Wet Hoger Onderwijs' (2025), existing educational programmes are reaccredited on a regular basis. During this accreditation, the accreditation panel assesses a couple of quality indicators, among which are a) the intended outcome level of the programme (as described in the intended learning outcomes), b) the assessment of students, and c) the realized outcome level at the programme. These quality indicators come together in the assessment of the intended learning outcomes.

In standardized programmes, where each student follows a similar educational trajectory, assessment of the intended learning outcomes is easily shown by comparing the intended learning outcomes to the courses and assessments the students have finalized during the programme. However, when a programme is not standardized, in the sense that the students follow highly diverse educational trajectories, the quality indicators of intended learning outcomes, realized learning outcomes, and assessment are not easily shown. An example is the Bachelor Liberal Arts and Sciences of the Erasmus University College (EUC). In this 180 EC educational programme, only 45 ECs are standardized through general courses (including a personal development track) which all students follow. Next to the general courses, students can choose one of 17 majors worth 75 EC, and can freely choose on- and off-campus courses worth a total of 60ECs. This means that out of the 180 ECs obtained during the Bachelor Liberal Arts and Sciences, a total of 135 ECs is filled with non-standardized, highly diverse courses. This makes it difficult for a programme such as the Bachelor Liberal Arts and Sciences to concretely show the assessment of the intended learning outcomes, as there is no consistent educational trajectory or assessment trajectory among students to rely upon.

This leads to the following question:

Given the highly diverse educational trajectories of its students, how can the Bachelor Liberal Arts and Sciences of the Erasmus University College show the relation between the intended outcome level (as described in the intended learning outcomes) and the realized outcome level within its programme? Or: how can the assessment of the intended learning outcomes can be made visible in this programme?

Approach

The aim of the project was to determine how the assessment of the intended learning outcomes, at the Bachelor Liberal Arts and Sciences of EUC, could be shown in an overview. This overview should be understandable for students and other readers, and feasible for institutional management and staff to create and update.

In order to determine what information regarding the assessment of intended learning outcomes was already available, the official assessment documentation of EUC was consulted and interviews were held with relevant stakeholders at EUC. To devise a model to visualize the assessment of the intended learning outcomes, inspiration was obtained by consulting assessment experts and documents from other faculties inside and outside of the Erasmus University Rotterdam. After a model had been devised, this model was tested for a subgroup of students of the Bachelor Liberal Arts and Sciences with the help of multiple examiners at EUC.

Main Findings

From EUC's official documentation and interviews with relevant stakeholders, it became clear that the Bachelor Liberal Arts and Sciences entails 14 intended learning outcomes. These learning outcomes are formulated clearly but broadly, to make sure they fit the diverse educational trajectories of the students. In addition, the official documentation stated how each course within the Bachelor Liberal Arts and Sciences links to the intended learning outcomes. However, as the courses the students follow are highly diverse, this documentation did not create a clear overview of the assessment of the intended learning outcomes in all students. Therefore, even though the assessment was present, it was not perceivable.

When consulting official documents and assessment experts from other faculties inside and outside of EUR, it became clear that an overview of the assessment of intended learning outcomes generally tends to be structured on the basis of categories of courses (e.g. foundational courses), categories related to chronology (e.g. first year content), categories related to assessment method (e.g. written exam), student cohort (e.g. 2023-2024), and / or categories of programme outcomes (e.g. Dublin descriptors). In addition, each visualization done by other faculties seemed to be based on checkboxes, in which intended learning outcomes are related to specific parts of the curriculum.

As the curriculum of EUC's Bachelor Liberal Arts and Sciences tests its learning outcomes on course level, its curriculum has a stable set of course categories. The model devised to visualize EUC's learning outcome assessment was structured by courses, creating a division between general courses and mandatory major courses.

Upon being tested for a subgroup of EUC students (who have all chosen 1 particular major), it turned out the model was able to concretely visualize the assessment of the intended learning outcomes within the students.

When evaluating the finalized overview, some interesting observations were made.

Firstly, it was noticeable that the level at which each intended learning outcome was examined (e.g. knowledge or application) could not be seen. This led to the question whether it would be possible to include the level or complexity at which the intended learning outcomes are tested.

Secondly, it became clear that a significant amount of the students' educational trajectory, the freely chosen on- and off-campus courses, could not be included in the model due to the high amount of diversity in content and assessment. This resulted in the question, what could be changed to include these courses in the model?

Recommendations

To concretely show the assessment of the intended learning outcomes within EUC's Bachelor Liberal Arts and Sciences, the first recommendation is to fill out the devised model for each of the majors within the programme (taking into account both the general courses and the mandatory major courses). This way, the assessment of the intended learning outcomes can be visualized for all (subgroups of) students at EUC's Bachelor Liberal Arts and Sciences.

Furthermore, as it is currently difficult to show at which level the intended learning outcomes are assessed, the second recommendation is to divide the existing intended learning outcomes into multiple levels of difficulty or complexity (e.g. knowledge level, application level). This creates the possibility to assess the intended learning outcomes with increasing complexity, and to clearly visualize the level at which each intended learning outcome is assessed in each part of the programme. In addition, this creates the opportunity to include a learning curve within the students' educational trajectories.



Lastly, it was remarkable that the relation between a significant part of the students' educational trajectory (the freely chosen courses) and the intended learning outcomes could not be made tangible, as the chosen courses were too diverse in content and assessment. Therefore, it is recommended to set requirements for the freely chosen courses, for example in terms of targeted learning outcomes. As a result, the goal of the freely chosen courses will be more uniform among students, creating a more standardized programme. Doing so, it will be easier to connect the freely chosen courses to the intended learning outcomes and include these in the overview of the assessment of the intended learning outcomes.



INVASION OF THE STUDENT-LEARNING SNATCHERS: THE IMPACT OF GENERATIVE AI ON ASSESSMENT AND STUDENT PERFORMANCE

JAY LEE

Erasmus School of
History, Culture and
Communication



Introduction

The advent of generative artificial intelligence (referred to as simply 'AI' in this report), particularly large language models (LLMs), has raised multiple challenges and opportunities for educators. In the past few years – since the emergence of ChatGPT in late 2022 (OpenAI, 2022) – educators and policy makers at ESHCC (Erasmus School of History, Culture and Communication at Erasmus University Rotterdam), and EUR at large, have been grappling with development of appropriate policies for AI use by students. Alarming concerns of quality of student work, assessment challenges, and fraud emerged with each new generation of ChatGPT at EUR, also echoed by educators globally, e.g., Rudolph et al. (2023) and Luth-Morgan et al. (2024). Discussions, workshops, and desk research into AI use undertaken by various sub-units at EUR – including the CLI (Community for Learning & Innovation) and the Learning & Innovation Hub and Examination Board at ESHCC – culminated into several guidelines for AI use policies in education. Much of the guidelines now appear prominently on EUR webpages, notably the pages “Generative AI in education at ESHCC” (Learning & Innovation Hub, 2025), “Generative AI (GenAI)” (Erasmus University Library, 2025), “Generative AI: Impact on teaching and assessment” (Erasmus University Library, 2025).[1]

In 2023, ESHCC instituted a policy for including a “Use of AI” Canvas page in every course, to provide students with how AI may or may not be used for their coursework. The several options a course’s AI use policy may be found in ESHCC’s Learning & Innovation Hub’s page “Use of AI” (Learning & Innovation Hub, 2025). Thus, courses have been afforded the flexibility to implement their own AI policies; a template Canvas page is provided for each course. However, uptake of this policy is still ongoing (as shown in results below).

The AI upheaval begs the question of how and to what extent various course AI policies have impacted assessment. While “impacted” can refer to many different aspects of assessments, this project report focuses on courses’ AI policies and they may have impacted student performance, namely grades, thus indirectly exposing the challenges for assessing.

Jay about the SEQ:

“I found the SKE quite valuable. It challenged me to think about assessment, policies, and teaching in ways I hadn't encountered in my normal duties. And it also provided me with opportunities to look further into some issues my faculty has been facing with an eye towards the future of our education.”

