

Inez Tandy Ortega

MSc Computer Science Dissertation

Inez Tandy Ortega

MSc Computer Science Dissertation

Home Page

Cohort Statistics

Student Statistics

Question Analysis

Teaching Summary

Feedback Generator

Comparison Analysis

Data Sources

Teaching Summary

This page provides an overview of cohort and student performance, highlighting areas that may require further teaching before the summative assessment.

Selected assessment format: Learning Management System

Cohort Overview

Total Students Analysed	Cohort Average
56	68.9%

Teaching Summary

Generate a Performance Summary and Actionable Recommendations for the Cohort.

Questions Requiring Attention

The table below shows the five lowest-performing questions in the formative assessment.

RED: High Priority | YELLOW: Medium Priority | GREEN: Lower Priority

Question	Average Percentage	Priority
Question 18	28.6	High

Student Performance Overview

View students who may require additional support, identify the highest-performing students, or search for an individual student.

Select student group

☒ All Students
 ☐ Scored 0%
 ☐ Struggling Students
 ☐ High-Performing Students
 ☐ Perfect Scorers

☒ Search for Student for Individual Performance Breakdown

Enter a student ID or student number

20

Selected Student	Overall Percentage
Student 20	63.2%

Student Group

Time Utilisers

Generate a brief overview for the selected student

Question Comparison

Compare the selected student's score for each question against the cohort average.

Feedback Generator: Programming Grades File

This page is designed to allow you to select students and receive feedback regarding what the students did well and areas that they should improve. It will take 1-2 minutes to produce each report.

Assessment Name

Programming Grades

Tutor Signature

Your Name

Select a Student ID or Student Number:

Choose options

Generate Feedback

Feedback Management

The table below summarises the current status of feedback for each student, including whether it has been generated, edited, moderated and exported.

No feedback has been generated for this assessment yet.

Research and Development Process

OFS | Jones (2023)
Original Feedback System

EFS | This Study
Enhanced Feedback System

1 Literature & Prior Findings

- Formative Feedback
- AI-Powered Feedback
- Identified key user needs for the OFS

2 Preliminary User Evaluation

- 6 Participants: 3 Students, 3 Educators
- Validated and redefined priorities
- Defined Enhancement Requirements

3 Developing the EFS

- Evidence - led redesign
- Implemented ER's
- Human-in-the-loop workflow

6 Conclusions

- Overall findings
- Future research

5 Analysing the Results

- Quantitative and qualitative analysis
- Strengths, concerns, and limitations

4 Evaluating the EFS

- 74 Participants: 37 Students, 37 Educators
- Mixed-methods evaluation

Research Context and Aim

1 Feedback at Scale

- Personalised, timely formative feedback is challenging to provide consistently across large cohorts

2 AI Opportunity

- Generative AI can produce personalised feedback from assessment data more efficiently

3 Core Challenge

- Efficiency / automation does not guarantee accuracy, contextual appropriateness, or educational value

Research Aim

To determine and evaluate whether evidence- and user-informed enhancements to the OFS can provide effective support for personalised formative assessment feedback.

Students in Higher Education

Educators in Further and Higher Education

Problems to Priorities

Literature
Review

Jones' Study 2023
48 Student
Participants

Carver's Workshop 2024
Only Educator
Participants

Preliminary User Evaluation
3 Student & 3 Educator
Participants

1 Actionability Gap

Literature

Jones 2023

PUE

- Clearer next steps requested
- More targeted learning resources
- Practice questions wanted
- Need cohort teaching priorities

ER1

Feedback Actionability

Improvement guidance, resources, and practice

2 Workflow Friction

Carver 2024

PUE

- Information spread across pages
- Key findings hard to identify
- Time-consuming navigation
- Need a clearer educator summary

ER2

Workflow Clarity & Efficiency

Reduced navigation that is concise and visual

3 Control and Verification

Literature

Carver 2024

PUE

- Review and edit AI feedback
- Track moderation before release
- Time-consuming to verify and review
- Educator review required before release

ER3

Educator Control & Moderation

Editable, reviewable, and trackable feedback

4 Quality and Consistency

Literature

Jones 2023

Carver 2024

- Generic or robotic tone
- Verbose output
- Inaccurate or unsupported content
- Inconsistent structure and detail

