

250606 Welke toetsvorm is AI-proof?

Visie op Toetsing en Examinering in het tijdperk van AI

Two-lane approach:

- Baan 1: Vaststellen van (kwaliteit) van leren en opbrengsten (Assessment of Learning)
- Baan 2: Ontwikkelingsgerichte toetsing (Assessment for & as Learning)

Ieder toetsprogramma zou idealiter bij elke leerlijn, onderwijseenheid of cursus zowel toetsing volgens baan 1 als baan 2 moeten bevatten (Liu & Bridgeman, 2023)

Cruciale leerdoelen worden geborgd in het eerste spoor (waardoor de diplomawaarde gegarandeerd is), terwijl toetsing volgens het tweede spoor ervoor zorgt dat studenten worden voorbereid op een beroepspraktijk waarin AI-gebruik de realiteit is.

4 Tips

- 1) **Andere toetsvormen.** Waarbij je rekening houdt met de consequenties voor validiteit en betrouwbaarheid op toetsprogramma-niveau. Maak een onderscheid tussen de leerzone en de presteerzone (en wellicht tussen bovenstaande Baan 1 en Baan 2).
Denk na over toetsvormen waarin het niet uitmaakt of een student AI gebruikt, omdat het gebruik van AI óf geen voordeel oplevert, óf juist onderdeel wordt van de leeractiviteit zelf. Dat gaat over **validiteit**. Toets je wat je denkt dat je toetst? Qua inhoud en qua vorm. Of test je eigenlijk vooral hoe goed iemand prompts kan intypen? Als we willen dat diploma's ook in een AI-tijdperk iets betekenen, moeten we niet harder roepen wat er niet mag. We moeten beter ontwerpen wat er toe doet.
- 2) **Ken je student.** Een student die de hele leerperiode niks laat zien en dan ineens een fantastisch product, daar moeten de alarmbellen gaan rinkelen. Eén keer is toeval, twee keer is opmerkelijk, drie keer is een patroon. Maar investeer ook in de relatie met je student en de studenten onderling. Studenten die onder de radar blijven en zich ongezien wanen, kunnen eerder dit gedrag vertonen, dan studenten die actief meedoen, betrokken zijn. Leren zichtbaar maken.
- 3) **Leerproces centraal.** We moeten over naar een ander systeem van toetsing. Waarin niet het afrekenen centraal staat maar het leerproces. Dat kan bijvoorbeeld door meer nadruk te leggen op formatief handelen of bijvoorbeeld een systeem van programmatisch toetsen, maar dat is geen garantie voor succes. Bijvoorbeeld als je nog steeds opdrachten via AI kunt genereren en er verandert niks aan de leercultuur. AI inzetten voor kwaliteitsbesef is goed. Aantoonbaar aanvullen met kennis die je hebt -> leerprocessen duidelijk maken. Leren kritische vragen te stellen. Leren zichtbaar maken.
Wat kan wel? Proces volgen, dialoog, leercultuur, leeruitkomsten, leren omgaan met AI... jouw leerproces staat centraal.
- 4) Zorg ervoor dat je als **team op 1 lijn** zit. Net als kinderen heel snel doorhebben bij welke ouder ze moeten zijn voor een snoepje (ondanks dat de regel is dat ze vlak voor het eten net meer mogen snoepen), zo weten studenten al heel snel bij wie ze moeten zijn om over de grenzen te gaan, zeker als er geen duidelijkheid is over de regels. O bij die docent mag je wel werken met AI / die docent weet niet eens wat AI inhoud. Maak teamafspraken. Hoe gaan we om met het gebruik van AI. Zorg ervoor dat zowel mensen met kennis van AI als mensen met toetskennis meedenken. Spreek met elkaar af wat we verwachten met betrekking tot AI. AI-geletterdheid is hiervoor cruciaal.

Risico's van generatieve AI, betrouwbaarheid en validiteit per toetsvorm

De afgelopen tijd is er per toetsvorm een risico-indeling gemaakt voor het gebruik van generatieve AI. In hoeverre zou de toets gemaakt kunnen worden met behulp van generatieve AI? Hiervoor zijn alle toetsvormen die binnen de hogeschool Leiden gebruikt worden geanalyseerd, deze staan per categorie vermeld in het document kwaliteitscriteria toetsing (Hogeschool Leiden, 2023).

Om de risico's vast te stellen zijn de kwalificaties *hoog*, *middel* en *laag* gebruikt. *Laag* wil zeggen dat het niet mogelijk is om de toets door middel van generatieve AI te maken. *Hoog* wil zeggen dat de toets met de juiste prompts geheel of gedeeltelijk te maken is met behulp van generatieve AI. In sommige gevallen hangt het van het type product af of er *hoge* of *middel* hoge risico's aan kleven. Een beroepsproduct die je digitaal kan inleveren kent vaak hoge risico's, een beroepsproduct die je fysiek moet inleveren, waarvan je de ontwikkeling hebt kunnen volgen, kent wellicht lagere risico's. De ontwikkelingen op het gebied van AI gaan zeer snel. Een toets die nu de classificatie 'laag' of 'middel' krijgt kan over enkele maanden een andere classificatie hebben.

