INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 15, Issue 2, (2026) E-ISSN: 2226-6348

Data-Driven Differentiated Biology Instruction in a High-Stakes Secondary Classroom: An Action Research Study of Grade Progressions from Trial Examinations to SPM 2025

Norshamsiah Samsudin*

Tunku Kurshiah College, Negeri Sembilan, Malaysia

*Corresponding Author Email: norshamsiah@tkc.edu.my

Shaharudin Jakpar

Faculty of Economics, Business and Accounting, i-CATS University College, Sarawak, Malaysia

DOI Link: http://dx.doi.org/10.6007/IJARPED/v15-i2/28425

Published Online: 29 June 2026

Abstract

Improving Biology achievement in high-stakes secondary examinations requires pedagogical decisions that are both responsive to student readiness and defensible through evidence. This action research study examined the achievement trajectory of 147 Form Five Biology students across three assessment points: Trial SPM SBP, Trial SPM JPNS, and SPM 2025. The practitioner intervention was conceptualised as a continuous differentiated Biology improvement cycle involving diagnostic grade profiling, tiered remediation, targeted feedback, active learning, peer explanation, and repeated verification through assessment evidence. Supplementary qualitative evidence from teacher reflective notes and informal student exit slips was used to interpret possible mechanisms of change. Because grades were ordinal, each grade was converted into a grade quality point (GPK) using A+ = 0, A = 1, A− = 2, B+ = 3, B = 4, C+ = 5, C = 6, D = 7, E = 8, and G = 9; lower scores therefore indicated stronger achievement. Descriptive statistics, bootstrapped confidence intervals, Friedman's repeated-measures test, Kendall's W, Wilcoxon signed-rank tests with Holm adjustment, class-level comparisons, and baseline risk-group analyses were used. Mean GPK improved, decreasing from 5.97 in Trial SBP to 4.51 in Trial JPNS and 2.39 in SPM 2025. A-range grades increased from 7.5% to 59.9%, while pass attainment increased from 81.0% to 100.0%. Friedman's test indicated a statistically significant and practically large shift, χ^2_F(2, N = 147) = 223.67, $p < .001$, Kendall's W = 0.761. All 147 students improved between Trial SBP and SPM 2025, with a mean improvement of 3.58 GPK points, 95% bootstrap CI [3.35, 3.81]. The findings suggest that a structured, data-guided, differentiated action research cycle can support substantial Biology performance gains, although causal interpretation is limited by the absence of a control group and lesson-level fidelity records.

Keywords: Biology Education, Action Research, Differentiated Instruction, Formative Assessment, Grade Trajectory, Spm, Secondary Science, Malaysia, Non-Parametric Statistics

Introduction

Biology is a conceptually dense science subject because students must connect observable phenomena with abstract processes, symbolic representations, microscopic structures, ecological interactions, genetics, physiology, and evidence-based reasoning. In examination-oriented secondary systems, the challenge is intensified because students are expected not only to remember biological facts but also to interpret diagrams, explain mechanisms, use scientific language accurately, and apply concepts in unfamiliar contexts. These demands make Biology a suitable subject for action research because teachers can collect classroom evidence, identify learning barriers, intervene quickly, and use assessment results to refine instruction.

Recent educational research increasingly argues that teachers require practical models that connect evidence, instructional adaptation, and implementation quality. Differentiated instruction is particularly relevant because learners enter the same Biology classroom with unequal prior knowledge, confidence, conceptual vocabulary, and examination fluency. Systematic reviews and meta-analyses indicate that differentiated instruction can improve learning outcomes when it is planned around readiness, content, process, product, and classroom context (AM et al., 2023; Kahmann et al., 2022; Langelaan et al., 2024). However, research also shows that differentiation is difficult to implement unless teachers receive clear decision structures, assessment information, and reflective feedback loops (Langelaan et al., 2024; Meuleners et al., 2025).

Formative assessment and feedback provide the second foundation of this study. Meta-analytic evidence demonstrates that formative assessment can positively influence K–12 learning, especially when assessment information is used to guide subsequent instruction rather than merely to record performance (Sortwell et al., 2024; Xuan et al., 2022). Feedback research further suggests that feedback is most powerful when it gives students actionable information about the task, the process of improvement, and self-regulated learning strategies (Morris et al., 2021; Panadero et al., 2023; Wisniewski et al., 2020; Yan et al., 2022). For Biology teachers, this means that grade evidence should be translated into specific learning repair actions, such as re-teaching misconceptions, strengthening diagram interpretation, building scientific explanation skills, and rehearsing response structures.

Science education literature also supports active and technology-enhanced learning when such tools are embedded in cognitively meaningful tasks. Active learning has been associated with stronger STEM outcomes and narrower achievement gaps (Freeman et al., 2014; Theobald et al., 2020). Digital tools, augmented reality, and mobile visualisation can strengthen science learning when they help students manipulate representations, visualise invisible processes, or connect classroom learning with authentic phenomena (Anđić et al., 2025; Czok et al., 2023; Faria & Miranda, 2024; Hillmayr et al., 2020; Hsu et al., 2023; Krug et al., 2023; Marini et al., 2022; Nurdin et al., 2025; Permana et al., 2024; Sattar et al., 2025; Stanič & Špernjak, 2025). The present study does not evaluate a single technology tool; instead, it positions digital and visual resources as part of a broader differentiated Biology improvement cycle.

Action research is appropriate for this context because it treats the teacher as a practitioner-scholar who systematically examines a problem of practice, implements an improvement

cycle, observes outcomes, and reflects on future refinement. Contemporary discussions of practitioner action research emphasise reflection, professional learning, evidence use, and the translation of research into classroom practice (Gaias et al., 2023; Ryan et al., 2024; Støren, 2024; Tsaliki et al., 2024; Ventista & Brown, 2023). The present manuscript contributes to that agenda by showing how a real Biology grade dataset can be transformed into a defensible action research narrative and a reproducible statistical analysis using ordinal assessment evidence.

The study was conducted in a high-performing Malaysian residential secondary school context with five Form Five Biology classes. The lead practitioner-researcher is an experienced Biology educator whose professional profile includes more than 27 years of secondary education experience, student-centred learning innovation, sustainability integration, and use of digital tools in Biology instruction. Rather than repeating a prior manuscript style, the present paper adopts a new outcome-focused action research design: it foregrounds anonymised student achievement trajectories, non-parametric statistical analysis, publication-ready black-and-white visuals, and a transferable model for Biology educators.

