

Dominasi Model CIPP dalam Evaluasi Program Pendidikan Indonesia: Studi Literatur Sistematis terhadap Model-model Evaluasi

Sofia Edriati

Universitas Negeri
Padang;
Universitas PGRI
Sumatera Barat

Ambiyar

Universitas Negeri
Padang

Fifi Yasmi

Universitas PGRI
Sumatera Barat

Liza Husnita

Universitas PGRI
Sumatera Barat

Alamat

Korespondensi

sofiaedriati81@gmail.com

ABSTRAK Evaluasi program sangat krusial untuk menjamin kualitas pendidikan di Indonesia, namun sintesis komprehensif mengenai pola penggunaan berbagai model evaluasi program dalam konteks pendidikan Indonesia masih terbatas. Studi literatur sistematis ini bertujuan menganalisis penerapan model CIPP, Kirkpatrick, Logic Model, dan CIRO dalam evaluasi program pendidikan di Indonesia periode 2015–2025. Mengikuti pedoman PRISMA 2020, pencarian dilakukan pada basis data Scopus dan menghasilkan 19 studi yang memenuhi kriteria inklusi. Hasil analisis menunjukkan bahwa model CIPP merupakan model evaluasi yang paling dominan digunakan dalam evaluasi program pendidikan di Indonesia. Model Kirkpatrick digunakan secara terbatas, terutama pada evaluasi pelatihan profesional dan capacity building dengan fokus pada evaluasi perubahan perilaku dan hasil pelatihan individu. Logic Model tidak ditemukan sebagai model evaluasi formal dalam artikel yang dianalisis, meskipun sebagian logika evaluasinya tercermin melalui pendekatan Theory of Change. Selain itu, model CIRO tidak ditemukan digunakan secara eksplisit sehingga belum teradopsi dalam studi yang dianalisis. Sebaliknya, beberapa studi menggunakan model alternatif seperti CIPPO, Bradley, Stake's Countenance, dan DEM. Hasil sintesis juga menunjukkan adanya kecenderungan integrasi model CIPP dengan teknik kuantitatif maupun pendekatan hibrid bersama Kirkpatrick dalam beberapa studi yang dianalisis. Temuan ini menegaskan bahwa CIPP merupakan model evaluasi yang paling dominan digunakan dalam publikasi yang dianalisis, meskipun penguatan evaluasi outcome jangka panjang masih diperlukan untuk mendukung pengambilan keputusan berbasis bukti.

Kata Kunci CIPP, Kirkpatrick, CIPPO, Theory of Change, Bradley.

1. Pendahuluan

Pendidikan merupakan investasi strategis yang menentukan kualitas sumber daya manusia dan pembangunan berkelanjutan suatu bangsa. Sistem pendidikan Indonesia, dengan keragaman geografis dan sosioekonomis yang luas, memerlukan evaluasi sistematis untuk memastikan pencapaian target pendidikan nasional (Wandi et al., 2024). Berbagai kajian menunjukkan bahwa pendidikan Indonesia menempatkan evaluasi program sebagai instrumen strategis untuk menjamin mutu, akuntabilitas, dan relevansi kebijakan dengan kebutuhan siswa serta tuntutan global abad ke-21 (Afriadi et al., 2024). Dalam konteks perubahan kurikulum yang cepat, perluasan akses pendidikan dan tekanan untuk menghasilkan lulusan berkarakter Pancasila sekaligus kompetitif secara internasional, kerangka evaluasi program yang kokoh menjadi prasyarat bagi perbaikan berkelanjutan sistem pendidikan nasional.

Evaluasi program pendidikan melampaui sekadar pengukuran outcome semata. Evaluasi merupakan proses komprehensif yang mempertimbangkan konteks implementasi, ketersediaan sumber daya, kualitas proses, dan hasil yang dicapai (Arikunto & Cepi, 2018). Kompleksitas ini menuntut penggunaan model evaluasi yang mampu menangkap aspek multi-dimensi dari sebuah program. Model evaluasi yang telah dikembangkan secara internasional termasuk CIPP (Stufflebeam, 1971), Kirkpatrick (Kirkpatrick & Kirkpatrick, 2008), Logic Model (Pratt et al., 2000), dan CIRO.

Literatur akademik mengenai evaluasi program pendidikan di Indonesia mengalami pertumbuhan signifikan pasca-2015, khususnya sejak implementasi Kurikulum 2013 dan reformasi pendidikan berkelanjutan (Turmuzi et al., 2022). Berbagai model evaluasi program telah dikembangkan dan digunakan dalam literatur evaluasi pendidikan internasional, antara lain CIPP, Kirkpatrick, Logic Model, dan CIRO, dengan karakteristik, fokus, serta kedalaman

analisis yang berbeda-beda (Stufflebeam & Coryn, 2014; Fitzpatrick et al., 2022; Gough et al., 2017; Kirkpatrick & Kirkpatrick, 2008). Dalam konteks Indonesia, hasil telaah penelitian ini menunjukkan bahwa penerapannya tidak merata, dengan dominasi CIPP dan penggunaan terbatas Kirkpatrick, sedangkan Logic Model dan CIRO tidak ditemukan sebagai model evaluasi formal dalam studi yang dianalisis.