Summary findings

For the reader's convenience, a high-level summary is first offered:

- **AI use permissibility has generally increased** in M&C courses between the 2023-24 and 2024-25 academic years.
- However, course grades are **not necessarily skyrocketing**, as the author initially feared. In fact overall, grades are diminishing for BA level despite increasing permissibility of AI use policies. However, this is largely due to decreasing exam grades.
- Thus, one could infer that students – especially BA students – need to use **AI more judiciously** and could receive guidance/training in critical AI use.
- An **explicit AI use policy for a course is necessary**, while the extent of permissibility should be considered with caution. While uptake of the policy page has been increasing, it is still not near 100%.

More summary details are:

- Courses with more permissive AI use do indeed exhibit higher grades, especially for non-exam assessments. While this pattern is statistically significant and possibly alarming at first blush, the increase is not drastic and qualified, as the reader will see.
- However, the dominant trend is that ESHCC course grades in general have actually dropped throughout the AI era (defined as after January 1, 2023 when indications of student use of ChatGPT emerged). But this is largely due to the downturn of grades at the BA level. My current speculation is that the detrimental effect from the COVID pandemic on student knowledge and learning lingers, more for BA students. An additional explanatory factor, with some evidence below, is that over-reliance on AI could be reducing the quality in student coursework, again seen for BA students, as well as their performance on exams.
- Permissibility of AI use has both increased and decreased between 2023-24 and 2024-25. It increased in the sense that AI policy pages have become more permissive between those two academic years. However, the number of courses without any AI policy pages was much higher in 2023-24 than in 2024-25. The absence of a page is tantamount to complete or at least uncertain permissibility. These changes to permissibility across years are statistically significant. This would at least partly explain the improvement in grade observed from more permissive AI policies.

Methodology

The data employed comprise:

1. "Use of AI" policy page on Canvas obtained via the Canvas API for all 2023-24 and 2024-25 courses in all three departments of ESHCC: a) Art & Culture Studies (ACS), b) History, and c) Media & Communication (M&C).² The sample is a full population, $N_{AIpages} = 442$.
2. AI-assisted quantitative content analysis (QTCA) of these policy pages – via Erudite – for only M&C courses, producing two AI scores: a) permissiveness of AI use, scored from 0-10, labeled 'AI' in tables far below, and b) the number of different ways AI could be used in the course, labeled '#'. See Appendix 1 for the actual prompt used. The sample size is $N_{AIscores} = 284$ (across both academic years, 2023-24 and 2024-25).

3. Course grades obtained via OSIRIS (9.2.88 report) for a sample of M&C courses across all available academic years for those courses. These courses include both the thesis courses, some mandatory courses, and non-mandatory courses, at both the BA and MA levels and spanning the range of AI scores produced by the aforementioned QTCA. So the sampling was only minimally random. The Research Masters (SCMA) and Pre-Master (PreMA) courses have been excluded from selection. The sample size of courses is $N_{\text{Courses}} = 22$ (thus far); the total number of individual student course grades is $N_{\text{StudentCourseGrades}} = 36131$; the number of course-avg-years is $N_{\text{AvgGradePerCoursePerYear}} = 352$. Assessments/assignments were flagged whether or not they were an examination (exam) or not.

Some of the analysis below has been abridged for the SEQ magazine.

Analysis

In each of the subsections below, a quick summary is first presented in ***bold italicized orange***.

Summaries of abridged sub-sections

Uptake of AI policy:

Not all departments in ESHCC are equal in their uptake of the AI use policy Canvas page, which is still ongoing.

- The percentages of AI use policy pages were compared across the three departments at ESHCC and reveals that uptake is 1) still incomplete and 2) varies across departments, with the highest level's being 84% (M&C) and lowest 67% (History).

Variation of course grades over time (i) – Selection of courses of interest:

Grade averages vary across courses, BA/MA programmes, and time, even in the recent AI era: no obviously singular effect in the AI era for this small selection of 6 courses.

- The full report presents a visual of grade trajectories for a handful of courses of interest to the author. What is revealed is that for this subsample of courses, there is not overarching positive or negative trends of grades in the AI era.

Time periods: Pre-COVID, COVID, and the AI era:

Grades for the larger selection of 22 courses have diminished in the AI era, largely at the BA level.

- In the full report, time periods were separated into pre-COVID, COVID/pre-AI, and AI era. The grade comparisons (via ANOVA) reveal significant downturn of grades for the sample of 22 courses in the AI era.

The remaining analysis subsections are presented in full.

AI policy comparisons

AI use permissibility increased slightly between 2023-24 and 2024-25, and clearly higher for MA courses than BA courses.

We now turn to differences in AI permissibility across years (2023 vs. 2024) and programmes (BA vs. MA) within only the M&C department. Single apostrophe ' indicates when a variable accounts for when there was no policy page, which was scored as maximum permissibility (AI=10,#=1,wc=0). Small note: Within the M&C department, there are more BA courses (~88) than MA courses (~55):

AI use	BA	MA	Δ	p	Interpretation
AI	4.23	5.31	1.08	0.017	AI permissibility higher for MA courses ...
AI'	5.38	6.64	1.26	0.004	... esp. when accounting for courses w/o AI policy.
#	1.58	1.97	0.39	0.188	Number of permitted uses of AI higher for MA courses (ns).
#'	1.46	1.70	0.23	0.293	
	2023	2024			
wc	2.40	2.50	0.10	0.495	2024 policy pages are slightly more verbose than 2023 pages (ns).
AI'	5.38	4.94	-0.44	0.076	Permissibility decreased in 2024 due to lack of policy pages in 2023.
AI	4.44	4.75	0.31	0.494	For existing pages, permissibility slightly increased between 2023 and 2024 (ns).
#'	1.44	1.67	0.23	0.226	Permitted ways of use also slightly increased (ns)
#	1.65	1.78	0.23	0.607	

- Mean values are reported (columns 2 and 3); Δ = difference in means between 3rd and 2nd columns; p = p-value from a two-sample t-test; red indicates p < .05 (statistically significant); pink indicates p < .10 (weakly significant); ns indicates p > .10 (non-significant or insignificant)
- AI = AI permissibility score; # = number of distinct ways AI can be used; wc = word count of AI page (verbosity)

Findings: As noted in the table itself, permissibility has increased in 2024, and has generally been higher in MA courses, which is not surprising given fewer exams at the MA level.

Impact of AI use (policy) on course grades

AI use permissibility tends to confer higher grades. But the findings in this section are tentative due to mixed effects in BA and MA levels (shown farther below).

And here we finally establish a direct link between AI use permissibility and course grades for the 2023-24 and 2024-25 academic years, for the sample of selected courses within the M&C department: N = 22 courses, but the full data is N = 6151 student course grades. Both Pearson correlation r and simple regressions were conducted. Here, only significant (p < .05) and weakly significant (p < .10) results are reported (in ascending sorted order of the p-value).

- For a simple regression, the standardized coefficient b^* is equivalent to the Pearson correlation r , thus the column appears as “ r, b^* ” A bigger positive value means the IV (independent variables) contributes or correlates more with the DV (dependent variable).
- The B is the unstandardized regression coefficient that indicates how much a unit increase in the IV (e.g., one point in the 0 to 10 AI permissibility range) would contribute to the DV (e.g., course grade).
- The 23 and 24 indicate the academic years: 2023-24 and 2024-25.

DV	IV	r, b^*	p_r	B	p_B	Adj- R^2	Interpretation
g_{23}	AI'_{23}	0.118	0.000	0.032	0.004	0.014	AI permissibility contributed to grade increase in 2023.
g_{2324}	AI'_{2324}	0.066	0.000	0.021	0.004	0.004	Less strong effect when considering both years
g_{23}	$\#_{23}$	-0.065	0.000	-0.021	0.006	0.004	More ways of use however oddly corresponded to decrease in grade.
g_{24}	wc_{24}	0.107	0.000	0.063	0.012	0.011	More verbose policy pages conferred higher grades in 2024, when a page was present (collinear with #).
g_{2324}	wc_{2324}	0.058	0.000	0.038	0.009	0.003	Similar but less strong effect considering both years
g_{23}	wc'_{23}	-0.099	0.000	-0.071	0.012	0.009	Absent policy pages increased grades in 2023
g_{24}	$\#_{24}$	0.054	0.006	0.030	0.011	0.003	Similarly to above wc , more ways of use conferred higher grades in 2023-24.
g_{23}	$\#_{23}$	-0.048	0.021	-0.014	0.006	0.002	Again, absent pages dominate to obscure the positive impact
g_{2324}	$\#_{2324}$	-0.024	0.062	-0.010	0.005	0.000	Similar to above

Findings: Firstly, the Adjusted R^2 – i.e. how much of the variation in the DV is explained by the IV – is unsurprising given that course grades are determined but many other factors than use AI use policy.

