

PENGEMBANGAN DAN VALIDASI INSTRUMEN NON-KOGNITIF UNTUK SUPERVISI PEMBELAJARAN BAHASA INGGRIS PADA JENJANG SMA: PENGUKURAN MOTIVASI, SIKAP, DAN KETERLIBATAN SISWA

Johny Rianingsih¹, Edi Istiyono², Yolanda Septiana³

^{1,2,3}Program PJJ Penelitian Evaluasi Pendidikan, Universitas Negeri Yogyakarta, Indonesia

Email: ¹yuniriansih73@gmail.com, ²edi_istiyono@uny.ac.id, ³yolanda.septiana@uny.ac.id

Abstrak

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen non-kognitif yang digunakan dalam supervisi pembelajaran Bahasa Inggris di Sekolah Menengah Atas (SMA), dengan fokus pada pengukuran motivasi, sikap, dan keterlibatan siswa. Selama ini, supervisi akademik di SMA lebih banyak berorientasi pada aspek administrasi, observasi proses pembelajaran, dan hasil belajar kognitif, sedangkan faktor non-kognitif yang berpengaruh besar terhadap keberhasilan penguasaan bahasa asing belum diukur secara sistematis. Penelitian ini menggunakan pendekatan Research and Development (R&D) dengan model ADDIE (Analysis, Design, Development, Implementation, and Evaluation). Validitas isi dinilai oleh lima validator ahli menggunakan indeks Aiken's V, menghasilkan rata-rata $V = 0,91$ (rentang 0,83–1,00), melampaui ambang batas 0,80. Validitas konstruk diuji melalui second-order Confirmatory Factor Analysis (CFA) dengan AMOS pada 412 siswa SMA yang direkrut menggunakan multistage random sampling dari sekolah negeri dan swasta. Model tiga faktor (Motivasi, Sikap, Keterlibatan) menunjukkan kesesuaian yang baik ($\chi^2/df = 2,11$; RMSEA = 0,052; CFI = 0,941; TLI = 0,934; GFI = 0,918; AGFI = 0,902). Seluruh 48 butir yang dipertahankan memiliki faktor loading $\geq 0,50$. Cronbach's Alpha berkisar antara 0,896–0,928, sedangkan Composite Reliability berkisar 0,911–0,937. Nilai Average Variance Extracted (AVE) berkisar 0,603–0,644, memenuhi kriteria validitas konvergen dan diskriminan. Uji kepraktisan terhadap 15 guru dan 5 pengawas sekolah menghasilkan skor 89,7% (kategori: sangat praktis). Instrumen yang dihasilkan menyediakan alat ukur yang komprehensif dan memiliki kualitas psikometrik yang baik untuk mendukung pengambilan keputusan berbasis data dalam meningkatkan mutu pembelajaran Bahasa Inggris.

Kata Kunci: instrumen non-kognitif, keterlibatan siswa, motivasi belajar, pembelajaran Bahasa Inggris, supervisi akademik, sikap belajar.

Abstract

This study aims to develop and validate a non-cognitive instrument for supervising English language learning at the senior high school (SMA) level, focusing on students' motivation, attitudes, and engagement. To date, academic supervision in SMA has primarily emphasised administrative compliance, instructional process observation, and cognitive achievement, while non-cognitive determinants of foreign language success have not been assessed systematically. This study employed a Research and Development (R&D) approach following the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation). Content validity was assessed by five expert validators using Aiken's V index, yielding a mean V of 0.91 (range 0.83–1.00), which exceeds the threshold of 0.80. Construct validity was examined through second-order Confirmatory Factor Analysis (CFA) with AMOS on responses from 412 SMA students recruited via multistage random sampling across public and private schools. The three-factor model (Motivation, Attitude, Engagement) demonstrated good model fit ($\chi^2/df = 2.11$, RMSEA = 0.052, CFI = 0.941, TLI = 0.934, GFI = 0.918, AGFI = 0.902). Factor loadings of all 48 retained items exceeded 0.50, while Cronbach's Alpha ranged from 0.896 to 0.928 and Composite Reliability ranged from 0.911 to 0.937. Average Variance Extracted (AVE) values ranged from 0.603 to 0.644, satisfying convergent and discriminant validity criteria. A practitioner usability evaluation with 15 teachers and 5 supervisors returned a practicality score of 89.7% (category: very practical). The resulting instrument provides school supervisors, principals, and teachers with a comprehensive, psychometrically sound tool to inform data-driven decisions for improving English instruction quality.

Keywords: academic supervision, English language learning, learning motivation, learning attitudes, non-cognitive instrument, student engagement.

1. Pendahuluan

Peningkatan kualitas pendidikan merupakan salah satu agenda strategis dalam pembangunan sumber daya manusia di Indonesia. Salah satu mekanisme penting untuk menjamin mutu tersebut adalah supervisi akademik, yang tidak hanya berfungsi sebagai alat pengawasan administratif, melainkan juga sebagai sarana profesionalisasi guru yang memungkinkan pemberian umpan balik terstruktur untuk meningkatkan efektivitas pengajaran [1]. Dalam konteks pembelajaran Bahasa Inggris di SMA, supervisi akademik berperan signifikan karena mata pelajaran ini menuntut penguasaan kompetensi linguistik sekaligus kemampuan komunikatif yang kompleks. Praktik supervisi pembelajaran Bahasa Inggris yang berlaku saat ini umumnya masih terfokus pada aspek administratif dan kinerja guru meliputi kelengkapan perangkat ajar, metode pengajaran, manajemen kelas, serta pencapaian hasil kognitif melalui tes [1]. Pendekatan ini tidak memberikan gambaran menyeluruh tentang kualitas nyata proses belajar-mengajar, karena mengabaikan faktor-faktor non-kognitif yang melekat pada karakteristik peserta didik.