Problem Statement and Research Questions

The instructional problem addressed in this study was the uneven Biology readiness shown by students in trial examination results and the need to convert those results into a structured improvement cycle before the final national examination. Although trial examinations are commonly used to diagnose student readiness, teachers often require a systematic method for translating grade distributions into differentiated instructional decisions and for demonstrating whether subsequent changes are educationally meaningful.

The study was guided by four research questions:

RQ1: How did the Biology grade distribution change from Trial SPM SBP to Trial SPM JPNS and SPM 2025?

RQ2: Was the within-student change in ordinal Biology achievement statistically significant and practically meaningful?

RQ3: How did improvement vary by class and by baseline risk category?

RQ4: What qualitative mechanisms of change were reflected in teacher observations and student exit-slip evidence?

RQ5: What data-guided action research model can be derived for future Biology educators?

Action Research Intervention Logic

The intervention logic was reconstructed as a continuous differentiated Biology improvement cycle. The cycle began with diagnostic profiling of trial results, followed by differentiated instructional planning, targeted learning activities, feedback-based error repair, and verification using the next assessment. This structure is consistent with research on differentiated instruction, formative assessment, teacher professional learning, and implementation-oriented educational improvement (Gaias et al., 2023; Kahmann et al., 2022; Langelaan et al., 2024; Ryan et al., 2024; Sortwell et al., 2024; Støren, 2024; Ventista & Brown, 2023).

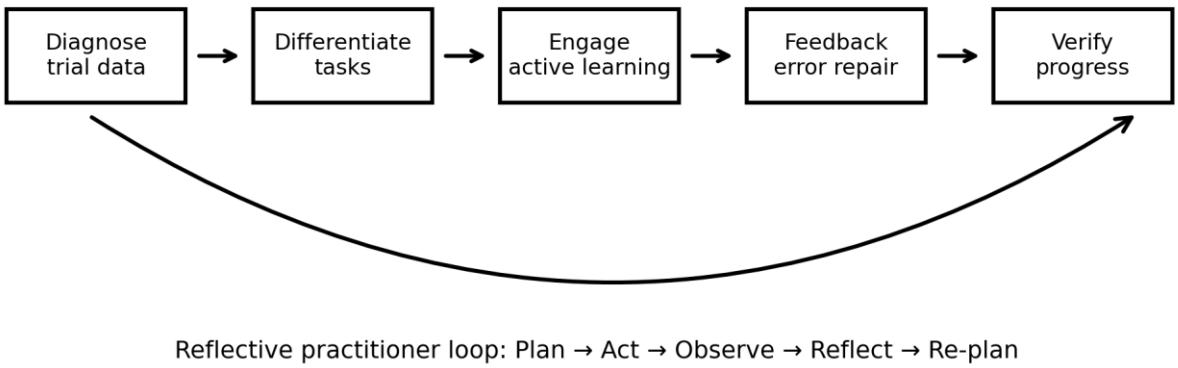

Figure 1. Biology differentiated action research cycle. The model is presented in black and white for academic journal reproduction.

Methodology

Research Design

This study used a practitioner action research design with retrospective analysis of classroom achievement evidence. The research design was not a randomised experiment; it was a real-world classroom improvement inquiry in which the teacher used trial examination evidence to plan differentiated instructional support before the final examination. The design therefore prioritises ecological validity, pedagogical usefulness, and transparent statistical documentation, while acknowledging that causal claims must be treated cautiously.

The study adopted an outcome-focused quantitative strand supported by a reflective action research interpretation. The quantitative strand examined ordinal grade progressions at three time points. The reflective strand translated those results into a transferable Biology action research model. This mixed logic follows the view that classroom action research can generate local professional knowledge when evidence is systematically collected, analysed, and connected to practical improvement decisions (Støren, 2024; Tsaliki et al., 2024; Ventista & Brown, 2023).

Participants, Data Sources, and Qualitative Materials

The dataset contained 147 Form Five Biology students from five classes: 5C, 5E, 5K, 5N, 5T. The data file included student identification numbers, names, class labels, gender labels, and Biology grades for three assessment points. Student names were excluded from all analyses and are not reproduced in this manuscript. The gender column showed a single category for all students; therefore, no gender comparison was conducted.

Supplementary qualitative materials comprised the teacher's reflective journal notes and informal student exit slips collected during the differentiated intervention period. These materials were used to illuminate the mechanisms behind the quantitative grade movement, including students' perceptions of tiered grouping, peer explanation, targeted feedback, and scaffolded Biology tasks. They were not treated as a standalone qualitative dataset for formal thematic saturation; rather, they served as practitioner evidence to support interpretation of the action research cycle.

Table 1
Participant structure and class-level mean achievement trajectory

Class	n	Trial SBP mean GPK	Trial JPNS mean GPK	SPM 2025 mean GPK	Mean improvement SBP→SPM
5C	29	6.90	5.07	3.38	3.52
5E	30	4.47	3.37	1.17	3.30
5K	29	7.66	4.72	3.72	3.93
5N	30	3.23	3.87	0.63	2.60
5T	29	7.72	5.59	3.14	4.59

Note. GPK = ordinal grade quality point; lower GPK indicates stronger achievement. Improvement is calculated as Trial SBP GPK minus SPM 2025 GPK; positive values indicate improvement.

Operationalisation of Biology Achievement
Biology grades were treated as ordinal categories. To allow repeated-measures analysis while respecting grade order, each grade was converted to a grade quality point (GPK) using the following mapping: A+ = 0, A = 1, A− = 2, B+ = 3, B = 4, C+ = 5, C = 6, D = 7, E = 8, and G = 9. Under this scale, lower values represent stronger achievement. The same transformation was applied consistently across all assessment points.

The within-student improvement score was defined as $\Delta_i = GPK_{i,\text{Trial SBP}} - GPK_{i,\text{SPM 2025}}$. A positive Δ_i means that a student moved to a stronger grade by the final examination. This operationalisation is transparent, reproducible, and appropriate for ranking improvement across ordinal achievement categories.

