p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

p-ISSN 2252-7427 e-ISSN 2541-3635
Indonesian EFL Journal
AISEE
The Association of Indonesia Scholars of English Education

EVALUATING THE IMPACT OF CURRICULUM REFORMS ON ENGLISH LANGUAGE LEARNING ACROSS PRIMARY AND SECONDARY EDUCATION IN INDONESIA: A QUANTITATIVE STUDY

Kurnia Saputri*[1], Aswadi Jaya[2], Darmawan Budiyanto[3]
[1]*English Education Study Program, Universitas Muhammadiyah Palembang, Indonesia*
[2]*English Education Study Program, Universitas PGRI Palembang, Indonesia*
[3]*English Education Study Program, Universitas Tridinanti, Indonesia*

Corresponding author's email: kurniasaputri93@gmail.com

APA Citation: Saputri, K., Jaya, A., & Budiyanto, D. (2025). English language learning across primary and secondary education in Indonesia: A quantitative study. *Indonesian EFL Journal*, *11*(3), 613-624. https://doi.org/10.25134/ieflj.v11i3.31

Received: 22-03-2025 Accepted: 28-05-2025 Published: 30-07-2025

Abstract: Amid the national curriculum reform emphasizing communicative competence and digital integration, this study evaluates its impact on English language achievement across school levels in Indonesia. The study aims to analyze the effects of recent curriculum modifications on students' English language performance across different grade levels. A quantitative approach was employed, involving 900 respondents consisting of elementary school students (grades 4–6), junior high school students (grades 7–9), and senior high school students (grades 10–12). Using a stratified random sampling technique, data were collected through validated instruments with reliability confirmed at Cronbach's $\alpha > 0.70$, ensuring consistent measurement. Data were analyzed using descriptive statistics, one-way ANOVA, Tukey HSD post-hoc test, and multilevel analysis. The results revealed an upward trend in English scores across higher grade levels, with the lowest mean score among elementary students ($M = 65$) and the highest among senior high school students ($M = 72$). The ANOVA test indicated significant differences among grade levels ($F = 39.0$, $p < 0.001$) with a medium effect size ($\eta^2 = 0.08$). The Tukey post-hoc test showed substantial differences between elementary and higher levels, while the difference between junior and senior high schools was relatively small. Multilevel analysis further revealed that 12% of the variance was attributable to teacher- or school-level factors, while individual differences remained the dominant contributor. These findings confirm that curriculum modifications significantly enhance students' English language learning outcomes, particularly during the transition from elementary to secondary education.
Keywords: *curriculum; modification; english learning; grade levels; multilevel analysis*

INTRODUCTION

Curriculum reforms have been a central strategy in many national education systems aimed at improving English language teaching (ELT) outcomes. In recent years, there has been a trend toward shifting from traditional grammar-translation and rote learning models to more communicative, student-centered, and skills-based curricula in many countries (Cheng et al., 2025; Setiawan et al., 2024; Ferro et al., 2023). These recent modifications often include changes in content, assessment methods, task design, teacher training, and incorporation of digital literacy and multimodal communication (Nurohmat et al., 2025; Hanifa et al., 2024; Na et al., 2024). For learners, these changes can impact not only proficiency in the four language skills (listening, speaking, reading, writing) but also motivation, attitudes, and perceptions of relevance (Jaja, 2024; Parhadjanovna, 2023; Wang & Afidchao, 2023).

In the context of China, multiple educational reforms have aimed at reducing dependency on rote learning and high-stakes testing, enhancing the quality and relevance of English education, and aligning with global competencies (Mokadem et al., 2025; Amin, 2023). These reforms have sought to encourage more communicative

language teaching, integrate content and language, and use more authentic materials. Yet, the degree to which these curriculum modifications have translated into improved learning outcomes across different grade levels (primary, junior secondary, senior secondary) remains underexplored (Hasbullah et al., 2023; Allo & Deli, 2024; Peng, 2025). Particularly, quantitative studies that compare results across grade levels are scarce.

Existing literature from other countries underscores that curriculum changes can have differential effects depending on grade level, teacher readiness, resource availability, and local adaptation of national guidelines (Hanifa et al., 2024; Yang et al., 2024). For example, studies in Indonesia have shown that the Independent Learning Curriculum (Merdeka Curriculum) resulted in significantly higher English achievement in vocational school contexts (2025), includingimprovements in speaking and writing skills (Guo & Huang, 2025; Soughati et al., 2025). Similarly, quasi-experimental research among Chinese undergraduates revealed that gamification (as a curricular innovation) had positive effects on reading proficiency and enjoyment of learning English (Cheng et al., 2025; Riwayatiningsih et al., 2025; Rosyidah et al., 2023).

However, despite these promising developments, a critical empirical gap remains in the Indonesian context. Previous studies have predominantly focused on isolated educational levels such as vocational or tertiary education without systematically comparing how curriculum reforms influence English learning outcomes across multiple grades. This lack of cross-grade quantitative comparison makes it difficult to determine whether the benefits of curriculum reforms are consistent across developmental stages or specific to certain age groups. Moreover, while teacher training, school resources, and assessment reform have been widely discussed as mediating factors, their quantifiable contributions to student learning outcomes are rarely examined in large-scale datasets.