- The obvious finding is that more AI permitting courses contributes to a higher grade for the student for the given course-year, compared to less permitting courses that year. This point is revisited below.
- The negative correlations/regression coefficients are due to the manner in which the wordcount variable (wc) and the number of ways of use ($\#$) were coded. For courses without AI policy pages in 2023-24, the wordcount for course without a policy page was set to 0 and the $\#$ was set to 1 (a subjective decision made by the author). The negative values we see above indicate that these dominated the correlation, meaning they indicate the reverse: these courses exhibited relatively higher grades.

The above finding appears to contradict the downward trend of grades. So, now we examine if effects might be distinguished, between BA and MA level courses.

Variation of course grades overtime (ii) - Distinguishing BA and MA courses

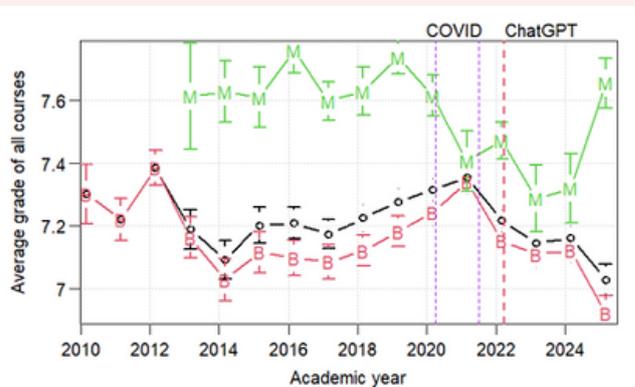
Newsflash! Grades for MA courses drastically increased between 2023-24 and 2024-25 but decreased for BA courses!

However, given possible the dominance of the higher number of students in M&C's BA programme, it behooves us to separate the analysis for BA and MA-level courses. If we look at course averages for the N = 22 sample of courses, also distinguishing between BA and MA courses, we find a striking difference in recent years.

- NBA = 12 courses; NBAgrades = 30040 ; NMA = 10 courses; NMAgrades = 6091
- The black line depicts the overall average, which is dominated by the preponderance of BA grades. The x-axis points were selected as 6 months into the academic year, i.e. February of the subsequent year. A more precise depiction would employ smoothing taking into account more precise timing of grade establishment.

Findings: While grades remained stable in both BA and MA in the early years of ChatGPT, these later rose sharply for MA students while declining for BA students.

BA students use AI worse than MA students?



If we conduct separate correlation/regression analyses, we find an explanation for the seemingly contradictory results thus far of AI permissibility conferring higher grades and yet grades from the selected courses decreasing since the end of COVID.

The row ids below refer to the full tables found in Appendix 4. Separate cross-year analyses may be found in Appendix 3.

	DV	IV	r,b*	p	B	P _B	Adj-R ²
	BA						
6	g ₂₃	AI'23	0.088	0.000	0.020	0.004	0.007
10	g ₂₄	AI ₂₄	-0.069	0.001	-0.029	0.009	0.004
11	g ₂₄	AI'24	-0.069	0.001	-0.029	0.009	0.004
12	g ₂₃₂₄	AI ₂₃₂₄	-0.046	0.003	-0.016	0.005	0.002
	MA						
7	g ₂₃	AI'23	0.202	0.000	0.100	0.018	0.040
9	g ₂₃₂₄	AI'2324	0.173	0.000	0.082	0.014	0.029
12	g ₂₃₂₄	AI ₂₃₂₄	0.163	0.000	0.116	0.025	0.025
14	g ₂₄	AI ₂₄	0.148	0.004	0.056	0.019	0.019
15	g ₂₄	AI'24	0.148	0.004	0.056	0.019	0.019

Findings: The above table tells two distinct stories for BA and MA students. For BA students in 2023-24, AI use permissibility – and lack thereof given the higher number of courses with absent pages in 2023 – positively impacted their grades. However, in 2024-25, when AI use policy pages appeared for all of the selected courses, permissibility appears to have diminished grades, which could partly explain the downturn of grades. For MA students, AI permissibility conferred higher grades in both (and across) years, absent policy pages notwithstanding. Two possible explanations are:

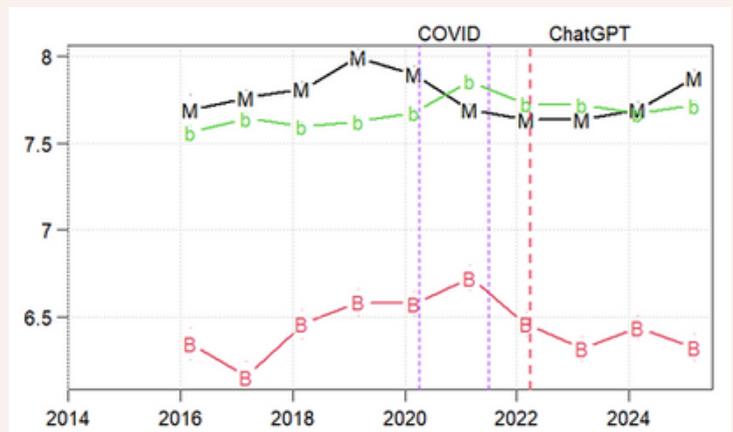
1. BA lecturers are more strict than MA lecturers in assessing AI-assisted coursework; this seems unlikely since many of M&C lecturers teach at both the BA and MA levels.
2. BA students have been using AI less intelligently or less judiciously – in light of increasing permissibility – leading to the lessening quality in their coursework. This is further confirmed in the next analysis below. Conversely, MA students could be using AI more intelligently/judiciously.

Still, the positive effect for MA students is concerning. While the grade average's returning to preCOVID levels for MA students may appear to be a desirable outcome, one might be concerned that over-reliance on AI hampers the development of their own academic abilities. The story becomes more nuanced when we consider the striking downturn of grades for the MA thesis in particular shown in the first figure of the full report – for which AI use is permitted with strict conditions. But as we all know, many students will use AI in a manner that can violate a course's policy conditions. These last two points should mute the claim that MA students easily use AI better than BA students.

AI use impacts non-exam assessments, improving their grades, more than exams: not entirely surprising!

Finally, exam and non-exam assessment grade averages are distinguished in the figure below, where 'b' indicates non-exam assessment for the selected BA courses and 'B' indicates exam assessments for the courses; 'M' indicates MA courses, none of which had exams. The confidence intervals lie close to the points so are not visible.

We can see that during in recent years the non-exam assessment grade average increased between 2023-24 and 2024-25 while grades for exam assessments decreased, suggesting that AI permissibility primarily impacted non-exam assessments. This is somewhat but not wholly expected, since students could use AI to augment their learning even in exam-heavy courses.



Looking more precisely via a regression model, we confirm the divergent grade trajectories in the AI era. Here, the DV (dependent variable) is individual student course grades in 2023-24 and 2024-25. The robust AI use permissibility variable (AI' in Model #1) and number of ways of AI use (#' in Model #2) are tested separately due to their collinearity. The independent variable (IV) of Exam Weight encodes the sum of exam weights for each course/year. The interaction terms capture the joint or moderated effect of EW on each of the AI use variables (AI' and #').

- The statistical significance here is presented using the standard notation: '^' means $p < .10$ (weak significance); '*' means $p < .05$ (standard significance) ; and '***' means $p < .001$ (highly significant).

