Reviewer's comments and responses

We thank the reviewer for meticulously reading through the manuscript and for the constructive comments given. The feedback has assisted us to substantially improve the manuscript, especially regarding the research design, the instructional model, and the interpretation of the results. All corrections, clarifications, and language edits suggested have been incorporated into the revised version of the manuscript. Detailed responses have been given to each comment made in the table below.

Comments	Response to Reviewer Comments
First, the authors claim to be using qualitative methods for at least part of the data collection.	We appreciate this significant methodological clarification. We concur that the study does not constitute a pure qualitative statistic investigation in the traditional interpretive sense. The goal of the study was to describe a case study of curriculum implementation using multiple sources of evidence, integrating students' feedback, expert reviews, and students' performance indicators. In order to address these drawbacks, we have revised the manuscript in the following ways: 1. The methodology section has been rewritten to explicitly describe the study as a descriptive case study using mixed evidence, rather than a qualitative approach. The revised manuscript now suggests that the study utilised: – Institutional evaluation data (Fig. 7) – Students' feedback surveys (including open-ended comments), see Figure 8. – External examiners' reports – Peer review evaluations (Fig. 5). – Students' performance (Fig. 6). These data sources were utilised mainly for descriptive analysis instead of interpretive qualitative analysis. 2. The manuscript is now explicitly explaining that: – Open-ended students' comments were not subjected to formal qualitative coding or thematic analysis. – Nonetheless, the comments were used illustratively to contextualize quantitative indicators of course implementation and students' engagement. 3. References implying that the study was conducted as a qualitative research design have been removed or revised. 4. Following the reviewer's suggestion, the reference to Lincoln and Guba's trustworthiness criteria has been removed, since the study does not apply a formal qualitative analysis.
This is not a qualitative investigation.	We agree with the reviewers' assessment that this study does not constitute a qualitative investigation. The manuscript has therefore been revised to clarify the descriptive nature of the study drawing on multiple sources of programme evaluation data, but not qualitative research approach.

The authors may have used some qualitative (open-ended) data, but only to create bins for completing some descriptive statistics (how many said this or said that).	Yes qualitative responses were used illustratively to contextualize quantitative indicators of course implementation and students' engagement. Table 2 has been added to demonstrate participants gain.
I saw no interpretation of the qualitative data.	We acknowledge this observation, as the study does not employ a qualitative research method, open-ended responses were used only to provide contextual support and were not subjected to systematic qualitative analysis. Nonetheless, Table 2 has been included to add clarification of the hybrid model for teaching and learning as well as assessment contributes to students' performance.
I don't even know how much was collected.	There were 13 participants and this has now been expanded to add clarification.
Was it just survey data, or were there interviews?	Survey data coupled with assessment data
Qualitative research is mainly about investigating how the participants make meaning. This requires a LOT of data (writing, interviews, conversations, etc.) to try and understand how the participants are developing their understanding of the content material (in this case).	We acknowledge the reviewers' observations. In this study, in-depth qualitative data were not collected or analysed, and the use of open-ended responses was limited to providing contextual support rather than exploring participants' meaning-making processes.
Qualitative research is NOT about whether an instructional intervention was effective or not and how much. It would be more about how the participants experience this instruction and how they learn from it.	We recognise that although this current study does attempt to measure the effectiveness of an instructional intervention in terms of magnitude or outcomes, it rather explores how participants experienced the instruction and how they developed their understanding. Accordingly, this study did not adopt a qualitative research design, and any qualitative data included are used only to provide contextual insights rather than to interpret participants' meaning-making processes.
Qualitative research is not supposed to be used for generalizing to larger populations. It is about the individual and hearing the voice of the participants or observing the thinking of the participants.	This has been corrected to show participants onion. However, Table 2 is added to illustrate students' performance.
Second, the authors are exploring the efficacy of a novel instructional structure. This is fine, but they don't describe (AT ALL) the design of the instruction. What is novel about the instruction?	In the revised manuscript, we have significantly expanded the instructional design and sufficiently described.
Why is it considered a hybrid?	The hybrid design is now described more explicitly (Fig. 3). The hybrid model incorporated: Face-to-face components – Interactive lectures – Laboratory practical sessions – Field excursions – Group discussions Online components – Pre-recorded lectures – Digital learning materials