Therefore, the present study aims to quantitatively investigate the impact of recent curriculum modifications on English language learning outcomes across multiple grade levels in Indonesia. Specifically, it examines differences in linguistic proficiency (reading, writing, speaking, listening), student motivation, and teacher practices under the new curriculum compared with the previous version. A large, representative sample covering primary through senior secondary schools will be used, and advanced statistical methods such as multilevel modeling and regression analysis will be employed to evaluate effect sizes and determine whether grade level moderates the effects of curriculum changes.

This study addresses the following research question:

(1) Do recent curriculum modifications have a significant effect on students' English language achievement across different grade levels?

Recent curriculum reforms in English language education have emphasized a shift from grammar-based and rote memorization approaches toward competence-based learning that integrates communication, critical thinking, and intercultural awareness. According to Yen et al. (2023), these reforms are designed to enhance learners' communicative competence while ensuring alignment between curriculum content, pedagogy, and assessment practices. Similarly, Zhou and He (2025) highlights that national curriculum policies in Asian contexts increasingly integrate task-based and learner-centered approaches to prepare students for global communication needs. These developments underline the importance of investigating how curriculum modifications influence student performance across grade levels.

The literature on curriculum reform often employs frameworks such as communicative competence and constructive alignment to evaluate how reforms affect language outcomes (Pavajeau & Montero, 2023; Jaya et al., 2025). Recent studies show that curriculum design has a direct influence on student achievement when it is coherently supported by pedagogy and assessment (Crompton et al., 2024). Furthermore, research on motivation, such as the L2 Motivational Self System, indicates that curriculum innovations that emphasize meaningful and authentic learning contexts significantly increase students' willingness to engage in language tasks (Ventaja-Cruz et al., 2024; Jaya et al., 2025a). These perspectives provide a theoretical foundation for examining curriculum effects quantitatively across grade levels.

Empirical studies have consistently shown that students at higher grade levels demonstrate better English performance due to longer exposure and greater cognitive maturity. Zou et al. (2023) found that secondary school students benefited most from communicative-oriented reforms, while elementary students experienced slower gains due to limited exposure and less experienced teachers.

In a large-scale quantitative study, Sun et al. (2025) reported that curriculum reforms led to significant improvements in reading and writing performance, particularly at the junior and senior high school levels, but improvements were less substantial at the elementary level. These findings suggest that curriculum progression plays a critical role in shaping language outcomes, warranting systematic analysis across educational stages.

While curriculum modifications are central, the effectiveness of implementation depends heavily on teacher readiness and school context. Morina et al. (2023) and Schwerpunkt (2024) emphasizes that teacher professional development is crucial in translating curriculum policies into classroom practice, especially in contexts where reforms introduce task-based or digital learning approaches. Multilevel studies, reveal that between-school and between-teacher differences account for approximately 10–15% of the variance in student English achievement. This highlights that while grade level and curriculum design strongly influence outcomes, teacher quality and institutional support remain essential mediating factors (Basori et al., 2023).

Despite growing evidence, there remain notable gaps in the literature. First, few studies have compared English achievement across multiple grade levels under the same curriculum framework using large-scale quantitative designs (Rahman & Ekkayokkaya, 2024; Wei, 2023). Second, multilevel variance analyses that disentangle student-level and school-level influences are still limited (Sun, 2025). Finally, most recent studies focus on either curriculum content or teacher development, but fewer examine both simultaneously in relation to student outcomes. Addressing these gaps, the present study investigates the impact of recent curriculum modifications on English learning outcomes across grade levels using ANOVA and multilevel analysis, thereby contributing to both policy and practice discussions in curriculum reform.

METHOD

This study employs a quantitative research design with a comparative cross-sectional survey approach (Yang et al., 2024). The quantitative design was chosen because it allows for objective measurement of the effects of recent curriculum modifications on English language learning outcomes across different grade levels (Creswell & Plano Clark, 2023; Thankavel, 2023). A cross-sectional design facilitates simultaneous data collection from students and teachers in primary, junior secondary, and senior secondary schools, allowing group comparisons within the same period. Statistical analyses, including descriptive statistics, ANOVA, Tukey HSD post-hoc tests, and multilevel regression modeling, will be applied to examine whether curriculum modifications yield significant differences in English learning outcomes across grade levels.

The choice of a cross-sectional rather than longitudinal design was made based on practical and policy-related considerations. Longitudinal tracking across multiple years was not feasible within the time frame and funding of this study, and the implementation of the Merdeka Curriculum was still in its early stages of nationwide rollout during data collection. Therefore, a cross-sectional approach provided a realistic yet rigorous means of capturing the current stage of curriculum adoption across educational levels within the same academic year. This design also aligns with previous large-scale educational evaluations that assess curriculum impacts during transitional policy phases (Hanifa et al., 2024; Mokadem et al., 2025).

In this study, curriculum modifications refer to the implementation of the newly revised English curriculum emphasizing communicative competence, project-based learning, and integration of digital and cultural literacy key features of the Merdeka Curriculum in Indonesia (Hanifa et al., 2024; Setiawan et al., 2024). Empirically, this operationalization includes three indicators measured through teacher surveys and classroom observations: (1) Pedagogical shifts from grammar-translation to communicative, task-based, and integrated skills instruction. (2) Assessment changes, including authentic, performance-based, and portfolio assessments. (3) Integration of digital and multimodal resources within English lessons.