Data Analysis Strategy
The analysis used descriptive and non-parametric inferential statistics because the dependent variable was ordinal and repeated within the same students. Descriptive analysis included grade frequencies, percentages, means, standard deviations, medians, quartiles, and bootstrapped 95% confidence intervals for mean GPK. Although means are reported for readability and to align with school grade-point reporting, medians and non-parametric tests were emphasised because ordinal achievement categories do not possess equal-interval measurement properties (Fagerland, 2012; Sullivan & Artino, 2013).

Friedman's test was used to evaluate whether the repeated grade distributions differed across the three assessments. Kendall's W was reported as a within-cohort effect-size index. Pairwise changes were examined using Wilcoxon signed-rank tests, with Holm adjustment to reduce the risk of familywise Type I error across multiple comparisons (Holm, 1979). Effect size r was computed as $|z|/\sqrt{N_{\text{nonzero}}}$ for Wilcoxon comparisons. Class-level differences in total improvement were explored using the Kruskal–Wallis test. Reporting effect sizes alongside significance tests follows recommendations that educational studies should provide practical magnitude, not merely p values (Tomczak & Tomczak, 2014).

Qualitative observations were reviewed interpretively to identify classroom mechanisms that could plausibly explain the observed GPK movement. The interpretation focused on recurring practitioner observations and student comments related to learner confidence, scaffolded

response construction, peer teaching, error repair, and active diagram-based reasoning. Direct student remarks are presented anonymously and only as illustrative evidence.

Ethical and Research Integrity Considerations

The dataset contained identifiable names in the original school record; therefore, anonymisation was essential. Student names were removed before analysis, and results are reported only at cohort, class, and grade-band levels. No individual student is identifiable in the manuscript. Because the study uses secondary achievement records from a school setting, authors should obtain institutional permission and, where required, ethics exemption or approval before journal submission. Any use of artificial intelligence or statistical software in manuscript preparation should be declared according to the selected journal's research integrity policy.

Results

Overall Grade Distribution

The grade distribution showed a substantial movement from lower achievement categories in Trial SBP toward higher achievement categories by SPM 2025. In Trial SBP, 28 students obtained G and 19 obtained E. By SPM 2025, no student remained in G, only two students obtained E, and 76 students obtained A+ or A. This distributional movement is summarised in Table 2.

Table 2

Biology grade distribution across three assessment points

Grade	Trial SBP	Trial JPNS	SPM 2025
A+	0	0	12
A	1	6	64
A-	10	28	12
B+	19	21	21
B	13	22	16
C+	20	22	7
C	16	20	6
D	21	12	7
E	19	8	2
G	28	8	0

Note. Grades are reported as raw counts. A+ is the highest grade and G is the lowest grade category in the dataset.

For publication clarity, Figure 2 presents the mean GPK trajectory and Figure 3 groups grades into interpretable bands. The largest visual change occurred between Trial JPNS and SPM 2025, when A-range attainment rose sharply and the fail category disappeared.

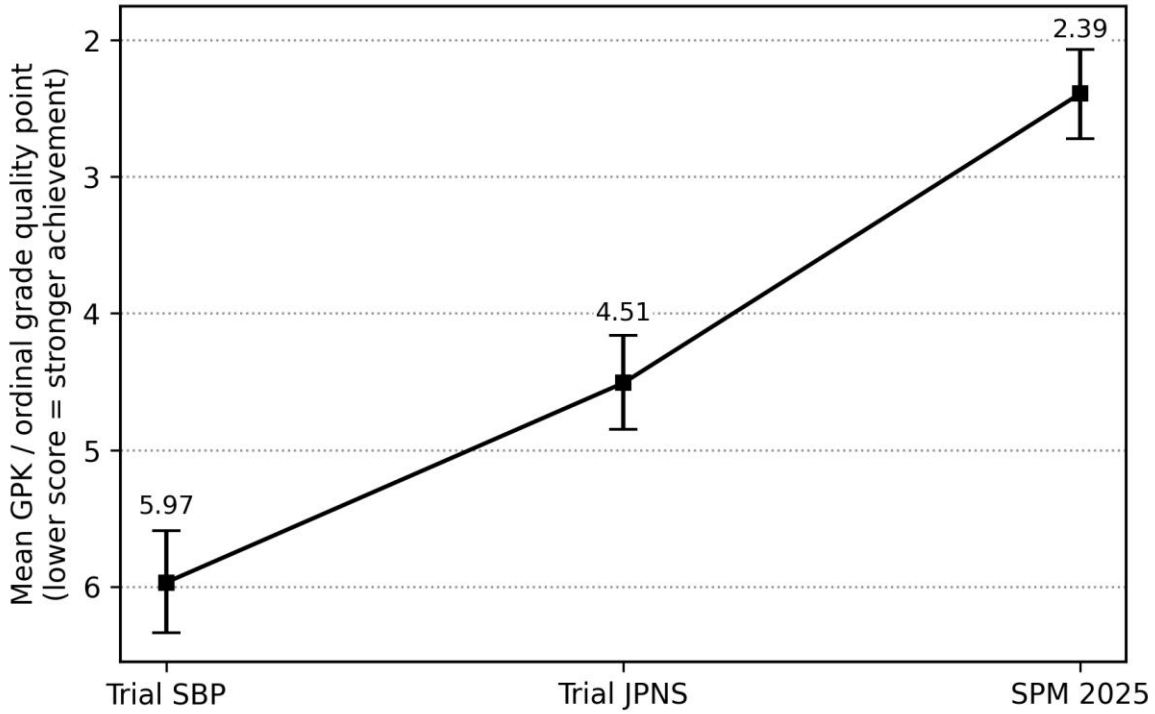

Figure 2. Mean Biology GPK trajectory with bootstrapped 95% confidence intervals.

