Jurnal Pendidikan Malaysia 51(2) (2026): 45-57
DOI: http://dx.doi.org/10.17576/JPEN-2026-51.02-04

Biology Distinction by Design: Continuous Differentiated Pedagogy Across Four Malaysian SPM Cohorts

(Kecemerlangan Biologi Melalui Reka Bentuk: Pedagogi Terbeza Berterusan Merentas Empat Kohort SPM Malaysia)

NORSHAMSIAH SAMSUDIN*, FATHIIYAH MOHD POZI, SHAHARUDIN JAKPAR

ABSTRACT

Biology achievement in high-stakes examinations requires more than content coverage because learners must connect concepts, interpret evidence, and produce precise scientific explanations. This action research examined whether continuous differentiated Biology pedagogy was associated with achievement improvement across four Malaysian SPM cohorts: Estelle 2023, Viatrix 2024, Gladius 2025, and Audrea 2026. The study used a retrospective quantitative design involving anonymised aggregate examination records from Form 4, Form 5, Trial SPM SBP, and SPM. Grade distributions, GPMP/GPK, pass rate, G-risk, A-band, and A-B band indicators were reconstructed and analysed using descriptive statistics, chi-square redistribution tests, Cramer's V, Friedman analysis, Kendall's W, and trend diagnostics. The strongest completed-cycle evidence appeared in the 2025 Gladius cohort, where G-risk declined from 43.15% to 0.00%, A-band representation increased from 1.37% to 57.82%, and GPK improved from 7.22 to 2.46. The completed cohorts reached zero SPM failure, while Audrea 2026 showed favourable interim movement. The findings provide practical implications for Biology panels by showing how evidence-informed differentiation can support early risk detection, targeted remediation, and enrichment for higher achievement.

Key Words: action research; Biology education; differentiated pedagogy; formative assessment; SPM achievement

ABSTRAK

Pencapaian Biologi dalam peperiksaan berisiko tinggi menuntut lebih daripada liputan kandungan kerana murid perlu menghubungkan konsep, mentafsir bukti, dan menghasilkan penjelasan saintifik yang tepat. Kajian tindakan ini meneliti sama ada pedagogi Biologi terbeza secara berterusan berkait dengan peningkatan pencapaian merentas empat kohort SPM Malaysia: Estelle 2023, Viatrix 2024, Gladius 2025, dan Audrea 2026. Kajian ini menggunakan reka bentuk kuantitatif retrospektif berdasarkan rekod peperiksaan agregat tanpa nama daripada Tingkatan 4, Tingkatan 5, Percubaan SPM SBP, dan SPM. Taburan gred, GPMP/GPK, kadar lulus, risiko G, jalur A, dan jalur A-B dibina semula serta dianalisis menggunakan statistik deskriptif, ujian khi kuasa dua, Cramer's V, analisis Friedman, Kendall's W, dan diagnostik trend. Bukti kitaran lengkap paling jelas terdapat pada kohort Gladius 2025, apabila risiko G menurun daripada 43.15% kepada 0.00%, jalur A meningkat daripada 1.37% kepada 57.82%, dan GPK bertambah baik daripada 7.22 kepada 2.46. Kesemua kohort lengkap mencapai sifar kegagalan SPM, manakala data interim Audrea 2026 menunjukkan perkembangan awal yang positif. Dapatan ini memberi implikasi praktikal kepada panitia Biologi dengan menunjukkan bagaimana pedagogi terbeza berasaskan bukti dapat menyokong pengesanan risiko awal, pemulihan bersasar dan pengayaan pencapaian tinggi.

Kata Kunci: kajian tindakan; pendidikan Biologi; pedagogi terbeza; pentaksiran formatif; pencapaian SPM

INTRODUCTION

Biology achievement in upper secondary schooling is shaped by more than the memory of factual content. Students are expected to connect molecular, cellular, physiological, ecological and genetic ideas; interpret diagrams; reason from data; and write precise scientific explanations within the constraints of high-stakes examinations. In the Malaysian SPM context, these demands become especially visible because a single class may contain learners who need basic remediation, students who require consolidation of core concepts and high-potential candidates who must be challenged to produce distinction-level responses.

The research problem addressed in this study concerns the uneven movement of Biology performance across repeated school assessments. In practice, examination data are often treated as terminal records rather than diagnostic

signals for instructional redesign. The 2025 Gladius cohort showed a severe early-warning profile, with 63 of 146 valid graded candidates in the G category at PPT F4 and only two candidates in the A-band. However, the same cohort later reached zero G at SPM and 85 candidates in the A-band. This steep movement from risk to stronger attainment provides a useful site for practice-based inquiry.

The specific gap addressed by this study is the limited evidence on how routine high-stakes examination records can be used longitudinally within school-based action research to guide differentiated Biology intervention across consecutive SPM cohorts. Previous studies have established the value of differentiated instruction, formative assessment and active learning, but fewer studies show how these approaches are connected to repeated GPMP/GPK monitoring, grade-band redistribution and recursive classroom decision-making in a real Malaysian examination setting.

The study is situated within action research because the instructional problem emerged from classroom practice and the evidence was used recursively to refine learning support. The intervention was not a commercial package; it was a disciplined pattern of data interpretation, differentiated support, active Biology learning, and formative feedback. The contribution of the study is therefore practical and methodological: it shows how school examination data can be reconstructed mathematically, analysed transparently and used to guide evidence-informed differentiated pedagogy.

The study was guided by a set of objectives and research questions designed to examine Biology achievement improvement across four cohorts. Specifically, the objectives of the study are listed as follows:

1. To document changes in GPMP/GPK, pass rate, A-band representation, A-B band representation and G-risk across the available assessment points.

2. To test whether grade-category and collapsed grade-band distributions changed significantly within the complete 2023, 2024 and 2025 cohorts.

3. To examine the extent to which the 2025 Gladius cohort moved from examination risk to stronger SPM attainment.

4. To interpret what the 2026 Audrea interim evidence suggests about the sustainability of the data-rich differentiated pedagogy cycle.

In line with these four objectives, the study was also structured around four research questions. These research questions are stated as RQ1, RQ2, RQ3 and RQ4 to ensure that each objective is directly connected to a specific analytical focus:

RQ1: How did GPMP/GPK, pass rate, A-band representation, A-B band representation and G-risk change across the available assessment points in the four cohorts?

RQ2: Was there statistically significant redistribution across grade categories and collapsed grade bands within the complete 2023, 2024 and 2025 cohorts?

RQ3: To what extent did the 2025 Gladius cohort demonstrate a movement from examination risk to stronger SPM attainment?

RQ4: What can the 2026 Audrea evidence suggest about the sustainability of the data-rich differentiated pedagogy cycle?