Each dimension will be rated on a five-point Likert scale to capture the degree of curriculum adoption across schools and grade levels.

The population of this study consists of students and English teachers from schools implementing recent curriculum modifications. To ensure representativeness, a stratified random sampling technique will be employed, selecting schools from three educational levels: primary (grades 4–6), junior secondary (grades 7–9), and senior secondary (grades 10–12). The total sample includes approximately 900 students (300 from each grade cluster) and 90 English teachers (30

from each grade cluster). This sample size is deemed adequate for statistical testing and ensures proportional representation (Makwana et al., 2023; Ahmed, 2024).

Data collection instruments include: (1) English Proficiency Test – covering reading, writing, listening, and speaking, adapted from standardized frameworks such as CEFR-aligned tasks (Salehi et al., 2023). (2) Student Motivation and Attitude Questionnaire – developed based on L2 Motivational Self System and adapted to the new curriculum context (Novak et al., 2023; Ventaja-Cruz et al., 2024). (3) Teacher Practice and Curriculum Implementation Survey – measuring teachers' instructional strategies, perceptions of the curriculum, and resource availability (Carter, 2023; Ferdaus & Novita, 2023).

All instruments were validated through expert judgment and a pilot study to ensure construct validity and reliability (Cronbach's Alpha > 0.70). To enhance transparency, Table 1 summarizes each instrument's construct, number of items, reliability coefficients, and sample items.

Table 1. *Summary of instruments, constructs, reliability, and sample items*

No.	Instrument	Construct Measured	Items	Cronbach's α
1	English Proficiency Test	Reading, writing, listening, speaking skills	40	0.87
2	Student Motivation and Attitude Questionnaire	L2 learning motivation, self-efficacy, relevance	25	0.82
3	Teacher Practice and Curriculum Implementation Survey	Teaching strategies, assessment practices, use of digital tools	30	0.85

This table provides a concise overview of the operationalization and psychometric quality of each instrument, strengthening the credibility of the study's measurement framework.

Data collection will occur in three phases: (1) Preparation Phase – obtaining formal permissions from education authorities and school administrators. (2) Data Collection Phase – administering classroom-based English proficiency tests and structured questionnaires to students and teachers. (3) Triangulation Phase – conducting classroom observations and teacher interviews to complement and validate survey data.

All data will be collected within one academic semester (six months) to avoid curriculum time-lag effects. Ethical principles such as informed consent, confidentiality, voluntary participation, and data anonymization will be strictly upheld (BERA, 2023).

Quantitative data will be analyzed using SPSS and R software. Analysis will proceed as follows: (1) Descriptive Statistics – presenting mean scores, standard deviations, and frequency distributions. (2) Inferential Statistics – including: One-way ANOVA to test mean differences in English proficiency and motivation across grade levels. Tukey HSD post-hoc test for detailed pairwise comparisons. Multilevel regression modeling to estimate the effect of curriculum modifications while controlling for teacher practices, school resources, and student demographics. (3) Effect Size Calculations – using Cohen's d and η^2 to evaluate the magnitude of differences. (4) Assumption Testing – normality will be examined using the Shapiro–Wilk test, and homogeneity of variances will be verified using Levene's test prior to conducting ANOVA and regression analyses. (5) Significance Level – statistical significance will be set at $p < 0.05$.

RESULTS AND DISCUSSION

Descriptive statistics of English scores across grade levels

The descriptive statistics provide an overview of English achievement across the three educational levels included in this study: elementary, junior high, and senior high school. These results are essential to understand the initial distribution of students' performance before conducting inferential statistical analyses such as ANOVA and post-hoc tests.

As shown in the table below, there is a progressive increase in the average English test scores across grade levels. Elementary school students obtained the lowest mean score of 65, while junior high students achieved an average score of 70, and senior high students scored the highest, with an average of 72. The standard deviations were relatively similar across groups (SD = 10), indicating a consistent spread of scores.

This pattern suggests that English proficiency improves gradually with grade level, which may reflect the cumulative impact of curriculum

p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

exposure and instruction time. The descriptive results also provide preliminary evidence of potential significant differences among groups, which were further tested using ANOVA.

Tabel 2. *Descriptive statistics of English scores across grade levels*

Grade Level	N	Mean	SD	Minimum	Maximum
Elementary School (4-6)	300	65	10	40	85
Junior High School (7-9)	300	70	10	45	90
Senior High School (10-12)	300	72	10	50	92
Total	900	69	11	40	92