Om risico's op het gebruik van AI te verkleinen worden er vaak toetsvormen gekozen die andere risico's met betrekking tot validiteit of betrouwbaarheid in zich dragen. Bijvoorbeeld het omzetten van een verslag naar een mondelinge presentatie is goed om het gebruik van AI te voorkomen, maar kent zeer grote risico's voor de betrouwbaarheid van je oordeel, omdat mondelinge toetsen zeer lastig betrouwbaar te maken zijn (Van Berkel et al., 2023). Bij grote studentenaantallen zit je daarnaast ook met de uitvoerbaarheid. Een mondeling afnemen bij 200 studenten met 2 beoordelaars is praktisch vaak niet haalbaar. Een ander risico heeft te maken met de validiteit van de toets. Een beroepsproduct omzetten naar een schriftelijke meerkeuzetoets is goed om de risico's op het gebruik van AI te verkleinen en ook verhoogt het de betrouwbaarheid, maar verlaagt vaak de validiteit van deze toets of het toetsprogramma als geheel.

Om aanpassingen te doen aan individuele toetsing dien je altijd rekening te houden met betrouwbaarheid, validiteit en uitvoerbaarheid. In tabel 1 hebben we de verschillende risico's naast elkaar gezet. Ook voor de risico's voor betrouwbaarheid en validiteit geldt dat deze tabel een indicatie geeft, afhankelijk van hoe het toetsproces is ingericht kunnen de toetsvormen een hoger of lager risico kennen. Weeg daarom altijd mee welke activiteiten zijn ondernomen om de betrouwbaarheid en validiteit van de toetsing te verhogen. We adviseren om aanpassingen altijd te doen op toetsprogramma niveau.

Tabel 1: risico's op het gebruik van generatieve AI, betrouwbaarheid en validiteit per toetsvorm.

Toetsvorm	Hoofdcategorie	Risico van generatieve AI	Risico's betrouwbaarheid	Risico's validiteit
Advies	Product-opdracht	Hoog	Middel/Hoog	Middel
Beroepshandelingen stageplek	Product-opdracht	Laag	Hoog	Laag
Beroepsproduct	Product-opdracht	Middel/Hoog	Middel/Hoog	Laag
Beroepsproduct	Performance assessment	Laag	Middel/Hoog	Laag
Casustoets	Schriftelijke toets	Laag	Middel	Middel
Criteriumgericht interview (CGI)	Performance assessment	Laag	Hoog	Middel
Demonstratie	Performance assessment	Laag	Middel/Hoog	Middel
Eindgesprek	Product-opdracht	Middel	Hoog	Middel
Eindproduct	Product-opdracht	Hoog	Middel/Hoog	Middel
Essay	Product-opdracht	Hoog	Middel/Hoog	Middel
Experiment	Performance assessment	Laag	Middel/Hoog	Laag
Game	Performance assessment	Laag	Middel/Hoog	Middel
Gedragssassessment	Performance assessment	Laag	Hoog	Middel
Geschreven beroepsproduct	Schriftelijke toets	Laag	Middel/Hoog	Middel
Gesloten items (selected response)	Schriftelijke toets	Laag	Laag	Hoog
Half-open vragen toets	Schriftelijke toets	Laag	Laag	Hoog
Invul	Schriftelijke toets	Laag	Laag	Hoog
Kennisclip	Product-opdracht	Middel	Middel/Hoog	Hoog
Kennistoets	Schriftelijke toets	Laag	Laag/Middel	Hoog
Korte antwoord	Schriftelijke toets	Laag	Laag	Hoog
Luistertoets	Schriftelijke toets	Laag	Laag	Hoog
Matching	Schriftelijke toets	Laag	Laag	Hoog
Meerkeuze	Schriftelijke toets	Laag	Laag	Hoog
Observatie	Performance assessment	Laag	Hoog	Middel
Onderzoek	Product-opdracht	Middel	Middel/Hoog	Middel
Ontwerp	Product-opdracht	Middel	Middel/Hoog	Middel
Open boek toets	Schriftelijke toets	Laag	Middel	Hoog
Open items (constructed response)	Schriftelijke toets	Laag	Middel	Hoog
Open-vragen toets	Schriftelijke toets	Laag	Middel	Hoog
Paper	Product-opdracht	Hoog	Middel/Hoog	Middel
Pitch	Performance assessment	Laag	Middel/Hoog	Hoog
Plan van aanpak/werkplan	Product-opdracht	Hoog	Middel/Hoog	Middel
Portfolio - ontwikkel	Product-opdracht	Middel	Middel	Middel
Portfolio - product	Product-opdracht	Hoog	Hoog	Middel
Portfolio assessment	Performance assessment	Laag	Middel/Hoog	Laag
Posterpresentatie	Performance assessment	Laag	Middel/Hoog	Hoog
Practicumtoets	Performance assessment	Laag	Middel/Hoog	Laag/Middel
Praktijktoets	Performance assessment	Laag	Hoog	Laag
Presentatie	Performance assessment	Middel	Middel/Hoog	Middel/Hoog
Proces	Product-opdracht	Middel	Middel	Middel
Product	Product-opdracht	Hoog	Hoog	Middel
Project	Product-opdracht	Middel	Middel	Middel
Rapport	Product-opdracht	Hoog	Middel/Hoog	Middel
Reflectie	Product-opdracht	Hoog	Middel/Hoog	Middel
Reflectiegesprek	Product-opdracht	Middel	Hoog	Middel
Resultaatformulier	Product-opdracht	Hoog	Middel	Middel
Rollenspel	Performance assessment	Laag	Middel/Hoog	Middel
Script-Concordance-Test (SCT)	Schriftelijke toets	Laag	Laag	Middel
Simulatie	Performance assessment	Laag	Middel/Hoog	Laag
Simulatie-patient-toets	Performance assessment	Laag	Middel/Hoog	Laag
Skillsstoets	Performance assessment	Laag	Hoog	Laag
Skillsstoets	Schriftelijke toets	Laag	Middel/Hoog	Laag
Spreektoets	Performance assessment	Laag	Hoog	Middel
Stationstoets	Performance assessment	Laag	Middel/Hoog	Middel
Take home toets	Product-opdracht	Hoog	Hoog	Middel
Team-based-learning	Product-opdracht	Middel	Middel/Hoog	Middel
Theorietoets	Schriftelijke toets	Laag	Laag/Middel	Hoog
Vaardighedentoets	Schriftelijke toets	Laag	Middel/Hoog	Middel
Verslag	Product-opdracht	Hoog	Middel/Hoog	Middel
Voortgangstoets	Product-opdracht	Middel	Middel	Hoog
Vorm in overleg met student(en)	Vorm in overleg met student(en)	?	?	?
Waar-niet waar	Schriftelijke toets	Laag	Laag	Hoog