IV	Model #1	Model #2	Interpretation
Intercept	7.34***	7.42***	
AI'	0.01		Main effect of AI permissibility <i>n.s.</i>
#'		-0.004	Same with number of ways
Exam Weight (EW)	-0.53***	-0.66***	Higher exam weight lowers grade
AI' x EW	-0.03*		Higher AI use less effective for exam heavy courses
#' x EW		-0.05^	Similarly with number of ways
Adj- R^2	0.04	0.04	

Unsurprisingly, higher exam weights confers lower course grade. However, the AI effects (AI' and #') mingle (or interact) with the extent of exam weight such that at least AI permissibility is muted, the higher the exam weight and conversely lower exam weight could make AI permissibility have a significant effect on grades, possibly a negative one, further suggesting the less effective use of AI by BA students.

Discussion

It has become clear that AI is indeed impacting student performance, and likely impacting the processes of assessment, which can manifest as adaptation of rubrics, adjustment of course content and assignment conditions that could be less vulnerable to AI, or fraud checking. If grades were to recently exceed expectations (e.g., very high course averages), this could indicate a serious problem in the effectiveness of assessing. However, recent grade fluctuations are not alarming (yet), but perhaps we need further investigation into MA courses. Even so, AI's impact to grades should make us educators question the extent to which students are actually meeting course learning goals, a concern echoed in the literature (e.g., Dawson et al., 2024).

As for judicious AI use – which this report claims is higher for MA students than BA students – one stakeholder (ESHCC's Vice Dean of Education Tonny Krijnen) mentions that a new course in responsible AI use is being developed for first year BA students in M&C. She also noted that MA students may be more motivated in their learning than BA students, hence, another reason for their possibly more judicious use of AI. Another stakeholder (ESHCC's Secretary of Examination Board Alex Kortekaas) rightly notes BA courses' having more AI-proof exams, while MA courses often do not include exams. However, one counter-argument could be that AI can be used by students as a supplementary learning tool for course material, thus indirectly improving their outcomes in assessments.

One final (last minute) note relates to the need for explicit AI policy, point #4 in the high-level summary earlier in this report. Buried within the myriad analysis (in Appendix 3) is indication that some courses and their assessing or assessments may not have fully adapted to their changed AI policies. This ought to be further looked into and also has implications on the course quality assurance responsibility of the ESHCC Examination Board, which has recently included AI policy check in its course quality assessment.

Limitations and future directions

There are of course various limitations which could point to possible future directions for this research. Some of these limitations were raised earlier in this report.

1. Some of the courses which did not have an explicitly AI use policy may still have announced one within the course (within another Canvas page or in a lecture slide).
2. Students' submission of the use of AI declaration form – often required by ESHCC's AI policies may or may not have been impactful to some of the outcomes. Obtaining these forms is theoretically possible.
3. Erudite's QTCA may harbor some errors, and its outcomes could use accuracy verification. However, if its accuracy errors were high, we would not likely see many of the statistically significant results.
4. The analysis currently distinguishes exams vs. non-exam assessments only at a surface level and could be looked into further.
5. Since the 2024-25 academic year is not yet over, one finding in section 4.6 – the striking distinction between BA and MA trends – is tentative. Possibly BA students perform worse in the first half of the academic year. This can be tested using grades from earlier years.

Endnotes

1. These pages are accessible only to members of EUR.
2. Not all AI policy Canvas pages were titled exactly as "Use of AI", so variations of the title were accounted for in the scraping.

References

- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005-1016. <https://doi.org/10.1080/02602938.2024.2386662>
- Erasmus University Library (n.d.). *Generative AI (GenAI)*. Retrieved April 11, 2025 from <https://my.eur.nl/en/library-employee/education-support/student/generative-ai>
- Erasmus University Library (n.d.). *Generative AI: Impact on teaching and assessment*. EUR. Retrieved April 11, 2025 from <https://my.eur.nl/en/library-employee/education-support/lecturer/generativeai-impact-teaching-and-assessment>
- Learning and Innovation Hub. *Use of AI*. EUR. Retrieved April 11, 2025 from <https://my.eur.nl/en/eshcc-employee/learning-innovation-hub/generative-ai/policy>
- Learning and Innovation Hub (n.d.), *Generative AI in education at ESHCC*. Retrieved April 11, 2025 from <https://my.eur.nl/en/eshcc-employee/learning-innovation-hub/generative-ai>
- Luth-Morgan, M., Valter, C., & Poort, L. (2024). ChatGPT als katalysator voor meer authentieke toetsing: praktijkvoorbeelden uit het ethiekonderwijs. *Ars Aequi*, September(2024), 790-794.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>



INCLUSIVE ASSESSMENT OR ASSESSMENT FOR INCLUSION?

International
Institute of
Social Studies



MARIJN FALING

The challenge of acknowledging and accommodating diversity in assessment

The International Institute of Social Studies (ISS) prides itself on a diverse student body, characterized by differences in cultural backgrounds, professional experiences, academic disciplines, and skill levels. This diversity forms the cornerstone of ISS' didactical approach, as stipulated in the ISS Draft Assessment Plan, 2024-2025. An inclusive vision on education aims not only to accommodate but also to strengthen and cherish this diversity at ISS. This squarely aligns with Erasmus University Rotterdam's (EUR) Educational Vision, which promotes inclusivity and the dismantling of barriers to equal participation in academia ([EUR educational Vision](#)). Beyond ISS and EUR, inclusive education has broad relevance as a means to support therealization of every individual's potential.

While inclusiveness is operationalized in ISS' education policy — such as through promoting participatory lectures, small-group collaboration, and critical reflection on students' own positionality and bias — there is room to strengthen the integration of inclusiveness in assessment. The Draft Assessment Plan 2024-2025 mentions the importance of a multitude of modes of assessment, through scaffolding demonstrating increased assessment of critical thinking as well as a shift to higher levels of Bloom's taxonomy from reproduction of knowledge to discuss and question theories learned. This could be a useful starting point to further incorporate inclusive assessment in assessment policy and practice at ISS.

As the ISS strives to transition to a new one-year master's program from 2026 onwards, this moment presents a critical opportunity to rethink and strengthen assessment.

This ensures that assessments not only measure knowledge acquisition but also reflect the inclusive values of the institution, fostering fairness and equity in student evaluation. To truly consider the diverse student population at ISS, inclusion needs to be explicitly designed into assessment at different levels (McArthur, 2016). Consequently, it is imperative to operationalize and implement inclusive assessment in the various assessment entities at ISS[1].

This article focuses on inclusive assessment by unpacking what this means and how this may be more strongly operationalized and integrated into ISS' Assessment Plan, as well as other assessment entities throughout the Institute.

Defining inclusive assessment

Inclusive assessment has gained traction in educational discourse, though interpretations vary widely.[2] Broadly, inclusive assessment aims to design assessment methods that prevent exclusion and disadvantage while enabling all students to demonstrate their competencies effectively (Oxford, n.d.). Hockings defines inclusive assessment as “the design and use of fair and effective assessment methods and practices that enable all students to demonstrate to their full potential what they know, understand and can do.” (Hockings, 2010, p.34).

[1] An overview of assessment entities is visualized in the quality pyramid of assessment, which incorporates assessment organization, assessment literacy, assessment policy, assessment program, assessment, and assessment tasks ((Sluijsmans and Struyven, 2014).

[2] For this assignment, the focus is on inclusiveness relating to summative assessment or assessment of learning. Inclusive assessment could be equally relevant to assessment for learning and assessment as learning. See Schellekens et al. (2021) for an overview.

Historically, inclusive assessment emerged to accommodate students with disabilities, often under a medical model that required a process of individual diagnosis, followed by appropriate accommodation practices, such as extra time, or separated space for testing (Nieminen, 2022). Over time, a shift toward the social model of disability has reframed assessment as an institutional responsibility (Hanesworth, 2019). Rather than placing the burden on students to seek accommodation - whereby assessment design may actually coproduce a notion of disability - assessment design came to be understood as proactively incorporating diverse formats to minimize the need for individualized adjustments (Nieminen, 2022). The social model acknowledges broader systemic barriers in assessment related to race, gender, class, and their intersections (McArthur, 2016). It promotes inclusion through diversity in assessment practices to offer multiple modalities to assess the level of knowledge uptake and hence reduces the necessity of accommodating individual circumstances (Nieminen, 2022). Inclusive assessment in this interpretation contributes to transformations towards more just societies (McArthur, 2016).