Model CIPP yang dikembangkan oleh Stufflebeam memandang evaluasi sebagai proses sistematis untuk mendukung pengambilan keputusan melalui analisis context, input, process, dan product. Dalam praktik mutakhir, CIPP berkembang menjadi CIPPO dengan penambahan *outcome* sebagai dimensi penting keberlanjutan program (Stufflebeam & Coryn, 2014; Hamid et al., 2025). CIPP digunakan secara luas di Indonesia untuk menilai program pelatihan vokasional, pendidikan inklusif, program sekolah, evaluasi kurikulum, dan bahkan program-program inovatif seperti kurikulum merdeka, karena kemampuannya menangkap kesesuaian kebutuhan, kesiapan sumber daya, mutu implementasi, dan hasil program secara holistik (Hamid et al., 2025; Dewi et al., 2025). Adaptasi model ini menjadi CIPPO menunjukkan kesadaran evaluator Indonesia akan pentingnya melacak dampak jangka panjang program pasca implementasi dan efektivitas dalam konteks dunia kerja (Stufflebeam & Zhang, 2017). meskipun secara konseptual dimensi outcome diakui penting, pelacakan sistematis melalui *tracer study* dan *employability* masih sangat terbatas.

Sementara itu, model empat level Kirkpatrick yang terdiri atas reaksi, pembelajaran, perilaku, dan hasil (Kirkpatrick & Kirkpatrick, 2008) banyak diadaptasi untuk mengevaluasi program pelatihan profesional. Penelitian Abraham et al. (2024) menunjukkan bahwa model Kirkpatrick lebih sesuai dalam menilai kepuasan peserta (reaction), peningkatan kompetensi (learning), perubahan perilaku kerja (behavior), dan dampak organisasional (result) dalam evaluasi program pelatihan terpadu, khususnya untuk pelatihan aparatur negara, pelatihan guru, dan program *capacity building* bagi pendidik profesional. Model ini memberikan ketajaman dalam mengukur perubahan individual dan dampak implementasi pada level organisasi. Namun, model ini relatif terbatas dalam menggambarkan dinamika konteks kebijakan, keselarasan dengan tujuan nasional, dan kesiapan input program sebagai aspek yang menjadi kekuatan utama CIPP.

Selain CIPP dan Kirkpatrick, Logic Model dan CIRO merupakan dua kerangka evaluasi yang banyak dibahas dalam literatur internasional. Logic Model memberikan representasi visual mengenai hubungan antara sumber daya, aktivitas, output, dan outcome program sehingga membantu evaluator memahami asumsi dan jalur perubahan yang mendasari suatu intervensi (LeCroy, 2021). Dalam konteks pendidikan Indonesia, hasil telaah menunjukkan bahwa Logic Model tidak muncul secara eksplisit sebagai model evaluasi yang digunakan dalam studi yang dianalisis. Namun demikian, sebagian logika evaluasinya tercermin melalui penggunaan pendekatan Theory of Change, yang memetakan hubungan antara intervensi, mekanisme perubahan, dan outcome program.

Model CIRO (Context, Input, Reaction, Outcome) dikembangkan oleh Warr et al. (1970) sebagai kerangka evaluasi pelatihan yang menekankan analisis konteks dan input sebelum menilai reaksi serta outcome peserta. Secara konseptual model ini berada di antara pendekatan komprehensif CIPP dan orientasi hasil yang menjadi ciri model Kirkpatrick (Fitzpatrick et al., 2022). meskipun memiliki potensi untuk diterapkan dalam evaluasi program pendidikan dan pelatihan di Indonesia, hasil telaah literatur ini tidak menemukan penggunaan CIRO secara eksplisit pada artikel yang memenuhi kriteria inklusi. Temuan tersebut menunjukkan bahwa model CIRO belum teradopsi dalam publikasi evaluasi program pendidikan Indonesia yang terindeks Scopus selama periode 2015–2025.

Keempat model tersebut pada dasarnya menawarkan lensa berbeda terhadap fenomena yang sama. Kajian-kajian terkini memperlihatkan bahwa di Indonesia, model CIPP cenderung mendominasi evaluasi program pendidikan formal dan vokasional namun

penggunaannya yang terpisah-pisah dalam literatur Indonesia menyebabkan peta pengetahuan menjadi terfragmentasi dan sulit dibandingkan secara sistematis lintas jenjang, jenis program, dan sektor pendidikan (Gough et al., 2017; Fitzpatrick et al., 2022).

Kajian-kajian terkini memperlihatkan bahwa di Indonesia model CIPP cenderung dominan pada evaluasi program pendidikan formal dan vokasional. Kirkpatrick lebih banyak digunakan dalam evaluasi program pelatihan profesional, sedangkan Logic Model tidak ditemukan sebagai model evaluasi formal dan CIRO belum ditemukan digunakan secara eksplisit dalam publikasi yang dianalisis (Hamid et al., 2025; Dewi et al., 2025; Abraham et al., 2024). Namun, hingga kini belum tersedia sintesis sistematis yang menganalisis secara komprehensif bagaimana dominasi model CIPP tercermin dalam praktik evaluasi program pendidikan di Indonesia, bagaimana model-model evaluasi yang ditemukan diimplementasikan secara empiris beserta kekuatan dan keterbatasannya dalam konteks kebijakan nasional, serta bagaimana posisi Logic Model dan CIRO tercermin dalam literatur yang dianalisis (Afriadi et al., 2024).