TABLE 1. Research Hypotheses and Alternative Hypotheses

Null hypothesis	Statement	Alternative hypothesis	Statement
H_{01}	There is no observable improvement in GPMP/GPK, pass rate, G-risk, A-band or A-B band indicators across the assessment points.	H_{A1}	The assessment indicators improve across the cycle, shown by lower GPMP/GPK, higher pass rate, lower G-risk and higher A-band/A-B band representation.
H_{02}	Grade-category and collapsed grade-band distributions do not differ significantly across assessment points in the complete cohorts.	H_{A2}	Grade-category and collapsed grade-band distributions differ significantly across assessment points in the complete cohorts.
H_{03}	The 2025 Gladius cohort does not demonstrate meaningful movement from examination risk to stronger SPM attainment.	H_{A3}	The 2025 Gladius cohort demonstrates meaningful movement from examination risk to stronger SPM attainment.
H_{04}	The 2026 Audrea interim evidence does not suggest a sustainable positive trajectory for the differentiated pedagogy cycle.	H_{A4}	The 2026 Audrea interim evidence suggests a sustainable positive trajectory for the differentiated pedagogy cycle, although final effectiveness requires Trial SBP and SPM evidence.

LITERATURE REVIEW

ACTION RESEARCH AND DATA-INFORMED SCIENCE TEACHING

Action research is well suited to classrooms where educators seek to improve classroom learning problems while producing evidence that can be scrutinised by others. In science education, contemporary reviews emphasise that action research is most useful when the problem is clearly defined, the action is documented, evidence is interpretable and the consequences of the intervention are critically reflected upon (Ilgaz et al. 2024; Kamarudin & Mat Noor 2024). Participatory and collaborative forms of science action research also support professional judgement because they connect instructional decisions with observed learner responses (Cortes et al. 2024; Esparza et al. 2022).

The present study adopts this tradition by positioning routine examination data as a source of pedagogical intelligence. However, data-informed teaching is not equivalent to data accumulation. The value of assessment evidence depends on whether it leads to better decisions, including who requires reteaching, which concepts need re-explanation, which students are ready for enrichment and which feedback routines help learners regulate their revision.

DIFFERENTIATED INSTRUCTION AND RESPONSIVE PEDAGOGY

Differentiated instruction is a responsive design orientation in which content, process, product, grouping, questioning and feedback are adjusted according to learners' readiness, learning needs and progress evidence. Recent systematic review and meta-analytic evidence indicates that differentiated instruction can improve learning outcomes when it is structured, aligned with learning objectives and supported by professional expertise (AM et al. 2023; Kahmann et al. 2022; Langelaan et al. 2024). In secondary classrooms, differentiation also has affective importance because students who experience repeated difficulty may disengage unless support is visible, attainable and non-stigmatising (Pozas et al. 2020; Pozas et al. 2021).

In examination classes, differentiation must avoid two weaknesses. The first is over-remediation, where attention is concentrated only on failing students and potential distinction candidates are neglected. The second is over-enrichment, where advanced learners receive demanding tasks while struggling learners continue to accumulate misconceptions. A balanced differentiated model therefore needs dual movement: rescuing learners from G-risk while simultaneously expanding the A-band.

FORMATIVE ASSESSMENT, FEEDBACK AND SELF-REGULATED LEARNING

Formative assessment is powerful when evidence about learning is used to change what learners and educators do next. Recent meta-analyses and umbrella reviews show positive effects of formative assessment on K-12 achievement while cautioning that implementation quality, feedback specificity and student uptake strongly influence outcomes (Sortwell et al. 2024; Yao et al. 2024). In science classrooms, formative assessment requires the elicitation of learners' conceptual reasoning, interpretation of errors and selection of the next instructional response rather than merely counting marks (Dayal 2021; Dini et al. 2020).

Further feedback research clarifies that information alone is insufficient. Feedback becomes educative when it clarifies standards, locates the gap between current and desired performance and guides students towards feasible next actions (Lipnevich & Panadero 2021; Panadero & Lipnevich 2022). Peer assessment, self-assessment and self-regulated learning routines can deepen this process because learners begin to monitor their own response quality and revision strategies (Double et al. 2020; Mejeh et al. 2022; Xu et al. 2023; Yan et al. 2020; Yan et al. 2022; Zheng et al. 2022).

ACTIVE BIOLOGY LEARNING, RETRIEVAL AND INQUIRY

Biology learning requires students to reconstruct explanations, not merely recognise correct terms. Active learning in Biology has been associated with deeper conceptual engagement when students explain mechanisms, draw processes, retrieve concepts, analyse evidence and discuss scientific reasoning (Driessen et al. 2020; Pratiwi et al. 2021; Wang et al. 2023). Wider STEM evidence supports active learning as a means of improving achievement, although methodological critiques remind researchers to describe interventions precisely and avoid simplistic claims (Freeman et al. 2014; Martella et al. 2023; Theobald et al. 2020).

Inquiry-based and Science-Technology-Society approaches are relevant because Biology examination questions often require transfer to unfamiliar contexts. Recent synthesis evidence suggests that inquiry-oriented teaching can strengthen critical thinking and science learning outcomes when inquiry is guided, purposeful and linked to assessment criteria (Acut & Antonio 2023; Arifin et al. 2025; Ozturk et al. 2022).

CONCEPTUAL FRAMEWORK

Figure 1 presents the conceptual framework. The framework links diagnostic evidence to differentiated action, active Biology learning, formative feedback and measurable outcomes. The arrow returning from outcomes to diagnostic evidence is essential because action research is not a one-time intervention but a recursive cycle in which new evidence reopens the next instructional decision.

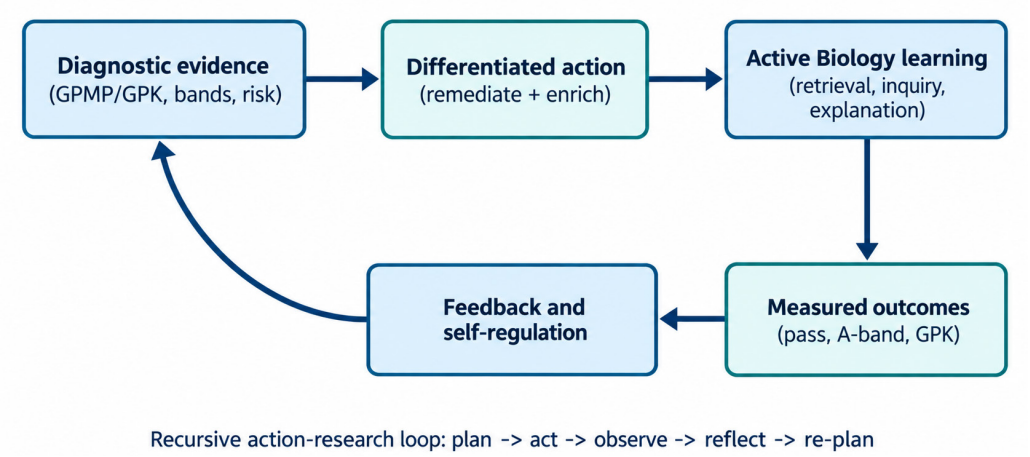


FIGURE 1. Data-rich differentiated Biology pedagogy as a recursive action-research framework

METHODOLOGY

RESEARCH DESIGN

The study used a retrospective action research design with quantitative outcome evaluation. The action-research component captured the iterative cycle of planning, acting, observing and reflecting, while the quantitative component analysed aggregate grade distributions and GPMP/GPK outcomes across repeated assessment points. Because the data were not randomised, did not include a control group and were not available at matched individual-student level, the design was interpreted as practice-based improvement evidence rather than experimental proof.