This assignment builds upon the social model's interpretation of inclusiveness, as such aligning with the diversity and inclusion aimed for at ISS.[3] Inclusive assessment thus refers to assessment that fosters inclusion through the acknowledgement of all students, including those marginalized or disabled and with various socio-cultural, historical, and political positioning, as 'agentic members of higher education communities (Nieminen, 2022). From this perspective, the translation of inclusive assessment into practice should strive towards assessment enabling judgments that align with what a task is designed to assess, rather than evaluating student attributes that are unrelated to the intended outcomes. (Tai et al., 2023).

The importance of inclusive assessment

There are three key – interrelated - reasons for integrating inclusive assessment, both at ISS as well as higher education institutions more broadly.

The first reason is to safeguard program-level quality assurance. Since assessment steers student's learning, it is important to design quality assessments that are well-aligned with curricular goals and policies. An assessment plan is an important aspect in this regard.[4] A well-designed and coherent assessment plan serves four functions. It informs consistency and transparency as it promotes an open and understandable standardized approach to assessment; it ensures alignment and coherence as it aligns ILOs, teaching methods and assessment in a systematic way; and promotes continuity to facilitate sustainability in program delivery; and it informs quality, under the condition that methods of assessment are purposefully and carefully selected and organized (SEQ session 4; Sluijsman and Struyven, 2014). Thus, the assessment programme needs to be implementable to safeguard quality of education. It is acknowledged that inclusive assessment is an inherent part of inclusive education and hence it should be adequately conceptualized to inform implementation throughout assessment entities.

[3] This is not to replace specific accommodations for students with disabilities and/or special needs, but rather to complement and reduce specific individual accommodations where possible.

[4] According to Sluijsman & Struyven (2014), assessment quality is determined by the quality of six entities: assessment tasks, assessments, assessment programme, assessment policy, and assessment literacy and assessment organization. This assignment concerns recommendations for assessment programme or assessment plan, whereby 'methods of assessment are purposefully and carefully selected and organized, aiming at an optimal positive effect on learning'.

The second reason relates to ensuring validity of assessment. Traditionally, poor assessment performance has been attributed to individual students rather than systemic flaws in assessment design. However, certain formats may disadvantage students based on factors unrelated to their actual knowledge. For instance, an oral presentation that disproportionately penalizes students with anxiety jeopardizes the validity of the assessment (Tai et al., 2023). Ensuring that assessments accurately measure intended learning outcomes is essential to fairness and academic integrity.

Third, there is a student-related advantage to inclusive assessment. By recognizing and valuing the diversity of students' backgrounds, abilities, and ways of demonstrating knowledge, inclusive assessment supports student empowerment. It shifts the focus from 'fitting into a system' to creating a system that fits the learners, which empowers students in several key ways, including acknowledging different strengths – helping them to feel more capable and confident, as well as encouraging their agency – by giving them autonomy in ways of assessment, as well as acknowledging their lived experiences as part of the learning process. This may build students' agency as well as motivation.

Approach to this project

The research process for this project involved a review of academic literature as well as scoping of other universities' practices regarding inclusive assessment to understand current best practices and theoretical frameworks. This was complemented by a thorough review of existing assessment documentation within EUR and ISS to evaluate how inclusive education and inclusive assessment are currently incorporated and operationalized. Additionally, an analysis of ISS courses listed in the academic calendar was conducted to identify how assessment practices are embedded across the curriculum. Informal conversations with colleagues alongside my own experiences as examiner furthermore provided valuable insights and practical perspectives on the implementation and perception of inclusive assessment strategies within the institution.

Recommendations to strengthen inclusive assessment

ISS acknowledges the importance of inclusive assessment in its policies. This is an essential first step towards embracing and operationalizing inclusive assessment. In this section I highlight areas to improve operationalization of inclusive assessment within ISS assessment entities (following Sluijsmans & van Struyven, 2014), particularly in the following areas.

First, it is advised that inclusive assessment is approached from a holistic perspective, to inform assessment quality (Sluijsmans & Struyven, 2014). This includes assessment program and policy. While the importance of inclusive education is acknowledged in ISS' Draft Assessment Plan 2024-2025, with a predominant focus on inclusive classrooms the elaboration of inclusive assessment in assessment policy is too general and underdeveloped to clearly inform assessment practice. Especially considering the multiple various ways in which inclusive assessment may be understood as well as operationalized (Kleinlein, 2025), the Assessment Plan leaves room for a more systematic and clear elaboration of inclusive assessment. Concrete suggestions include to define inclusive assessment and detail how assessment should be designed to accommodate diverse learners. Although inclusive assessments could be simply operationalized to incorporate a balance of assessment tasks (including essays, case studies, and presentations) undertaken in different conditions (such as invigilated, and non-invigilated) and with timings that are distributed across the terms and year (see for instance University of Oxford, n.d.), the actual interpretation of inclusiveness critically hinges upon the norms and values prevalent within the institution. To make sense of operationalizing inclusive assessment in the context of ISS, it would be helpful to launch a peer group of examiners (and students?) to systematically explore a meaningful operationalization of inclusive assessment within ISS.

Based on analysis of current practices and preferences of both students and examiners, this working group could advise the Education Committee of the new master's program on embedding inclusive assessment throughout the assessment entities. Involvement of students in this exercise is essential as inclusive assessment opportunities are unlikely to be identified by examiners and education management, as they have a bias towards their own positionality and perspective (Leathwood, 2005). The subsequent recommendations are points to consider for this peer group.

Second, ISS could identify and map current inclusive assessment practices and ensure to embed these in assessment plans and policy. Currently there are various assessment practices at ISS that to a large extent already cater to inclusiveness. Throughout courses and assessments, individual examiners aim to widely incorporate authors as well as case studies from various contexts throughout the globe. Furthermore, students are frequently encouraged to incorporate their own experiences and insights as empirical cases in exercises and assessment. Recognition of these practices into policy – such as the Assessment Plan – would promote continuity and sustainability of inclusive assessment, while enabling to communicate these practices to fellow institutes and students.

Third, group work has the potential to enhance inclusive assessment (Nieminen, 2024). It allows students to contribute in different and unique ways, recognizing multiple intelligences and forms of participation, making space for students who might not perform as well in standard tests or individual written tasks. At ISS, collaboration is recognized as important to the programme. One of the ILOs of the MA program indicates the objective to “work collaboratively within a multidisciplinary and multicultural context [...]”. At least 1/5th of the courses have a group work component, judging from the self-identified assessment types indicated in the Academic Calendar.

As such, collaborative work is incorporated in policy and practice. However, group work is a topic of considerable debate in higher education, including at ISS. Currently, as per the ISS Student Charter examiners may base a maximum of 15% of the final course grade on group work. This is to curb challenges relating to e.g. fairness and validity. More importantly and related to inclusive assessment, group dynamics may disproportionately affect outcomes, with students from marginalized or less dominant backgrounds being excluded or undervalued in group settings (McKay & Sridharan, 2023). At ISS, there are indications that grades resulting from group work tend to be consistently higher as compared to grades based on individual assessment. Therefore, the Institute Board could consider to raise the percentage of the final course grade that is based on group work, meanwhile introducing safeguards to ensure fairness and validity, including for instance individual reflections or peer evaluations (Davies, 2009).

Fourth, to advance inclusive assessment an organization must be capable to monitor, oversee and inform assessment throughout the course portfolio. At ISS, the way in which insights is gained into assessment at the course level leaves room for improvement. Examiners identify and categorize the mode of assessment per course using open and self-defined categories. This leads to a non-exhaustive list of non-distinctive assessment modes, including overlapping categories such as essay assignment, take-home assignment, individual assignment, take-home essay, and reflection paper. This obstructs the Education Committee's role to safeguard the quality of assessment as well as alignment with the assessment plan, while making it difficult if not impossible for the Education Committee to advise examiners on modes of assessment to advance inclusive assessment throughout the organization.