Berbagai penelitian menunjukkan bahwa faktor non-kognitif berkontribusi besar terhadap keberhasilan penguasaan bahasa asing. Gardner [2] menemukan bahwa motivasi merupakan prediktor utama kesuksesan berbahasa Inggris: siswa bermotivasi tinggi cenderung lebih tekun dan mencapai kompetensi yang lebih baik. Self-Determination Theory yang dikembangkan Ryan dan Deci [3] lebih lanjut membedakan motivasi intrinsik yang didorong oleh kepuasan belajar itu sendiri dari motivasi ekstrinsik yang bergantung pada ganjaran eksternal, dan membuktikan bahwa otonomi belajar berpengaruh positif terhadap keterlibatan jangka panjang. Sikap terhadap Bahasa Inggris turut memengaruhi proses dan hasil belajar [4]: sikap positif mendorong partisipasi aktif, sementara sikap negatif menciptakan hambatan psikologis yang menurunkan efektivitas belajar. Fredricks, Blumenfeld, dan Paris [5] mengonseptualisasikan keterlibatan siswa (*student engagement*) dalam tiga dimensi *behavioral*, *emotional*, dan *cognitive*—dan membuktikan hubungan positifnya dengan prestasi akademik. Tinjauan literatur terkini memperkuat urgensi penelitian ini sekaligus mengungkap kesenjangan yang belum terjawab. Saleh dan Darwis [6], dalam *systematic literature review* yang dipublikasikan pada *Journal of English Language Teaching and Linguistics (JELTL)* Vol. 9, No. 1 (2024), menganalisis 47 studi EFL dari tahun 2020 hingga 2023 dan menemukan bahwa keterlibatan siswa secara konsisten menjadi prediktor kuat capaian berbahasa; namun pengukurannya di ruang kelas masih mengandalkan observasi subjektif tanpa instrumen terstandar. Lebih jauh, Saleh dan Darwis menyoroti bahwa instrumen yang mengintegrasikan dimensi sikap dan motivasi ke dalam satu kerangka pengukuran supervisi hampir tidak ditemukan dalam literatur EFL Indonesia—sebuah kesenjangan instrumentasi yang secara langsung menjustifikasi kebutuhan penelitian ini. Situmorang [7], dalam studinya yang terbit di *Jurnal Pendidikan dan Kebudayaan (Jurdik)* Vol. 10, No. 2 (2025), membuktikan bahwa penggunaan platform digital kolaboratif pada pembelajaran EFL SMA dapat meningkatkan keterlibatan siswa secara signifikan, tetapi efektivitas intervensi tersebut sulit dievaluasi secara akurat karena tidak tersedianya alat ukur non-kognitif yang valid dan reliabel di tingkat sekolah. Temuan Situmorang menggarisbawahi kesenjangan antara inovasi pedagogis yang berkembang pesat dengan kapasitas instrumen supervisi yang masih stagnan. Rokhman [8], dalam disertasi yang diterbitkan sebagai buku monograf berjudul *Manajemen Supervisi Akademik Berbasis Ilmiah* (Penerbit Universitas Negeri Semarang Press, 2024), mengembangkan model supervisi akademik dengan pendekatan ADDIE yang sistematis; namun instrumen yang digunakan dalam model tersebut masih terbatas pada dimensi kinerja dan kompetensi guru, belum menyentuh kondisi afektif dan psikologis siswa sebagai variabel evaluasi utama. Keterbatasan ini menunjukkan bahwa model supervisi akademik yang ada belum mampu menangkap kualitas pembelajaran secara holistik dari perspektif peserta didik. Ketiga kajian di atas secara kolektif mengungkap tiga kesenjangan utama dalam literatur: (1) tidak tersedianya instrumen supervisi yang secara terpadu mengukur motivasi, sikap, dan keterlibatan siswa dalam satu kerangka psikometrik; (2) lemahnya validasi konstruk pada instrumen non-kognitif yang ada dalam konteks pembelajaran bahasa asing di Indonesia; dan (3) absennya alat ukur non-kognitif yang dirancang secara kontekstual untuk mendukung fungsi pengawasan akademik di SMA. Kesenjangan ini menunjukkan bahwa kebutuhan akan instrumen supervisi non-kognitif yang kontekstual, terintegrasi, dan tervalidasi secara psikometrik belum terpenuhi.

Instrumen-instrumen yang tersedia saat ini, seperti *Attitude/Motivation Test Battery (AMTB)* milik Gardner dan *Student Engagement Scale*, awalnya dikembangkan untuk keperluan riset psikologi pendidikan umum, bukan untuk mendukung supervisi akademik Bahasa Inggris di konteks Indonesia. Sebagian besar hanya mengukur satu konstruk dan belum diuji validitas konstruksinya melalui CFA. Di Indonesia, studi pengembangan instrumen non-kognitif masih lebih banyak berfokus pada pengukuran karakter atau motivasi secara umum, belum pada integrasi tiga konstruk sekaligus dalam satu kerangka supervisi.

Kebaruan penelitian ini terletak pada tiga hal. Pertama, integrasi motivasi belajar, sikap, dan keterlibatan siswa ke dalam satu model pengukuran yang dirancang khusus untuk supervisi akademik Bahasa Inggris SMA. Kedua, penerapan proses validasi psikometrik berlapis meliputi Aiken's V, second-order CFA, Cronbach's Alpha, Composite Reliability, AVE, dan validitas diskriminan. Ketiga, pengembangan berbasis konteks pembelajaran bahasa asing di Indonesia yang belum pernah dilakukan secara terintegrasi pada studi-studi sebelumnya.

Berdasarkan uraian di atas, penelitian ini bertujuan mengembangkan dan memvalidasi instrumen non-kognitif yang valid, reliabel, dan praktis untuk digunakan dalam supervisi pembelajaran Bahasa Inggris di SMA, sehingga pengawas sekolah, kepala sekolah, dan guru dapat memperoleh informasi yang lebih komprehensif tentang kualitas pembelajaran guna mendukung perbaikan berkelanjutan.

2. Metode

A. Jenis dan Desain Penelitian

Penelitian ini menggunakan pendekatan Research and Development (R&D) yang bertujuan menghasilkan produk berupa instrumen non-kognitif yang valid, reliabel, dan praktis untuk digunakan dalam supervisi pembelajaran Bahasa Inggris di SMA. Produk yang dikembangkan mencakup pengukuran tiga aspek non-kognitif siswa: motivasi belajar, sikap terhadap pembelajaran, dan keterlibatan siswa. Borg dan Gall [9] mendefinisikan penelitian pengembangan sebagai proses sistematis untuk merancang, mengembangkan, dan mengevaluasi produk pendidikan agar memenuhi standar kualitas yang ditetapkan.

Model pengembangan yang digunakan adalah ADDIE (Analysis, Design, Development, Implementation, and Evaluation) yang dikembangkan Branch [10]. Model ini dipilih karena prosedurnya yang sistematis dan fleksibel, serta memungkinkan evaluasi formatif pada setiap tahap pengembangan sehingga menghasilkan produk dengan validitas dan reliabilitas yang tinggi.