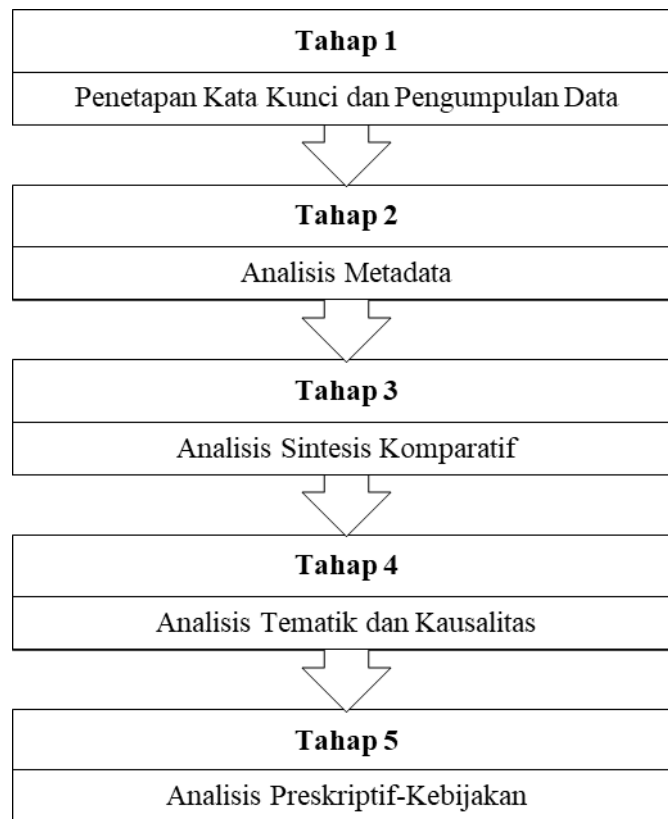
Kesenjangan pengetahuan ini menghambat pembuat kebijakan dan evaluator dalam memahami pola adopsi, karakteristik, serta konteks penggunaan berbagai model evaluasi serta dalam memilih atau mengkombinasikan pendekatan evaluasi yang paling sesuai dengan karakteristik program pendidikan vokasional, kewarganegaraan hingga pendidikan inklusif dan pendidikan nilai (Hamid et al., 2025; Dewi et al., 2025; Istiyani et al., 2024).

Studi ini bertujuan melakukan telaah literatur sistematis terhadap penerapan model CIPP, Kirkpatrick, Logic Model, dan CIRO dalam evaluasi program pendidikan di Indonesia periode 2015–2025, serta mengidentifikasi pola penggunaan, karakteristik, dan kecenderungan adopsi berbagai model evaluasi yang muncul dalam literatur. Secara spesifik, tujuan penelitian ini adalah: (1) mengidentifikasi model evaluasi program pendidikan yang digunakan dalam publikasi terindeks Scopus di Indonesia periode 2015–2025; (2) menganalisis karakteristik, kekuatan, keterbatasan, serta pola penerapan model-model evaluasi yang ditemukan, termasuk mengidentifikasi posisi model yang tidak muncul secara eksplisit dalam literatur; (3) mengidentifikasi dimensi evaluasi yang paling banyak dikaji serta faktor-faktor yang mendukung maupun menghambat implementasi model evaluasi program; dan (4) merumuskan rekomendasi strategis bagi pengembangan praktik evaluasi program pendidikan di Indonesia.

2. Metode

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan Systematic Literature Review (SLR) untuk mengidentifikasi, menyeleksi, menilai kualitas, dan mensintesis bukti ilmiah mengenai penerapan model evaluasi program pendidikan di Indonesia. Pendekatan SLR dipilih karena menyediakan prosedur yang sistematis, transparan, dan dapat direplikasi dalam proses identifikasi, seleksi, penilaian kualitas, serta sintesis literatur. Pelaksanaan penelitian mengacu pada pedoman Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 (Page et al., 2021), yang mencakup tahap identifikasi, penyaringan (screening), penilaian kelayakan (eligibility), penilaian kualitas artikel (quality assessment), dan sintesis hasil penelitian. Alur penelitian secara umum disajikan pada Gambar 1.



Gambar 1. Bagan Alur Prosedur Penelitian

2.2 Strategi Pencarian Literatur

Strategi pencarian literatur disusun melalui identifikasi kata kunci yang merepresentasikan model evaluasi, konteks evaluasi program, lokasi penelitian, dan sektor pendidikan. Pencarian literatur dilakukan pada 14 Januari 2026 menggunakan basis data Scopus, yang dipilih karena memiliki cakupan jurnal internasional bereputasi yang luas pada bidang pendidikan dan ilmu sosial. Proses pencarian menggunakan sintaks sebagai berikut:

TITLE-ABS-KEY (("CIPP model" OR "Context Input Process Product" OR "Kirkpatrick model" OR "Kirkpatrick four levels" OR "Logic model" OR "program theory" OR "theory of change" OR "CIRO model" OR "evaluation model*" OR "evaluation framework*") AND ("program evaluation" OR "program assessment" OR "curriculum evaluation" OR "training evaluation" OR "course evaluation" OR "intervention evaluation" OR "policy evaluation") AND (indonesia* OR indonesian OR "Bahasa indonesia") AND (education* OR school* OR university OR "higher education" OR vocational OR training OR curriculum OR "professional development")) AND PUBYEAR > 2014 AND PUBYEAR < 2026 AND (LIMIT-TO (LANGUAGE , "English") OR LIMIT-TO (LANGUAGE , "Indonesian")).