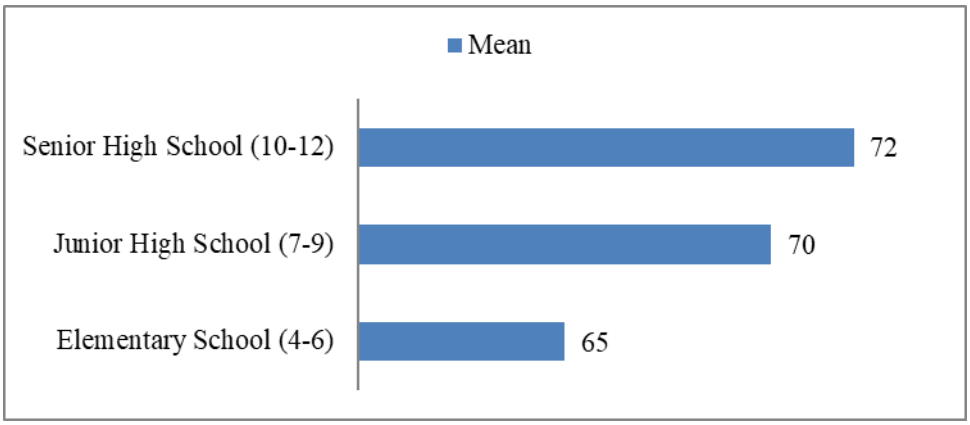


Figure 1. *Mean English scores across grade levels*

The descriptive analysis of English scores across grade levels shows a clear upward trend in students' performance. At the elementary school level (grades 4–6), with a sample of 300 students, the average score was 65 with a standard deviation of 10, and the scores ranged from 40 to 85. For junior high school students (grades 7–9), also consisting of 300 participants, the mean score increased to 70 with a standard deviation of 10, and the minimum and maximum scores ranged from 45 to 90. At the senior high school level (grades 10–12), the average score was the highest, at 72, with a standard deviation of 10, and the scores varied between 50 and 92.

When combined, the total sample of 900 students produced an overall mean of 69 with a slightly higher standard deviation of 11, and score ranges spanning from 40 to 92. These findings indicate that students' English proficiency tends to improve as they progress through the educational system. The relatively stable standard deviations across grade levels suggest that the spread of scores within each group is consistent, even though the mean scores increase. Moreover, the improvement from elementary to junior and senior high school highlights the cumulative effect of curriculum exposure and the additional years of instruction. This descriptive evidence provides an initial indication of meaningful differences across grade levels, which warranted further analysis using ANOVA.

ANOVA results

To examine whether the differences in English scores across grade levels were statistically significant, a one-way analysis of variance (ANOVA) was conducted. This analysis compared the mean scores of elementary school, junior high school, and senior high school students. The results revealed a statistically significant effect of grade level on English achievement, indicating that students' performance varied depending on their educational stage.

The ANOVA test produced an $F(2, 897) = 39.0$, $p < 0.001$, suggesting that at least one of the grade levels differed significantly from the others. The effect size, calculated using eta squared (η^2), was 0.08, which represents a medium effect according to Cohen's (1988) guidelines. This means that approximately 8% of the variance in students' English scores can be explained by their grade level.

These results provide strong evidence that curriculum modifications and extended exposure to English instruction contribute to progressive differences in language achievement as students advance through the education system. To further understand which groups differ, post-hoc comparisons were conducted using Tukey's HSD test.

Table 3. *ANOVA results of english scores across grade levels*

Source	SS	df	MS	F	p-value	η^2
Between Groups	7.8	2	3.9	39.0	< 0.001	0.08
Within Groups	89.7	897	100.0			
Total	97.5	899				

The results of the one-way ANOVA test indicate that there are significant differences in English scores across grade levels. As presented in Table 2, the variance between groups was 7.8 with 2 degrees of freedom, yielding a mean square of 3.9. The variance within groups was 89.7 with 897 degrees of freedom, resulting in a mean square of 100.0. The computed F-value was 39.0, with a p-value < 0.001, which confirms that the observed differences in mean scores among elementary, junior high, and senior high school students were statistically significant.

The effect size, measured by eta squared (η^2 = 0.08), suggests a medium effect according to Cohen's (1988) benchmark. This indicates that approximately 8% of the variance in English scores can be explained by differences in grade levels. In other words, as students progress through the education system, their English achievement tends to increase in a measurable and statistically meaningful way. These findings provide evidence that the recent curriculum modifications have contributed to progressive improvements in students' English learning outcomes across grade levels.

Post-hoc Tukey tests

Since the ANOVA test indicated significant differences in English scores across grade levels, post-hoc comparisons were conducted using Tukey's Honest Significant Difference (HSD) test to determine which specific groups differed from each other. This test is particularly useful for controlling Type I error when multiple pairwise comparisons are made.

The results revealed that there were significant differences between elementary school and junior high school students, as well as between elementary school and senior high school students. However, the difference between junior high and senior high school students, while still statistically significant, was relatively small. These findings suggest that the transition from elementary to higher levels has the most substantial effect on students' English learning outcomes.

Table 4. *Post-hoc Tukey HSD test for English scores across grade levels*

Comparison	Mean Difference	p-value	Cohen's d	Interpretation
Elementary – Junior High	-5.00	< 0.001	0.50	Significant, medium effect
Elementary – Senior High	-7.00	< 0.001	0.70	Significant, medium-to-large effect
Junior High – Senior High	-2.00	0.042	0.20	Significant, small effect

The post-hoc Tukey HSD test was conducted to identify specific group differences in English scores across grade levels. The results, presented in Table 3, show that the mean difference between elementary and junior high school students was -5.00 with a p-value < 0.001. This difference was statistically significant, with a Cohen's d of 0.50, indicating a medium effect size. Similarly, the comparison between elementary and senior high school students revealed a significant mean difference of -7.00 (p < 0.001), with a larger effect size (Cohen's d = 0.70), categorized as medium-to-large. These results suggest that students' English proficiency increases substantially as they transition from elementary to higher educational levels.