	– Online assessments – Technology-supported learning resources
What parts are hybrid?	The revised manuscript now specifies: – That both theory and practical sessions were represented in a hybrid style.
Why those parts?	Pre-recorded videos introduced students to the conceptual content. Virtual field trip helped students visualise real world geology before attending the actual practical. Both approaches are imperative in developing geological knowledges, observation and interpretation skills.
What content material was presented?	Table 1 is now added to outline the themes taught.
Who were the students?	The manuscript now specifies that: – The cohort consisted of 13 undergraduate geology students – Students were enrolled in a newly established geology programme for the very first time.
Were they majors?	The cohort was a mixture of major and elective (Section 3.3).
Did they have any experience?	All students were first time geology enrolments with no prior knowledge or field experience (Section 3.4).
What were the assessments of the students?	The study now clarifies that the curriculum implementation was evaluated over a full academic teaching cycle, including: – Semester-based theoretical instruction – Laboratory practicals – Field-based instruction – Examination assessment
How did the authors use the assessments to come to their conclusions?	Field performance following prior virtual preparation was compared with performance in the absence of a virtual field trip.
How long was the intervention?	The course ran over a semester (Section 3.4).
Did they spend every day or every other day?	Two field trips were organised on two separate days (Section 4.4).
How many field excursions?	Two field visits were undertaken over two separate days (Section 4.4).
How were they different from any regular geoscience course?	Session 1 started with virtual field trip prior to actual field excursion (Fig. 6). Session 2 was carried directly in the field without prior virtual preparation (Actual field trip).
How did AI play a role?	The manuscript now clarifies the role of digital tools including AI-tools used primarily for learning support, including: – Assistance with visualisation of geological model using Visible Geology web application – Interpretation support for geological information vs human insights. AI was therefore used as a supportive educational technology rather than the central instructional mechanism.
Was the remote part of the hybrid aspect synchronous or asynchronous?	The remote component was primarily asynchronous, allowing students to access recorded lectures and supporting materials before classroom discussions, laboratory and field reports (Section 4.4).
Basically, what was the design of the course?	The course was designed as a hybrid instructional model integrating face-to-face lectures, laboratory practical sessions, field-based experiential learning, and digital learning resources (Fig. 3). The structure followed

	principles of constructive alignment, whereby theoretical instruction introduced core geological concepts, laboratory activities developed analytical skills, and field excursions enabled the application of knowledge in real-world contexts (Fig. 1). Online and AI-supported resources were used to support preparation, reinforce learning, and provide flexible access to course materials.
What effective teaching strategies did instructors use? Why those particular ones?	A combination of interactive lectures, laboratory-based practical training, field-based experiential learning, and guided problem-solving activities were implemented (Table 1). These strategies were selected to align with the epistemological nature of geoscience, where understanding is developed through observation, interpretation, and application of concepts in real-world contexts. Fieldwork and laboratory activities were particularly emphasised to develop students' observational and analytical skills, while digital and AI-supported resources were used to enhance accessibility, support independent learning, and reinforce conceptual understanding.
Was there more traditional instruction? What parts? Why those parts?	The revised manuscript now explains that the instructional design was informed by: – Constructive alignment principles – Experiential learning through fieldwork – Blended learning approaches Teaching strategies included: – Guided field work – Problem-based geological interpretation – Laboratory sample analysis Online preparatory learning
Third, this is an investigation into the efficacy of an innovative instructional design. However, if the authors do not know where the participants' knowledge was in the beginning (I did not see any pre-intervention measurements), how can they say anything about the effects of students' participation in the instructional intervention?	We acknowledge the reviewer's concern regarding the use of the term "efficacy." The purpose of the study was to evaluate curriculum implementation and alignment in a newly established geology programme, the study has attempted to measure causal learning gains (Table 2). 1. To address this issue, the following revisions have been made: The manuscript no longer claims to demonstrate instructional efficacy in a causal sense. Instead, the study now states that it: – Evaluates the implementation of a curriculum model – Examines alignment between curriculum goals, instruction, and assessment – Reports student engagement indicators and performance outcomes 2. The manuscript now explains that evaluation was based on: – Student pass rates (Table 2) – Performance in moderated assessments (Fig. 6a) – External examiner evaluations – Peer reviews (Fig. 5) – Institutional quality assurance instruments (Fig 7) – Student feedback (Fig. 6-8)