ER4

Feedback Structure & Consistency

Consistent structure with personalisation

From Requirements to Implementation

Python | Streamlit | Pandas | NumPy | SQLite | GPT-4 API

ER 1 Feedback Actionability

Controlled Generation Logic

Feedback Generator

Teaching Summary

- Boolean branches in `build_prompt()`
- Resources / practice questions only when evidence supports them
- Python detects patterns → GPT-4 generates grounded actions
`include_further_reading` | `include_practice_questions`

ER 2 Workflow Clarity & Efficiency

Dynamic Cohort Analysis

Teaching Summary

- `st.radio()` / `st.selectbox()` control dynamic views
- Pandas filters the same `student_summary`
- `np.percentile()` → adaptive cohort-relative thirds
- Conditional colour coding → Visual prioritisation

ER 3 Educator Control & Moderation

Persistent Human-in-the-Loop State

Feedback Generator

Teaching Summary

- Generated → Edited → Moderated → Exported
- SQLite stores `original_feedback` + educator-managed `feedback_text`
 - Tracks separate `is_edited` | `is_moderated` | `exported` states
 - Editing / regeneration invalidates outdated moderation and export states

Composite Primary Key: `assessment_name` + `exam_type` + `student_id`

ER 4 Feedback Structure & Consistency

Shared Prompt Logic

Feedback Generator

- `get_group_instruction()` → shared `build_prompt()`
- Performance-sensitive guidance + fixed core structure
- Centralised length + evidence constraints
- Format-specific processing → Shared generation logic

Live Demonstration of the EFS

Home Page

Cohort Statistics

Student Statistics

Question Analysis

Teaching Summary

Feedback Generator

Comparison Analysis

Data Sources

Select relevant page

Home Page

Welcome to the home page of the formative assessment feedback system. Here you can upload the data you wish to analyse and gain feedback for.

What program was the test marked in?

- ☒ Learn
☐ Gradescope
☐ Other

Upload the response .CSV file here: ?

 Upload 200MB per file

Upload the marks .CSV file here: ?

 Upload 200MB per file

Please navigate to the sidebar and select the desired page.

Cohort Statistics Page: Visualises the performance of all students who completed the assessment.

Student Statistics Page: Analyses an individual student's performance and compares their responses with the expected answers.

Questions Page: Analyses the performance of grouped students across each question in the assessment.

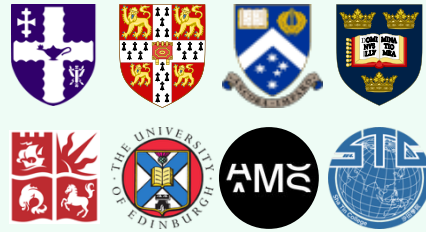
Teaching Summary Page: Provides a complete overview of the cohort, the assessment, and student performance.

Feedback Generator Page: Generates personalised feedback reports for students.

Comparison Page: Compares student performance across two different years.

Evaluation Design

Criterion-based Purposive Sampling



74
Participants

37 Students in Higher Education
37 Educators
20 Further Education + 17 Higher Education

Evaluation Criteria

Feedback Quality

Personalisation | Actionability | Consistency

Usability

Clarity | Workflow | Ease of Use

Trust

Confidence | Traceability | Verifiability

Usefulness & Intention to Use

Practical Value | Adoption

Synchronous Sessions

Remote Completion

Consent

Standardised Demonstration Video

Focus Group / Interview

**Anonymous Questionnaire
Aligned to Evaluation Criteria**

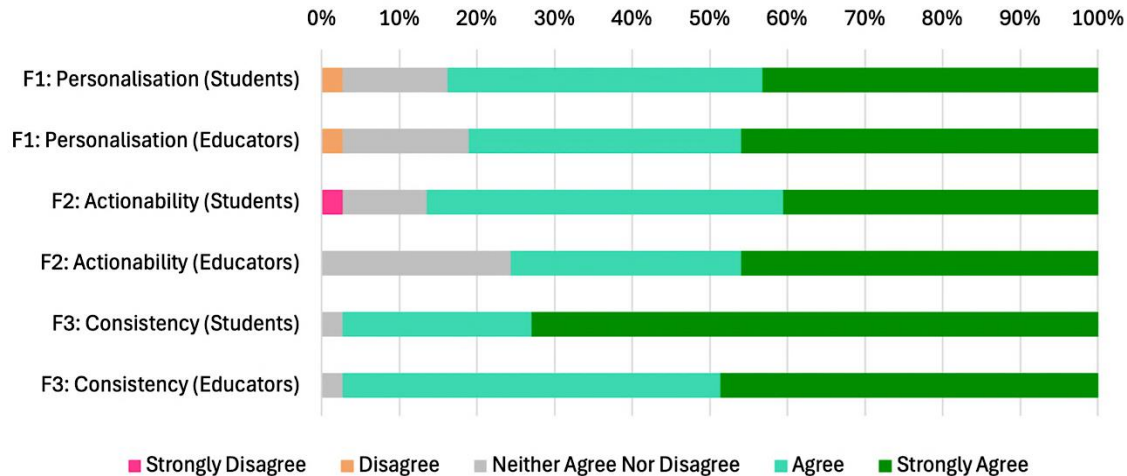
- Quantitative: 5-Point Likert-scale statements
 - Qualitative: Open-ended questions