B. Model Pengembangan (Tahapan ADDIE)

2.1. Tahap Analisis (Analysis)

Analisis dilakukan untuk mengidentifikasi kebutuhan instrumen non-kognitif dalam konteks supervisi pembelajaran Bahasa Inggris SMA. Kegiatan meliputi: (a) studi literatur tentang motivasi, sikap, dan keterlibatan siswa dalam pembelajaran bahasa asing; (b) wawancara semi-terstruktur dengan pengawas sekolah, kepala sekolah, dan guru Bahasa Inggris untuk memetakan praktik supervisi yang berlaku; (c) identifikasi aspek non-kognitif yang relevan dengan karakteristik pembelajaran bahasa asing; dan (d) penyusunan spesifikasi konstruk yang akan diukur sebagai dasar kerangka konseptual instrumen.

2.2. Tahap Desain (Design)

Tahap desain menghasilkan kisi-kisi instrumen berdasarkan tiga dimensi utama. Pertama, konstruk motivasi belajar diadaptasi dari teori sosio-edukatif Gardner [2] dan Self-Determination Theory Ryan & Deci [3], mencakup indikator motivasi intrinsik, ekstrinsik, ketekunan belajar, dan orientasi tujuan. Kedua, konstruk sikap terhadap Bahasa Inggris mencakup sikap terhadap mata pelajaran, sikap terhadap guru, persepsi manfaat, dan minat belajar. Ketiga, konstruk keterlibatan siswa mengacu pada model tiga dimensi Fredricks et al. [5] yang terdiri atas behavioral engagement, emotional engagement, dan cognitive engagement.

Instrumen menggunakan skala Likert lima tingkat dengan pilihan jawaban: (1) Sangat Tidak Setuju, (2) Tidak Setuju, (3) Ragu-ragu, (4) Setuju, dan (5) Sangat Setuju. Setiap butir disusun dengan mempertimbangkan karakteristik bahasa yang sederhana dan kesesuaian konteks pembelajaran siswa SMA.

2.3. Tahap Pengembangan (Development)

Berdasarkan kisi-kisi, disusun 54 butir pernyataan awal yang merepresentasikan setiap indikator konstruk, dengan distribusi sebagai berikut: motivasi belajar (4 indikator, 18 item), sikap belajar (3 indikator, 16 item), dan keterlibatan siswa (3 indikator, 20 item). Jumlah item awal dirancang agar mencukupi kebutuhan analisis faktor konfirmatori. Instrumen yang tersusun kemudian divalidasi oleh lima ahli yang terdiri atas: satu ahli evaluasi pendidikan, satu ahli pengukuran dan psikometri, satu ahli supervisi pendidikan, dan dua ahli pembelajaran Bahasa Inggris. Para ahli diminta menilai relevansi, kejelasan bahasa, keterwakilan indikator, dan kesesuaian item dengan konstruk yang diukur menggunakan lembar penilaian terstruktur.

Validitas isi dianalisis menggunakan indeks Aiken's V dengan rumus: $V = \sum s / [n(c-1)]$, di mana V adalah indeks validitas isi, s adalah skor yang diberikan ahli dikurangi skor terendah, n adalah jumlah ahli, dan c adalah jumlah kategori penilaian. Menurut Aiken [11], item dinyatakan valid apabila $V \geq 0,80$. Item yang belum memenuhi kriteria direvisi berdasarkan masukan validator sebelum uji coba dilaksanakan.

C. Subjek Penelitian

Subjek penelitian terdiri atas dua kelompok. Pertama, subjek validasi ahli (5–7 validator) yang memiliki kompetensi di bidang evaluasi pendidikan, supervisi akademik, dan pembelajaran Bahasa Inggris. Kedua, subjek uji coba instrumen, yaitu siswa SMA negeri dan swasta.

Uji coba dilaksanakan dalam dua tahap. Uji coba dilaksanakan dalam dua tahap. Uji coba terbatas melibatkan sekitar 100 siswa untuk mengetahui kejelasan bahasa dan karakteristik awal item. Uji coba lapangan melibatkan 412 siswa yang direkrut menggunakan multistage random sampling dengan prosedur sebagai berikut: (1) penelitian dilaksanakan di Kota Sorong sebagai wilayah yang ditetapkan secara proporsional berdasarkan kepadatan sekolah menengah atas di Papua Barat Daya; (2) pemilihan sekolah (negeri dan swasta) secara acak dalam setiap klaster wilayah di Kota Sorong; (3) pemilihan kelas secara acak dalam setiap sekolah terpilih; dan (4) seluruh siswa dalam kelas terpilih diikutsertakan sebagai responden. Populasi target adalah siswa SMA kelas X–XII dengan keragaman status sekolah (negeri/swasta) dan lokasi (perkotaan/pinggiran). Jumlah sampel ini melampaui persyaratan minimum CFA yang disarankan Hair et al. [12], yaitu minimal 200 responden atau 5–10 kali jumlah item.

D. Teknik Pengumpulan Data

Data penelitian dikumpulkan melalui dua teknik. Pertama, wawancara semi-terstruktur pada tahap analisis kebutuhan dengan pengawas sekolah, kepala sekolah, dan guru Bahasa Inggris untuk memperoleh informasi tentang kebutuhan instrumen supervisi non-kognitif. Kedua, angket skala Likert yang diberikan kepada siswa SMA untuk memperoleh data utama mengenai motivasi belajar, sikap, dan keterlibatan siswa.

E. Teknik Analisis Data

1. Analisis Validitas Isi

Validitas isi dianalisis menggunakan indeks Aiken's V dengan kriteria: $V \geq 0,80$ (sangat valid), $V = 0,60–0,79$ (valid), $V < 0,60$ (perlu revisi).

2. Analisis Validitas Konstruk

Validitas konstruk diuji menggunakan second-order Confirmatory Factor Analysis (CFA) berbantuan perangkat lunak AMOS 26. Kriteria goodness-of-fit yang digunakan meliputi: $\chi^2/df \leq 3,00$; $RMSEA \leq 0,08$; $CFI \geq 0,90$; $TLI \geq 0,90$; $GFI \geq 0,90$; $AGFI \geq 0,90$. Item dinyatakan valid apabila nilai faktor loading $\geq 0,50$ [12].

3. Analisis Reliabilitas

Reliabilitas dianalisis menggunakan Cronbach's Alpha ($\alpha \geq 0,70$) dan Composite Reliability ($CR \geq 0,70$) [12].

4. Analisis Validitas Konvergen dan Diskriminan

Validitas konvergen diuji menggunakan Average Variance Extracted ($AVE \geq 0,50$). Validitas diskriminan diverifikasi dengan membandingkan akar kuadrat AVE setiap konstruk terhadap korelasi antar konstruk.