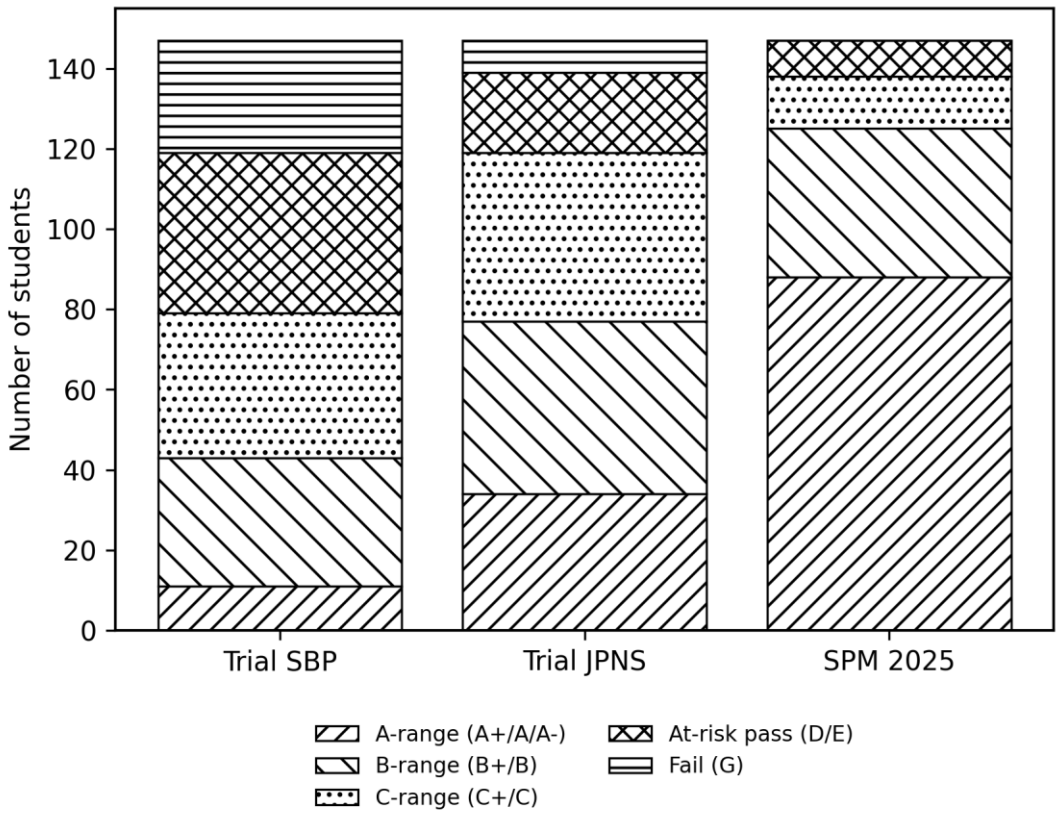

Figure 3. Distributional shift in Biology grade bands from Trial SBP to SPM 2025.

Descriptive Achievement Indicators

Table 3

Descriptive indicators of Biology achievement

Indicator	Trial SBP	Trial JPNS	SPM 2025
Mean GPK	5.97	4.51	2.39
SD	2.31	2.17	1.99
95% bootstrap CI for mean	[5.59, 6.34]	[4.16, 4.85]	[2.07, 2.72]
Median	6	4	1
Q1–Q3	4–8	3–6	1–4
A-range, n (%)	11 (7.5%)	34 (23.1%)	88 (59.9%)
B and above, n (%)	43 (29.3%)	77 (52.4%)	125 (85.0%)
C+ and above, n (%)	63 (42.9%)	99 (67.3%)	132 (89.8%)
Pass, n (%)	119 (81.0%)	139 (94.6%)	147 (100.0%)

Note. Lower GPK values indicate stronger achievement. A-range = A+, A, and A−. B and above = A+ through B. C+ and above is treated as credit-level or stronger for descriptive purposes.

Mean GPK decreased from 5.97 in Trial SBP to 4.51 in Trial JPNS and 2.39 in SPM 2025. Because lower GPK represents stronger achievement, the direction of change indicates improvement. The A-range rate increased from 7.5% to 23.1% and then to 59.9%. Pass attainment also improved from 81.0% to 94.6% and finally to 100.0%.

Qualitative Observations of the Mechanisms of Change

Although the quantitative GPK trajectories showed substantial achievement gains, qualitative evidence from the teacher's reflective journal and informal student exit slips helped explain how those gains may have occurred. The Trial SBP profile indicated that a conventional whole-class pace risked leaving D-to-G grade students behind, particularly in abstract Biology topics such as genetics. This observation became the basis for tiered grouping, scaffolded practice, and targeted concept repair.

During the intervention, the teacher used tiered grouping to differentiate both task complexity and instructional support. A-band students worked independently and collaboratively on complex, higher-order application essays (KBAT), whereas at-risk students joined small-group sessions using scaffolded Punnett square templates. Reflective notes recorded an immediate change in classroom atmosphere: when students were freed from the pressure of performing in front of higher-achieving peers, they began asking foundational questions that they had previously been reluctant to voice.

Post-intervention feedback suggested that peer teaching also benefited higher-achieving students. One A-band student reflected, "I thought I understood the mechanism of urine formation, but when I had to explain the loop of Henle to my group members who were struggling, I realized I had gaps in my own knowledge. Teaching them actually helped me secure my A+ in the final SPM paper." This reflection suggests that explanation-based peer support may have strengthened metacognitive monitoring and conceptual precision among stronger learners.

Students who moved from a failing grade in Trial SBP to a passing grade in SPM 2025 frequently described the differentiated environment as a turning point. A student who improved from G to C+ wrote, "In standard classes, the teacher goes too fast to cover the syllabus. In our targeted group, the teacher gave us specific checklists for the subjective questions. I felt less stupid making mistakes in the small group, and I finally understood how the marks are awarded." The comment indicates that the intervention may have supported not only content understanding but also examination literacy and learner confidence.

Observational notes documented a shift in classroom dynamics. Before the targeted feedback intervention, students in the intermediate bands (C to B+) tended to engage in passive note-copying. During the differentiated cycle, these students were observed spontaneously drawing biological diagrams, including synapse diagrams, on mini-whiteboards to justify their answers during peer discussion. This change points to more active knowledge construction, as students moved from receiving explanations to representing, defending, and revising biological reasoning.

These qualitative observations should be interpreted as explanatory and contextual rather than as independent causal proof. Nevertheless, they strengthen the action research interpretation by showing that the numerical improvement was accompanied by visible changes in learner participation, peer explanation, confidence to ask questions, and willingness to repair errors.

Repeated-Measures Inferential Analysis

Friedman's repeated-measures test showed a statistically significant difference in Biology achievement across the three assessment points, χ^2_F(2, N = 147) = 223.67, $p < .001$. Kendall's $W = 0.761$, indicating a large practical shift in ordinal grade ranks across time. Pairwise Wilcoxon signed-rank tests confirmed significant improvement from Trial SBP to Trial JPNS, from Trial JPNS to SPM 2025, and from Trial SBP to SPM 2025 after Holm adjustment.

