AI assisted Grade Justification

Findings from a small-scale pilot at BI

February 2026

UNIwise

AI Assisted Grade justification in WISEflow at BI Norwegian Business School

Findings from a SmallScale assessor pilot (quality-first sanity test)

Executive summary

Grade justification is a high-stakes and time-bound part of assessment practice in Norway, where institutions operate under a strict delivery expectation (commonly referenced as a two-week / 14-day timeline).

This report summarises findings from a small-scale pilot at BI Norwegian Business School focused exclusively on assessors' use of UNIwise' AI-assisted Grade Justification module in WISEflow. The pilot was designed as a quality-first sanity test of AI-assisted grade justification snippets – testing whether generated arguments are coherent, suitable, and aligned with academic standards, rather than optimising for speed in a mature, production-ready workflow.

Two pilot conditions are important for interpreting the findings:

Quality-first prompting

Assessors were encouraged to explore and experiment with the assistant during testing, while still grounding grade justification in realistic use cases ("retrospective benchmarking").

Manual transfer of suggestions

The prototype required manual copy/paste of AI-generated suggestions into an editor. BI has highlighted that the next maturity step is one-click insertion integrated in the normal workflow, which is expected to strengthen usability and adoption.

Despite these conditions, assessors provided notably positive feedback on usefulness and output quality, reporting that the assistant often produced a strong foundation that could be used with limited edits, while retaining academic ownership of the final justification.

Across the three assessors and three BI exam flows, the pilot indicates that AI-assisted grade justification can:

- Raise the baseline quality and consistency of grade justifications, particularly by providing a structured starting point and examples of what is missing relative to criteria.
- Support a “feed-forward” style of justification (i.e., not only defending the grade, but helping students understand what would need to change to reach the next grade level).
- Deliver strongest value when assessors can rely on good marking guidance / rubrics being present on the flow (a prerequisite in the test framework

Background and rationale

Grade justification sits at the intersection of academic judgement, transparency, and compliance. In Norway, the process is shaped by strict turnaround expectations, and BI Norwegian Business School’s broader framing of the initiative emphasises the operational burden associated with delivering justifications at scale.

This creates a clear need for solutions that can help educators:

- Produce consistent, criteria-aligned justifications
- Maintain academic ownership and accountability
- Meet institutional turnaround expectations in busy periods

UNIwise’ approach is grounded in human-in-the-loop principles: the assistant drafts suggestions, while the educator reviews, edits, and remains responsible for the final wording and judgement.

Aim of the pilot

The pilot was positioned as an early-stage sanity test of AI-generated grade justification snippets with BI assessors. The test framework defines the aim as evaluating whether generated snippets:

- demonstrate coherent and well-founded reasoning aligned with the assigned grade,
- adhere to accepted academic standards of style, structure, and tone, and
- meet expectations for clarity and suitability in assessment contexts.

In addition, the testing guidance framed the activity as retrospective benchmarking, comparing AI-assisted work with manual grade justifications previously provided.

Pilot context and design

Setting and test environment

The pilot was conducted on the WISEflow Stage (sandbox) environment using existing materials from previous courses. The sandbox environment secured assessors could use WISEflow normally, while students did not see or get justifications and notifications.

Participants and exam flows

Three assessors participated, each testing the assistant on one exam flow and returning structured feedback. The three flows referenced in the returned feedback forms were:

- Anvendt samfundsøkonomi - SØK3550
- Business communication - EXC34011
- International tax law - ELE38001

Inputs and generation approach

The test framework describes that grade justification snippets are generated based on: available marking material on the exam flow (rubric/marking guidance), the assignment (if provided), student submissions, grading scale definition (Norwegian A–F), and the awarded grade. The output is “text snippets” intended to be composed into a grade justification finalised by the assessor.

Pilot condition

For clarity, BI’s prototype feedback was from the outset prior to pilot that assessors should be able to insert the AI-generated text into the Grade Justification editor with a single click. They argue rightly, that copy/paste increases friction and introduces risk of small errors. The pilot prototype however, relied on manual transfer at this stage, and thus introduced some friction that later would be removed.

Planned improvements: workflow integration and prompt re-use

Two improvement directions are particularly relevant:

One-click insertion

and full workflow integration inside standard WISEflow processes

Prompt re-use

Described as the ability to "remember earlier prompts" so users can select a prompt again next time, supporting consistent practice at assessor level without removing flexibility.

Findings: output quality and academic usefulness

Argument quality was generally strong and usable

Across the three feedback forms, assessors evaluated the output quality positively. One assessor characterised the arguments as "strong". Another described the arguments as "towards the strong side". A third described them as "strong and precise," while also noting that users should expect to review AI-generated content against disciplinary standards (which is consistent with human oversight expectations).

