AI-RESILIENT ASSESSMENT
A toolkit for analyzing assessments

OPEN UP NEW HORIZONS.

HAN_UNIVERSITY
OF APPLIED SCIENCES

THIS DOCUMENT WAS CREATED WITH CONTRIBUTIONS FROM:

Name	Role
Sabrina Bloemen	Assessment expert and educational advisor - Services: Education, Research and Quality Assurance
Jessica Koning	Assessment expert - School of Business and Communication
Minke Zeillemaker	Education specialist - School of Business and Communication
Mirjam Raanhuis	Education specialist - School of Social Studies
Jochem ten Böhmer	Innovation Lead Artificial Intelligence - Services: Education, Research and Quality Assurance
Jorn Bunk	Education specialist and AI expert - Services: Education, Research and Quality Assurance
Martijn Hendriks	Lecturer – School of Law and School of IT and Media Design, member of assessment committee
Maartje Gijselhart-Keijsers	Educational advisor - School of Organisation and Development
Sara Struik	Education specialist – School of Organisation and Development
Dirk Weijers	Assessment expert - School of Health Studies
Frank Vonk	Policy officer for Services: Education, Research and Quality Assurance
Marjoleine Dobbelaer	Researcher in teacher quality
Marion van de Wijdeven	Educational advisor and trainer - Teacher Quality research team
Edwin Buijs	Researcher in teacher quality
Mariëlle Wisselink	Educational policy officer - International School of Business
Rob Warmershoven	Senior lecturer - International School of Business
Ingeborg van den Heuvel	Lecturer - International School of Business
Wout Brink	Lecturer - Entrepreneurship & Retail Management

Version 1 was adopted on 10 October 2024 by the AI & Assessment theme group of the HAN Learning Network for Assessment.
Revised Version 2 by Sabrina Bloemen, Jorn Bunk, Jessica Koning and Minke Zeillemaker (10 March 2025).

CONTENTS

1 INTRODUCTION

AI systems, and specifically generative AI systems, have a clear effect on how students learn, what they learn (or should learn), and how degree programs assess whether students have mastered the desired learning outcomes. This requires degree programs to take a critical look at their education and assessment and make adjustments as needed. This toolkit is a guide to reviewing a program's assessments and evaluating to what extent they are AI-resilient.

> **AI-resilient education:** Education with constructive alignment, where AI is used only to improve, enrich, or support learning process and/or assessment.

AI-resilient education can include or exclude AI. When it includes AI, certain AI use may be permitted while other AI use is not. Effective education is grounded in solid constructive alignment, where learning outcomes, assessment and learning activities form a logically consistent whole in terms of both content and format.

For AI-resilient education, this means, for example, that if the learning outcome requires students to possess writing skills, the use of AI in the assessment is undesirable, whereas this may not be a problem in education that is focused on other learning outcomes.

Note:
1. The purpose of this toolkit is to provide insight into the AI-resilience of assessments in degree programs. The results do not speak to the validity or reliability of the assessments themselves.
2. As AI continues to evolve at lightning speed, we encourage degree programs to critically examine - and continue re-examining - their assessment in light of their overall exam program. This means including the rationale behind the risk indications and the intended measures in the analysis.

Figure 1: Constructive Alignment (Biggs & Tang, 2011)

1.1 Notes to the toolkit

This document is a toolkit designed to assist degree programs in initiating conversations about the AI-resilience of their assessment practices. The toolkit consists of a flowchart, accompanied by appendices that delve deeper into the questions posed in the flowchart. By working through the questions in the flowchart, degree programs can identify potential risks in their exam program related to

AI-resilience and use the resultsf the analysis to set priorities for ongoing development in education and/or assessment strategies.

Ideally, the toolkit should be used collaboratively with members of the board of examiners, assessment committee, curriculum committee (and degree committee) to analyze the entire exam program and discuss follow-up actions. This includes providing feedback to the teaching team and involving lecturers/examiners in implementing adjustments or improvements. The toolkit can also be used for individual assessments with or by the relevant lecturer(s)/examiner(s).

1.2   Key points and tips for the discussion

When conducting the conversation guided by the flowchart, we offer the following tips:

* Involve the various relevant actors from the degree program in the discussion.

* Involve your academy's assessment or education specialist(s) or contact the HAN Learning Network for Assessment. The Teams channel: HAN Leernetwerk Toetsing | Algemeen | Microsoft Teams[1]

* The purpose of the discussion is not to ban assessment formats, but to work together to find the right, suitable solutions for assessments that are vulnerable to AI-related fraud / AI-sensitive. Because AI-resilience depends on several factors, it makes sense for the questions in the flowchart to lead to answers that start with "yes, provided" or "no, unless" rather than a hard "yes" or "no." This allows for conscious choices to be made and gives options for taking action.