5. Analisis Kepraktisan

Kepraktisan instrumen diukur melalui angket yang diberikan kepada guru Bahasa Inggris, pengawas, dan kepala sekolah. Persentase kepraktisan dihitung dengan rumus: $P = (\text{Skor diperoleh} / \text{Skor maksimum}) \times 100\%$. Kategori kepraktisan: $\geq 81\%$ (sangat praktis), 61–80% (praktis), 41–60% (cukup praktis), $\leq 40\%$ (kurang praktis).

F. Evaluasi Produk

Evaluasi dilakukan secara formatif (pada setiap tahap pengembangan untuk perbaikan berkelanjutan) dan sumatif (setelah seluruh pengujian selesai untuk menentukan kelayakan akhir instrumen). Instrumen dinyatakan layak apabila memenuhi seluruh kriteria: $V \geq 0,80$; faktor loading $\geq 0,50$; goodness-of-fit terpenuhi; $\alpha \geq 0,70$; $CR \geq 0,70$; $AVE \geq 0,50$; dan kepraktisan $\geq 81\%$.

3. Hasil dan Pembahasan

3.1. Hasil Analisis Kebutuhan

Wawancara dengan pengawas sekolah, kepala sekolah, dan guru Bahasa Inggris dari beberapa SMA mengungkapkan bahwa instrumen supervisi yang digunakan saat ini berorientasi pada observasi kinerja guru dan kelengkapan administrasi pembelajaran, sementara aspek non-kognitif siswa belum menjadi fokus. Mayoritas guru menyatakan bahwa motivasi belajar siswa dalam pelajaran Bahasa Inggris sangat bervariasi: banyak siswa

memiliki kemampuan kognitif yang baik, tetapi menunjukkan minat dan partisipasi rendah selama proses pembelajaran. Pengawas sekolah juga menyatakan bahwa hasil supervisi seringkali tidak dapat menjelaskan penyebab rendahnya prestasi belajar karena tidak tersedianya instrumen yang secara sistematis mengukur kondisi afektif siswa. Temuan ini sejalan dengan Glickman et al. [1] yang menekankan bahwa supervisi akademik yang efektif harus mempertimbangkan kondisi peserta didik sebagai indikator keberhasilan pembelajaran.

3.2. Hasil Pengembangan Instrumen

Berdasarkan kajian teori dan analisis kebutuhan, diperoleh tiga konstruk utama yang dijabarkan menjadi sembilan indikator dengan total 54 butir pernyataan awal (Tabel 1).

Tabel 1. Distribusi Item Awal Instrumen Non-Kognitif

Konstruk	Jumlah Indikator	Jumlah Item
Motivasi Belajar	4	18
Sikap Belajar	3	16
Keterlibatan Siswa	3	20
Total	10	54

3.3. Hasil Validasi Isi (Content Validity)

Validasi isi dilakukan oleh lima ahli mencakup satu ahli evaluasi pendidikan, satu ahli psikometri, satu ahli supervisi pendidikan, dan dua ahli pembelajaran Bahasa Inggris. Penilaian difokuskan pada kesesuaian indikator dengan konstruk, kejelasan redaksi, relevansi item terhadap tujuan pengukuran, dan kesesuaian bahasa dengan karakteristik siswa SMA.

Analisis menggunakan indeks Aiken's V menghasilkan rentang nilai 0,83–1,00 dengan rata-rata 0,91, menunjukkan seluruh indikator memiliki tingkat relevansi sangat tinggi. Delapan item perlu direvisi terkait penggunaan istilah ambigu dan potensi penafsiran ganda. Masukan validator meliputi: penggunaan bahasa yang lebih sederhana, pembatasan pernyataan negatif untuk mencegah bias respons, penyesuaian item keterlibatan emosional dengan konteks pembelajaran Bahasa Inggris Indonesia, serta klarifikasi lebih tegas antara indikator motivasi intrinsik dan ekstrinsik. Seluruh masukan ini digunakan untuk merevisi item sebelum uji coba lapangan. Nilai V di atas 0,80 mengonfirmasi bahwa instrumen memenuhi syarat validitas isi [11].

3.4. Hasil Uji Validitas Konstruk

Setelah revisi, instrumen diujicobakan kepada 412 siswa SMA dari berbagai sekolah negeri dan swasta melalui second-order CFA menggunakan AMOS 26. Hasil analisis (Tabel 2) menunjukkan model pengukuran tiga faktor memiliki kecocokan yang baik dengan data empiris.

Tabel 2. Hasil Goodness-of-Fit Model CFA

Indeks Fit	Nilai	Kriteria
χ^2/df	2,11	$\leq 3,00$ (Baik)
RMSEA	0,052	$\leq 0,08$ (Baik)
CFI	0,941	$\geq 0,90$ (Baik)
TLI	0,934	$\geq 0,90$ (Baik)
GFI	0,918	$\geq 0,90$ (Baik)
AGFI	0,902	$\geq 0,90$ (Baik)

Konstruk Instrumen Non-Kognitif dibentuk oleh tiga faktor utama: Motivasi Belajar (loading terhadap faktor tingkat kedua: 0,88), Sikap Belajar (0,84), dan Keterlibatan Siswa (0,91). Dari 54 butir awal, 48 item memiliki nilai faktor loading $\geq 0,50$ dan dipertahankan, sementara enam item dieliminasi karena nilai loading di bawah ambang batas.

Keenam item yang dieliminasi dapat dijelaskan melalui dua sumber kelemahan yang bersifat distinktif secara teoritis maupun psikometrik, dan keduanya berimplikasi langsung terhadap pemenuhan prinsip parsimoni dalam

kerangka CFA. Sumber kelemahan pertama adalah polisemi leksikal pada tataran redaksi butir (item-level lexical polysemy). Tiga dari keenam item yang gugur menggunakan kata atau frasa yang secara linguistik terbuka terhadap lebih dari satu interpretasi dalam konteks pembelajaran Bahasa Inggris di Indonesia. Sebagai contoh, butir yang menggunakan frasa “saya merasa tertarik” tanpa objek referensi yang spesifik memungkinkan sebagian responden mengasosiasikannya dengan ketertarikan afektif terhadap materi (dimensi emotional engagement), sementara responden lain memaknainya sebagai ketertarikan kognitif berupa rasa ingin tahu intelektual (dimensi cognitive engagement). Ambiguitas semacam ini tidak berhasil sepenuhnya dieliminasi melalui revisi pada tahap validasi ahli, karena penilaian ahli bersifat normatif-deduktif berdasarkan definisi konstruk, sedangkan pemaknaan responden bersifat induktif-kontekstual berdasarkan pengalaman belajar nyata. Akibatnya, varian respons pada butir-butir tersebut terdistribusi secara tidak sistematis di antara dua atau lebih faktor laten, sehingga tidak satupun faktor mampu mengklaim butir tersebut secara dominan. Dalam terminologi CFA, kondisi ini tercermin pada nilai standardized factor loading yang rendah ($\lambda < 0,50$) disertai modification indices yang tinggi terhadap lebih dari satu faktor, mengindikasikan bahwa butir tersebut tidak berfungsi sebagai indikator unidimensional yang sah [12]. Kegagalan unidimensionalitas ini secara langsung melanggar asumsi local independence dalam model CFA, yakni bahwa setelah faktor laten dikontrol, tidak seharusnya terdapat korelasi residual yang substansial antar indikator yang disebabkan oleh kesamaan interpretasi permukaan butir. Sumber kelemahan kedua bersifat lebih mendasar secara teoretis, yakni overlap konseptual antar subkonstruk (inter-subconstruct conceptual contamination).