To strengthen the action research characteristics, each major assessment point was treated as the beginning of a planning-acting-observing-reflecting-replanning cycle. During planning, GPMP/GPK values, pass rate, G-risk, A-band representation and topic weaknesses were reviewed to identify learners requiring remediation, consolidation or enrichment. During action, differentiated tasks, targeted clinics, retrieval practice, inquiry prompts and structured feedback were implemented. During observation, the next assessment results and grade-band movements were examined. During reflection and re-planning, the next instructional cycle was adjusted according to persistent misconceptions, stagnant G-risk or unrealised A-band potential.

Ethical safeguards were observed by using anonymised aggregate school assessment records only. No student names, demographic identifiers, individual marks or interview data were collected. Permission to conduct the analysis was granted by the school administration where the study was conducted.

RESEARCH SAMPLE AND PARTICIPANTS

The study context was Form 4 and Form 5 Biology at Tunku Kurshiah College, Negeri Sembilan, Malaysia. The dataset consisted of anonymised aggregate examination records for four cohorts: Estelle 2023, Viatrix 2024, Gladius 2025 and Audrea 2026. Candidate numbers reported in the summary dashboard ranged from 147 to 149. The 2023 to 2025 cohorts had complete outcome sequences through SPM, while the 2026 Audrea cohort was still in progress and therefore included only interim early-cycle evidence at the time of analysis.

The assessment labels were PPT F4 or PPT T4, PAT F4 or PAT T4, PPT F5, Trial SPM SBP and SPM. The detailed grade table supplied candidate counts across A+, A, A-, B+, B, C+, C, D, E and G. The comparison dashboard supplied GPMP/GPK indicators, ETR targets, SPM outcomes and SPM failure counts.

DATA COLLECTION METHOD AND INSTRUMENTATION

The principal instrument was the school examination record, reconstructed from aggregate grade distributions and GPMP/GPK comparison dashboards. The intervention model was named Iterative Data-Rich Differentiated Biology Pedagogy (IDDBP). It converted examination

evidence into immediate instructional action at different readiness levels. Table 2 summarises the model.

The classroom implementation included concrete differentiated strategies that were selected after each diagnostic review. For remediation groups, the intervention used short reteaching clinics, guided diagram reconstruction, vocabulary repair, worked examples, micro-quizzes and correction of frequent misconceptions. For consolidation groups, the intervention emphasised retrieval practice, peer explanation, structured response rehearsal and comparison with model answers. For A-band enrichment, learners engaged with higher-order data interpretation, mechanism-writing tasks, transfer questions and timed SPM-style responses. Formative assessment was implemented through feedback cycles, self-checklists, peer review and post-assessment reflection so that students could identify why an answer was incomplete and how to improve it in the next cycle.

TABLE 2. Iterative Data-Rich Differentiated Biology Pedagogy Intervention Structure

Mechanism	Instructional action	Student learning function	Expected evidence
Diagnostic evidence	Analyse GPMP/GPK, grade bands, G-risk, A-band potential and topic weakness after each major assessment.	Recognise current achievement position and immediate target.	Sharper risk identification and more accurate grouping.
Flexible grouping	Create temporary groups for remediation, consolidation and enrichment; adjust membership after new evidence.	Receive support or challenge matched to readiness rather than fixed labels.	Movement out of G and D/E bands; stronger A-B band participation.
Tiered remediation	Re-teach difficult concepts through diagrams, guided examples, vocabulary repair and short checks.	Rebuild prerequisite Biology knowledge and reduce misconception load.	Lower G-risk and improved pass rate.
A-band enrichment	Use higher-order structured-response tasks, data interpretation, mechanism explanation and transfer questions.	Practise distinction-level reasoning under examination conditions.	A-band and A-B band expansion.
Retrieval and inquiry	Use low-stakes recall, concept maps, drawing, oral explanation, peer discussion and inquiry prompts.	Strengthen recall, conceptual links and scientific explanation.	Improved GPK and response quality.
Feedback and self-regulation	Provide specific feedback, model answers, correction cycles, peer review and self-monitoring checklists.	Identify gaps, revise responses and plan next revision.	Sustained improvement across assessment cycles.

DATA ANALYSIS METHOD

The main variables were valid graded candidate count (N), grade count (n_g), pass count, G-risk count, A-band count, A-B band count, GPMP/GPK, ETR target and SPM failure count. A-band was defined as A+, A and A-. The A-B band was defined as A+, A, A-, B+ and B. G-risk referred to candidates in the G category. Because the Malaysian grade-point scale assigns lower values to stronger grades, improvement is reflected by declining GPK.

The GPK reconstruction used the formula GPK_t = Sigma_g(n_gt w_g) / N_t, where n_gt is the number of candidates in grade g at assessment t, w_g is the grade-point weight and N_t is the valid graded candidate count. The weights were A+ = 0, A = 1, A- = 2, B+ = 3, B = 4, C+ = 5, C = 6, D = 7, E = 8 and G = 9. The main proportion formulas were Pass_t = ((N_t - G_t) / N_t) x 100; G-risk_t = (G_t / N_t) x 100; A-band_t = ((A+_t + A_t + A-_t) / N_t) x 100; and A-B_t = ((A+_t + A_t + A-_t + B+_t + B_t) / N_t) x 100.

Descriptive statistics reported GPMP/GPK, A-band, A-B band, G-risk and pass-rate trajectories. Chi-square tests of independence examined whether full grade-category distributions changed within each complete cohort. Collapsed grade-band tests used five bands as a sensitivity analysis to reduce sparse-cell concerns. Cramer's V was interpreted as the distributional effect size. Friedman analysis tested whether the three complete cohorts showed systematic GPMP/GPK movement across five repeated assessment points, with Kendall's W reported as a non-parametric effect size.