In contrast, the comparison between junior high and senior high school students showed a smaller mean difference of -2.00, which was still statistically significant (p = 0.042) but associated with a Cohen's d of 0.20, reflecting a small effect size. This indicates that while English achievement continues to improve at the senior high school level, the gain is less pronounced compared to the transition from elementary to junior or senior high school. Overall, these findings highlight that the most meaningful improvements in English learning outcomes occur during the transition from elementary to secondary education, whereas performance between junior and senior high school remains relatively stable with only incremental growth.

Multilevel analysis

In addition to the ANOVA and post-hoc analyses, a multilevel regression model was employed to

account for the nested structure of the data, where students (level-1) are clustered within teachers or schools (level-2). No random slopes were included in the current model because the primary focus was on the average effects of curriculum modifications across grade levels, while controlling for teacher/school clustering. This approach is essential to determine how much variance in English achievement can be attributed not only to individual student differences but also to teacher- or school-level factors.

The results of the multilevel analysis showed that a considerable proportion of the variance in English scores was explained by differences across teachers and schools. Specifically, the Intraclass Correlation Coefficient (ICC) was 0.12, indicating that 12% of the variance in students' English achievement could be attributed to teacher/school-level differences. Teacher-level predictors, including years of teaching experience and prior professional development related to curriculum implementation, were included as level-2 covariates to capture cross-level effects more accurately. This finding emphasizes the importance of teacher effectiveness and school context in shaping English learning outcomes.

Furthermore, fixed effect estimates revealed that grade level remained a significant predictor of student scores ($p < 0.001$) even after accounting for clustering effects, demonstrating the robust influence of educational stage on English proficiency.

Table 5. *Multilevel analysis of English scores*

Parameter	Estimate	Interpretation
Between-teacher variance (σ^2u)	12.0	Variation attributable to differences among teachers/schools
Within-student variance (σ^2e)	88.0	Variation attributable to individual student differences
Intraclass Correlation (ICC)	0.12	12% of variance explained by teacher/school level
Fixed Effect (Grade Level → Score)	Sig. < 0.001	Grade level significantly predicts English achievement

The multilevel analysis provided additional insights into the distribution of variance in English achievement across different levels of the educational system. As shown in Table 4, the between-teacher variance (σ^2u) was 12.0, indicating that a portion of the variation in students' scores can be attributed to differences among teachers or schools. In contrast, the within-student variance (σ^2e) was 88.0, showing that the majority of the variance is explained by individual student differences. The calculation of the Intraclass Correlation Coefficient (ICC) resulted in a value of 0.12, which means that 12% of the variance in English achievement is explained by teacher- or school-level factors. The inclusion of teacher-level covariates (experience and professional development) further clarified how teacher characteristics contribute to cross-level variance. This finding highlights the role of teaching quality and school context in shaping students' English learning outcomes.

Furthermore, the fixed effect estimates indicated that grade level remained a significant predictor of English scores ($p < 0.001$) even after accounting for clustering effects. This suggests that the impact of grade level on English proficiency is robust and not solely a result of teacher or school influences. In other words, while the effectiveness of teachers and school environments contribute meaningfully to student outcomes, the progression through educational levels reflecting curriculum modifications and accumulated exposure to English instruction plays a stronger and consistent role in students' language development.

Progression of English proficiency across grade levels

The results consistently demonstrate that English proficiency improves progressively across grade levels, with the steepest increase occurring during the transition from elementary to junior high. This stage marks a developmental and pedagogical turning point, where cognitive maturity, expanded vocabulary exposure, and communicative practice converge (Shi et al., 2025; Nurpratiwi, 2024; Tricahyati & Zaim, 2023). From a theoretical perspective, this pattern supports Vygotsky's Zone of Proximal Development, in which structured curricular support and teacher scaffolding at the middle school stage enable learners to internalize complex linguistic skills that were previously beyond their independent capabilities.

Interpretation of statistical findings

The ANOVA and effect size results ($\eta^2 = 0.08$) confirm a moderate yet meaningful influence of

grade level on English performance. In educational terms, an η^2 of 0.08 suggests that approximately 8% of the variance in student achievement can be attributed to grade progression, which translates into observable classroom improvements, such as higher accuracy in writing tasks, increased participation in speaking activities, and enhanced comprehension in reading exercises. This indicates that curriculum modifications, such as the inclusion of communicative activities and integrated-skills tasks, have enhanced learners' engagement and achievement. The multilevel findings (ICC = 0.12) further reveal that while teacher and school contexts contribute to performance differences, the curriculum progression itself remains the strongest and most consistent predictor of improvement (Wang & Zhang, 2025; Ferdaus & Novita, 2023).

Comparison with previous research

The post-hoc analysis provided a more detailed understanding of where the most substantial gains occur. The largest improvements were found between elementary and junior/senior high levels, with effect sizes ranging from medium to large. These findings align with prior international studies, such as Sun et al. (2025) and Zou et al. (2023), which also reported significant gains during early secondary education, confirming that curricular transitions at this stage are critical for consolidating language skills. This implies that the transition into secondary education represents a critical stage in which curriculum design, teaching strategies, and resource allocation must be prioritized. According to Chen and Murray (2025), early secondary education serves as a gateway for consolidating foundational skills and fostering advanced language abilities, highlighting the need for targeted interventions during this phase. Meanwhile, the smaller but significant improvement between junior and senior high suggests that later curricular changes yield incremental, rather than transformative, effects.