Key Results

81.1% Overall Satisfaction

60 / 74 Participants

Feedback Quality



- Consistency was the clearest strength
- Actionability was positive; educators were more cautious
- Assessment-specific ≠ contextual personalisation

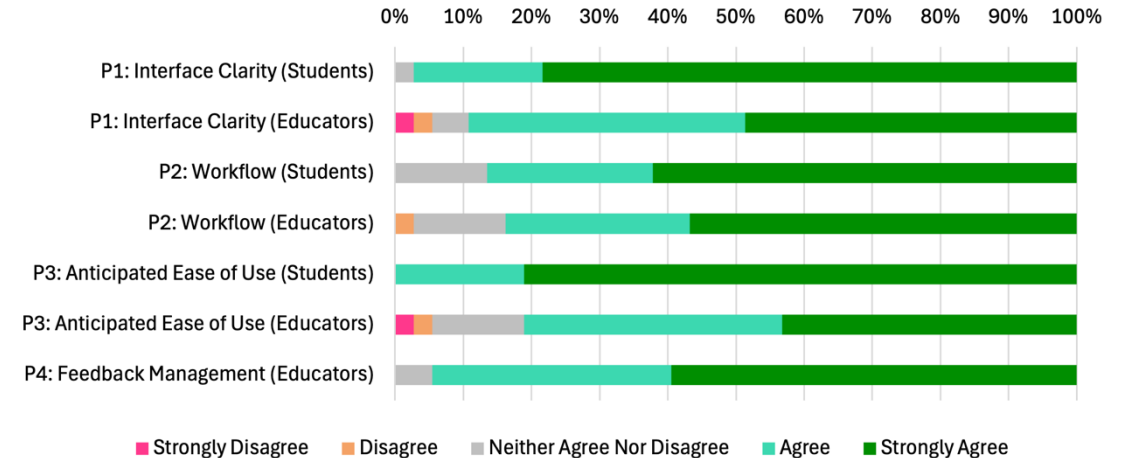
Students

- Valued clear next steps
- Resources + practice questions supported improvement

Educators

- Valued clear strengths → improvements → actions
- Wanted rubric / module-specific context

Usability



- Interface clarity was strongly rated
- Navigation was viewed as logical
- Feedback Management was a key educator strength

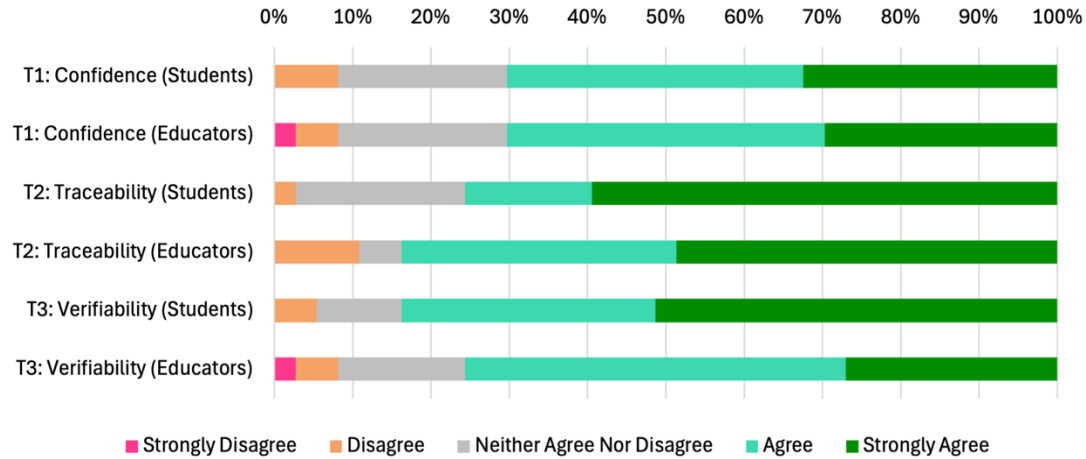
Students

- Clear navigation + visual organisation
- Optional verbal feedback + voice-to-text suggested