The dataset was first reconstructed and verified using Microsoft Excel. Excel was used to organise the aggregate grade distributions, confirm valid candidate counts, classify

grades into A-band and A-B band categories, calculate pass rate, G-risk, GPMP/GPK indicators, ETR comparisons and percentage changes across assessment points. The reconstructed dataset was then imported into Stata for inferential statistical analysis. In Stata, chi-square tests of independence were conducted to examine whether full grade-category distributions and collapsed grade-band distributions differed significantly across assessment points within the complete cohorts. Cramer's V was computed to determine the strength of distributional change. Friedman analysis was used to evaluate repeated GPMP/GPK movement across the five assessment points for the completed cohorts, while Kendall's W was reported as the non-parametric effect-size measure. The combined use of Excel and Stata strengthened the analysis by allowing transparent data reconstruction, systematic statistical testing and reproducible interpretation of cohort-level achievement trajectories.

FINDINGS AND DISCUSSION

RESEARCH QUESTION 1

RQ1 asked how GPMP/GPK, pass rate, A-band representation, A-B band representation and G-risk changed across the available assessment points. Table 3 shows the GPMP/GPK trajectory across the four cohorts. Lower values indicate stronger grade quality. The completed cohorts all reached zero SPM failure, while the 2026 Audrea cohort remained incomplete but showed stronger interim values than the corresponding Gladius values at matching early stages.

TABLE 3. GPMP/GPK Comparison Dashboard Across Four Cohorts

Cohort	Candidates	PPT T4	PAT T4	PPT T5	Trial SBP	ETR	SPM	SPM fail	PPT T4 to SPM delta
2023 Estelle	149	5.79	6.69	5.44	5.09	1.47	1.68	0	4.11
2024 Viatrix	147	4.91	5.87	4.45	4.37	1.31	1.59	0	3.32
2025 Gladius	147	7.22	6.69	5.93	5.97	1.52	2.46	0	4.76
2026 Audrea	148	5.34	5.31	4.82	-	1.58	-	-	-

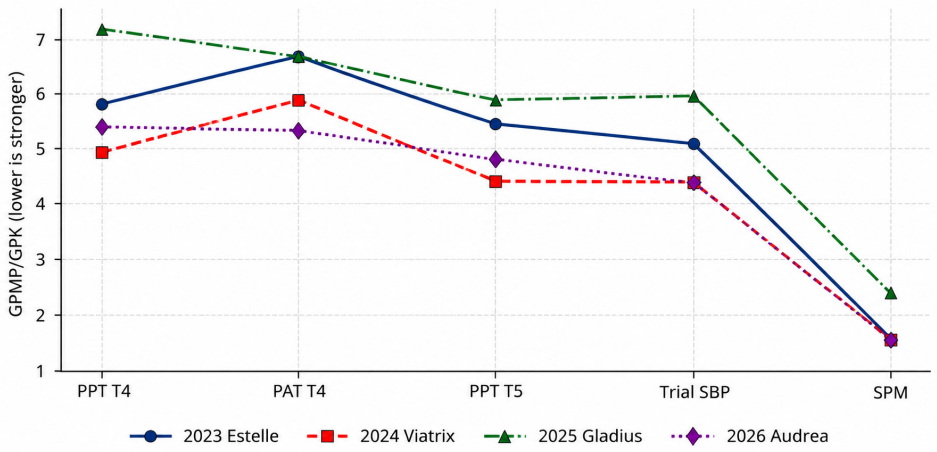


FIGURE 2. GPMP/GPK trajectory by cohort, with lower values indicating stronger grade quality

GPMP/GPK improved across the completed cohorts by the final SPM outcome, and SPM failure was eliminated in Estelle, Viatrix and Gladius. The 2025 Gladius cohort showed the largest improvement from PPT T4 to SPM because it started from the weakest early position. The Audrea 2026 data also moved favourably at the interim stage, with GPMP/GPK improving from 5.34 at PPT T4 and 5.31 at PAT T4 to 4.82 at PPT T5.

H_{01} is rejected and H_{A1} is accepted. The direction of the evidence is consistent with lower GPMP/GPK, elimination of SPM failure in the complete cohorts, and stronger grade-band movement. This decision is based on descriptive cohort trajectories and the significant repeated GPMP/GPK result reported under RQ2.

Table 4 provides the grade-band evidence for the most important completed-cycle and interim comparisons. The Gladius cohort shifted from 43.15% G-risk at PPT F4 to 0.00% at SPM. At the same time, its A-band representation increased from 1.37% to 57.82%, and its A-B band increased from 13.01% to 84.35%. Audrea also showed favourable interim movement: G-risk fell from 14.77% to 10.27%, while A-band representation increased from 19.46% to 26.71%.

TABLE 4. Grade-Band Outcomes for Key Assessment Points

Cohort	Assessment	Valid N	GPK	Pass n (%)	G-risk n (%)	A-band n (%)	A-B band n (%)
2025 Gladius	PPT F4	146	7.22	83 (56.85)	63 (43.15)	2 (1.37)	19 (13.01)
2025 Gladius	PAT F4	147	6.69	101 (68.71)	46 (31.29)	8 (5.44)	25 (17.01)
2025 Gladius	PPT F5	147	5.93	110 (74.83)	37 (25.17)	17 (11.56)	52 (35.37)
2025 Gladius	Trial SBP	147	5.97	119 (80.95)	28 (19.05)	11 (7.48)	43 (29.25)
2025 Gladius	SPM	147	2.46	147 (100.00)	0 (0.00)	85 (57.82)	124 (84.35)
2026 Audrea	PAT F4	149	5.32	127 (85.23)	22 (14.77)	29 (19.46)	65 (43.62)
2026 Audrea	PPT F5	146	4.82	131 (89.73)	15 (10.27)	39 (26.71)	72 (49.32)

RESEARCH QUESTION 2

RQ2 examined whether there was statistically significant redistribution across grade categories and collapsed grade bands within the complete cohorts. Table 5 shows that all complete cohorts recorded statistically significant redistribution in both the full 10-grade and collapsed five-band analyses. The 2025 Gladius cohort produced significant redistribution in the full grade table, chi-square(36, N = 734) = 407.20, $p < .001$, V = .372, and in the collapsed grade-band table, chi-square(16, N = 734) = 302.47, $p < .001$, V = .321. The repeated GPMP/GPK trajectory across completed cohorts was also significant, Friedman chi-square(4, n = 3) = 10.93, p = .027, Kendall W = .91.