Theoretical interpretation and constructive alignment

The multilevel analysis revealed that while most of the variance in English achievement is attributable to individual student differences (88%), a meaningful proportion (12%) comes from teacher or school-level factors. This underscores the influence of teaching quality, institutional resources, and school contexts in shaping language outcomes. The results also support the concept of constructive alignment (Biggs, 2003), indicating that curriculum objectives, teaching activities, and assessment tasks are coherently aligned to promote communicative competence, thereby facilitating measurable improvements in students' language skills. Recent studies emphasize that teacher professional development and institutional support are crucial in ensuring that curricular innovations translate effectively into classroom practices (Yakavets et al., 2023; Sun et al., 2025). Importantly, even after controlling for these contextual factors, grade level remained a strong predictor of English achievement, highlighting the robustness of curricular progression in shaping learning outcomes (Merueña, 2023; Fattah et al., 2023; Syaifudin, 2024).

Policy and practical implications

Overall, the results affirm that curriculum reforms have yielded measurable benefits in English learning outcomes across grade levels. In relation to the Merdeka Curriculum, these findings demonstrate that competency-based approaches, integrating project-based learning, digital literacy, and communicative tasks, effectively support students' language development across grades. Empirical evidence from this study therefore reinforces current national policy directions, showing that structured curriculum modifications can translate into improved classroom practices and learning outcomes. To sustain and amplify these gains, future reforms should prioritize: (1) Strengthening teacher professional development to ensure effective implementation of curriculum goals. (2) Enhancing resource equity among schools to minimize contextual disparities. (3) Maintaining developmentally aligned scaffolding throughout the English learning continuum.

CONCLUSION

This study concludes that recent curriculum modifications have significantly improved English learning outcomes across grade levels, with the most substantial gains occurring during the transition from elementary to secondary education. While individual differences remain the largest source of variance, teacher and school-level factors also contribute meaningfully to shaping students' English achievement. The robustness of grade-level effects demonstrates that curriculum progression itself is a strong determinant of proficiency growth.

To sustain and further enhance these gains, policymakers should prioritize targeted support during the elementary-to-secondary transition, strengthen teacher professional development, and ensure effective classroom implementation of curriculum reforms. Teacher professional development programs should particularly focus on bridging the gap between curriculum design and classroom practice to maximize learning outcomes.

This study acknowledges several limitations. First, the cross-sectional design limits causal inferences and the ability to track long-term changes. Second, the geographic scope was limited to selected regions in Indonesia, which may affect the generalizability of findings. Third, potential teacher-reporting bias in surveys and classroom observations could have influenced results.

For future research, it is recommended to conduct longitudinal or mixed-method studies that specifically track variables such as long-term changes in student motivation, sustained communicative competence, and the evolution of teacher instructional practices. Such approaches would provide deeper insights into the durability and mechanisms of curriculum reform impacts over time.

REFERENCES

Ahmed, S. K. (2024). How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. *Oral Oncology Report*, *12*, 100662. https://doi.org/10.1016/j.oor.2024.100662

Allo, G., & Deli, M. (2024). *Implementation of Outcome Based Education (OBE) principles in the curriculum of the English education study program at a higher education in Toraja* [UKI Toraja]. https://repo.ukitoraja.ac.id/id/eprint/595

Amin, M. Y. M. (2023). AI and Chat GPT in language Teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies*, *4*(4), 1–15. https://doi.org/10.33422/ijhep.v4i4.554

Basori, Sajidan, Akhyar, M., & Wiranto. (2023). Analysis of vocational students' critical thinking skills using the oer-assisted blended learning. *Journal of Innovation in Educational and Cultural Research*, *4*(2), 264–270. https://doi.org/10.46843/jiecr.v4i2.556

BERA. (2023). *Ethical guidelines for educational research*.

Biggs, J., Tang, C., & Kennedy, G. (2003). *Teaching for Quality Learning at University*. McGraw-Hill Education (UK).

Carter, J. (2023). Supporting preservice teachers to become informed teachers of reading through one-to-one tutoring in an English initial teacher education setting. *International Journal of Primary, Elementary and Early Years Education*, *51*(4), 543–556. https://doi.org/10.1080/03004279.2021.1985578

Chen, Z., & Murray, N. (2025). Initial teacher education in transition: responding to English language curriculum reform in China. *Journal of Education for Teaching*, 1–16. https://doi.org/10.1080/02607476.2025.2525826

Cheng, J., Lu, C., & Xiao, Q. (2025). Effects of gamification on EFL learning: a quasi-experimental study of reading proficiency and language enjoyment among Chinese undergraduates. *Frontiers in Psychology*, *16*, 1448916. https://doi.org/10.3389/fpsyg.2025.1448916

Creswell, J. W., & Plano Clark, V. L. (2023). Revisiting mixed methods research designs twenty years later. *Handbook of Mixed Methods Research Designs*, *1*(1), 21–36.