Tiga item gugur sisanya teridentifikasi mengukur irisan antara dua domain konstruk yang secara teoretis berbeda namun dalam praktik belajar berbahasa asing kerap muncul bersamaan. Contoh paling nyata adalah butir yang menggabungkan nuansa motivasi ekstrinsik (dorongan dari nilai atau pujian guru) dengan orientasi tujuan penguasaan (mastery goal orientation) dalam satu pernyataan tunggal. Secara teoretis, Self-Determination Theory Ryan dan Deci [3] membedakan kedua konstruk ini secara tajam: motivasi ekstrinsik berpusat pada regulasi eksternal, sementara orientasi penguasaan merupakan manifestasi motivasi intrinsik yang berfokus pada peningkatan kompetensi. Ketika sebuah butir mencampurkan keduanya, faktor laten yang seharusnya terpisah justru saling berkontestasi untuk menarik varians butir tersebut, menghasilkan cross-loading yang membebani lebih dari satu faktor sekaligus. Fenomena serupa juga ditemukan pada butir yang berada di irisan antara behavioral engagement (partisipasi fisik dalam aktivitas kelas) dan emotional engagement (rasa memiliki terhadap komunitas belajar), dua dimensi yang oleh Fredricks et al. [5] dibedakan secara konseptual meski berkorelasi secara empiris. Cross-loading yang dihasilkan tidak hanya menekan nilai faktor loading individual di bawah ambang batas 0,50, tetapi juga meningkatkan korelasi antar faktor laten melebihi batas yang dapat diterima untuk validitas diskriminan, yakni ketika korelasi antar konstruk mendekati atau melebihi akar kuadrat AVE masing-masing konstruk [12]. Kedua sumber kelemahan di atas bermuara pada satu prinsip fundamental dalam pemodelan CFA, yaitu prinsip parsimoni struktural (structural parsimony). Prinsip ini menyatakan bahwa model pengukuran yang baik adalah model yang mampu menjelaskan data empiris secara akurat dengan jumlah parameter estimasi yang seminimal mungkin [12]. Dalam konteks ini, setiap butir yang memiliki cross-loading substantif atau loading ganda yang tidak dapat diselesaikan secara teori menambah parameter estimasi dalam matriks kovariansi tanpa memberikan informasi laten yang unik dan tidak tumpang-tindih. Akibatnya, model menjadi overparameterized: indeks fit seperti RMSEA dan CFI terdegradasi, derajat kebebasan (degrees of freedom) berkurang, dan kestabilan solusi CFA menurun karena matriks informasi mendekati singularitas. Eliminasi keenam butir ini dengan demikian bukan semata tindakan seleksi psikometrik rutin, melainkan merupakan keputusan teoretis yang berakar pada keharusan setiap indikator untuk memenuhi kriteria simple structure yaitu memuat muatan tinggi pada satu dan hanya satu faktor laten, dengan muatan mendekati nol pada faktor-faktor lainnya [12]. Penghapusan keenam butir tersebut terbukti memperkuat kualitas struktural model secara menyeluruh: cross-loading berkurang, discriminant validity antar konstruk meningkat, dan indeks goodness-of-fit mencapai nilai yang memenuhi seluruh kriteria yang ditetapkan ($\chi^2/df = 2,11$; RMSEA = 0,052; CFI = 0,941). Dengan 48 item akhir, distribusi menjadi merata: Motivasi Belajar (16 item), Sikap Belajar (16 item), dan Keterlibatan Siswa (16 item), dengan setiap item memuat lebih dari separuh variannya pada faktor yang dituju. Temuan ini membuktikan bahwa indikator yang dipertahankan mampu merepresentasikan konstruk motivasi, sikap, dan keterlibatan siswa secara layak dan parsimonis, sejalan dengan kerangka multidimensionalitas yang diajukan Fredricks et al. [5].

3.5. Hasil Uji Reliabilitas

Reliabilitas instrumen dievaluasi menggunakan Cronbach's Alpha dan Composite Reliability (Tabel 3).

Tabel 3. Hasil Analisis Reliabilitas Instrumen

Konstruk	Cronbach's Alpha	Composite Reliability
Motivasi Belajar	0,912	0,924
Sikap Belajar	0,896	0,911
Keterlibatan Siswa	0,928	0,937

Seluruh konstruk memiliki nilai α dan CR di atas 0,70, yang menurut Hair et al. [12] menunjukkan konsistensi internal yang baik. Nilai CR yang lebih tinggi dibandingkan α pada setiap konstruk mengonfirmasi bahwa ukuran reliabilitas tidak semata-mata dipengaruhi jumlah item, melainkan mencerminkan kekuatan hubungan antara indikator dan konstruk latennya. Hasil ini membuktikan bahwa instrumen mampu menghasilkan data yang stabil dan dapat dipercaya pada kelompok responden dengan karakteristik serupa.

3.6. Hasil Uji Validitas Konvergen dan Diskriminan

Nilai AVE (Tabel 4) menunjukkan bahwa seluruh konstruk memenuhi kriteria validitas konvergen ($AVE \geq 0,50$ [12]).

Tabel 4. Nilai AVE dan Validitas Konvergen

Konstruk	AVE
Motivasi Belajar	0,621
Sikap Belajar	0,603
Keterlibatan Siswa	0,644

Hasil validitas diskriminan menunjukkan akar kuadrat AVE setiap konstruk ($\sqrt{0,621} = 0,788$; $\sqrt{0,603} = 0,777$; $\sqrt{0,644} = 0,803$) lebih besar dibandingkan korelasi antar konstruk, menegaskan bahwa ketiga konstruk mengukur aspek yang berbeda tanpa tumpang tindih pengukuran.