A recurring positive theme was that the assistant can provide concrete examples of what is missing relative to expectations – particularly when responding to the common student question: "Why was the awarded grade not one step higher?"

Assessor ownership and willingness to use again

A strong adoption signal in the feedback is that assessors reported continued ownership of the final justification (with appropriate review) and expressed clear willingness to use the tool again. Comments included that the tool was "very helpful" and that the assessor would likely try it again; another assessor stated they would love to use it again while emphasising the normal need to monitor the output; and a third framed future use as particularly valuable in certain contexts (e.g., language support).

Prompts were helpful even before standardisation

Although the pilot encouraged exploration, assessors still found the predefined prompt questions helpful – one assessor explicitly noted he/she liked trying several questions to see which produced the best answer, and another found the predefined prompts to provide a good foundation.
This supports the planned prompt re-use direction ("remember earlier prompts") as a path to more consistent practice and reduced setup effort in future testing.

Constructive vaceats - Useful guidance for the next iteration

Assessors also provided constructive feedback that helps define where future refinement should focus. One assessor highlighted that lower-grade cases require more critical review because the assistant can otherwise risk producing a too supportive narrative. Another assessor noted occasional output inconsistencies (e.g., placeholders), reinforcing the need for review and editing before finalising.

These caveats are aligned with the pilot's purpose as a sanity test: to identify strengths and limitations early and guide improvements before wider use.

Findings: workflow, time and compliance vlue

Efficiency signals were mixed and are best interpreted through the pilot's setup: assessors were encouraged to explore prompting strategies, and manual copy/paste was used for transferring text into the editor. However, it is worth mentioning that the most credible time-saving pathway is not "zero-effort automation", but a combination of:

1. Good marking guidance on the flow,
2. A reliable submission summary for fast recall, and
3. Draft snippets that reduce blank-page time and promote consistent structure.

This matters because Norway's statutory timelines create institutional risk if justifications cannot be produced consistently within deadline.

Within this setup, assessors still reported meaningful process value. One assessor found the effort comparable to manual writing, but more comfortable because the assistant helps draft a starting point that can then be refined. Another noted that response time and workflow steps matter in practice.

From an institutional perspective, the compliance-driven value proposition remains clear: solutions that improve consistency and reduce effort in drafting can support delivery within strict turnaround expectations, provided workflow integration reduces friction and supports everyday use.

Appeals and escalation dynamics

One assessor explicitly noted that as long as information is accurate, richer justifications will be useful, yet the same detail could also provide a basis for further appeals if students disagree.

This does not argue against quality; rather it highlights the need for:

- Good marking guidance on the flow
- Institutional alignment on wording norms
- Clear guidance on what justification must (and must not) include
- Auditability and consistent structure across assessors

Outlook and conclusion

This BI pilot provided early evidence that AI-assisted grade justification in WISEflow can generate usable, criteria-aligned draft arguments that assessors find helpful and appropriate as a startingpoint. In the collected feedback, assessors described the overall argument quality positively (e.g., “strong” / “towards the strong side” / “strong and precise”) and reported that outputs could often be used with limited edits, while still retaining full academic ownership of the final wording and judgement.

The pilot was explicitly framed as a sanity test of AI-generated grade justification snippets, focused on quality, appropriateness, and constructive value. This makes it a solid baseline for continued iteration: it clarifies what already works well, and it pinpoints the product steps that will unlock even stronger adoption and efficiency in the next test phase.

Overall conclusion

Taken together, the pilot results and the already-identified improvements create strong momentum for the next test cycle. The assistant demonstrates early academic usability and positive assessor reception, while the roadmap focuses precisely on the two factors that typically drive adoption at scale: low-friction workflow and repeatable, consistent prompting.

External references used for this report

- Laws and regulations (Grade Justification: section 5-13) – from BI https://www.bi.no/en/programmes-and-individual-courses/admissions/laws-and-regulations/#section-5-13-justification-of-a-grade-and-appeal-against-a-grade
- UNESCO (2023; updated 2026). Guidance for generative AI in education and research. https://unesco.org.uk/site/assets/files/10375/guidance_for_generative_ai_in_education_and_research.pdf
- European Commission (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- EU AI Act. Article 14: Human oversight (overview pages). https://artificialintelligenceact.eu/article/14/
- GDPR / data protection principle on purpose limitation (Article 5(1)(b)), practical guidance. https://gdpr-info.eu/art-5-gdpr/
- Hattie, J. & Timperley, H. (2007). The “feed up / feed back / feed forward” model of feedback. https://www.moedu-sail.org/wp-content/uploads/2016/03/feedback-model.pdf

Acknowledgement

UNIwise would like to thank BI Norwegian Business School for its participation and cooperation, and especially the three professors participating and willingly sharing their feedback to us. Additionally, we want to thank the administrative staff at BI for supporting the setup and running of the pilot together with us.