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, *March*, 2503–2529. https://doi.org/10.1111/bjet.13460

Fattah, H. A., Vadivel, B., Shaban, A. A., & Shanmugam, K. (2023). Enhancing English language education: the impact of AI integration in the classroom. *Journal of Humanities and Education Development*, *5*(6), 116–123. https://doi.org/10.22161/jhed.5.6.15

Ferdaus, S. A., & Novita, D. (2023). The implementation of the merdeka curriculum in English subject at a vocational high school in Indonesia. *Briliant: Jurnal Riset Dan Konseptual*, *8*(2), 297. https://doi.org/10.28926/briliant.v8i2.1201

Ferro, E. F., Cid, F. M., Cárcamo, F. P., Gajardo, E. P., Ibarra- Mora, J., Vargas, C. M., & Muñoz, B. Q. (2023). Self-perception of the level of achievement of the disciplinary and pedagogical standards in physical education students of two universities in Santiago de Chile. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, *48*, 197–206. https://doi.org/10.47197/retos.v48.94589

Guo, F., & Huang, X. (2025). A practical study of university curriculum reform from the perspective of outcome-based education: A case study of the course of academic English writing and research methods. *Journal of Humanities and Social Sciences Studies*, *7*(5), 18–30. https://doi.org/10.32996/jhsss.2025.7.5.3

Hanifa, R., Yusuf, F. N., Yusra, S. R., & Suherdi, D. (2024). Adapting EFL materials and its influences on Indonesia secondary school students' language learning. *Asian-Pasific Journal of Second and Foreign Language Education*, *9*(1), 69. https://link.springer.com/article/10.1186/s40862-024-00297-7

Hasbullah, Wekke, I. S., Dabamona, M., & Nanning. (2023). Developing curriculum component to upgrade students' English language acquisition. *Langkawi Journal of The Association for Arabic and English*, *9*(2), 145–157. https://doi.org/10.3133 2/lkw.v0i0.6292

Jaja. (2024). Research trends on students' writing skills: A Bibliometric analysis using scopus database. In *Journal of Language and Education* (Vol. 10, Issue 3, pp. 161–182). National Research University, Higher School of Econoimics. https://doi.org/10.17323/jle.2024.18806

Jaya, A., Hartono, R., Wahyuni, S., & Yulianto, H. J. (2025a). From silent to supreme: The transformative power of project-based learning on language learners. *Multidisciplinary Reviews*, *8*(8), 2025258. https://doi.org/10.31893/multirev.2025258

Jaya, A., Hartono, R., Wahyuni, S., & Yulianto, H. J. (2025b). Los efectos de la estrategia de aprendizaje basado en proyectos con actividad física en función del género sobre el rendimiento escolar y la confianza en sí mismos de los estudiantes. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, *66*, 349–360. https://doi.org/10.47197/retos.v66.110067

Makwana, D., Engineer, P., Dabhi, A., & Chudasama, H. (2023). Sampling methods in research: A review. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, *7*(3), 762–768. www.ijtsrd.com

Merueña, J. A. A. (2023). Navigating synergy: Outcome-Based Education in English language curriculum. *ERIC Journal*, *8*(12), 115–118. https://eric.ed.gov/?id=ED636802

Mokadem, M. E. A., Arshad, D., & Saoula, O. (2025). Organizational culture and its impact on effectiveness in Algerian higher education institutions. *Multidisciplinary Reviews*, *8*(4), 1–15. https://doi.org/10.31893/multirev.2025129

Morina, F., Futterer, T., Hubner, N., Zitzmann, S., & Fishcher, C. (2023). *Effects of online teacher professional development on the teacher, classroom, and student level: a meta-analysis*.

Na, M., Jill, L. S. S., Noor, H. M., Qi, F. J., & Ying, W. (2024). A pre-service art teacher digital literacy framework for digital literacy in pre-service art teacher education in China. *Asian Journal of University Education*, *20*(2), 235–247. https://doi.org/10.24191/ajue.v20i2.27007

Novak, E., McDaniel, K., & Li, J. (2023). Factors that impact student frustration in digital learning environments. *Computers and Education Open*, *5*(October), 100153. https://doi.org/10.1016/j.caeo.2023.100153

Nurohmat, Dian, & Gurning, S. T. (2025). The Influence of the independent learning curriculum on students' English learning achievement. *JEELL (Journal of English Education, Linguistics and Literature)*, *12*(1), 97–105. https://doi.org/10.32682/jeell.v1201/21

Nurpratiwi, F. I. (2024). Designing ict competences-integrated lesson planning course teaching model for English language education. *Journal of Literature Language and Academic Studies*, *3*(02), 52–63. https://doi.org/https://doi.org/10.56855/jllans.v3i02.1159

Parhadjanovna, S. S. (2023). Teaching foreign languages in the context of sustainable development: Best Practices, problems and opportunities. *International Scientific and Current Research Conferences*, 482–485. https://www.orientalpublication.com/index.php/iscrc/article/view/1348

Pavajeau, C. A., & Montero, J. J. (2023). Podcasting in efl: Fostering speaking skills through a cooperative learning experience cesar. *SAGE Journal*. https://doi.org/https://doi.org/10.1177/13621688231109845

Peng, Q. (2025). Research on the reform of public English curriculum in higher vocational colleges based on Outcome-Based Education. *International Journal of Education and Economics*, *8*(2), 95–99. http://www.acadpubl.com

Rahman, A. N. I., & Ekkayokkaya, M. (2024). The use of contextual teaching and learning approach on students' analytical exposition writing skills. *Indonesian Journal of Applied Linguistics*, *13*(3), 455–467. https://doi.org/10.17509/ijal.v13i3.66955