Sintaks tersebut menggunakan fitur TITLE-ABS-KEY untuk menelusuri kata kunci pada judul, abstrak, dan kata kunci artikel. Kata kunci dikelompokkan ke dalam empat komponen utama, yaitu: (1) model evaluasi (CIPP, Kirkpatrick, Logic Model, CIRO, evaluation model, dan evaluation framework); (2) konteks evaluasi (evaluasi program, kurikulum, pelatihan, intervensi, dan kebijakan); (3) konteks geografis (Indonesia); serta (4) sektor pendidikan (sekolah, pendidikan tinggi, vokasi, pelatihan, dan pengembangan profesional). Selain itu, filter tahun publikasi (2015–2025) dan bahasa (Inggris atau Indonesia) diterapkan untuk memperoleh artikel yang relevan dengan tujuan penelitian.

2.3 Seleksi Artikel dan Kriteria Inklusi–Eksklusi (PRISMA)

Seleksi artikel dilakukan mengikuti pedoman PRISMA 2020 yang meliputi tahap identifikasi (identification), penyaringan (screening), penilaian kelayakan (eligibility), dan

penetapan artikel yang memenuhi kriteria inklusi (included). Proses screening dilakukan berdasarkan judul, abstrak, dan naskah lengkap menggunakan kriteria inklusi dan eksklusi yang telah ditetapkan. Kriteria tersebut disajikan pada Tabel 1.

Tabel 1. Kriteria Inklusi dan Eksklusi Data

Kriteria	Inklusi	Eksklusi
Populasi	Program pendidikan di Indonesia (semua tingkatan).	Program di luar Indonesia atau program non-pendidikan.
Intervensi	Menggunakan model atau kerangka evaluasi program yang dinyatakan secara eksplisit, baik model evaluasi yang telah mapan maupun model alternatif yang relevan dengan evaluasi program pendidikan.	Tidak menggunakan model atau kerangka evaluasi program secara eksplisit, atau hanya berfokus pada pengembangan program, kurikulum, atau instrumen tanpa melakukan evaluasi program.
Desain Studi	Studi empiris (metode kuantitatif, kualitatif, campuran).	Makalah konseptual, ulasan, studi non-empiris.
Jangka waktu	Diterbitkan 2015-2025.	Diterbitkan sebelum 2015 atau setelah 2025.
Bahasa	Bahasa Inggris atau Bahasa Indonesia	Bahasa lain.
Jenis Publikasi	Artikel jurnal peer-review.	Abstrak konferensi, disertasi, buku, sastra abu-abu.
Fokus Konten	Evaluasi program.	Penilaian siswa saja, studi non-evaluasi.

Seleksi artikel dilakukan mengikuti pedoman PRISMA 2020 yang meliputi tahap identifikasi (identification), penyaringan (screening), penilaian kelayakan (eligibility), dan inklusi (included). Proses seleksi dilakukan secara independen oleh tiga orang penulis melalui penelaahan judul, abstrak, dan naskah lengkap (full-text review) berdasarkan kriteria inklusi dan eksklusi yang telah ditetapkan. Perbedaan hasil penilaian diselesaikan melalui diskusi dan konsensus dengan melibatkan penelaah akhir yang tidak terlibat dalam proses screening independen. Kriteria inklusi dan eksklusi yang digunakan dalam proses seleksi disajikan pada Tabel 1. Artikel yang memenuhi seluruh kriteria inklusi selanjutnya menjalani proses penilaian kualitas sebelum dilakukan sintesis data.

2.4 Penilaian Kualitas Artikel

Seluruh artikel yang memenuhi kriteria inklusi selanjutnya menjalani proses penilaian kualitas (quality assessment) untuk memastikan bahwa studi yang disintesis memiliki kualitas metodologis yang memadai. Penilaian kualitas dilakukan secara independen oleh tiga orang penulis menggunakan checklist yang disusun berdasarkan prinsip critical appraisal dalam Systematic Literature Review dan disesuaikan dengan tujuan penelitian. Perbedaan hasil penilaian diselesaikan melalui diskusi dan konsensus dengan melibatkan penelaah akhir yang tidak terlibat dalam proses penilaian independen.

Checklist penilaian kualitas terdiri atas lima kriteria, yaitu: (1) kejelasan tujuan penelitian, (2) kejelasan metode penelitian, (3) validitas instrumen atau teknik pengumpulan data, (4) kelengkapan penyajian data dan hasil penelitian, serta (5) relevansi penggunaan model evaluasi terhadap tujuan penelitian. Setiap kriteria diberi skor 1 apabila memenuhi indikator penilaian dan skor 0 apabila tidak memenuhi indikator, sehingga setiap artikel memperoleh skor maksimum 5. Artikel dengan skor minimal 4 dikategorikan memiliki kualitas metodologis yang memadai dan dipertahankan dalam proses sintesis, sedangkan artikel dengan skor di bawah kriteria tersebut didiskusikan kembali hingga diperoleh keputusan berdasarkan konsensus penulis.