* The guiding factor in analyzing the assessment formats (see Appendix A) is what examiners really want to see from students, not the OSIRIS code and/or the name of the assessment format (in this document). Practice shows that the one does not have to coincide with the other. For example, the assessment format may be a portfolio, but in practice the student also has to discuss and defend the portfolio in an oral exam or criterion-based interview.

* An assessment should not be viewed as a separate entity. Assessments are derived from learning outcomes and, together with learning activities, make up the complete curriculum. Also, a single medium-risk assessment (see Appendix A) need not be a problem if the overall quality of the exam program as a whole is high and the student's learning process is visible.

* To ensure that the exam program is AI-resilient, analyze the risks for each assessment setting in addition to the assessment format. The actual level of risk of an assessment format depends on several factors, such as how the student's development is made visible and how assessments are administered.

---

[1] https://teams.microsoft.com/l/team/19%3Aa52986629b804e3db124edbf0bdee00e%40thread.tacv2/conversations?groupId=8ac74beb-bde3-4a2d-8f85-b03952141a0b&tenantId=5d73e7b7-b3e1-4d00-b303-056140b2a3b4

2   ANALYSIS OF AI-RESILIENT ASSESSMENTS

The goal is to get an initial indication for each assessment as to whether the assessment is vulnerable to AI-related fraud or is AI resilient. We do this based on the assessment formats. We define assessment formats that are vulnerable to AI-related fraud and AI-resilient assessment formats as follows:

Assessment formats vulnerable to AI-related fraud / AI-sensitive: All assessment products produced <u>without the direct supervision</u> of an examiner or supervisor in which one or more learning outcomes must be demonstrated without the use of AI. These formats carry a medium to high risk of AI-related fraud.

AI-resilient assessment formats: All assessment products created or administered <u>under the direct supervision</u> of an examiner or supervisor, as well as assessments in which the use of AI is (or has become) part of the learning outcome(s) or does not compromise the demonstration of the outcome(s).

The flow chart in Figure 2 was created to help work with these definitions. In Appendix A, for the first question in the flowchart, we explain for each assessment format what supervision is or can be involved in that assessment format and when it is AI resilient. n Appendix B, we provide examples of activities that can be performed using a low-threshold approach to AI. These examples can help answer the second question in the flow chart. The flow chart gives an initial indication of vulnerability to AI-related fraud. This indication does not consider the entire context of the assessment. This context, including learning activities and thus visibility of the learning process, also affects the actual vulnerability of an assessment to AI-related fraud. So use the result from the flowchart as a discussion starter, not a final answer!
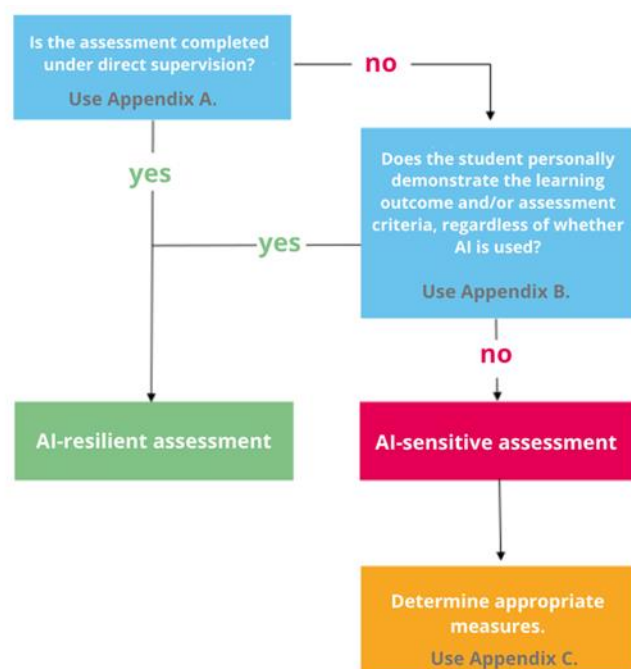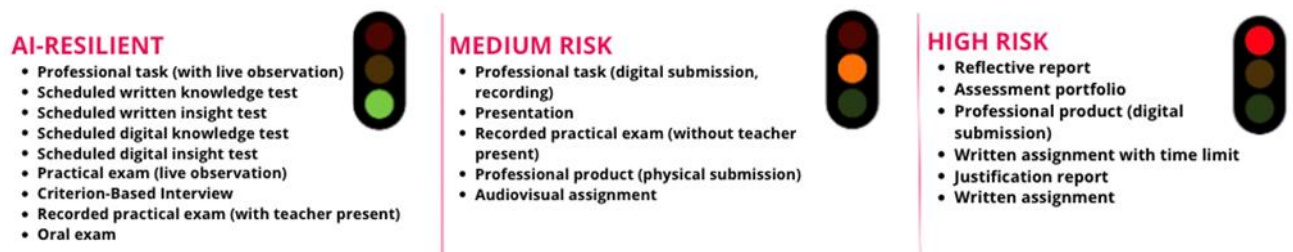