Table 4
Pairwise Wilcoxon signed-rank analyses of Biology grade improvement

Comparison	**Mean Δ**	**Median Δ**	**Improved / Same / Declined**	**W**	**|z|**	**r**	p_Holm
Trial SBP vs Trial JPNS	1.46	1	104 / 25 / 18	1340.0	6.20	0.561	< .001
Trial JPNS vs SPM 2025	2.12	2	128 / 7 / 12	1109.5	8.01	0.677	< .001
Trial SBP vs SPM 2025	3.58	4	147 / 0 / 0	0.0	10.57	0.872	< .001

The strongest change was observed between Trial SBP and SPM 2025. All 147 students improved, the mean improvement was 3.58 GPK points, the median improvement was 4 points, and the bootstrapped 95% CI for mean improvement was [3.35, 3.81].

Class-Level Outcomes

Every class demonstrated statistically significant within-class improvement across the three assessment points. The class-level pattern showed that students in classes with weaker Trial SBP starting points tended to record larger absolute GPK improvements, while classes already

closer to the A-range had less room for numerical improvement. This pattern is educationally important because it indicates both achievement recovery among initially weaker groups and excellence consolidation among stronger groups.

Table 5
Class-level repeated-measures improvement

Class	n	Trial SBP mean	SPM 2025 mean	Mean Δ	Friedman χ^2	p	Kendall W
5C	29	6.90	3.38	3.52	38.97	< .001	0.672
5E	30	4.47	1.17	3.30	52.90	< .001	0.882
5K	29	7.66	3.72	3.93	45.49	< .001	0.784
5N	30	3.23	0.63	2.60	48.50	< .001	0.808
5T	29	7.72	3.14	4.59	49.78	< .001	0.858

Note. Mean Δ = Trial SBP mean GPK minus SPM 2025 mean GPK. Positive values indicate improvement.

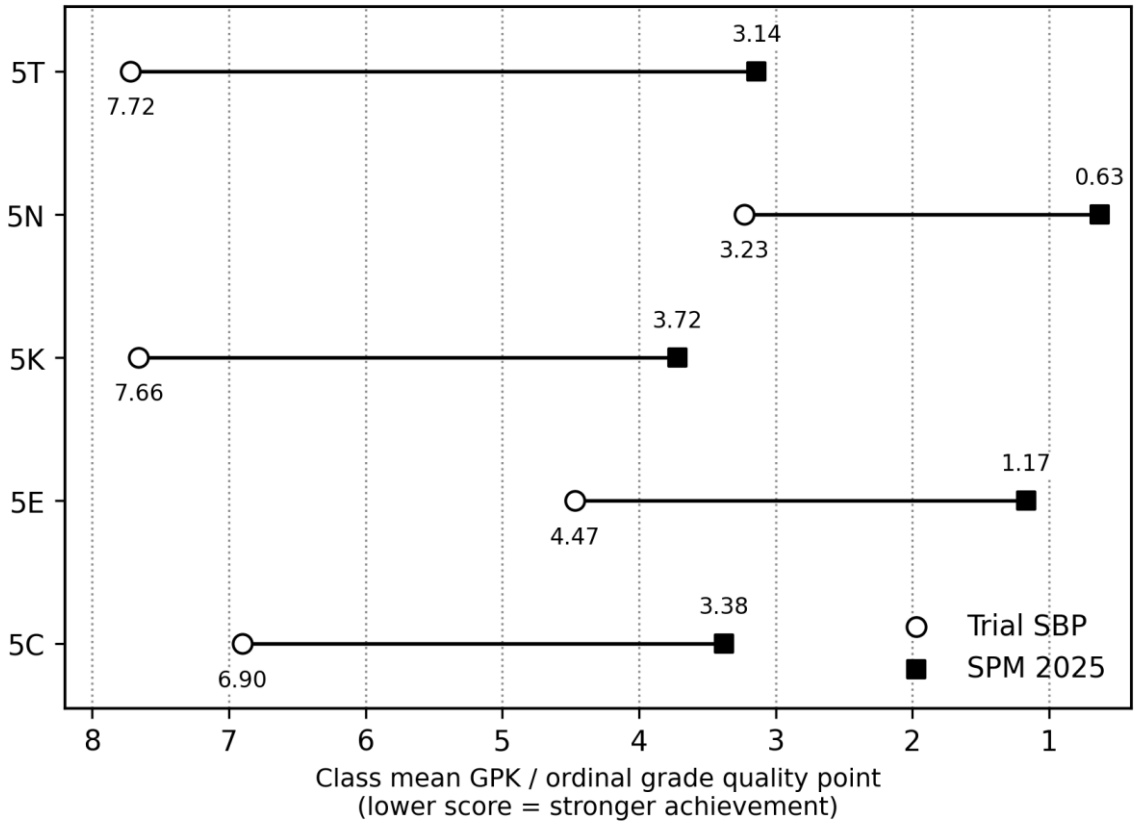

Figure 4. Class-level movement from Trial SBP to SPM 2025.

A Kruskal–Wallis test indicated that total improvement differed by class, $H(4) = 30.83$, $p < .001$. This result should not be interpreted as a competition among classes because baseline readiness differed substantially. Instead, the class-level findings support differentiated planning: high-performing classes may require enrichment and precision feedback, whereas higher-risk classes may benefit from diagnostic re-teaching and foundational concept repair.

Baseline Risk-Group Outcomes

Students were grouped according to their Trial SBP starting point. This analysis is useful for action research because it shows whether improvement was concentrated only among

already strong students or whether initially weaker students also benefited. The results indicate that all baseline groups improved, including the at-risk baseline group.

Table 6
SPM 2025 outcomes by Trial SBP baseline risk group

Trial SBP baseline group	n	SPM mean GPK	SPM A-range n	SPM B+ and above n	Mean Δ	Improved n
A-range baseline (A+, A, A-)	11	0.45	11	11	1.45	11
B-range baseline (B+, B)	32	0.88	31	32	2.53	32
C-range baseline (C+, C)	36	1.28	34	36	4.17	36
At-risk baseline (D, E, G)	68	4.00	12	30	4.10	68

Note. At-risk baseline refers to students who obtained D, E, or G in Trial SBP. B+ and above includes A+, A, A−, and B+.