Tabel 2. Kriteria Penilaian Kualitas Artikel

Kriteria	Indikator Penilaian	Skor
Kejelasan tujuan penelitian	Tujuan penelitian dirumuskan secara jelas dan sesuai dengan fokus evaluasi program	0-1
Kejelasan metode penelitian	Desain penelitian, subjek, dan prosedur penelitian dijelaskan secara memadai	0-1
Validitas instrumen atau teknik pengumpulan data	Instrumen atau teknik pengumpulan data dijelaskan dan sesuai dengan metode penelitian	0-1
Kelengkapan penyajian data dan hasil	Hasil penelitian disajikan secara sistematis serta didukung oleh data yang memadai	0-1
Relevansi model evaluasi	Model evaluasi yang digunakan relevan dengan tujuan dan konteks program yang dievaluasi	0-1

2.5 Analisis Data

Artikel yang memenuhi kriteria inklusi selanjutnya dianalisis menggunakan beberapa pendekatan. Analisis metadata dilakukan untuk mengidentifikasi tren publikasi, konteks penelitian, dan distribusi model evaluasi yang digunakan. Selanjutnya dilakukan sintesis komparatif untuk membandingkan karakteristik, kekuatan, dan keterbatasan setiap model evaluasi. Analisis tematik digunakan untuk mengidentifikasi dimensi evaluasi yang dominan serta faktor pendukung dan penghambat implementasi model evaluasi. Berdasarkan hasil sintesis tersebut, dilakukan analisis preskriptif untuk merumuskan rekomendasi pengembangan sistem evaluasi program pendidikan di Indonesia.

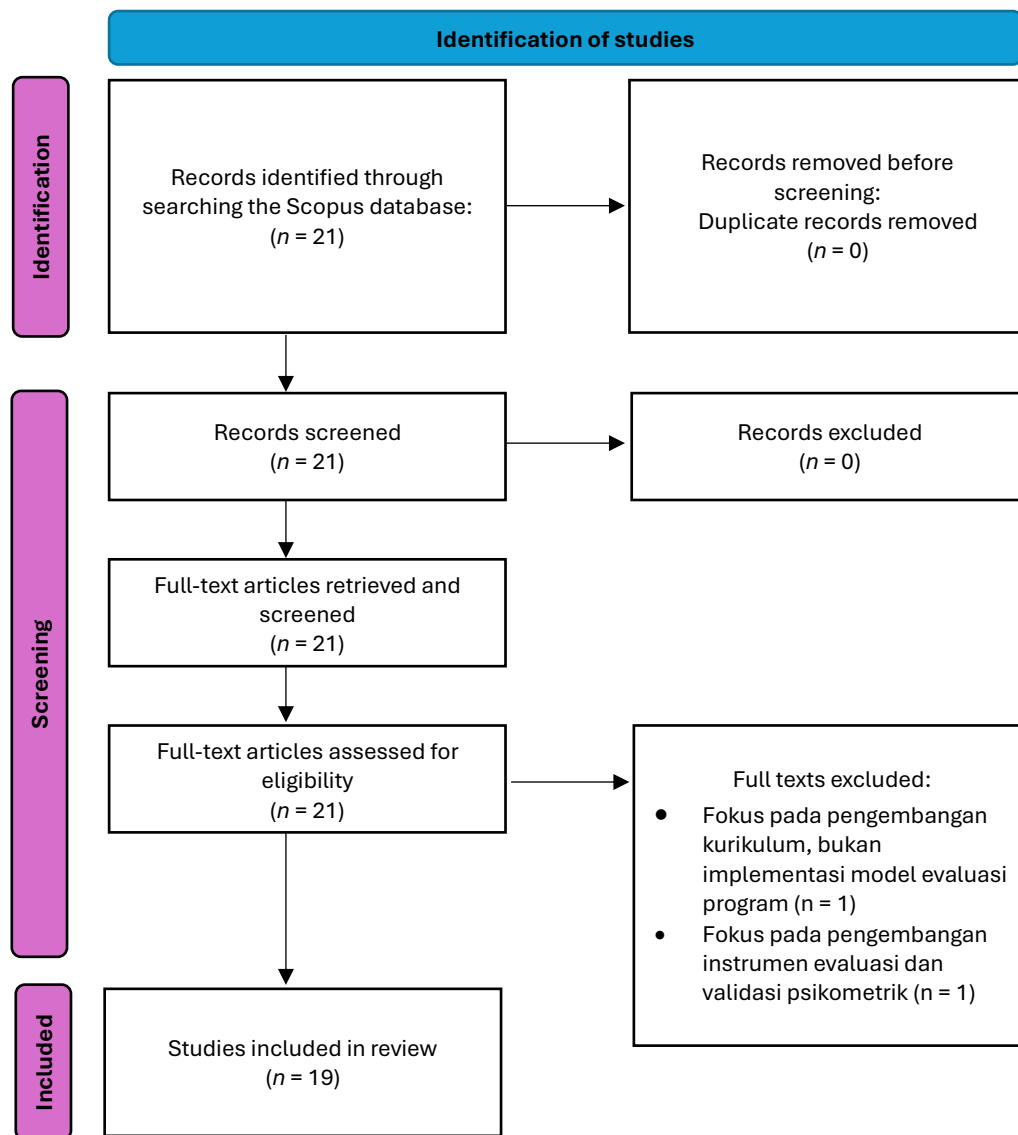
3. Hasil

3.1 Hasil Seleksi Artikel

Pencarian literatur dilakukan secara sistematis pada 14 Januari 2026 menggunakan basis data Scopus dan menghasilkan 21 publikasi yang memenuhi hasil pencarian awal berdasarkan strategi pencarian yang telah ditetapkan. Proses seleksi mengikuti pedoman PRISMA 2020 yang meliputi tahap identifikasi (identification), penyaringan (screening), penilaian kelayakan (eligibility), dan inklusi (included). Tidak ditemukan artikel duplikat sehingga seluruh publikasi yang diperoleh dilanjutkan ke tahap penyaringan.