	These indicators were used as descriptive measures of curriculum functioning, rather than experimental evidence of learning gains. 3. The revised manuscript explicitly acknowledges that: – The cohort size (n = 13) limits statistical generalisation. – The study therefore functions as a programme-level case study rather than a population-level evaluation.
Efficacy should be a measure of growth. This manuscript does not show growth, only the endpoint.	Table 2 is now added and exhibits the status of participants at the beginning and at the end of the course theory assessments, what improvements have been made.
And with an n of 13, making any kind of generalization is dubious.	We concur that n = 13 is insufficient for a broad generation.
I mean, there are some statistical methods (and I am not a statistician) that can be useful in showing some difference between pre- and post-intervention.	We acknowledge that statistical methods such as paired-sample comparisons can be used to examine differences between pre- and post-intervention performance. However, this study did not include baseline (pre-intervention) measurements at the individual student level, and therefore the dataset limits inferential statistical testing of learning gains (Table 2). Instead, the analysis focuses on descriptive indicators of student performance progression and programme-level evaluation data to provide insight into curriculum implementation.
Mainly, what I saw was participant self-report (Did you like the instruction? Were you engaged? etc.). While these kinds of metrics are important, they are not measurements of the efficacy of instruction.	We agree that student feedback alone cannot demonstrate instructional effectiveness. The revised manuscript now clarifies that student feedback was used only as one component of evaluation, alongside: – Moderated assessments (Table 2; Fig. 6a) – External examiners' reports – Peer review (Fig. 5) – Institutional programme evaluation processes These multiple sources were used to provide triangulated evidence of curriculum implementation quality, rather than instructional impact.
For this, you would need to assess what they understand after instruction and how they understand it.	We acknowledge that evaluating instructional impact requires assessing not only what students understand after instruction but also how they develop and apply that understanding. In this study, assessment focused primarily on demonstrated performance in tests and practical tasks, rather than on probing students' underlying conceptual development or reasoning processes in depth.
The authors talk about how there is alignment among the goals, the content material, the assessments, and the teaching strategies, yet this is never described or displayed. The reader must just take the authors' words for it. I am sure the authors are honorable, but as a reader, I like to see some empirical evidence so I can also make some judgments about the success of the project, to check my interpretation of some of the data against the interpretation of the authors. This will give me, the	We agree that this relationship needed to be presented more clearly. In the revised manuscript we have added: – Table 1 illustrating constructive alignment, linking: ▪ Programme learning outcomes ▪ Teaching strategies ▪ Learning activities ▪ Assessment methods This table demonstrates how course objectives were aligned with: – Field exercises – Laboratory analysis – Online learning activities – Theoretical assessments

reader, more confidence in the results.	
Oh, and since this is not qualitative research, the use of Lincoln and Guba (etc.) for the reliability assurances is not warranted. Even if it was warranted, I would shy away from Lincoln and Guba, anyway.	We appreciate this observation, considering that the study does not adopt a qualitative research design. Accordingly, Lincoln and Guba reference has been removed, and the manuscript has been revised to reflect a descriptive case study approach supported by multiple sources of programme evaluation evidence.
Lastly, there are a bunch of typographical errors, additional words, missing words, peculiar phrasing, incomplete sentences, etc. that also need to be addressed.	We have carefully reviewed the entire manuscript and corrected: – Grammatical errors – Typographical issues – Incomplete sentences – Awkward phrasing In addition, the manuscript has undergone substantial language editing to improve clarity and readability.
I am attaching a PDF of the manuscript with my comments added. This will show the exact locations of the issues I had with the paper.	We are grateful of the reviewers' provision of the annotated manuscript and for the detailed, location-specific feedback. These comments have been meticulously considered, and the manuscript has been revised accordingly to address the identified issues in clarity, methodology, and presentation.

We thank the reviewers over again for the thoughtful critique. The comments have significantly improved the manuscript, particularly in the areas of:

1. Methodological clarity
2. Instructional design description
3. Interpretation of evaluation data

We believe the revised manuscript now provides a clearer and more transparent account of the curriculum implementation case study.