Bronnenlijst

- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005-1016.
- Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 1-11.
- Furze, L. (2024). The AI Assessment Scale (AIAS) in action.
- Iacono, C. (2025) 22 Lessons from the GenAI Shadows. Gedownload op 28 mei 2025 van <https://hybridhorizons.substack.com/p/22-lessons-from-the-genai-shadows>
- Kolvoort, I. & Edwards, J. (2025) AI-bestendige toetsing ontwerpen. Gedownload op 28 mei 2025 van https://toetsbekwaamheid.nl/wp-content/uploads/2024/10/UvA-handout_AI-bestendige-toetsing-ontwerpen-NL.pdf
- Lafeber, J. Uitvoer gevraagd van: ChatGPT (OpenAI). AI gebruikt voor: aanvullen. 2025 mei 28. (Afbeeldingen op slide
- Liu, D., Fawns, T., Cowling, M., & Bridgeman, A. (2023). Responding to generative AI in Australian higher education, DOI:10.35542/osf.io/9wa8p
- NPULS (2025) Visie op Toetsing en Examinering in het tijdperk van AI. Gedownload op 28 juni 2025 van <https://npuls.nl/kennisbank/visie-op-toetsing-examinering-en-ai/>
- Online Learning Consortium (2025) Student Engagement and AI: Research Overview and Findings. Gedownload op 28 mei 2025 van <https://www.d2l.com/resources/assets/student-engagement-and-ai-research/>
- Roberto, M. (2024) Help students think critically in the age of AI. Gedownload op 28 mei 2025 van <https://hbsp.harvard.edu/inspiring-minds/enhance-critical-thinking-students-ai-assignment-strategies>.
- Vernieuwonderwijs (2024). De Wijsneuzen – aflevering 27: Betekenisvol toetsen met AI. . Gedownload op 29 mei 2024 van <https://vernieuwonderwijs.nl/de-wijsneuzen-aflevering27-betekenisvol-toetsen-ai/>
- Winstone, N., & Carless, D. (2020). Designing effective feedback processes in higher education: A learning-focused approach. Londen: Routledge.

Tabel 1: risico's op het gebruik van generatieve AI, betrouwbaarheid en validiteit per toetsvorm.