Proses screening artikel dilakukan secara independen oleh tiga orang penulis berdasarkan kriteria inklusi dan eksklusi yang telah ditetapkan. Perbedaan hasil penilaian diselesaikan melalui diskusi dan konsensus dengan melibatkan penelaah akhir, yang tidak terlibat dalam proses screening independen. Seluruh artikel yang lolos tahap screening kemudian ditelaah secara penuh (full-text review) pada tahap eligibility. Berdasarkan hasil penilaian terhadap naskah lengkap, dua artikel dikeluarkan karena tidak berfokus pada implementasi model evaluasi program, melainkan membahas pengembangan kurikulum serta pengembangan instrumen evaluasi beserta validasi psikometriknya. Dengan demikian, sebanyak 19 artikel memenuhi seluruh kriteria inklusi dan selanjutnya dianalisis dalam studi ini. Alur proses seleksi artikel berdasarkan pedoman PRISMA 2020 disajikan pada Gambar 2.

Sebanyak 19 artikel yang memenuhi kriteria inklusi dianalisis untuk mengidentifikasi karakteristik penelitian, konteks implementasi, model evaluasi yang digunakan, pendekatan metodologis, dimensi evaluasi, serta temuan utama masing-masing studi. Karakteristik masing-masing artikel disajikan untuk memberikan gambaran mengenai konteks penelitian, model evaluasi, pendekatan metodologis, dimensi evaluasi, serta temuan utama sebagai dasar sintesis pada bagian berikutnya. Ringkasan karakteristik 19 artikel yang dianalisis disajikan pada Tabel 2.



Gambar 2. Diagram Alur Seleksi Studi Menggunakan PRISMA 2020

Selanjutnya dilakukan penilaian kualitas terhadap 19 artikel yang memenuhi kriteria inklusi. Hasil penilaian menunjukkan bahwa seluruh artikel memperoleh skor minimal 4 dari skor maksimum 5 sehingga dinilai memiliki kualitas metodologis yang memadai dan layak untuk disintesis pada tahap berikutnya.

3.2 Karakteristik Studi

Sebanyak 19 artikel yang memenuhi kriteria inklusi selanjutnya dianalisis untuk mengidentifikasi karakteristik penelitian, meliputi penulis, tahun publikasi, konteks penelitian, metode penelitian, model evaluasi yang digunakan, serta temuan utama masing-masing studi. Penyajian karakteristik studi bertujuan memberikan gambaran umum mengenai variasi konteks dan pendekatan evaluasi program pendidikan di Indonesia sekaligus menjadi dasar bagi sintesis temuan pada subbagian berikutnya. Ringkasan karakteristik 19 artikel yang dianalisis disajikan pada Tabel 3.