TABLE 5. Inferential Statistical Summary

Analysis	Statistic	df/design	p value	Effect size	Interpretation
2023 full 10-grade distribution	chi-square = 475.04	df = 36	< .001	V = 0.399	Large significant redistribution
2023 collapsed five-band distribution	chi-square = 293.12	df = 16	< .001	V = 0.314	Sensitivity-confirmed redistribution
2024 full 10-grade distribution	chi-square = 472.13	df = 36	< .001	V = 0.401	Large significant redistribution
2024 collapsed five-band distribution	chi-square = 272.08	df = 16	< .001	V = 0.304	Sensitivity-confirmed redistribution
2025 full 10-grade distribution	chi-square = 407.20	df = 36	< .001	V = 0.372	Large significant redistribution
2025 collapsed five-band distribution	chi-square = 302.47	df = 16	< .001	V = 0.321	Sensitivity-confirmed redistribution
2026 interim collapsed five-band distribution	chi-square = 3.04	df = 4	.551	V = 0.102	Descriptive but not conclusive
Completed-cohort GPMP/GPK trajectory	Friedman chi-square = 10.93	k = 5; n = 3	.027	Kendall W = 0.91	Strong repeated trajectory with small-n caution

There was statistically significant redistribution across grade categories and collapsed grade bands in the complete 2023, 2024 and 2025 cohorts. The finding is robust because the full 10-grade tests and collapsed five-band sensitivity tests produced the same substantive conclusion.

H_{02} is rejected and H_{A2} is accepted. The p values for all complete-cohort grade redistribution tests were below .001, and the Cramer's V values indicated meaningful distributional change. The 2026 interim test was not significant and is therefore not used as evidence of final redistribution.

RESEARCH QUESTION 3

RQ3 focused on the extent to which the 2025 Gladius cohort moved from examination risk to stronger SPM attainment. Figures 3 and 4 show a clear rescue-and-excellence pattern. G-risk declined steadily and reached zero at SPM, while A-band and A-B band outcomes expanded sharply by the final assessment. This indicates that the intervention did not merely move weak candidates across a minimum pass boundary; it also supported upward movement into stronger achievement bands.

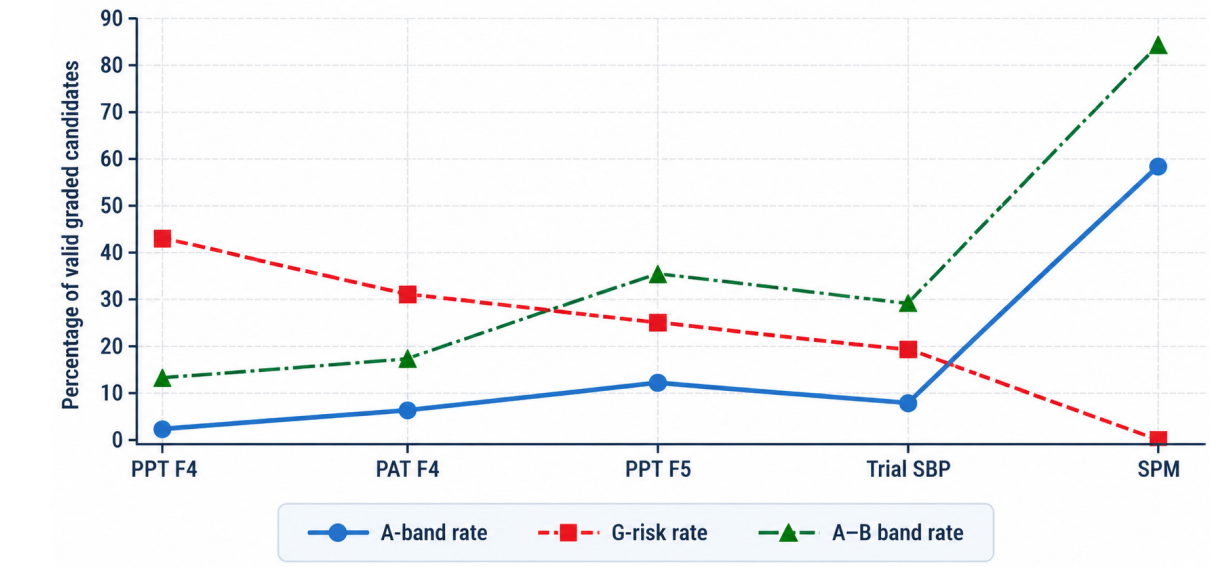


FIGURE 3. 2025 Gladius movement from G-risk to A-band and A-B band achievement

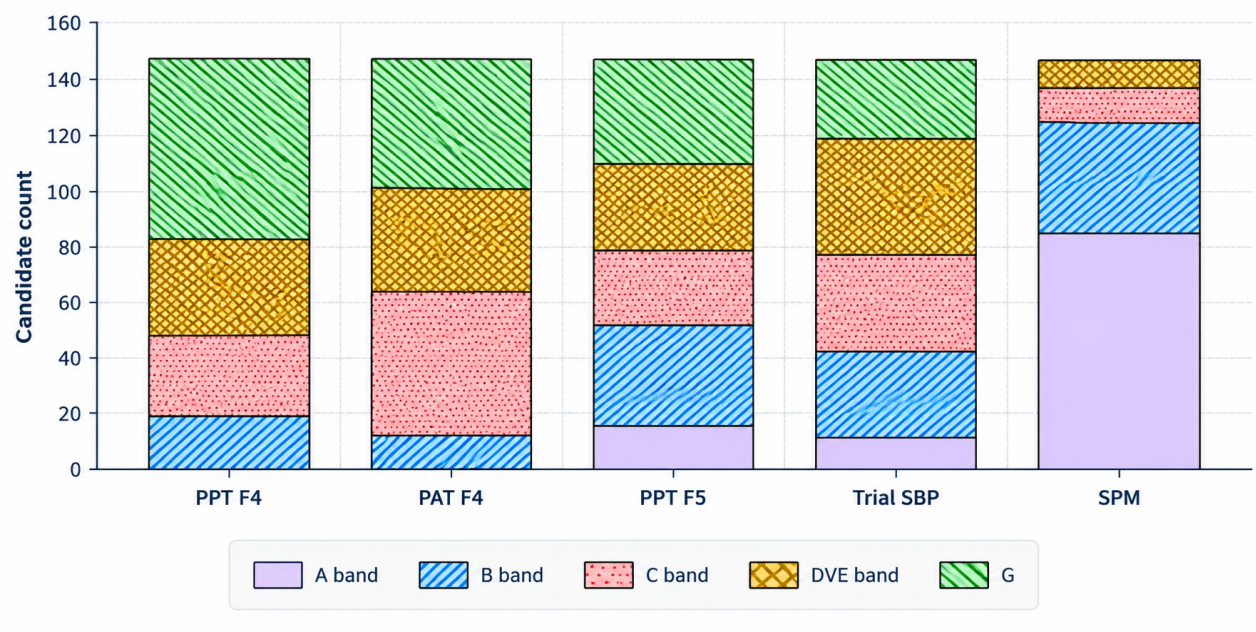


FIGURE 4. Grade-band redistribution for the 2025 Gladius cohort

The 2025 Gladius cohort demonstrated a substantial movement from examination risk to stronger SPM attainment. Its GPK improved from 7.22 at PPT F4 to 2.46 at SPM, the pass rate increased from 56.85% to 100.00%, G-risk fell from 43.15% to 0.00%, A-band representation rose from 1.37% to 57.82%, and the A-B band expanded from 13.01% to 84.35%.