Furthermore, the variety in assessment per ILO remains undiscernible, whereas from a viewpoint of inclusive assessment it could make sense to introduce variety in how ILOs are assessed throughout the program. To enhance monitoring and insight in the variety of assessment types offered, ISS could simply introduce a pre-categorization of assessment types from which examiners can pick.

Fifth and related, there are indications that despite the progression of assessment modes through the course of the program – from invigilated exams to the RP – there is limited variation in assessment, both in terms of assessment type as well as assessment timing. Regarding assessment type, based on informal conversations with various colleagues, there seems to be a predominant focus on essay-like assessment in most of the courses. Essay as assessment align well with ISS' focus on enhancing critical thinking, though simultaneously running the risk of disproportionately benefiting those with strong writing abilities and the skills to express arguments and ideas concisely and clearly. While important academic skills, these are often not integral to the objectives to be assessed, and these skills may not be equally present among the student population.

References

Barnartt, S. (2016). Role Theory and the Fluidity of Disability: World Perspectives in Culture and Society." In *Rethinking Disability*, edited by Patrick Devlieger, Beatriz Miranda-Galarza, S. E. Brown, and Megan Strickfaden, 47–57. Chicago, IL: Garant Publishers.

Davies, W.M. (2009). Groupwork as a form of assessment: common problems and recommended solutions. *Higher Education* 58, 563-584.

EUR (2023). Educational Vision Erasmus University Rotterdam. [online] last accessed 14-04-2025 <https://www.eur.nl/en/media/2023-09-education-vision-2023>

Hanesworth, P. (2019). Inclusive assessment: where next? [online] last accessed 11-03-2025 <https://www.advance-he.ac.uk/news-and-views/inclusive-assessment-where-next>

Hockings, C. (2010). *Inclusive learning and teaching in higher education: A synthesis of research*. EvidenceNet, Higher Education Academy. www.heacademy.ac.uk/evidencenet.

ISS (2025). Draft Assessment Plan 2024-2025 (internal document).

Kleinlein, E. (2025). Conceptualising inclusive assessment. A critical literature review at the nexus of education, inclusion, and assessment. *European Journal of Special Needs Education* <https://doi.org/10.1080/08856257.2025.2500137> .

Leathwood, C. (2005). Assessment policy and practice in higher education: Purpose, standards and equity. *Assessment and Evaluation in Higher Education*, 30(3), 307–324. <https://doi.org/10.1080/02602930500063876>

McArthur, J. (2016). Assessment for Social Justice: The Role of Assessment in Achieving Social Justice." *Assessment & Evaluation in Higher Education* 41(7): 967–981.

McKay, J. & Sridharan, B. (2023). Student perceptions of collaborative group work (CGW) in higher education. *Studies in Higher Education* 49(2), 221-234. <https://doi.org/10.1080/03075079.2023.2227677>

Nieminen, J. H. (2022). Assessment for Inclusion: rethinking inclusive assessment in higher education. *Teaching in Higher Education*, 29(4), 841–859. <https://doi.org/10.1080/13562517.2021.2021395>.

Nieminen, J.H. (2024). The paradox of inclusive assessment. *Assessment & Evaluation in Higher Education* 50(4), 564-576. [References](#)

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.

Sluijsmans, D. M., & Struyven, K. (2014). Quality assurance in assessment: An introduction to this special issue. *Studies in Educational Evaluation*, 43, 1-4. <https://doi.org/10.1016/j.stueduc.2014.08.003>

Tai, J., Ajjawi, R., Bearman, M., Boud, D., Dawson, P., & Jorre de St Jorre, T. (2022). Assessment for inclusion: rethinking contemporary strategies in assessment design. *Higher Education Research & Development*, 42(2), 483–497. <https://doi.org/10.1080/07294360.2022.2057451>

University of Oxford (n.d.) IncludED: A guide to designing inclusive assessments. [online] last accessed 14-04-2025 <https://wwwctl.ox.ac.uk/included-designing-inclusive-assessments>

NAVIGATING THE DIGITAL FRONTIER

GEORGINA GÓMEZ

International
Institute of
Social Studies



The integration of Artificial Intelligence (AI) into the academic landscape presents both unprecedented opportunities and challenges for teachers and students. The widespread use of generative AI in the last two years has significantly impacted research, teaching, and innovation. This has, in turn, challenged academic integrity, equity, and ethical conduct to the point where the academic community may need to redefine these terms. One key area affected is the definition of critical thinking – specifically, who is thinking?

Given that it's safe to assume all students are now using AI in some capacity, the crucial question becomes how to regulate its use feasibly to allow examiners to assess student work accurately. This text has been English-edited by Gemini AI, for example, considering that the author uses English as a Second Language, but the ideas are her own conception.

In response to these developments, the European Commission issued the "Living guidelines on the responsible use of generative AI in research" in March 2024 and has been updated several times since then. This document has inspired European and Dutch universities, including EUR, to develop their own adaptations. [1]

This article aims to address three main goals:

- Define "responsible" use of AI.
- Identify permissible and impermissible AI uses that might obscure the assessment of "only human" intelligence.
- Draft an action plan to build skills for responsible AI use.

Defining Responsible Use of AI

There is no universal consensus on "responsible use of AI" in postgraduate education. However, the concept consistently refers to the ethical, transparent, and accountable use of AI tools. Most sources indicate that AI exacerbates a tension between maintaining the quality of learning, research, assessment, and scholarly activities while upholding academic integrity, ensuring data privacy, and fostering critical human capacity.

Responsible AI use generally seeks a balance between enhancing human learning and limiting AI. While AI can significantly boost productivity and support learning, it also prompts a re-evaluation of what is uniquely human. Therefore, "responsible use of AI" is contingently defined by this balance: increasing human intellect and creativity without replacing them. It essentially questions the right balance between using AI and protecting human learning.

At the International Institute of Social Studies (ISS), the aim is to train postgraduate students for a future where AI is integral to their professional lives, assuming that skilled professionals will need to master AI tools.

Simultaneously, ISS is committed to encouraging academic integrity and penalizing false claims of student ownership of work.

Policies on the responsible use of AI have been issued by several universities, multilateral organizations, and the Graduate School at EUR. This article draws upon these documents, particularly the EC guidelines. The EC document emphasizes that responsible AI use must uphold the principles of reliability, transparency, respect, and accountability. [2]

Here's a summary of key guiding principles and guidelines:

- **Reliability:** Researchers must produce quality knowledge, verify results, and be aware of biases.
- **Transparency and Disclosure:** There is an obligation to disclose the use of AI and provide proper attribution.
- **Accountability:** This involves human control, critical engagement, and responsibility for one's academic work in relation to peers and other knowledge communities.
- **Skill Development:** While specific skills may be debated, critical thinking and analytical abilities are paramount for human learning processes.
- **Data Privacy and Security:** Users must adhere to institutional data protection policies to ensure confidentiality and security of sensitive information, including respect for informant privacy and intellectual ownership.
- **Energy Use:** Processing AI algorithms is computationally and electrically demanding.
- **Equity and Social Justice:** AI use can perpetuate biases (e.g., ethnicity, culture, gender, race), potentially marginalizing minorities unless their data is specifically and consciously retrieved.
- **Disciplinary Ethical Considerations:** Literature suggests nuances and differences across disciplines.