Tabel 3. Ringkasan Karakteristik 19 Artikel yang Dianalisis

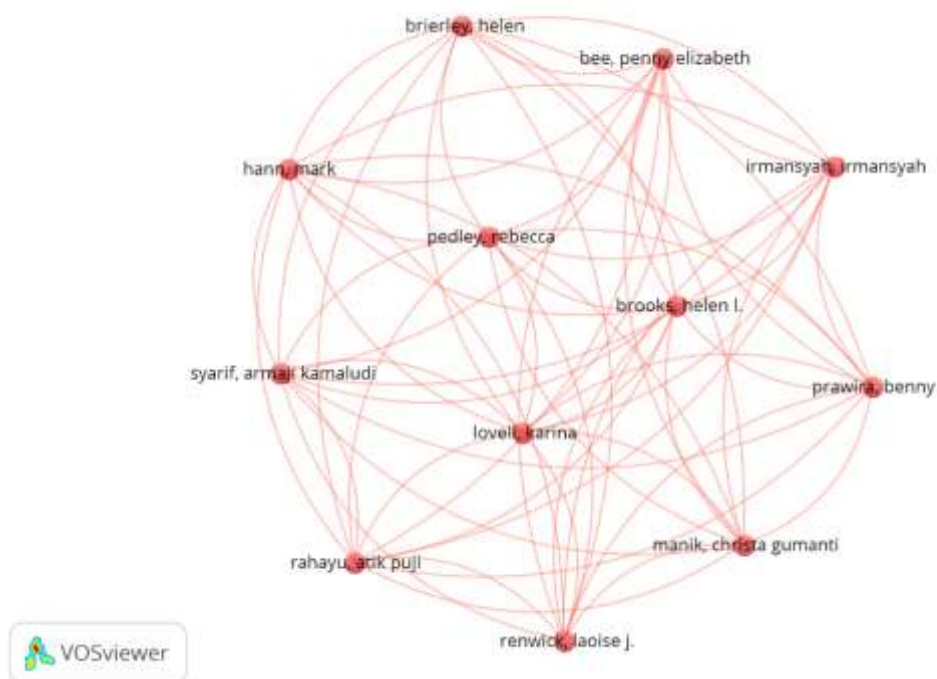
No.	Penulis (Tahun)	Judul	Negara (Konteks)	Metode Penelitian	Model Evaluasi (Dimensi)	Temuan Utama
1	Syam & Ishak (2025)	<i>Bridging Policy and Practice: Evaluating the Curriculum Effectiveness in Fostering Religious Moderation and 21st-Century Competencies</i>	Indonesia (Madrasah Sulawesi Selatan)	Mixed methods	CIPP (Context, Input, Process, Product)	Context sangat baik; input, proses, dan produk baik; implementasi memerlukan penguatan guru, sarana, dan mekanisme evaluasi.
2	Idrus et al. (2025)	<i>Evaluation of Education Program Implementation at PKBM, Jambi City</i>	Indonesia (Pendidikan Nonformal-PKBM Jambi)	Evaluasi program (kualitatif)	CIPP (Context, Input, Process, Product)	Program berjalan baik, tetapi pandemi menghambat implementasi praktik pembelajaran.
3	Hamid et al. (2025)	<i>Lesson Learned from Vocational Training Center as a School-to-Work Transition Program</i>	Indonesia (Pendidikan Vokasi-Balai Latihan Kerja)	Kualitatif	CIPPO (Context, Input, Process, Product, Outcome)	Context, proses, dan produk memuaskan; input dan outcome masih lemah.
4	Dewi et al. (2025)	<i>Inclusive Education for Students with Autism Spectrum Disorder...</i>	Indonesia (Sekolah Dasar-Bali)	Kualitatif	CIPP (Context, Input, Process, Product)	Program pendidikan inklusif berjalan baik dan layak direplikasi.
5	Angelina (2024)	<i>Evaluating English Pre-service Teacher Professional Education Program Implementation</i>	Indonesia (Pendidikan Tinggi-Universitas Sanata Dharma)	Mixed methods	CIPP (Context, Input, Process, Product)	Evaluasi digunakan sebagai dasar perbaikan implementasi PPG.
6	Abraham et al. (2024)	<i>Evaluating Integrated Training for Juvenile Criminal Justice System</i>	Indonesia (Pelatihan Profesional-Lemdiklat Polri)	Mixed methods	Kirkpatrick (Reaction, Learning, Behavior, Results)	Dampak pelatihan positif pada seluruh level Kirkpatrick.
7	Sari et al. (2024)	<i>English for Specific Purposes Curriculum Evaluation</i>	Indonesia (Pendidikan Tinggi-Politeknik Maritim)	Mixed methods	CIPP kombinasi ESP Framework (Context, Input, Process, Product)	Silabus menjadi kelemahan utama kurikulum ESP.
8	Afriadi et al. (2024)	<i>Evaluation of the Effectiveness of the Civic Education Curriculum...</i>	Indonesia (Pendidikan Dasar-Menengah)	Mixed methods	Bradley (Efektivitas kurikulum)	Kurikulum Merdeka dinilai paling efektif dibanding KTSP dan K13.
9	Kaewkumkong & Kaewkamkong (2024)	<i>Enhancing ASEAN Community Educational Cooperation...</i>	Thailand-Indonesia (Kemitraan Sekolah)	Kualitatif	CIPP kombinasi dengan Honig Model (Context, Input, Process, Product)	Merekomendasikan integrasi CIPP dengan model implementasi kebijakan.
10	Padli et al. (2023)	<i>Evaluation of the Sepak Takraw Training Program...</i>	Indonesia (Pendidikan Olahraga)	Mixed methods	CIPPO (Context, Input, Process, Product, Outcome)	Context baik, input dan proses lemah, outcome cukup efektif.
11	Brooks et al. (2023)	<i>Evaluating a Prototype Digital Mental Health Literacy Intervention...</i>	Indonesia (SMP-Jawa)	Mixed methods	Theory of Change (Jalur perubahan)	Intervensi layak diterapkan, namun outcome kuantitatif masih terbatas.

12	Rizal et al. (2022)	<i>Evaluation of Educator's Functional Training Program...</i>	Indonesia (Pendidikan Nonformal)	Kualitatif	CIPP (Context, Input, Process, Product)	Semua komponen pelatihan berada pada kategori baik.
13	Bandu et al. (2021)	<i>Context Evaluation on Implementation of English for Islamic Studies Program</i>	Indonesia (Pendidikan Tinggi-IAIN Palu)	Mixed methods	CIPP (Context)	Analisis kebutuhan dan tujuan program dinilai efektif.
14	Widoatmodjo et al. (2021)	<i>Assessment of CEFE Training Evaluation Model</i>	Indonesia (Pelatihan Kewirausahaan)	Survei	Model CEFE (Evaluasi pelatihan)	Model layak digunakan untuk evaluasi pelatihan kewirausahaan.
15	Lantu et al. (2021)	<i>Was the Training Effective?</i>	Indonesia (Pelatihan Kepemimpinan - Corporate University)	Mixed methods	Kirkpatrick (Behavior Level 3)	Kompetensi kepemimpinan peserta meningkat setelah pelatihan.
16	Yuliansari et al. (2019)	<i>Evaluation of English Subject Learning...</i>	Indonesia (SMA-Mataram)	Deskriptif Kuantitatif	Stake's Countenance (Antecedent, Transaction, Outcome)	Perencanaan sangat baik, proses baik, hasil belajar memenuhi standar.
17	Marisya et al. (2019)	<i>Education and Training Evaluation for Apparatus in Regional Public Hospital</i>	Indonesia (Pelatihan Aparatur)	Kualitatif	CIPP + Kirkpatrick (Context, Input, Process, Product + Behavior)	Program pelatihan berjalan sangat baik dan perlu dilanjutkan.
18	Sholeh et al. (2019)	<i>Evaluating the Implementation of Teacher Assignment Policy as School Principals</i>	Indonesia (Pendidikan Dasar-Menengah)	Kualitatif	DEM (Discrepancy-Gap)	Ditemukan kesenjangan implementasi kebijakan penugasan kepala sekolah.
19	Eryanto et al. (2019)	<i>Effectiveness of Entrepreneurship Practice Program: Using CIPP Program Evaluation</i>	Indonesia (praktik kewirausahaan-SMK)	Mixed Methods	CIPP (Context, Input, Process, Product)	Evaluasi menunjukkan implementasi program berjalan baik, tetapi aspek input dan proses masih menjadi tantangan utama dalam pelaksanaannya.

