

Independent Audit of Fobizz's 'AI Grading Assistant' — Osnabrück Study Finds Erratic Scores in a Tool Licensed by German States

AI in Education · Good practice · Germany

Quality score: 47/100

Evidence strength: Moderate

Summary

Mühlhoff and Henningsen (Osnabrück) tested Fobizz's AI grading tool in two test series: scores and feedback looked arbitrary, one text ranged 1–14 points across runs, and a 2026 follow-up on two other tools found similar variation.

Key dimensions

Dimension	Score
Evidence of learning impact	1/3
Equity & inclusion	1/3
Transparency, ethics & data protection	2/3
Teacher capability & pedagogy	2/3
Scalability & sustainability	1/3

Evaluated by C-NAPSE

Data sources

[C-NAPSE web research](#) — accessed 2026-10-09

[University of Osnabrück \(Mühlhoff & Henningsen\) — audit of Fobizz](#) — accessed 2026-10-09

[arXiv 2412.06651 \(Mühlhoff & Henningsen\)](#) — accessed 2026-10-09

[arXiv 2412.06651](#) — accessed 2026-10-09

[heise report on 38C3 talk](#) — accessed 2026-10-09

[Follow-up: FelloFish and Edaira](#) — accessed 2026-10-09

Cite this — Evidoria. *Independent Audit of Fobizz's 'AI Grading Assistant' — Osnabrück Study Finds Erratic Scores in a Tool Licensed by German States*. Persistent ID: d33bac8c-3ecf-4019-9291-814435253eb0.