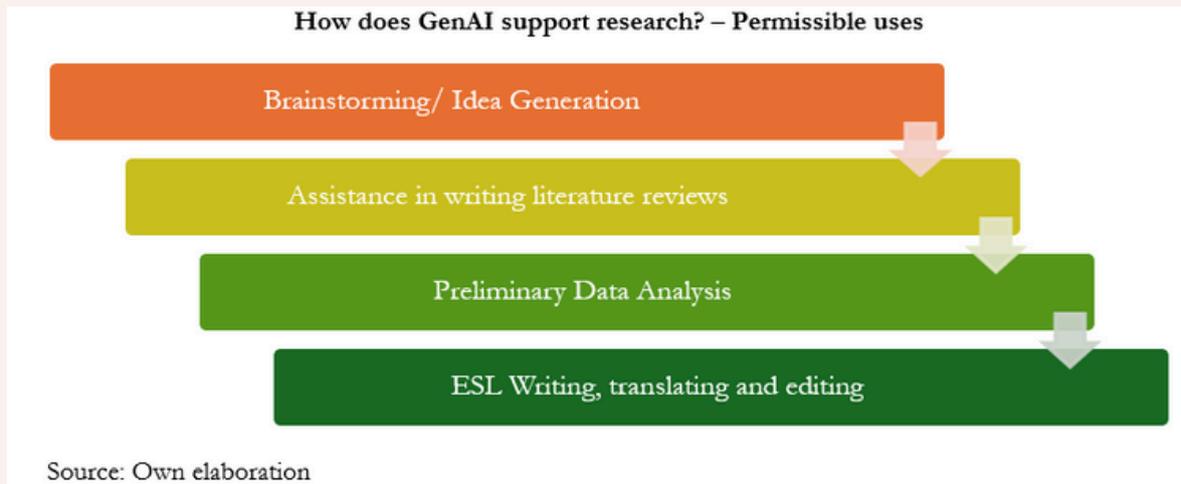
Uses of AI and the Research Cycle

"Cheating" is defined as any activity that obscures the examiners' ability to assess student progress (ISS Student Charter, 2024). Under this definition, any AI use that helps students achieve more than they would have without it is often treated similarly to other forms of cheating by Examiner Boards. See through the lens of the research cycle, there are steps in which the use of AI becomes evident.[3] The CLI (2025) document highlights the practical need for a more productive approach that embraces AI. Within the research cycle, some AI uses are permissible, while others are inappropriate:

Permissible Uses of GenAI (How GenAI supports research)

- **Brainstorming/Idea Generation:** AI can be used for initial topic exploration based on current research agendas and to generate multiple research questions for experimentation. These ideas must be subsequently developed and researched by students.
- **Assistance in Writing Literature Reviews:** AI can significantly enhance productivity by identifying, summarizing, and extracting relevant literature. Many popular readers include tools to summarize articles, helping students cover more research. AI can also help organize sources for future reference. However, summarizing is not the same as critically reviewing literature, a qualitative difference examiners look for.
- **Preliminary Data Analysis:** AI can help identify broad trends and reduce time for complex dataset analysis. Nevertheless, identifying only broad trends can bias and flatten the analysis, so teachers should support the development of skills to critically interpret main patterns in data.

- **Writing and Editing Support:** Non-native English speakers can benefit greatly from AI for grammar checks, style suggestions, or proofreading. However, AI introduces biases and often selects "most used words," leading to a loss of nuance and originality. When AI summarizes or retrieves highlights, it tends to focus on frequently mentioned content, not necessarily the main contributions, requiring the researcher to verify the output.



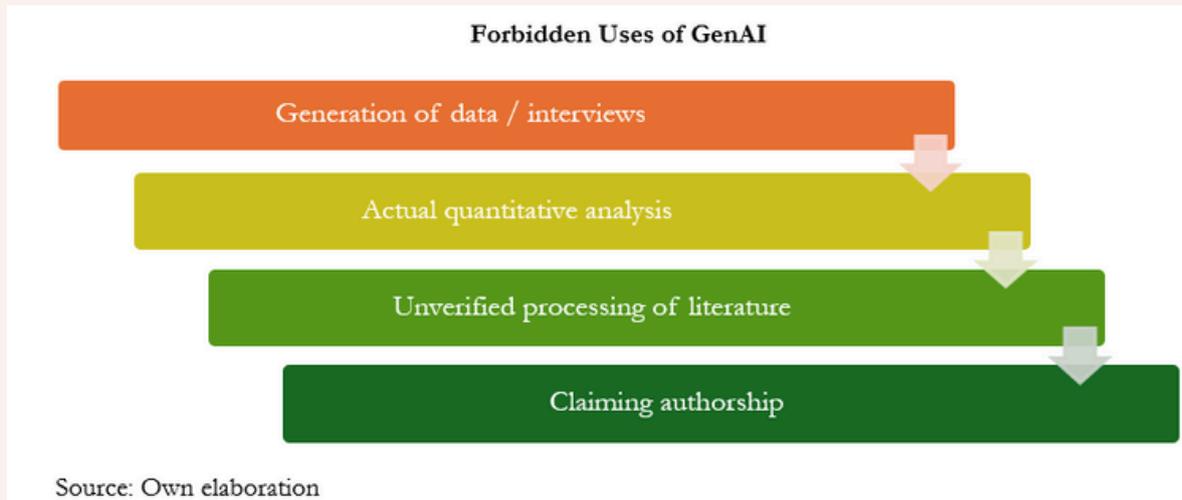
As these examples show, AI can assist researchers in searching, writing, and summarizing, but each activity implies a loss of control unless verified by a human. The limits of these permissible uses can only be mitigated by human intelligence. Students need to move beyond what they can obtain from generative AI, and to do so, they must first understand AI's limitations and gain the capacity to critique them. While more liberal permission for AI tools is emerging, greater awareness of its limitations is crucial.

Generative AI tools can still produce inaccuracies, biases, and "AI hallucinations" (false information from interpolating existing data). All AI-generated content must be critically evaluated and verified by humans with sufficient knowledge to assess its veracity.

Violations of academic integrity should continue to be penalized, and the guiding principles remain the same:

- Work that is not the author's creation should not be submitted as such.
- The principle of disclosure regulates what AI-generated text, ideas, or data can be included, provided the source is mentioned with proper attribution.
- Falsification or fabrication of data, which generative AI can also do, still constitutes a violation of academic integrity.

The principle of disclosure has been frequently emphasized in relevant documents. Consequently, examiners need to enhance tools and opportunities for disclosure and attribution. The CLI document suggests actions like repeated reminders before submissions, differentiating the steps in which AI was used, its purpose, and how AI-generated materials were controlled. Frequent reminders increase user awareness and create opportunities to specify AI tools used and the extent of engagement.



Proposed Recommendations for Examination Boards

- Publish clear communication regarding what is permitted or prohibited in courses and research papers.
- Prepare examples and explanations for decisions on what is allowed or forbidden.
- Be aware that AI detection tools show 15% false positives, and staff must understand these limitations.
- Stay informed about new detection tools. AI has led to an overuse of certain words previously less common in human-generated writing, a phenomenon termed "AI-overuse". James Welch updated a list of AI-overused words in May 2025. Another source lists words ChatGPT often overuses, such as: Delve, Tapestry, Vibrant, Landscape, Realm, Embark, Excels, Vital, Comprehensive, Intricate, Pivotal, Moreover, Arguably, Notably.
- Encourage examiners to design assessments that do not rely on summarizing, preparing literature reviews, translating existing knowledge, or managing data. Assignments should prioritize higher-order thinking skills, critical analysis, and original thought. This can include in-class discussions, oral presentations, or practical applications.
- Educate faculty on the responsible use of AI through workshops or newsletters, covering its ethical dimensions, potential impact, and implications for teaching and supervision.
- Facilitate discussions on maintaining and adapting academic standards.
- Include training and workshops on AI literacy, responsible AI use, and the ethical implications of AI for both students and staff, integrating these into regular curricula. For example, development studies students need to be particularly aware of algorithmic biases and their power implications.
- Include opportunities for disclosure in the TMS (Thesis Management System), allowing students to acknowledge authorship. This would also help ISS generate data on AI usage, enabling improvements and updates to training workshops.
- Repeat and control disclosure and verification opportunities.

References

Community for Learning and Innovation (12 February 2025) *Educational Challenges in an era of AI*, EUR

European Commission, Directorate General for Research and Innovation (March 2024) Living guidelines on the responsible use of Generative AI in research, ERA Forum stakeholders' document. EC: Brussels.

