

RESEARCH STUDY · CARDIFF BUSINESS SCHOOL

What Written Assessment Cannot See

How scalable oral assessment reveals the difference between recognising knowledge and demonstrating understanding.

Cardiff University · Spring 2026

Written tests can show what students recognise. Oral assessment reveals whether they can explain, justify and apply what they know. This controlled Cardiff study explored whether AI-supported vivas can deliver that depth consistently and at scale.

What the study revealed

Students entered the oral phase with comparable baseline knowledge, but oral assessment exposed differences the written test could not. The AI viva applied a consistent challenge across students, revealing the gap between recognised knowledge and independently explained understanding.

C O N T E X T

A controlled study with practical implications



The study at a glance

DESIGN

Three-condition controlled comparison

CONDITIONS

Control (n=7) · AI viva (n=7) · Traditional viva (n=8)

OUTCOME MEASURE

Viva performance, scored as a percentage

COHORT

24 postgraduate students

BASELINE MEASURE

MCQ administered prior to all viva activity

The Institution

Cardiff University is a Russell Group university with a strong research profile and a longstanding commitment to academically rigorous assessment. Cardiff Business School provided the setting for a controlled comparative study examining how AI-supported oral assessment performs against traditional viva practice and baseline written testing.

What Integrevise Does

After submitting written work, students take part in a short AI-supported dialogue, responding to questions about their work in their own words. They explain, justify and reflect on their submission. It does not grade. It reveals whether genuine understanding sits behind the work.

Integrevise then provides structured feedback to students and aggregated insight for educators, supporting assessment validity, student development and cohort-level understanding.

Why this design matters

The MCQ was administered before group allocation and before any viva activity. It therefore acts as a baseline measure of prior knowledge.

Since no statistically significant differences were found across groups at baseline, subsequent viva performance can be interpreted as reflecting differences in assessment process rather than student ability.

WHAT WE FOUND

Three findings, one assessment signal

01 Groups entered with comparable baseline knowledge $p = .142$

There were no statistically significant differences in MCQ performance across the three groups. Students entered the viva phase with similar prior knowledge, strengthening the validity of the comparison.

02 Oral assessment revealed deeper understanding **Near ceiling**

MCQ scores were uniformly high, but viva scores were lower and more varied. The viva required students to explain, justify and apply knowledge rather than simply recognise correct answers.

03 AI-supported viva applied consistent criteria $p = .006$

The study showed a significant difference between traditional and AI viva scores. Traditional examiners can scaffold and interpret responses, while the AI applied the same criteria to every student.

What this means

A written test can show recognition. A viva reveals whether students can explain what they understand. AI-supported oral assessment makes that deeper check repeatable across larger cohorts.

IMPLICATIONS

What this means for assessment

The Cardiff findings point to a clear role for AI-supported oral assessment: not as a replacement for written assessment or academic judgement, but as an added layer that tests whether students can explain, justify and apply what they know.

From findings to practice

01**Deeper measurement**

Tests explanation, justification and application, reaching cognitive processes written formats often miss.

02**Consistent evaluation**

Applies the same criteria across students, supporting more comparable assessment evidence.

03**Immediate feedback**

Gives students structured feedback quickly, while the learning moment is still live.

04**Cohort intelligence**

Shows patterns of misunderstanding across a cohort, helping module leaders target support earlier.

The broader assessment opportunity

Together, these capabilities move oral assessment from an occasional high-touch process to something institutions can apply more consistently across larger cohorts. The next question is not whether oral assessment adds value, but how it can be implemented responsibly, at scale and with the right academic safeguards.

INTERPRETATION

What the study also showed

AI viva scores were lower than traditional viva scores, and the difference was statistically significant. This should not be read simply as a weakness of AI assessment. Traditional vivas allow examiners to adapt questions, offer implicit guidance and exercise interpretive judgement. These features may help students demonstrate their understanding more effectively.

The AI-supported viva operated differently. It applied predefined criteria consistently across participants and required students to construct and articulate responses with less examiner scaffolding. The study therefore highlights a key assessment tension: interpretive richness versus standardised consistency.

Acknowledged limitations & publication status

The study involved 24 students in a single postgraduate module. MCQ scores showed a ceiling effect, limiting the test's ability to distinguish levels of prior understanding. Findings should therefore be treated as promising rather than conclusive. Further research with larger cohorts, varied disciplines and longitudinal performance data is needed. A peer-reviewed paper is currently in development.

From evidence to implementation

- Expand across larger cohorts, multiple modules and varied disciplines.
- Measure whether AI viva feedback improves later assessment performance.
- Evaluate how cohort-level insight supports teaching decisions and curriculum improvement.

The assessment opportunity

The Cardiff study shows the core assessment challenge: written tests can show recognition, but not whether students can explain, justify and apply what they know. AI-supported oral assessment enables that deeper level of evaluation in a way that is more consistent, scalable and instructionally useful, at a time when proving what students genuinely understand has never mattered more.

For follow-up discussion:

www.integrevise.com | contact@integrevise.com