H_{03} is rejected and H_{A3} is accepted. The magnitude and direction of change show a strong practical transformation. The distributional tests reported under RQ2 further support the conclusion that the Gladius movement was not a minor fluctuation across adjacent categories.

The Gladius outcome is educationally important because it shows dual movement. A purely remedial strategy might have reduced failure but would not necessarily have expanded distinction-level outcomes. Conversely, a purely enrichment strategy might have improved the top band while leaving weaker learners behind. The observed shift supports the instructional logic of combining diagnostic evidence, tiered remediation, A-band enrichment, retrieval practice, inquiry-based explanation and feedback cycles.

This pattern is consistent with differentiated instruction research showing that learning gains are more likely when support is structured, responsive and aligned with achievement evidence (AM et al. 2023; Kahmann et al. 2022; Langelaan et al. 2024). It also supports formative assessment literature, which argues that assessment becomes educationally powerful only when evidence changes subsequent teaching and learner action (Dayal

2021; Dini et al. 2020; Sortwell et al. 2024; Yao et al. 2024). In the present study, the plausible mechanism was not a single teaching activity but the accumulation of connected decisions: diagnostics made risk visible, differentiation matched support to readiness, retrieval and inquiry strengthened conceptual explanation, and feedback converted errors into specific next steps. This helps explain why the reduction of G-risk and the expansion of A-band attainment occurred together.

RESEARCH QUESTION 4

RQ4 asked what the 2026 Audrea evidence suggests about the sustainability of the data-rich differentiated pedagogy cycle. The Audrea cohort should not be judged as a completed outcome because Trial SBP and SPM data were not yet available. Nevertheless, the interim signal is favourable. Compared with Gladius at matching stages, Audrea recorded a GPMP/GPK advantage of 1.38 points at PAT T4/PAT F4 and 1.11 points at PPT F5. Its G-risk also decreased and its A-band representation increased between the two available assessment points.

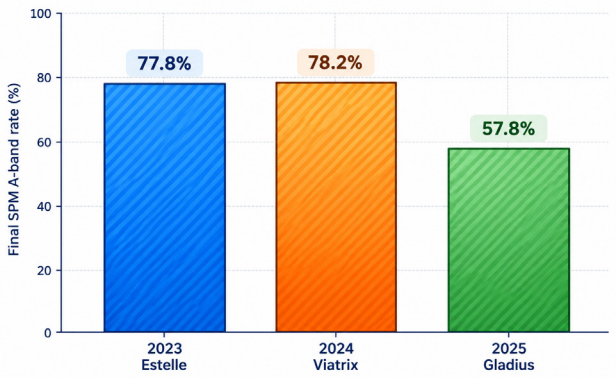


FIGURE 5. Final SPM A-band rates across the three completed cohorts

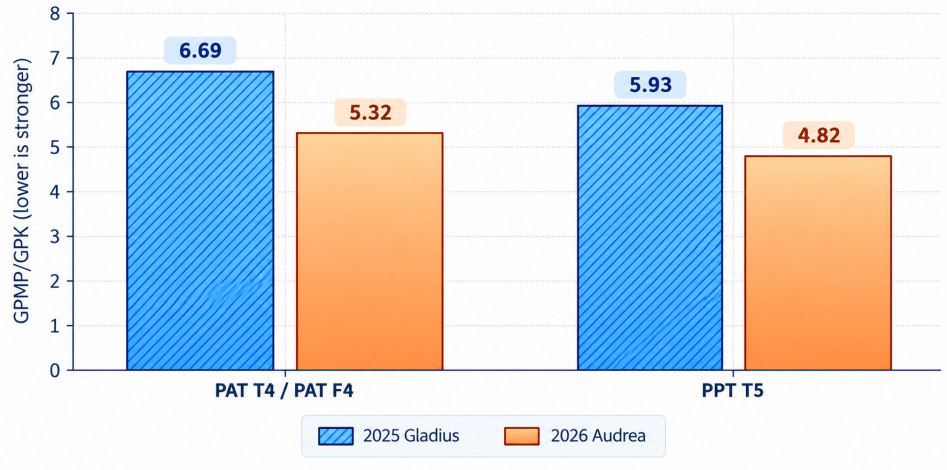


FIGURE 6. Interim 2026 Audrea GPMP/GPK compared with the 2025 Gladius cohort

A cautious scenario projection was also computed using historical calibration from the completed cohorts. The projection estimated Audrea's Trial SBP GPMP/GPK at 4.68 and SPM 2026 GPMP/GPK at 1.68. This is shown in Figure 7. The projection is not a confirmed outcome; it is a planning scenario that helps school leaders and Biology panels monitor whether the cohort remains on track.

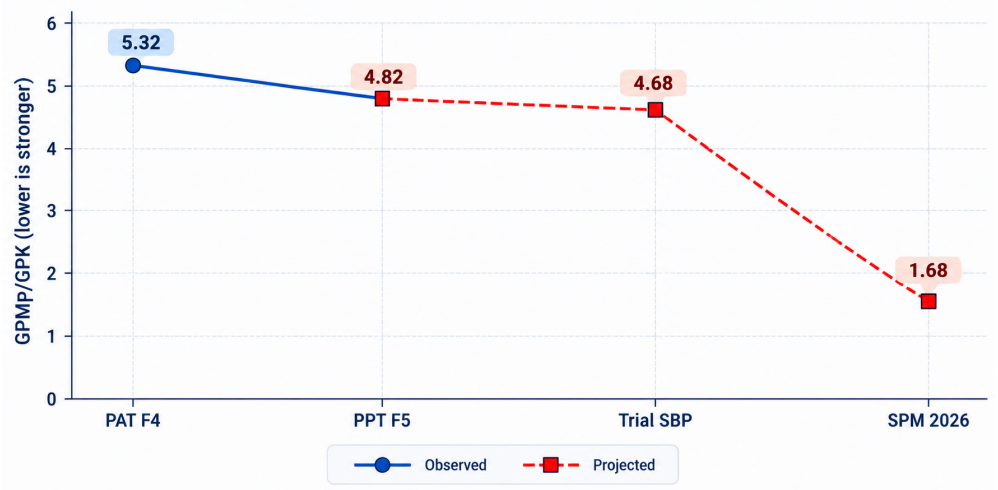


FIGURE 7. Observed and projected GPMP/GPK trajectory for the 2026 Audrea cohort

The Audrea evidence suggests early sustainability rather than final effectiveness. The available interim indicators move in the expected direction: GPMP/GPK improved, G-risk decreased, pass rate increased and A-band representation expanded. In addition, Audrea performed better than Gladius at comparable early stages. However, the interim collapsed-band chi-square result was not statistically significant, and the final Trial SBP and SPM outcomes are still required.