Figure 2: Flowchart for AI-resilient assessments

APPENDIX A: RISK CLASSIFICATION FOR ASSESSMENTS WITH AND WITHOUT SUPERVISION

In this appendix, we explain why assessments with direct supervision are AI-resilient and assessments without direct supervision are vulnerable to AI-related fraud. For this purpose, all HAN assessment formats have been evaluated and assigned a risk category with justification. Three risk categories were chosen for this purpose: AI-resilient assessment formats, medium-risk assessment formats and high-risk assessment formats. These three categories provide more space for the current state of AI and the context of the assessment (format). We would like to reiterate that the solution to vulnerability to AI-related fraud is not to simply offer AI-resilient assessment formats only. The solutions can be found throughout the constructive alignment, as can be seen in the measures in Appendix C.

AI-RESILIENT
- Professional task (with live observation)
- Scheduled written knowledge test
- Scheduled written insight test
- Scheduled digital knowledge test
- Scheduled digital insight test
- Practical exam (live observation)
- Criterion-Based Interview
- Recorded practical exam (with teacher present)
- Oral exam

MEDIUM RISK
- Professional task (digital submission, recording)
- Presentation
- Recorded practical exam (without teacher present)
- Professional product (physical submission)
- Audiovisual assignment

HIGH RISK
- Reflective report
- Assessment portfolio
- Professional product (digital submission)
- Written assignment with time limit
- Justification report
- Written assignment

A1 AI-resilient assessment formats

Performance of professional task (with live observation)
When observing the performance of a professional task, the examiner assesses what they see. Even if the professional task involves the use of a computer and/or AI, the examiner can see what is happening and the use of AI can be part of the assessment.

Oral assessment
The examiner conducts a live oral assessment with the student, examining the student's own knowledge and/or skills, such as their communication or reflection skills. Focus point: when assessing content, keep in mind that the content may have been generated by AI.

Scheduled written assessment of knowledge and insight
This takes place under supervision and there is no access to AI due to the secure environment.

Scheduled digital assessment of knowledge and insight

There is no risk due to the requirements of using a testing application such as ANS, fraud prevention software Schoolyear and supervisors. Of course, this requires that Internet access is disabled.

Practical (with live observation)

See "Performance of professional task (with live observation)"

Criterion-based interview

The examiner asks questions directly to the student. The student does not have access to a laptop, or if they do, their use is visible to the examiner. The criterion-based interview is often a supplement to a portfolio or product. The portfolio or product, however, may have been created by AI. Assessment validity also remains a major concern with this assessment format.

A2 Medium-risk assessment formats

Performance of professional task (digital submission, recording)

There are applications that can convert text into film and speech. These applications are improving and will become more accessible to a wider audience. This is not yet the case, but in the future, these tools will pose a greater risk to digitally submitted performances of professional tasks.

The risk with recordings is that students can make a montage of the recording. If you want to assess the performance of a professional task or a presentation as a whole, be mindful of this risk.

Presentation

The risk of using AI in a presentation depends on what you are assessing (content or presentation skills). When assessing content, keep in mind that the content may have been generated by AI. When assessing presentation skills, the assessment is done under direct supervision and there is no risk. Questions at the end of the presentation can lower the risk when assessing content.

Recording of practical, without lecturer presence

AI could be used for preparation. It is possible to use AI to fake a recording of a practical, but there is currently a threshold for doing so. See "Performance of professional task".

Professional product (physical submission)

In the case of a written report, the professional product has a high risk (see A.3 High Risk -> Professional product (digital submission)). Tangible products, such as drawings or a model, have lower risk. However, AI may have been used in the design or creation process.

Audiovisual assignment

See "Performance of professional task (digital submission, recording)"

A3   High-risk assessment formats (vulnerable to AI-related fraud):

Reflection report

A reflection report is usually created outside the examiner's view. This means a lack of additional information to ensure authenticity.

Assessment portfolio

The risk of an assessment portfolio depends on the type of products, learning outcome, assignments, assessment criteria and the educational process (formative performance, visibility of student and pathway to the portfolio, etc.). Many portfolio products, such as reflection reports, as well as visual materials are easily falsified by AI.