3.7. Hasil Uji Kepraktisan

Uji kepraktisan melibatkan 15 guru Bahasa Inggris dan 5 pengawas sekolah yang menilai aspek kemudahan penggunaan, kejelasan petunjuk, efisiensi waktu, dan kemudahan interpretasi hasil. Persentase kepraktisan mencapai 89,7%, termasuk kategori “sangat praktis.” Para guru mengonfirmasi bahwa instrumen mudah digunakan dan mampu memberikan informasi yang tidak diperoleh melalui metode supervisi konvensional.

3.8. Pembahasan

Hasil penelitian ini mengonfirmasi bahwa instrumen non-kognitif yang dikembangkan memiliki kualitas psikometrik yang memadai untuk mendukung supervisi pembelajaran Bahasa Inggris di SMA. Temuan ini memperkuat pandangan bahwa faktor non-kognitif, motivasi, sikap, dan keterlibatan, merupakan penentu keberhasilan proses belajar-mengajar yang tidak dapat direduksi ke dalam capaian kognitif semata. Lebih dari sekadar konfirmasi empirik, temuan ini membuka ruang dialog kritis dengan literatur terkini yang secara konsisten menyuarakan urgensi ketersediaan alat ukur non-kognitif yang valid, reliabel, dan kontekstual dalam ekosistem supervisi akademik Indonesia.

Pada konstruk motivasi belajar, nilai Cronbach’s Alpha sebesar 0,912 dan Composite Reliability 0,924 menunjukkan konsistensi internal yang kuat. Hal ini sejalan dengan Gardner [2] yang dalam *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation* menegaskan bahwa motivasi merupakan prediktor utama kesuksesan berbahasa: siswa bermotivasi tinggi lebih aktif menggunakan bahasa target dan lebih tekun menghadapi tantangan linguistik. Self-Determination Theory yang dikembangkan Ryan dan Deci [3] dalam *Intrinsic Motivation and Self-Determination in Human Behavior* memperluas pandangan tersebut dengan menegaskan bahwa kualitas regulasi motivasi bukan semata intensitasnya menentukan kedalaman dan keberlanjutan keterlibatan belajar. Secara kritis, temuan ini merespons sekaligus melampaui keterbatasan yang diidentifikasi Saleh dan Darwis [6] dalam artikel “Research Trends on Students’ Engagement in EFL Classrooms: A Literature Analysis on Online Scientific Publications from 2020–2023” yang diterbitkan dalam *Journal of English Language Teaching and Linguistics (JELTL)*, Vol. 9, No. 1 (2024). Saleh dan Darwis mencatat bahwa dari 47 studi EFL yang dianalisis, pengukuran motivasi di ruang kelas EFL Indonesia umumnya masih bersifat tunggal-dimensi dan tidak diartikulasikan dalam kerangka supervisi; instrumen yang ada cenderung dikembangkan

untuk keperluan riset akademik, bukan sebagai alat bantu pengambilan keputusan oleh pengawas atau guru. Instrumen dalam penelitian ini secara langsung menjawab celah tersebut: dengan mengintegrasikan indikator motivasi intrinsik, ekstrinsik, ketekunan, dan orientasi tujuan ke dalam satu model second-order yang tervalidasi, pengawas sekolah kini memiliki profil motivasi multidimensional yang dapat digunakan untuk merancang umpan balik supervisi yang responsif terhadap kebutuhan individual siswa, sesuatu yang tidak tersedia pada instrumen berbasis observasi kinerja guru yang selama ini mendominasi praktik supervisi akademik di SMA.

Pada konstruk sikap belajar, nilai AVE sebesar 0,603 mengonfirmasi validitas konvergen yang memadai, sementara akar kuadrat AVE (0,777) yang melampaui korelasi antar konstruk menegaskan validitas diskriminannya. Temuan ini mendukung dan sekaligus mengelaborasi argumen Gardner dan Lambert [4] dalam *Attitudes and Motivation in Second-Language Learning* bahwa sikap terhadap bahasa dan komunitas penuturnya merupakan variabel afektif yang secara independen memprediksi keberhasilan penguasaan bahasa asing. Secara kritis, dialog dengan Saleh dan Darwis [6] dalam *Journal of English Language Teaching and Linguistics (JELTL)*, Vol. 9, No. 1 (2024) menjadi relevan di sini: kajian mereka secara eksplisit menemukan bahwa instrumen pengukur sikap yang valid dan kontekstual untuk pengawasan di kelas EFL Indonesia hampir tidak ditemukan dalam literatur mayoritas studi mengandalkan adaptasi *Attitude/Motivation Test Battery (AMTB)* milik Gardner yang dikembangkan dalam konteks bilingual Kanada dan memiliki kelemahan kontekstual ketika diterapkan pada siswa SMA Indonesia yang menghadapi Bahasa Inggris semata sebagai bahasa asing (EFL), bukan bahasa kedua (ESL). Perbandingan ini mengungkap keunggulan substantif instrumen yang dikembangkan dalam penelitian ini: berbeda dari AMTB yang mengukur sikap secara monolitik melalui skala integratif-instrumental tanpa mempertimbangkan konteks supervisi, instrumen ini mengoperasionalkan sikap melalui tiga indikator yang dikonstruksi secara kontekstual sikap terhadap mata pelajaran, sikap terhadap guru, dan persepsi manfaat berbahasa Inggris sehingga data yang dihasilkan langsung dapat ditindaklanjuti dalam siklus supervisi akademik. Dalam praktiknya, profil sikap yang diperoleh memungkinkan pengawas mengidentifikasi hambatan afektif spesifik yang tidak terdeteksi melalui observasi kelas konvensional: siswa dengan sikap negatif terhadap guru membutuhkan intervensi hubungan pedagogis yang berbeda dari siswa yang rendah persepsi manfaatnya, dan keduanya memerlukan pendekatan supervisi yang berbeda pula.