The Audrea pattern should therefore be interpreted as a monitoring signal rather than proof of completed impact. Its early advantage over Gladius at comparable points indicates that the IDDBP cycle may have become more stable and better institutionalised, but the non-significant interim chi-square result requires caution. This interpretation is consistent with action research principles because each new dataset is used to guide the next instructional cycle rather than to make premature causal claims.

H_{04} is rejected at the descriptive and practical-monitoring level, and H_{A4} is cautiously accepted as an interim sustainability claim. However, the hypothesis is not treated as fully confirmed at the inferential level because the interim chi-square result for Audrea was non-significant, chi-square(4) = 3.04, p = .551, V = .102. The responsible conclusion is that Audrea is on a promising trajectory, but final effectiveness must be verified when Trial SBP and SPM 2026 results become available.

SUMMARY OF RESEARCH QUESTIONS AND HYPOTHESIS DECISIONS

TABLE 6. Direct Answers to Research Questions and Hypothesis Decisions

RQ	Direct answer	Null hypothesis decision	Alternative hypothesis decision
RQ1	GPMP/GPK, pass rate, G-risk, A-band and A-B band moved favourably across the cycle, especially in completed cohorts.	H_{01} rejected	H_{A1} accepted
RQ2	Complete cohorts showed statistically significant full-grade and collapsed-band redistribution; repeated GPMP/GPK trajectory was also significant.	H_{02} rejected	H_{A2} accepted
RQ3	Gladius 2025 moved strongly from examination risk to SPM attainment: G-risk 43.15% to 0.00%, A-band 1.37% to 57.82%, and GPK 7.22 to 2.46.	H_{03} rejected	H_{A3} accepted
RQ4	Audrea 2026 shows promising interim sustainability but not final effectiveness; inferential confirmation awaits Trial SBP and SPM data.	H_{04} rejected descriptively but not fully at inferential level	H_{A4} cautiously accepted

PRACTICAL IMPLICATIONS

Table 7 translates the findings into a practical dashboard that schools can adapt after major assessments, during professional learning community sessions or before Dialog Prestasi discussions. The dashboard is deliberately simple because an unsustainable model is unlikely to shape classroom learning meaningfully.

For replication, Biology panels should keep a brief intervention log after each assessment, recording the diagnostic issue, the learner group targeted, the differentiated strategy applied, the evidence collected and the next decision made. This log would make the action research cycle visible and would allow future studies to link quantitative outcomes with classroom implementation fidelity.

TABLE 7. Practical Biology Action-Research Dashboard for Future Implementation

Dashboard element	Guiding question	Action trigger	Suggested response
GPK/GPMP	Is overall grade quality improving?	GPK stagnates or increases.	Revisit topic weaknesses and intensify response practice.
G-risk count	How many candidates remain at failure risk?	G count remains high after one cycle.	Run targeted remediation clinics and micro-quizzes.
A-band count	Are potential distinction candidates being stretched?	A-band growth remains flat.	Introduce higher-order data interpretation and mechanism-writing tasks.

Dashboard element	Guiding question	Action trigger	Suggested response
A-B band count	Is the middle distribution moving upward?	B and C bands do not convert upward.	Use model answers, structured peer review and criteria-based correction.
Topic diagnostics	Which Biology topics produce repeated errors?	Same misconception appears across classes.	Re-teach using diagrams, analogy, retrieval and application questions.
Feedback uptake	Do students act on feedback?	Corrections are submitted but errors repeat.	Add self-checklists and require revision commentary.

CONCLUSION

The main objective of this study was to determine how routine examination evidence could be transformed into diagnostic action, differentiated support and measurable improvement. The main findings answer all four research questions by showing improved GPMP/GPK and grade-band outcomes across the completed cohorts, with all three completed cohorts achieving zero SPM failure and statistically significant grade redistribution. The 2025 Gladius cohort showed the strongest transformation, while the 2026 Audrea cohort provides a promising but still cautious early indication of sustainability pending Trial SBP and SPM 2026 results.

The study concludes that aggregate school examination records can become a rigorous source of pedagogical improvement when they are reconstructed carefully, analysed transparently and interpreted with methodological caution. The proposed dashboard can support earlier risk detection, targeted remediation, enrichment for high-potential learners and structured feedback cycles.

The study is limited by its aggregate, non-randomised design, absence of individual-level student trajectories and incomplete Audrea data. Future research should use anonymised longitudinal student-level data, topic-level diagnostics, qualitative reflections and matched comparison designs where ethically feasible.

REFERENCES

Acut, D.P. & Antonio, R.P. 2023. Effectiveness of Science-Technology-Society (STS) approach on students' learning outcomes in science education: Evidence from a meta-analysis. *Journal of Technology and Science Education* 13(3): 718-739. https://doi.org/10.3926/jotse.2151

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

Arifin, Z., Sukarmin, Saputro, S. & Kamari, A. 2025. The effect of inquiry-based learning on students' critical thinking skills in science education: A systematic review and meta-analysis. *Eurasia Journal of Mathematics, Science and Technology Education* 21(3): Article em2592. https://doi.org/10.29333/ejmste/15988

Cortes, S.T., Lorca, A.S., Pineda, H.A., Tubog, R. & Vilbar, A. 2024. Strengthening science education in basic education through a professional development program on participatory action research for science teachers. *Social Sciences & Humanities Open* 10: Article 101194. https://doi.org/10.1016/j.ssaho.2024.101194

Dayal, H. 2021. How teachers use formative assessment strategies during teaching: Evidence from the classroom. *Australian Journal of Teacher Education* 46(7): 1-21. https://doi.org/10.14221/ajte.2021v46n7.1

Dini, V., Sevian, H., Caushi, K. & Orduna Picon, R. 2020. Characterizing the formative assessment enactment of experienced science teachers. *Science Education* 104(2): 290-325. https://doi.org/10.1002/sce.21559

Double, K.S., McGrane, J.A. & Hopfenbeck, T.N. 2020. The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review* 32: 481-509. https://doi.org/10.1007/s10648-019-09510-3

Driessen, E.P., Knight, J.K., Smith, M.K. & Ballen, C.J. 2020. Demystifying the meaning of active learning in postsecondary biology education. *CBE-Life Sciences Education* 19(4): Article ar52. https://doi.org/10.1187/cbe.20-04-0068

Esparza, D., Wagler, A.E. & Olimpo, J.T. 2022. Empowering current and future educators: Using a scalable action research module as a mechanism to promote high-quality teaching and learning in STEM. *Frontiers in Education* 6: Article 754097. https://doi.org/10.3389/feduc.2021.754097