Professional product (digital submission)

High risk if it concerns a written product. Recorded performance (conversation - sound, video - presentation) have a medium risk. These recordings can be falsified with AI (e.g., by Deep Fakes), but these tools are less accessible and easy to use.

Written assignment with time limit

Written assignments are easily created with generative AI. The risk is eliminated if the assignment is done in a classroom with an assessment application such as ANS or fraud prevention software Schoolyear and the presence of one or more supervisors.

Accountability report

If a student provides context to a prompt, a language model can quickly generate a strong report.

Written assignment

See "Written assignment with time limit".

APPENDIX B: LOW-THRESHOLD AI USE

The following (assessment) activities can be performed using a low-threshold approach to AI. If these activities are relevant to the learning outcome(s) and/or assessment criteria of an assessment, the risk is that the student will use AI instead of demonstrating the learning outcome and/or assessment criteria themselves. The list of activities below gives a good idea of the types of activities, but will never be entirely complete due to the variety of ways in which AI can be used.

1. Generating and editing text (paper, report, plan, reflection, opinion, etc.).

- (re)formulating
- structuring
- summarizing
- editing
- translating

2. Creating and formatting audio-visual content. For example, creating:

- audio
- video
- animation
- poster
- flyer
- presentation
- image
- infographic

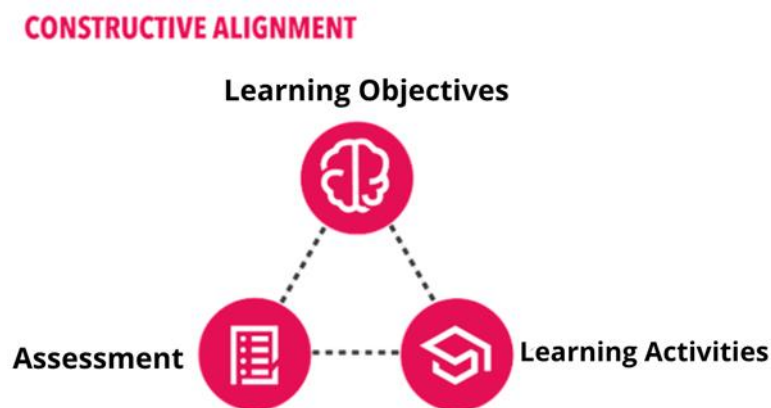**3. Contributing to (research) assignments**, like:

- brainstorming
- providing argumentation and justification
- searching for literature
- searching for sources
- analyzing
- drawing conclusions
- reflecting
- generating interviews (fictional interviews)

APPENDIX C: LIST OF MEASURES

As described earlier, effective education is grounded in strong constructive alignment. This involves intentional alignment of learning outcomes, assessment and learning activities.
The purpose of this triangular relationship is also to ensure that all elements of education contribute to AI-resilient assessment.

What measures are you taking in learning outcomes, assessment and learning activities to make the assessment AI-resilient?

Measures for learning outcomes

☐ Analyse and update learning outcomes, learning objectives and/or learning content
   ➔ Are the current learning outcomes, learning objectives and/or learning content still up-to-date and relevant, also considering the possibilities of AI?
      Action: Update learning outcomes, learning objectives, and/or learning content as needed to suit the profession, education and assessment. The focus here can shift to learning outcomes that are more focused on the student's learning process.
   ➔ Can the student demonstrate learning outcomes/learning objectives with AI use?
      Action: Check the extent to which the student can demonstrate the learning outcomes/learning objectives and/or learning content with the use of AI. Next, determine what learning outcomes students need to master and demonstrate without the use of AI. Type and importance of learning outcome determines the assessment format and setting.
   ➔ Should each competency/learning outcome be tested for AI resilience at least .... times at each level?
      Action: Discuss together (assessment committee and board of examiners) whether it is desirable and feasible to agree on a rule of thumb for the minimum number of times that a competency/learning outcome needs to be assessed at each level.

➔ When analyzing and updating learning outcomes and learning objectives, involve all relevant actors; such as the curriculum committee, editor of the degree statute/education and examination regulations, professional field, professional advisory committee and possibly degree committee, etc.

Action: Discuss with all actors the analysis and updating of learning outcomes, learning objectives and/or learning content.

Assessment measures

Before going into the measures, it is important to ask yourself the following questions:

- Why is the student being assessed?
- What knowledge/skills attitude do we want to assess and why? And what do we see when the student has a good command of the required knowledge/skill?
- What assessment format is suitable for this?
- Which assessment setting (controlled or uncontrolled environment) is appropriate and why?
- Does the assessment format suit the chosen learning activities?
- How do we ensure reliable and valid assessment?
- Is the product or the learning process the focus of the assessment?