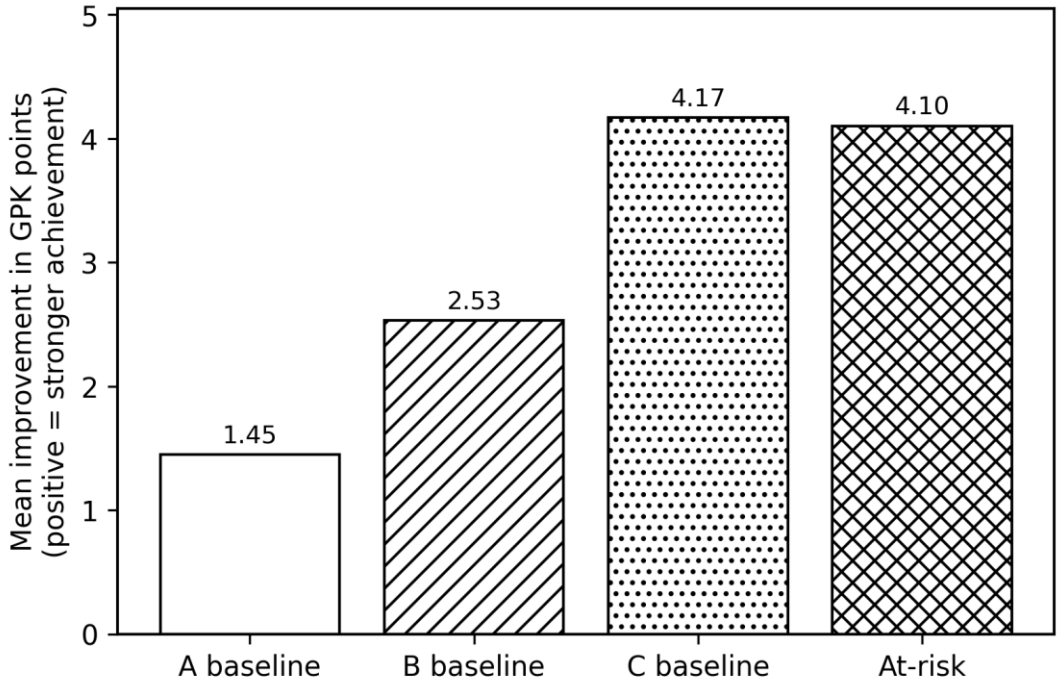

Figure 5. Mean improvement by Trial SBP baseline group.

The at-risk baseline group contained 68 students. All 68 improved by SPM 2025, with a mean gain of 4.10 GPK points. However, the final achievement distribution also shows that recovery did not mean identical outcomes: 12 at-risk baseline students reached the A range and 30 reached B+ or above. This finding reinforces the need for differentiated expectations: some

students require excellence extension, while others require sustained consolidation beyond pass-level recovery.

Correlation of Assessment Points

Spearman rank correlations suggested that assessment points were positively related but not redundant. Trial SBP correlated with Trial JPNS (Spearman rho = 0.412, $p < .001$), and Trial JPNS correlated with SPM 2025 (Spearman rho = 0.393, $p < .001$). The strongest association was between Trial SBP and SPM 2025 (Spearman rho = 0.856, $p < .001$). The correlations indicate that earlier achievement was informative, but the distributional shift also shows that many students changed position substantially over time.

Discussion

Interpretation of the Achievement Trajectory

The results show a large and consistent improvement in Biology achievement from Trial SBP to SPM 2025. The improvement was visible in several indicators: lower mean GPK, higher A-range attainment, disappearance of the G category, universal individual improvement between Trial SBP and SPM 2025, and statistically significant repeated-measures test results. The magnitude of Kendall's W suggests that the change was not merely statistically detectable because of sample size; it was practically substantial in the context of ordinal grade movement.

From an action research perspective, the most important finding is that assessment evidence can be converted into instructionally useful categories. Trial SBP identified substantial risk: nearly half of the cohort was in the D/E/G baseline category. The subsequent improvement suggests that the period between trial examinations and the final examination was not simply a waiting period; it functioned as an instructional recovery and consolidation window. This supports the argument that trial assessments should be treated as formative diagnostic opportunities, not only as summative predictors.

The qualitative observations help explain the mechanisms behind this movement. Tiered grouping appeared to reduce performance anxiety among at-risk learners, scaffolded Punnett square templates clarified procedural reasoning in genetics, peer teaching deepened conceptual monitoring among A-band students, and mini-whiteboard diagramming encouraged intermediate learners to externalise and defend their biological reasoning. These mechanisms are consistent with the quantitative pattern, in which improvement was not confined to already high-achieving students but was observed across baseline risk groups.

Contribution to Biology Pedagogy

The study contributes to Biology education in four ways. First, it demonstrates a practical method for analysing real school grade data without violating the ordinal nature of grades. Second, it provides a black-and-white action research model that can be reproduced in schools and journal publications. Third, it shows how differentiated instruction can be linked with grade-band diagnostics, rather than being discussed only as a general philosophy. Fourth, it highlights that Biology improvement requires attention to both conceptual understanding and examination performance, including scientific explanation, vocabulary, diagram interpretation, and application.

The findings align with meta-analytic and review evidence that differentiated instruction, formative assessment, feedback, self-assessment, peer assessment, and active learning can strengthen achievement when they are intentionally implemented (AM et al., 2023; Freeman et al., 2014; Panadero et al., 2023; Sortwell et al., 2024; Theobald et al., 2020; Wisniewski et al., 2020; Yan et al., 2022). In Biology classrooms, the practical implication is that teachers should not use one uniform revision strategy for all students. Learners at A-range baseline may need enrichment, application, and precision marking; B/C baseline learners may need structured conversion of partial understanding into high-quality responses; at-risk learners may need foundational concept repair, repeated retrieval practice, and tightly guided feedback.