Toetsvorm	Hoofdcategorie	Risico van generatieve AI	Risico's betrouwbaarheid	Risico's validiteit
Advies	Product-opdracht	Hoog	Middel/Hoog	Middel
Beroepshandelingen stageplek	Product-opdracht	Laag	Hoog	Laag
Beroepsproduct	Product-opdracht	Middel/Hoog	Middel/Hoog	Laag
Beroepsproduct	Performance assessment	Laag	Middel/Hoog	Laag
Casustoets	Schriftelijke toets	Laag	Middel	Middel
Criteriumgericht interview (CGI)	Performance assessment	Laag	Hoog	Middel
Demonstratie	Performance assessment	Laag	Middel/Hoog	Middel
Eindgesprek	Product-opdracht	Middel	Hoog	Middel
Eindproduct	Product-opdracht	Hoog	Middel/Hoog	Middel
Essay	Product-opdracht	Hoog	Middel/Hoog	Middel
Experiment	Performance assessment	Laag	Middel/Hoog	Laag
Game	Performance assessment	Laag	Middel/Hoog	Middel
Gedragssassessment	Performance assessment	Laag	Hoog	Middel
Geschreven beroepsproduct	Schriftelijke toets	Laag	Middel/Hoog	Middel
Gesloten items (selected response)	Schriftelijke toets	Laag	Laag	Hoog
Half-open vragen toets	Schriftelijke toets	Laag	Laag	Hoog
Invul	Schriftelijke toets	Laag	Laag	Hoog
Kennisclip	Product-opdracht	Middel	Middel/Hoog	Hoog
Kennistoets	Schriftelijke toets	Laag	Laag/Middel	Hoog
Korte antwoord	Schriftelijke toets	Laag	Laag	Hoog
Luistertoets	Schriftelijke toets	Laag	Laag	Hoog
Matching	Schriftelijke toets	Laag	Laag	Hoog
Meerkeuze	Schriftelijke toets	Laag	Laag	Hoog
Observatie	Performance assessment	Laag	Hoog	Middel
Onderzoek	Product-opdracht	Middel	Middel/Hoog	Middel
Ontwerp	Product-opdracht	Middel	Middel/Hoog	Middel
Open boek toets	Schriftelijke toets	Laag	Middel	Hoog
Open items (constructed response)	Schriftelijke toets	Laag	Middel	Hoog
Open-vragen toets	Schriftelijke toets	Laag	Middel	Hoog
Paper	Product-opdracht	Hoog	Middel/Hoog	Middel
Pitch	Performance assessment	Laag	Middel/Hoog	Hoog
Plan van aanpak/werkplan	Product-opdracht	Hoog	Middel/Hoog	Middel
Portfolio - ontwikkel	Product-opdracht	Middel	Middel	Middel
Portfolio - product	Product-opdracht	Hoog	Hoog	Middel
Portfolio assessment	Performance assessment	Laag	Middel/Hoog	Laag
Posterpresentatie	Performance assessment	Laag	Middel/Hoog	Hoog
Practicumtoets	Performance assessment	Laag	Middel/Hoog	Laag/Middel
Praktijktoets	Performance assessment	Laag	Hoog	Laag
Presentatie	Performance assessment	Middel	Middel/Hoog	Middel/Hoog
Proces	Product-opdracht	Middel	Middel	Middel
Product	Product-opdracht	Hoog	Hoog	Middel
Project	Product-opdracht	Middel	Middel	Middel
Rapport	Product-opdracht	Hoog	Middel/Hoog	Middel
Reflectie	Product-opdracht	Hoog	Middel/Hoog	Middel
Reflectiegesprek	Product-opdracht	Middel	Hoog	Middel
Resultaatformulier	Product-opdracht	Hoog	Middel	Middel
Rollenspel	Performance assessment	Laag	Middel/Hoog	Middel
Script-Concordance-Test (SCT)	Schriftelijke toets	Laag	Laag	Middel
Simulatie	Performance assessment	Laag	Middel/Hoog	Laag
Simulatie-patient-toets	Performance assessment	Laag	Middel/Hoog	Laag
Skillsstoets	Performance assessment	Laag	Hoog	Laag
Skillsstoets	Schriftelijke toets	Laag	Middel/Hoog	Laag
Spreektoets	Performance assessment	Laag	Hoog	Middel
Stationstoets	Performance assessment	Laag	Middel/Hoog	Middel
Take home toets	Product-opdracht	Hoog	Hoog	Middel
Team-based-learning	Product-opdracht	Middel	Middel/Hoog	Middel
Theorietoets	Schriftelijke toets	Laag	Laag/Middel	Hoog
Vaardighedentoets	Schriftelijke toets	Laag	Middel/Hoog	Middel
Verslag	Product-opdracht	Hoog	Middel/Hoog	Middel
Voortgangstoets	Product-opdracht	Middel	Middel	Hoog
Vorm in overleg met student(en)	Vorm in overleg met student(en)	?	?	?
Waar-niet waar	Schriftelijke toets	Laag	Laag	Hoog