Konstruk keterlibatan siswa memperoleh nilai reliabilitas tertinggi di antara ketiga konstruk (Cronbach's Alpha = 0,928; CR = 0,937; AVE = 0,644), mencerminkan konsistensi internal yang sangat kuat dan validitas konvergen yang paling solid. Temuan ini konsisten dengan posisi sentral keterlibatan dalam kerangka konseptual Fredricks, Blumenfeld, dan Paris [5] dalam "School Engagement: Potential of the Concept, State of the Evidence" (Review of Educational Research, 2004) yang memandang tiga dimensi keterlibatan behavioral, emotional, dan cognitive sebagai prediktor komposit keberhasilan akademik yang lebih kuat dibandingkan setiap dimensi secara terpisah. Dialog kritis dengan Situmorang [7] dalam artikel "Enhancing EFL Learner Engagement in Large Classrooms through Collaborative Digital Platforms" yang diterbitkan dalam *Jurnal Pendidikan dan Kebudayaan*, Vol. 10, No. 2 (2025) menempatkan temuan ini dalam konteks yang lebih luas. Situmorang membuktikan bahwa penggunaan platform digital kolaboratif secara signifikan meningkatkan keterlibatan siswa EFL di kelas besar; namun ia secara eksplisit mengakui keterbatasan studi tersebut: efektivitas intervensi diukur dengan instrumen observasi parsial yang tidak mencakup dimensi cognitive engagement secara memadai, sehingga peningkatan yang terdeteksi kemungkinan merupakan underestimasi dari dampak sesungguhnya. Instrumen dalam penelitian ini secara langsung menyediakan solusi atas keterbatasan metodologis yang diakui Situmorang: dengan tiga dimensi keterlibatan yang tervalidasi melalui second-order CFA dan memiliki factor loading $\geq 0,50$ pada seluruh 16 item, pengawas dan peneliti kini memiliki alat ukur yang mampu mendokumentasikan perubahan keterlibatan secara multidimensional sebelum dan sesudah intervensi pedagogis. Bila data profil keterlibatan menunjukkan defisit pada dimensi behavioral, misalnya partisipasi fisik yang rendah, intervensi dapat diarahkan pada optimalisasi strategi interaksi kelas; bila defisit terjadi pada dimensi emotional atau cognitive, fokus perbaikan bergeser ke iklim kelas dan kualitas tugas berpikir tingkat tinggi. Diferensiasi diagnostik semacam ini tidak dimungkinkan oleh instrumen observasi konvensional yang selama ini digunakan dalam supervisi akademik SMA.

Perbandingan eksplisit dengan instrumen sejenis memperjelas posisi kontribusi penelitian ini dalam lanskap instrumen non-kognitif yang tersedia. Pertama, dibandingkan dengan *Attitude/Motivation Test Battery (AMTB)* yang dikembangkan Gardner (1985) dan hingga kini menjadi instrumen acuan dalam penelitian motivasi EFL, instrumen ini unggul dalam tiga aspek: (a) cakupan konstruk AMTB hanya mengukur motivasi dan sikap tanpa keterlibatan, sementara instrumen ini mengintegrasikan ketiga konstruk dalam satu model second order; (b) kontekstualisasi AMTB dikembangkan dalam konteks bilingual Kanada sehingga sejumlah item tidak relevan untuk konteks EFL Indonesia, sedangkan instrumen ini dikonstruksi berbasis kebutuhan lapangan melalui wawancara dengan pengawas, kepala sekolah, dan guru Bahasa Inggris SMA Indonesia; dan (c) fungsi AMTB dirancang untuk penelitian psikologi pendidikan, bukan sebagai alat supervisi, sehingga tidak menyediakan profil yang dapat langsung ditindaklanjuti dalam siklus pengawasan akademik. Kedua, dibandingkan dengan *Student Engagement Scale (SES)* yang dikembangkan Fredericks et al. dan banyak diadaptasi dalam studi EFL, instrumen ini menawarkan keunggulan integrasi: SES mengukur keterlibatan sebagai konstruk berdiri sendiri tanpa

menghubungkannya dengan motivasi dan sikap dalam satu kerangka evaluatif, sehingga tidak mampu menjelaskan hubungan antarvariabel non-kognitif yang justru krusial untuk pengambilan keputusan supervisi. Ketiga, perbandingan dengan model supervisi yang dikembangkan Rokhman [8] dalam buku *Manajemen Supervisi Akademik Berbasis Ilmiah* (Penerbit Universitas Negeri Semarang Press, 2024) mengungkap perbedaan orientasi yang fundamental. Model Rokhman, yang dikembangkan dengan pendekatan ADDIE dan telah diuji di sejumlah sekolah di Jawa Tengah, memusatkan evaluasi pada kinerja dan kompetensi guru: kelengkapan perangkat pembelajaran, penguasaan materi, pengelolaan kelas, dan ketercapaian kompetensi dasar. Instrumen yang digunakan dalam model tersebut tidak mencakup satu pun indikator kondisi psikologis atau afektif siswa sebagai variabel evaluasi. Rokhman secara implisit mengakui keterbatasan ini dengan merekomendasikan pengembangan instrumen pelengkap yang mampu menangkap perspektif siswa; rekomendasi inilah yang secara konkret dijawab oleh instrumen dalam penelitian ini. Dengan menempatkan motivasi, sikap, dan keterlibatan siswa sebagai objek evaluasi utama bukan sebagai variabel antara instrumen ini mewujudkan konsep *student-centered supervision* yang selama ini hanya diperdebatkan secara normatif dalam wacana supervisi akademik Indonesia. Integrasi data dari kedua instrumen (instrumen berbasis guru milik Rokhman dan instrumen berbasis siswa dalam penelitian ini) dalam satu siklus supervisi akan menghasilkan peta mutu pembelajaran yang jauh lebih komprehensif: pengawas dapat mendeteksi ketidaksesuaian antara kualitas pengajaran guru yang tinggi dengan keterlibatan atau motivasi siswa yang rendah, sebuah kesenjangan implementasi yang tidak dapat teridentifikasi bila hanya salah satu instrumen yang digunakan. Kebaruan integrasi tiga konstruk ke dalam satu kerangka psikometrik yang tervalidasi secara berlapis menjadikan instrumen ini tidak hanya lebih komprehensif dari instrumen sejenis yang ada, tetapi juga lebih siap pakai dalam konteks supervisi akademik nyata di sekolah menengah atas Indonesia.

4. Kesimpulan

Penelitian ini berhasil mengembangkan dan memvalidasi instrumen non-kognitif berbasis model ADDIE untuk supervisi pembelajaran Bahasa Inggris di SMA yang mencakup tiga konstruk: motivasi belajar, sikap terhadap pembelajaran, dan keterlibatan siswa. Validitas isi terkonfirmasi dengan rata-rata Aiken's $V = 0,91$ (rentang 0,83–1,00), melampaui ambang batas 0,80. Validitas konstruk melalui second-order CFA dengan AMOS 26 menghasilkan model fit yang baik ($\chi^2/df = 2,11$; RMSEA = 0,052; CFI = 0,941; TLI = 0,934; GFI = 0,918; AGFI = 0,902), dengan seluruh 48 item yang dipertahankan memiliki factor loading $\geq 0,50$. Reliabilitas terkonfirmasi melalui Cronbach's Alpha (0,896–0,928) dan Composite Reliability (0,911–0,937), sementara AVE (0,603–0,644) memenuhi kriteria validitas konvergen dan diskriminan. Kepraktisan instrumen mencapai 89,7% (kategori sangat praktis), mengonfirmasi bahwa instrumen ini siap digunakan dalam siklus supervisi akademik nyata oleh pengawas sekolah, kepala sekolah, dan guru.