Data-Guided Differentiation Framework for Future Educators

The following framework is proposed for Biology educators who wish to replicate or adapt the action research cycle:

1. Convert grade evidence into an ordinal quality-point profile so that individual and class movement can be monitored.
2. Identify grade-band needs instead of treating the class as a single group. A-range, B-range, C-range, and at-risk learners require different support.
3. Design differentiated Biology tasks around concept repair, diagram interpretation, structured explanation, data-response practice, and application to unfamiliar contexts.
4. Use feedback cycles that require students to correct errors, justify revised answers, and explain why the correction improves scientific accuracy.
5. Verify progress through the next assessment and reclassify learners according to updated evidence.
6. Reflect on class-level patterns to decide whether the next cycle should emphasise recovery, consolidation, or excellence extension.

Digital and sustainability-oriented tools can be integrated where they strengthen scientific meaning, not as decorative additions. For example, animations, simulations, augmented reality, ecological datasets, and interactive quizzes may support Biology learning when they help students visualise dynamic processes and practise explanation (Czok et al., 2023; Faria & Miranda, 2024; Hillmayr et al., 2020; Hsu et al., 2023; Meuleners et al., 2025; Nurdin et al., 2025; Sattar et al., 2025; Stanič & Špernjak, 2025).

Caution Against Over-Claiming Causality

Although the outcome pattern is strong, the study should not be interpreted as definitive causal proof that the intervention alone produced the observed improvement. The dataset did not include a control group, random assignment, attendance data, lesson-level fidelity logs, student motivation measures, prior achievement covariates, or raw examination scores. Assessment difficulty may also have differed across Trial SBP, Trial JPNS, and SPM 2025. These limitations are common in school-based action research, but they must be reported transparently to satisfy journal expectations.

A stronger future design would include lesson observation records, intervention dosage, student work samples, pre-post concept inventories, student voice data, and teacher reflection logs. A comparative or quasi-experimental design could also examine whether classes receiving a more formalised differentiated cycle improve more than comparable

classes receiving ordinary revision. Implementation-science literature further suggests that future studies should document context, fidelity, adaptation, feasibility, and sustainability if the model is scaled beyond one school (Gaias et al., 2023; Ryan et al., 2024).

Implications for Practice

For classroom teachers, the study demonstrates that grade data can become a practical instructional map. Teachers should begin by identifying the students and classes furthest from target performance, then develop differentiated Biology learning tasks based on the specific type of weakness. Rather than interpreting D, E, or G merely as failure categories, those grades can be treated as diagnostic signals for concept repair and confidence rebuilding.

For school leaders, the findings suggest that trial examination data should be reviewed through improvement dashboards rather than only through ranking tables. A school-level Biology dashboard can include mean GPK, A-range rate, pass rate, class movement, at-risk recovery, and the number of students improving by two or more grade points. Such a dashboard can help departments decide where to allocate peer tutoring, small-group coaching, remedial modules, or teacher collaboration time.

For Biology curriculum leaders, the study highlights the importance of connecting scientific literacy with assessment literacy. Students must learn not only Biology content but also how to express causal mechanisms, interpret diagrams, justify biological relationships, and answer structured items with precision. Differentiated instruction should therefore include both conceptual scaffolding and answer-construction scaffolding.

Limitations and Future Research

Several limitations should be acknowledged. First, the study used one cohort from one school; therefore, generalisation to other school types should be cautious. Second, the dataset consisted of final grades rather than raw marks or item-level performance, limiting fine-grained analysis of topic-specific weaknesses. Third, all students shared the same gender label, preventing gender-based subgroup analysis. Fourth, the absence of a control group limits causal attribution. Fifth, the qualitative evidence was based on practitioner reflection and informal exit slips rather than a formally coded qualitative protocol with independent raters. Sixth, the action research cycle was reconstructed from the practitioner context and outcome records; future studies should include lesson plans, observation rubrics, student artefacts, and fidelity evidence.

Future research should examine whether the model can be replicated across schools with different achievement profiles. Studies could combine quantitative grade analysis with qualitative student interviews to understand how learners experience differentiated Biology support. Item-level analysis would also allow teachers to identify whether improvement is strongest in knowledge recall, application, experimental skills, data interpretation, or scientific explanation. Finally, a longitudinal design could evaluate whether students retain Biology understanding beyond the examination period.

Conclusion

This action research study provides evidence of substantial Biology achievement improvement across three assessment points in a cohort of 147 Form Five students. Mean

GPK improved from 5.97 to 2.39, A-range achievement rose from 7.5% to 59.9%, and every student improved between Trial SBP and SPM 2025. Supplementary qualitative observations suggest that the improvement was supported by tiered grouping, scaffolded concept repair, peer explanation, targeted feedback, and more active diagram-based reasoning. The statistical results support the educational value of a data-guided differentiated action research cycle, while the limitations remind readers that practice-based evidence must be interpreted with methodological care. For future Biology educators, the central lesson is clear: assessment evidence becomes powerful when it is transformed into targeted teaching, structured feedback, reflective re-planning, and repeated verification of student progress.

Theoretical and Contextual Contribution

Theoretically, this study advances action research and differentiated instruction by translating an integrated evidence–adaptation–feedback–verification cycle into a Biology classroom context. By using ordinal grade evidence and non-parametric repeated-measures analysis, and the teacher's reflection, the study demonstrates how classroom assessment data can be systematically converted into responsive pedagogical action. It extends current understanding by showing that grade-band diagnostics can guide tiered remediation, enrichment, targeted feedback, peer explanation, and repeated performance verification. In doing so, the study connects teacher decision-making, student readiness, and measurable achievement growth within a coherent and evidence-informed model of classroom improvement.

Contextually, the study is significant because it demonstrates how Trial SBP and Trial JPNS results can be converted into a practical improvement dashboard before SPM in a Malaysian high-stakes secondary science setting. The model offers Biology educators in comparable examination-oriented schools a reproducible and ethically anonymised framework for identifying at-risk learners, supporting excellence extension, documenting class-level movement, and strengthening Biology achievement without relying on costly intervention packages or experimental designs that are difficult to implement in real classrooms.