Educators

- Valued how the Teaching Summary consolidated cohort insight
- Status tracking supported feedback management

Trust



- Verification was stronger than direct AI confidence
- Trust depended on traceability + educator oversight

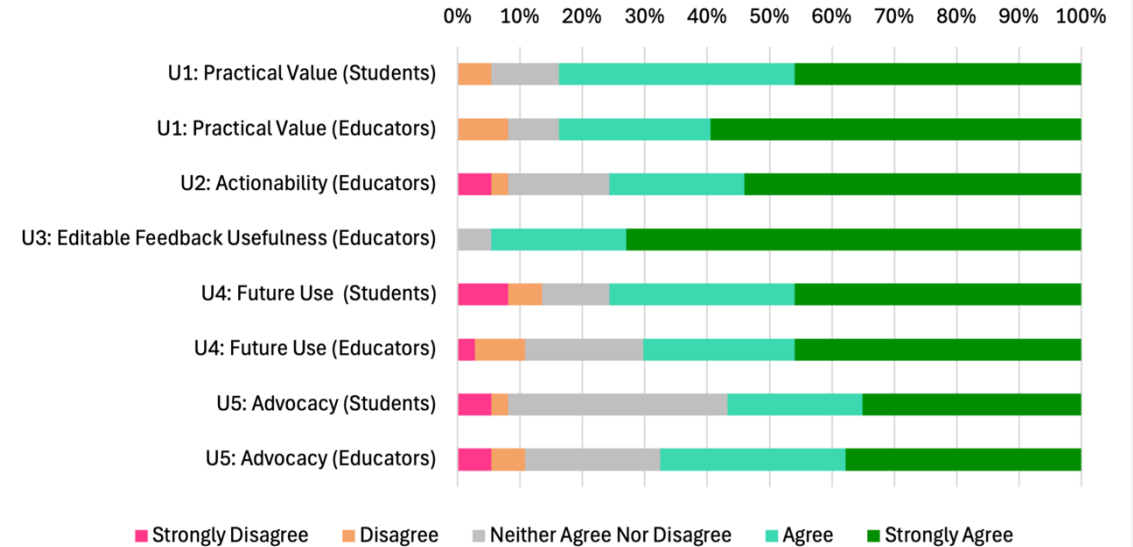
Students

- Educator review increased trust
- Wanted feedback linked to evidence

Educators

- Wanted prompt / evidence visibility
- Easier comparison with original student work

Usefulness & Intention to Use



- Practical value was strong
- Editable feedback was a key educator strength
- Personal use > institutional advocacy

Students

- Resources + practice questions supported next steps

Educators

- Teaching Summary + editable feedback supported practice
- More cautious about automated teaching recommendations

Exploratory Spearman

Students: Practical Value ↔ Future Use $\rho = .75$

Educators: Future Use ↔ Advocacy $\rho = .83$ | Satisfaction $\rho = .82$

AI Familiarity ↔ ratings: no clear association ($|\rho| \leq .26$)

Conclusion and Future Research

The EFS was perceived as effective as a human-supervised feedback-support tool, but not as a fully autonomous replacement for educator judgement.

What the Evaluation Revealed

1 Effectiveness is Context-Dependent

- Structured assessments were a clearer use case; essays / subjective tasks raised greater caution

Trust

Generalisability

2 Workflow Promising, Not Yet Proven

- Usability was positive, but real-world efficiency and adoption were not tested

Usability

Adoption

3 Trust Depends on Human Oversight

- AI was better accepted for identifying needs than determining the pedagogical response

Trust

Human Control

Implications for Future Development

Richer Assessment Context

Full responses

Rubrics

Module-specific prompts

Live Deployment & Refinement

Live assessment cycle

Configurable thresholds

Usage metrics

Stronger Human Control

Compare approval models

Audit trail

Visible edits