Riwayatiningsih, R., Prastikawati, E. F., Muchson, M., Haqiqi, F. N., Setyowati, S., & Kartiko, D. A. (2025). Empowering higher-order thinking skills in writing through gamification and multimodal learning within PBL. *Forum for Linguistic Studies*, *7*(2), 385–398. https://journals.bilpubgroup.com/index.php/fls

Rosyidah, U. J., Dewanti Laksmi, E., & Anugerahwati, M. (2023). Gamification in duolingo app on improving English listening proficiency of junior high school students. *ENGLISH FRANCA : Academic Journal of English Language and Education*, *7*(1), 119. https://doi.org/10.29240/ef.v7i1.5377

Salehi, H., Keshtiarast, B., & Rahimi, M. A. (2023). Relationship between Iranian intermediate EFL learners' interpersonal and intrapersonal

intelligences and their speaking skill. *Journal of Practical Studies in Education*, *5*(1), 7–16. https://doi.org/10.46809/jpse.v5i1.79

Schwerpunkt. (2024). Teachers' engagement in online professional development the interplay of online professional development quality and teacher motivation. *Zeitschrift Für Erziehungswissenschaft*, *27*, 739–768. https://link.springer.com/article/10.1007/s11618-024-01241-8

Setiawan, M. N. A., Suherman, W., Suharjana, Widiyanto, & Hardianto. (2024). Revitalizing traditional sports in physical education curricula a literature review on cultural heritage and educational philosophy. *Revista Iberoamericana de Psicología Del Ejercicio y El Deporte*, *19*(2), 141–146. https://psykebase.es/servlet/articulo?codigo=9578630

Shi, L., Li, S., & Xing, J. (2025). Exploring Chinese secondary EFL students' self-regulated learning and task engagement in ai-assisted classrooms: a latent growth curve modelling study. *Europan Journal of Education*, *60*(4), e70241. https://doi.org/10.1111/ejed.70241

Soughati, N., El Hadi, K., & Flyura, A. (2025). Technopedagogical making workshop and writing/ programming. *XLinguae*, *18*(1), 153–169. https://doi.org/10.18355/XL.2025.18.01.12

Sun, K., Li jiang, A., & Jin, Y. (2025). A multi-level moderation analysis of teacher immunity, school climate, and teaching quality among English as a foreign-language teachers in China. *Teaching and Teacher Education*, *159*, 105005. https://doi.org/10.1016/j.tate.2025.105005

Syaifudin, M. (2024). Needs analysis of english learning skills for students of hospitality study program. *Linguistic, English Education and Art (LEEA) Journal*, *7*(2), 444–455. https://doi.org/10.31539/leea.v7i2.8550

Thankavel, S. (2023). Artificial intelligence integrated language teaching to enhance creative speaking and creative writing of engineering learner. *Journal of Research Administration*, *5*(2). https://journalofresearchadmin.com/index.php/jra/article/view/5%0Ahttps://journalofresearchadmin.com/index.php/jra/article/download/5/6

Tricahyati, S., & Zaim, M. (2023). English Teachers' readiness in implementing of 'merdeka belajar' curriculum in teaching English at junior high school in Padang. *Journal of English Language Teaching*, *12*(1), 97–105.

Ventaja-Cruz, J., Rincón, J. M. C., Tejada-Medina, V., Martín-Moya, R., & Sánchez, R. P. (2024). Impact of the motivation, perception and self-concept of girls soccer players in multicultural contexts: a study in Melilla. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, *60*, 632–642. https://doi.org/10.47197/retos.v60.108908

Wang, S. C., & Afidchao, S. (2023). Attitudinal changes among high school students toward the English language following educational reforms in China. *Journal of Pedagogical Sociology and Psychology*, *5*(1), 1–11. https://doi.org/10.33902/jpsp.202315710

Wang, Y., & Zhang, J. (2025). From erosion to fluency: reversing language shift in Chinese Australian households. *Frontiers in Psychology*, *16*, 1553439. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2025.1553439/full

Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1261955

Yakavets, N., Winter, L., Malone, K., Zhontayeva, Z., & Khamidulina, Z. (2023). Educational reform and teacher agency in reconstructing pedagogical practices in Kazakhstan. *Journal of Educational Change*, *24*, 727–757. https://link.springer.com/article/10.1007/s10833-022-09463-5

Yang, F., Li, K., & Li, R. (2024). AI in language education: enhancing learners' speaking awareness through AI-supported training. *International Journal of Information and Education Technology*, *14*(6), 828–833. https://doi.org/10.18178/ijiet.2024.14.6.2108

Yang, H., Shao, Y., Liu, Y., Dong, J., Li, Q., & Chen, Y. (2024). The construction and application of a blended contextual teaching model for computational thinking development. *Proceedings of the 2024 15th International Conference on E-Education, E-Business, E-Management and E-Learning*, 99–106. https://doi.org/10.1145/3670013.3670027

Yen, P. H., Th, N. A., Thao, L. T., & Thuy, P. T. (2023). Assessment strategies in Outcome-Based Education: Preferences and practices among university lecturers in Vietnam. *International Journal of Learning, Teaching and Educational Research*, *22*(10), 416–432. https://doi.org/10.26803/ijlter.22.10.23

Zhou, A., & He, H. (2025). English Curriculum reform in China and teachers' preparedness to implement the latest curriculum standards. *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.1935

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability (Switzerland)*, *15*(4). https://doi.org/10.3390/su15042872