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* 111(23): 8410-8415. https://doi.org/10.1073/pnas.1319030111

Hsu, H.P. 2023. A case study of a secondary biology teacher's pedagogical reasoning and actions with augmented reality technology. *Education Sciences* 13(11): Article 1080. https://doi.org/10.3390/educsci13111080

Ilgaz, G., Vural, L. & Bagislayan, B. 2024. Action research in science education: A systematic review. *Journal for the Education of Gifted Young Scientists* 12(4):

193-203. https://doi.org/10.17478/jegys.1554538

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

Kamarudin, M.Z. & Mat Noor, M.S.A. 2024. What do we know about the selection of action research methodologies in primary science education? A systematic literature review. *Educational Action Research* 32(5): 825-847. https://doi.org/10.1080/09650792.2023.2261502

Kemmis, S., McTaggart, R. & Nixon, R. 2014. *The action research planner: Doing critical participatory action research*. Singapore: Springer. https://doi.org/10.1007/978-981-4560-67-2

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

Lipnevich, A.A. & Panadero, E. 2021. A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education* 6: Article 720195. https://doi.org/10.3389/feduc.2021.720195

Martella, A.M., Martella, R.C., Yatcilla, J.K., Newson, A., Shannon, E.N. & Voorhis, C. 2023. How rigorous is active learning research in STEM education? An examination of key internal validity controls in intervention studies. *Educational Psychology Review* 35: Article 107. https://doi.org/10.1007/s10648-023-09826-1

Mejeh, M., Wyss, C. & Schmitz, B. 2022. An intervention study to promote self-regulated learning in vocational schools. *Vocations and Learning* 15: 531-568. https://doi.org/10.1007/s12186-022-09298-4

Ozturk, B., Kaya, M. & Demir, M. 2022. Does inquiry-based learning model improve learning outcomes? A second-order meta-analysis. *Journal of Pedagogical Research* 6(4): 201-216. https://doi.org/10.33902/JPR.202217481

Panadero, E. & Lipnevich, A.A. 2022. A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review* 35: Article 100416. https://doi.org/10.1016/j.edurev.2021.100416

Pozas, M., Letzel, V. & Schneider, C. 2020. Teachers and differentiated instruction: Exploring differentiation practices to address student diversity. *Journal of Research in Special Educational Needs* 20(3): 217-230. https://doi.org/10.1111/1471-3802.12481

Pozas, M., Letzel, V., Lindner, K.-T. & Schwab, S. 2021. DI (Differentiated Instruction) does matter! The effects of DI on secondary school students' well-being, social inclusion and academic self-concept. *Frontiers in Education* 6: Article 729027. https://doi.org/10.3389/feduc.2021.729027

Pratiwi, E.D., Masykuri, M. & Ramli, M. 2021. Active learning strategy on higher education biology learning: A systematic review. *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah* 6(1): 75-86. https://doi.org/10.24042/tadris.v6i1.7345

Sangur, K., Zubaidah, S. & Sulisetijono. 2025. A systematic literature review of mobile learning trends in biology education over ten years. *Social Sciences & Humanities Open* 11: Article 101429. https://doi.org/10.1016/j.ssaho.2025.101429

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

Theobald, E.J., Hill, M.J., Tran, E., Agrawal, S., Arroyo, E.N., Behling, S., Chambwe, N., Cintron, 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. et al. 2020. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences* 117(12): 6476-6483. https://doi.org/10.1073/pnas.1916903117

Wang, M., Yang, M. & Kyle, W.C. Jr. 2023. Effect of retrieval practice and drawing on high school students' conceptual understanding of the carbon cycle. *Disciplinary and Interdisciplinary Science Education Research* 5: Article 15. https://doi.org/10.1186/s43031-023-00083-4

Xu, Z., Wijekumar, K., Ramirez, G., Hu, X. & Irey, R. 2023. A meta-analysis of the efficacy of self-regulated learning interventions on academic achievement in online and blended environments. *Behaviour & Information Technology* 42(16): 2911-2931. https://doi.org/10.1080/0144929X.2022.2151935

Yan, Z., Chiu, M.M. & Ko, P.Y. 2020. Effects of self-assessment diaries on academic achievement, self-regulation, and motivation. *Assessment in Education: Principles, Policy & Practice* 27(5): 562-583. https://doi.org/10.1080/0969594X.2020.1827221

Yan, Z., Lao, H., Panadero, E., Fernandez-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

Yao, Y., Amos, M., Snider, K. & Brown, T. 2024. The impact of formative assessment on K-12 learning: A meta-analysis. *Educational Research and Evaluation* 29(7-8): 452-475. https://doi.org/10.1080/13803611.2024.2363831

Zheng, X., Luo, L. & Liu, C. 2022. Facilitating undergraduates' online self-regulated learning: The role of teacher feedback. *The Asia-Pacific Education Researcher*. Advance online publication. https://doi.org/10.1007/s40299-022-00697-8

APPENDIX A

List of abbreviations and symbols used in the manuscript.

TABLE 8. List of Abbreviations and Symbols

Abbreviation/Symbol	Meaning in this manuscript
A-band	Combined distinction category comprising A+, A and A- grades.
A-B band	Combined high-achievement category comprising A+, A, A-, B+ and B grades.
ETR	Expected Target Result used by the school as an achievement benchmark.
F4 / F5	Form 4 and Form 5 in the Malaysian upper secondary context.
G-risk	Proportion or count of candidates in the G category.
GPK / GPMP	Grade-point average indicators where lower values indicate stronger grade quality.
IDDBP	Iterative Data-Rich Differentiated Biology Pedagogy.
PAT / PPT	End-of-year and mid-year examinations.
SBP	Sekolah Berasrama Penuh, or fully residential school context.
SPM	Sijil Pelajaran Malaysia, the national secondary school leaving examination.
V / W	Cramer's V and Kendall's W effect-size indices.
chi-square	Statistic used to test distributional differences.
Delta	Change score between two assessment indicators.

Norshamsiah Samsudin*
Tunku Kurshiah College,
Bandar Enstek, 71760 Seremban, Negeri Sembilan, Malaysia
E-mail: norshamsiah@tkc.edu.my

Fathiiyah Mohd Pozi
Tunku Kurshiah College,
Bandar Enstek, 71760 Seremban, Negeri Sembilan, Malaysia
E-mail: pengetua.ktk@moe.gov.my

Shaharudin Jakpar
Faculty of Economics, Business and Accounting,
i-CATS University College, Jalan Stampin Timur, 93350 Kuching, Sarawak, Malaysia
E-mail: shaharudin@icats.edu.my

*Corresponding Author: Norshamsiah Samsudin

Dihantar: 14 Jun 2026
Dinilai: 21 Julai 2026
Diterima: 30 Julai 2026
Diterbitkan: 30 November 2026