☐ Analysis of assessment criteria

➔ Are the current assessment criteria still up-to-date, relevant, unambiguously formulated and suitable for the learning outcomes, learning objectives and/or learning content?

Action: Update and/or sharpen current assessment criteria suitable for the learning outcomes, learning objectives and/or learning content. In doing this, ensure uniform interpretation of assessment criteria.

➔ Is the current cut-off point sufficient?

Action: Review the current cut-off point and adjust it, if necessary. The bar may be lower for students when using AI!

☐ Analysis of the assessment format

➔ What assessment format is appropriate to demonstrate learning outcomes? Is this assessment format AI-resilient or vulnerable to AI-related fraud?

The assessment format can sometimes be retained if the assessment process is designed differently. For example, in the case of a writing assignment, the lecturer may choose to have more insight into the student's writing process instead of having just one assessment moment. This can be done by viewing and discussing different types of products a student has made. You can verbally ask students critical questions about their written piece. [resources: Appendix A and PowerPoint AI in je toetsontwerp - The Hague University of Applied Sciences]

Action: Where appropriate, modify the assessment format, taking into account the

assessment setting (with or without direct supervision), how it suits the learning outcomes and learning activities of the degree program and to the program's assessment policy.

- Analysis of the assessment task
  - ➔ Can the assessment task be made more specific and/or authentic?
    Action: Relate the task to a specific context or practice/organization, have students look for and reflect on concrete, specific examples, incidents, scenarios and/or their own experiences.
  - ➔ Can the assessment task be made more challenging and/or difficult?
    Action: Review the assessment task and assess not only basic knowledge, but also specifically whether the student can think, analyze and evaluate critically. Challenge the student to use higher cognitive skills to complete the assessment task. In the assessment task, avoid assessing the lower cognitive skills. These can be more easily generated with AI.

- Assessment in controlled setting
  For medium and high-risk assessment formats, the risk is reduced if the assessment is conducted under the direct supervision of a supervisor.

- Tightening the assessment procedure
  Action: Tightening the assessment procedure can be done by adding more examiners and setting up calibrations.
  Action: Contact the board of examiners if improper use of AI is suspected. A reason may be: the quality of the work differs from previous work and/or the style/language is different from before. The examiner identifies a change in trend.

- Shifting from product to process
  Action: Place more emphasis on the student's work and learning process rather than the product. Plan a series of formative moments where feedback motivates and guides the student's control of their learning process.

- Transparent communication to students
  Action: Discuss the concerns that apply to the use of AI.
  [resource: Aandachtspunten gebruik Generatieve AI voor studenten.pdf]
  Action: Discuss and clarify at what learning and assessment moment AI may or may not be used and explain why these choices were made.

- Cooperation with relevant actors
  Action: Discuss analysis outcomes and actions with e.g. the assessment committee, curriculum committee and/or assessment expert. Ask them for advice about outstanding questions.
  Action: Submit a modified assessment to the assessment committee. Discuss the adjustments and the reason for these.

Action: Determine (in consultation with curriculum committee, assessment committee, editor of the degree statute/education and examination regulations (and school manager) whether or not the adjustments can be made in the current academic year

(within the established, applicable degree statute/education and examination regulations). The assessment format should be defined in the education and examination regulations. Interim changes to these require the necessary actions.

☐ Feasibility, organizability and affordability

Action: One or more changes in the assessment formats may result in a more time-intensive assessment methods. It is important to discuss with all actors and the school management the considerations, consequences and implications regarding the feasibility, organization and affordability of assessment.

Measures for learning activities:

☐ AI-literate professionals

Action: During learning activities, where appropriate and necessary, teach students to use digital tools, such as AI tools, responsibly.

☐ Insight into the student's work and learning process

Action: Improve insight into student learning by organizing learning activities that emphasize formative assessment. Create feedback moments where students request and receive information that provides direction for their development and progress towards learning outcomes.

☐ Attention to the critical attitude of the student

In the age of AI, learning to be critical is essential to being able to distinguish between truth and falsehood.

Action: Teach students to be critical of AI-generated output.

☐ Responsible use of AI in education

Action: During teaching and feedback moments, also pay attention to the sources used. Education and support may reduce the need for using AI.

☐ Privacy, ethics and integrity

Action: During teaching and feedback moments, pay attention to privacy, ethics and integrity.

OPEN UP NEW HORIZONS.
HAN_UNIVERSITY OF APPLIED SCIENCES