ISS (2023) *Programmatic assessment Plan, internal document*. The Hague. ISS (2024) Academic Calendar, internal document. The Hague

Smith, G., Luka, N., Osborne, M., Newman, J., Lattimore, B., Nonnecke, B., and Mittelstadt, B. (2025) *Responsible Generative AI Use by Product Managers: Recoupling Ethical Principles and Practices* ([link is external](#)). Bair and Berkeley Haas.

Sprung, G. (2025). *Estimating ai usage by comparing changes in wording and occurrence of phrases over time*. In EDULEARN25 Conference Proceedings (pp. 5252-5261). IATED. [10.21125/edulearn.2025.1317](#)

Swaak, T. (2024) *Adapting to AI*, *The Chronicle of Higher Education*. Washington DC (In [livershare](#), Accessed 30 may 2025)

Welch, J (2025), *A list of words that AI over-uses*. <https://embryo.com/blog/list-words-ai-overuses/>

West, H., Malcolm, G., Keywood, S., & Hill, J. (2019). Writing a successful essay. *Journal of Geography in Higher Education*, 43(4), 609–617. <https://doi.org/10.1080/03098265.2019.1655720>

MAY YOU LIVE IN INTERESTING TIMES – ASSESSMENT OF THE THESIS IN THE AI-AGE



UNIVERSITEIT VAN AMSTERDAM

PHILIPPE VERSIJP

While a writer's block may soon be consigned to the litter bin of history, the advent of Artificial Intelligence (AI) and Large Language Models (LLM, most known example is ChatGPT) is also having implications at universities that are less welcome. In particular, those concerned with safeguarding the assessment and the granting of credits and diploma's worry about the apparent ease with which complete essays and assignments can be produced in minutes – and without the need for the student to master the material. In these interesting times, examination boards, education management, thesis supervisors and yes, students too, need to find a way for the degree to remain meaningful. This article, and hence my SKE project, doesn't claim to have the solution, but hopes to highlight steps that can be taken to navigate these developments. The thesis will be the special focus of this project.

Goals

As a foundation of the analysis, it may be a good idea to first look at the goals of assessment as applied to the thesis. There's a wide choice, also because universities like to know if their job has been done well – if the student has learned something. Yet the external effects of assessment deserve consideration too: what does a diploma – the thesis tends to be the final step leading towards the degree certificate – mean to the outside world? While very little empirical data on this is available, a diploma may be assumed to demonstrate at least knowledge and abilities in line with the Intended Learning Outcomes (ILOs), both on a thesis (i.e. course) and programme level.

One of the goals therefore has to be that we can vouch for the student having met these ILOs (also a major consideration with accreditation and, last but not least, a task enshrined in law (WHW 7.10 to 7.12c).

Interestingly, these learning objectives are almost never formulated with AI use in mind, so a phrase like 'Report on the research findings in a clearly structured and well-written manner' can mean that the student has to have considerable writing and communicative skills him/herself, or that the student can let AI – or any other ghostwriter – transform some bullet points into smooth text. Implicitly, university lecturers always assumed that the first view is applicable, but with a staggering amount of greyscales in the form of AI assistance now available, the question needs a more explicit answer.

This article, and hence my SKE project, doesn't claim to have the solution, but hopes to highlight steps that can be taken to navigate these developments.

But it doesn't stop at the way the work is communicated. Core ILOs like 'Make a well-informed choice between theories and argue which parts are relevant to the research at hand' or 'Analyse and interpret empirical data' can just as easily be outsourced to AI. That cannot be the way it should be; otherwise prospective employers might just as well hire a good prompt-writer rather than a university graduate.

The external relevance of the degree and the knowledge and skills that implies is therefore directly threatened. If graduates can only work with the assistance of AI, they become unsuited for those jobs where reliability, confidentiality and energy usage – to name but three common concerns – prohibit the continuous outsourcing to AI. In fact, the more critical the task, the bigger the probability that a tool that is known to 'hallucinate' will be unacceptable. One can also wonder where an employer will find a competitive advantage if everyone uses similar AI. On top of this, a critical reflection on AI outcomes, so knowing when they might be used or not, also requires a solid foundation. If universities cannot validate that their graduates have that foundation and more, they'd be failing society at large as well as the individual student.

Measures

So what to do? The first recommendation seems obvious – but only after the best part of the SKE trajectory! : make sure that is explicitly stated which tools are allowed for which ILOs, and weigh the options carefully, also on a curriculum-level.

Secondly, clarity doesn't mean that everyone abides by the rules. The temptation can be enormous as the pressure under which a student finds him/herself mounts, nor can we exclude the scenario where the certificate is the sole consideration for the student and learning is willingly sacrificed at the altar of time-savings – or even deemed irrelevant. Both prevention and enforcement are necessary.

In many cases, moving the assessment to a controlled location (i.e. an exam hall, where access to unauthorized aides can be policed) or limiting the grade weight and checking students' mastery by asking critical, on-the-spot, questions, will provide a reasonable degree of protection. Yet assessment that by its very nature is too elaborate to have anything else than a major take-home component is more difficult. The thesis is the ultimate example of this challenge.

And it is a tough challenge: not only is the thesis too big a project to write in a controlled setting (even logging keystrokes in a digital learning environment is futile, apps exist to mimic this even though the text is copy-pasted from AI), checking only the final product is a thankless task. The distinction between student writing and AI writing is difficult (not impossible, but no detection package has as of yet been able to overcome the widespread reluctance attached to automated scanning; partly with the spurious argument that detecting should be 100% accurate on its own, so as a sole measure to decide innocence or guilt – an unreasonable standard).

Towards a safer thesis

A more fruitful approach would be to force the student into actions that allow a supervisor to perform checks during the construction phase already. This doesn't – probably shouldn't – be summative assessment, but rather a condition that needs to be met. One turns the major weakness of the thesis – its magnitude – into a strength: it also allows from many points where checks can be made. If a student can knowledgeably discuss the ideas at the heart of the thesis, follow the supervisor when suggestions are given, identify the stronger and weaker spots of their own analysis, then a thesis assessor will not need to fear AI's influence. The student has shown mastery, and that would check the box on the ILOs too.

Of course, a process like that needs to be documented to some extent. External checks, also on whether the supervisor is doing a good job, are necessary. Trust is nice, but gives little mileage if a student challenges the grade by formal appeal, or the examboard needs to explain how it remains in control to an accreditation board. A changelog is a good first step regarding such documentation, especially if students are not only keeping track of which major changes they make, but also explain why they do so – and why they changed their opinion compared to an earlier moment in the process.

As an additional benefit, a changelog can make it easy for an assessor to ask critical questions; not only is that type of reasoning harder to outsource to AI, it goes to the heart of the academic process: questioning assumptions, earlier results, making adjustments and seeing if they bring understanding.

Yet this intense supervision on ideas and process does come at a cost. It is much more labour-intensive, especially from the perspective of the supervisor. It seems neigh-on impossible if student numbers are large; if a single supervisor has a few dozen students for the thesis in any given year, meetings will be limited, discussion will be laser-focussed on the tasks ahead and documentation becomes a profound drag. Also, it assumes a student who is indeed not too far away from mastery and can clearly communicate with the supervisor. That is not necessarily a given.

It is conceivable that larger programmes will be forced to emphasise the mechanical aspects of control, rather than the focus on thought. That is a pity, but probably inevitable. Yet even in that case having a dot on the horizon might be very helpful indeed. Getting closer to the ideal is always worthwhile.

But a necessary condition

But in the end, in order to make progress in tackling the extraordinary influence AI has on assessment, even for a thesis, we need to decide which of the current ILOs need to be wholly safeguarded from AI use – the ones where we can say, hand on heart, “the student knows this” – and which ones are secondary. Maybe all fall in the first category, and none in the second. It would after all depend on the ILOs themselves, which differ for each program. Societal expectations matter. They may also shift, though often at a glacial pace compared to technical innovation. But quite likely not all of these ILOs are of primary concern. Or maybe they are in the context of the program, but have been amply assessed in previous courses. The fact is that, unless we take a proverbial magnifying glass, we may not know – and therefore take insufficient measures to safeguard what truly matters.

June 2025

COHORT 2024/2025



SEQ/SKE MAGAZINE

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