References

AM, M. A., Hadi, S., Istiyono, E., & Retnawati, H. (2023). Does differentiated instruction affect learning outcome? Systematic review and meta-analysis. *Journal of Pedagogical Research*, *7*(5), 18–33. https://doi.org/10.33902/JPR.202322021

Anđić, B., Weinhandl, R., Maričić, M., & Dünser, B. (2025). Pre-service biology teachers as developers of mobile augmented reality teaching materials: Enhancing, not replacing, science practices and connections with the natural environment. *Thinking Skills and Creativity*, *57*, Article 102019. https://doi.org/10.1016/j.tsc.2025.102019

Czok, V., Krug, M., Müller, S., Huwer, J., Kruse, S., Müller, W., & Weitzel, H. (2023). A framework for analysis and development of augmented reality applications in science and engineering teaching. *Education Sciences*, *13*(9), Article 926. https://doi.org/10.3390/educsci13090926

Fagerland, M. W. (2012). t-tests, non-parametric tests, and large studies—A paradox of statistical practice? BMC Medical Research Methodology, 12, Article 78. https://doi.org/10.1186/1471-2288-12-78

Faria, A., & Miranda, G. L. (2024). Augmented reality in natural sciences and biology teaching: Systematic literature review and meta-analysis. *Emerging Science Journal*, *8*(4), 1666–1685. https://doi.org/10.28991/ESJ-2024-08-04-025

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gaias, L. M., Cook, C. R., Brewer, S. K., Bruns, E. J., & Lyon, A. R. (2023). Addressing the "last mile" problem in educational research: Educational researchers' interest, knowledge, and use of implementation science constructs. *Educational Research and Evaluation*, *28*(7–8), 205–233. https://doi.org/10.1080/13803611.2023.2285440

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, *153*, Article 103897. https://doi.org/10.1016/j.compedu.2020.103897

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hsu, H.-P., Cheah, Y. H., & Hughes, J. E. (2023). A case study of a secondary Biology teacher's pedagogical reasoning and action with augmented reality technology. *Education Sciences*, *13*(11), Article 1080. https://doi.org/10.3390/educsci13111080

Kahmann, R., Droop, M., & Lazonder, A. W. (2022). Meta-analysis of professional development programs in differentiated instruction. *International Journal of Educational Research*, *116*, Article 102072. https://doi.org/10.1016/j.ijer.2022.102072

Krug, M., Thoms, L. J., & Huwer, J. (2023). Augmented reality in the science classroom—Implementing pre-service teacher training in the competency area of simulation and modeling according to the DiKoLAN framework. *Education Sciences*, *13*(10), Article 1016. https://doi.org/10.3390/educsci13101016

Langelaan, B. N., Gaikhorst, L., Smets, W., & Oostdam, R. J. (2024). Differentiating instruction: Understanding the key elements for successful teacher preparation and development. *Teaching and Teacher Education*, *140*, Article 104464. https://doi.org/10.1016/j.tate.2023.104464

Marini, A., Nafisah, S., Sekaringtyas, T., Safitri, D., Lestari, I., Suntari, Y., Umasih, Sudrajat, A., & Iskandar, R. (2022). Mobile augmented reality learning media with Metaverse to improve student learning outcomes in science class. *International Journal of Interactive Mobile Technologies*, *16*(7), 99–115. https://doi.org/10.3991/ijim.v16i07.25727

Meuleners, J. S., Lindermayer, C., Traub, D., Aufleger, M., Rutkowski, A., Kosiol, T., Reith, S., Arvaneh, B., Bannert, M., Ufer, S., & Neuhaus, B. J. (2025). Are digital tools more often implemented in high-quality lessons? Profiles of instructional quality and the use of digital tools in Biology and mathematics lessons. *International Journal of Science and Mathematics Education*, *23*, 1439–1465. https://doi.org/10.1007/s10763-024-10510-1

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), Article e3292. https://doi.org/10.1002/rev3.3292

Nurdin, A. M., Gofur, A., Sari, M. S., & Munzil. (2025). Technology-supported differentiated Biology education: Trends, methods, content, and impacts. *Eurasia Journal of Mathematics*, Science and Technology Education, 21(3), Article em2598. https://doi.org/10.29333/ejmste/16044

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review*, *35*, Article 113. https://doi.org/10.1007/s10648-023-09823-4

Permana, T. I., Husamah, Nurhamdani, M. I., Zaskia, A., Savitri, A., & Salsabila, D. A. (2024). Augmented reality in biology education: A systematic literature review. *Research and Development in Education (RaDEn)*, *4*(1), 630–652. https://doi.org/10.22219/raden.v4i1.32636

Ryan, A., Prieto-Rodriguez, E., Miller, A., & Gore, J. (2024). What can implementation science tell us about scaling interventions in school settings? A scoping review. *Educational Research Review*, *44*, Article 100620. https://doi.org/10.1016/j.edurev.2024.100620

Sattar, M. A., Maqbool, M. U., Zakir, F., & Billah, M. (2025). Enhancing student engagement through augmented reality in secondary Biology education. *Frontiers in Education*, *10*, Article 1628004. https://doi.org/10.3389/feduc.2025.1628004

Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K–12 students' learning: Toward sustainable quality education. *Sustainability*, *16*(17), Article 7826. https://doi.org/10.3390/su16177826

Stanič, K., & Špernjak, A. (2025). Augmented reality in biology education: A literature review. *Multimodal Technologies and Interaction*, *9*(12), Article 117. https://doi.org/10.3390/mti9120117

Støren, K. (2024). Action research and professional development in schools: Reflection as quality development and knowledge production. *Nordic Journal of Comparative and International Education*, *8*(1). https://doi.org/10.7577/njcie.5535

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., ... Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate STEM. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited: An overview of some recommended measures of effect size. *Trends in Sport Sciences*, *21*(1), 19–25.

Tsaliki, C., Papadopoulou, P., Malandrakis, G., & Kariotoglou, P. (2024). A long-term study on the effect of a professional development program on science teachers' inquiry. *Education Sciences*, *14*(6), Article 621. https://doi.org/10.3390/educsci14060621

Ventista, O. M., & Brown, C. (2023). Teachers' professional learning and its impact on students' learning outcomes: Findings from a systematic review. *Social Sciences & Humanities Open*, *8*(1), Article 100565. https://doi.org/10.1016/j.ssaho.2023.100565

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*, Article 3087. https://doi.org/10.3389/fpsyg.2019.03087

Xuan, Q., Cheung, A., & Sun, D. (2022). The effectiveness of formative assessment for enhancing reading achievement in K–12 classrooms: A meta-analysis. *Frontiers in Psychology*, *13*, Article 990196. https://doi.org/10.3389/fpsyg.2022.990196

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, *37*, Article 100484. https://doi.org/10.1016/j.edurev.2022.100484