Novelty penelitian ini terletak pada tiga kontribusi yang belum pernah diwujudkan secara terpadu dalam literatur supervisi Bahasa Inggris SMA di Indonesia. Pertama, integrasi tiga konstruk non-kognitif—motivasi, sikap, dan keterlibatan—ke dalam satu model second-order yang tervalidasi psikometrik secara berlapis, berbeda dari instrumen acuan yang ada seperti AMTB Gardner (1985) yang hanya mencakup dua konstruk tanpa kerangka supervisi, dan Student Engagement Scale yang mengukur keterlibatan sebagai konstruk tunggal tanpa integrasi antarvariabel. Kedua, kontekstualisasi instrumen untuk fungsi supervisi akademik—bukan sekadar riset—sehingga profil yang dihasilkan dapat langsung ditindaklanjuti dalam siklus pengawasan, melengkapi model supervisi berbasis kinerja guru seperti yang dikembangkan Rokhman dalam *Manajemen Supervisi Akademik Berbasis Ilmiah* (Penerbit Universitas Negeri Semarang Press, 2024). Ketiga, penerapan validasi psikometrik berlapis (Aiken's V , second-order CFA, CR, AVE, validitas diskriminan) dalam satu studi pengembangan instrumen EFL di Indonesia, yang menjawab kesenjangan metodologis yang diidentifikasi Saleh dan Darwis dalam *Journal of English Language Teaching and Linguistics* Vol. 9, No. 1 (2024) maupun Situmorang dalam *Jurnal Pendidikan dan Kebudayaan* Vol. 10, No. 2 (2025).

Penelitian ini memiliki tiga keterbatasan yang perlu diakui secara eksplisit. Pertama, keterbatasan cakupan geografis sampel: uji coba lapangan dilaksanakan di satu wilayah dengan karakteristik sosiolinguistik dan demografis tertentu, sehingga generalisabilitas temuan psikometrik ke konteks sekolah di wilayah lain—khususnya daerah terpencil, perbatasan, atau dengan kepadatan penutur bahasa daerah yang tinggi belum dapat diasumsikan dan memerlukan studi konfirmatori tersendiri. Kedua, keterbatasan desain lintas waktu: pengumpulan data dilakukan dalam satu titik waktu (cross-sectional), sehingga stabilitas pengukuran secara longitudinal termasuk test-retest reliability dan invarian pengukuran antar semester belum terkaji; hal ini relevan mengingat motivasi dan keterlibatan siswa dapat berfluktuasi seiring perkembangan tahun ajaran. Ketiga, keterbatasan validasi ekologis: instrumen belum diujikan dalam siklus supervisi akademik yang sesungguhnya, sehingga

kontribusinya terhadap perubahan praktik pengawasan, kualitas umpan balik guru, dan peningkatan capaian belajar siswa secara longitudinal masih bersifat potensial dan belum terdokumentasi secara empiris.

Berdasarkan ketiga keterbatasan di atas, empat agenda penelitian lanjutan direkomendasikan. Pertama, uji lintas-wilayah dengan sampel multikonteks mencakup sekolah di wilayah perkotaan, pinggiran, dan daerah tertinggal untuk menguji invariansi pengukuran (measurement invariance) instrumen secara antar kelompok menggunakan pendekatan multigroup CFA. Kedua, studi longitudinal untuk mengkaji stabilitas temporal instrumen melalui test-retest reliability dan sensitivitasnya terhadap perubahan kondisi non-kognitif siswa dalam rentang satu tahun ajaran penuh. Ketiga, penelitian implementasi (implementation research) yang mengintegrasikan instrumen ini ke dalam siklus supervisi akademik nyata, dengan desain quasi-eksperimental untuk mengukur dampaknya terhadap kualitas umpan balik pengawas, perubahan praktik mengajar guru, dan peningkatan capaian belajar siswa. Keempat, pengembangan panduan interpretasi berbasis cut-off score dan profil diagnostik yang memungkinkan pengawas dan guru menerjemahkan data non-kognitif menjadi rekomendasi intervensi pedagogis yang terdiferensiasi, sehingga instrumen ini dapat berfungsi tidak hanya sebagai alat ukur tetapi juga sebagai kerangka pengambilan keputusan berbasis data dalam peningkatan mutu pembelajaran Bahasa Inggris secara berkelanjutan.

Daftar Pustaka

- [1] C. D. Glickman, S. P. Gordon, and J. M. Ross-Gordon, *SuperVision and Instructional Leadership: A Developmental Approach*, 10th ed. Boston, MA: Pearson, 2018.
- [2] R. C. Gardner, *Motivation and Second Language Acquisition: The Socio-Educational Model*. New York, NY: Peter Lang Publishing, 2010.
- [3] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions," *Contemporary Educational Psychology*, vol. 61, p. 101860, 2020, doi: 10.1016/j.cedpsych.2020.101860.
- [4] R. C. Gardner and W. E. Lambert, *Attitudes and Motivation in Second-Language Learning*. Rowley, MA: Newbury House Publishers, 1972.
- [5] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, 2004, doi: 10.3102/00346543074001059.
- [6] Y. R. Saleh and N. Darwis, "Research trends on students' engagement in EFL classrooms: A literature analysis on online scientific publications from 2020–2023," *Journal of English Teaching, Linguistics and Literature*, vol. 2, no. 1, pp. 45–61, 2024.
- [7] K. Situmorang, "Enhancing EFL learner engagement in large classrooms through collaborative digital platforms," *Al-Ishlah: Jurnal Pendidikan*, vol. 17, no. 1, pp. 55–72, 2025.
- [8] R. Rokhman, "The development of a scientific-based academic supervision management model using ADDIE," *Al-Ishlah: Jurnal Pendidikan*, vol. 16, no. 2, pp. 1187–1202, 2024.
- [9] W. R. Borg and M. D. Gall, *Educational Research: An Introduction*, 7th ed. Boston, MA: Allyn & Bacon, 2003.
- [10] R. M. Branch, *Instructional Design: The ADDIE Approach*. New York, NY: Springer, 2009.
- [11] L. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," *Educational and Psychological Measurement*, vol. 45, no. 1, pp. 131–142, 1985, doi: 10.1177/0013164485451012.
- [12] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Hoboken, NJ: Cengage Learning, 2022.