Berdasarkan Tabel 3, seluruh artikel yang dianalisis diterbitkan pada periode 2019-2025 dan sebagian besar dilakukan di Indonesia, mencakup konteks pendidikan formal, pendidikan vokasi, pendidikan nonformal, serta pelatihan profesional. Pendekatan penelitian yang paling banyak digunakan adalah mixed methods dan kualitatif, sedangkan model evaluasi yang paling sering digunakan adalah CIPP. Selain itu, sebagian besar penelitian mengevaluasi program menggunakan dimensi context, input, process, dan product, sementara dimensi outcome hanya muncul pada sebagian kecil studi.

3.3 Analisis Bibliometrik

Analisis jaringan co-authorship menggunakan VOSviewer terhadap 19 artikel yang dianalisis mengidentifikasi 16 peneliti utama yang membentuk satu kluster kolaborasi. Sebagaimana ditunjukkan pada Gambar 3, jaringan kolaborasi didominasi oleh beberapa peneliti dengan tingkat keterhubungan tinggi (central nodes), yaitu Helen Brierley, Karina Lovell, dan Rebecca Pedley, yang berkolaborasi dengan peneliti Indonesia.



Gambar 3. Jaringan Co-Autorship Evaluator Program Pendidikan

3.4 Trend Publikasi

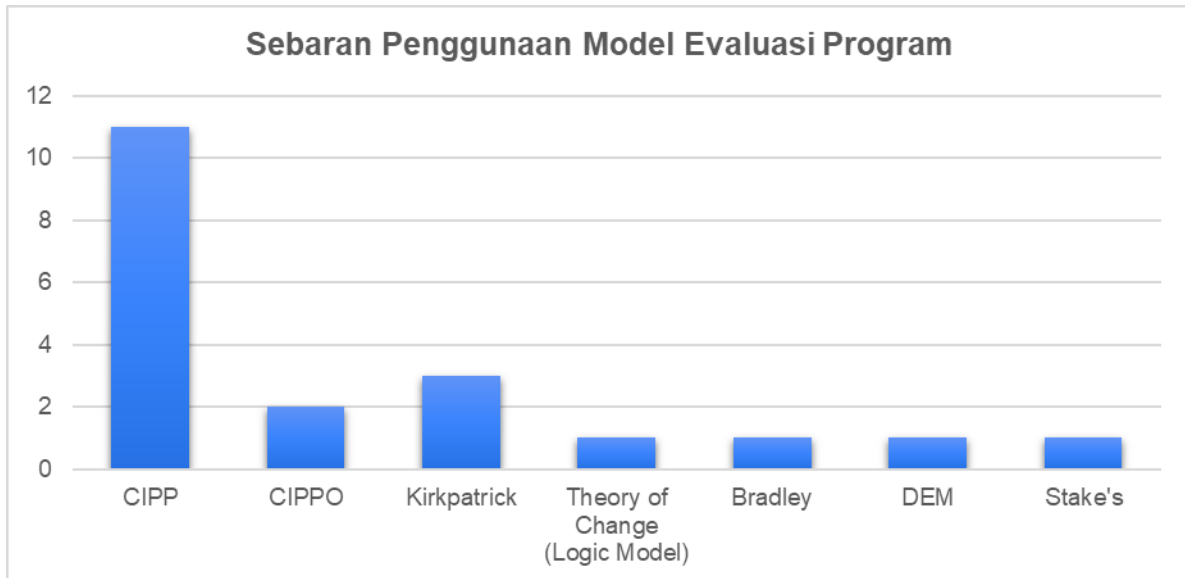
Tren publikasi artikel evaluasi program pendidikan di Indonesia yang terindeks Scopus selama periode 2015–2025 disajikan pada Gambar 4. Jumlah publikasi menunjukkan kecenderungan meningkat sejak tahun 2022 dan mencapai puncaknya pada tahun 2024.



Gambar 4. Trend Publikasi dalam 10 tahun Terakhir

3.5 Distribusi Model Evaluasi

Distribusi model evaluasi yang digunakan dalam 19 artikel dianalisis untuk menggambarkan kecenderungan pemilihan model evaluasi program pendidikan di Indonesia selama periode 2015–2025. Sebaran penggunaan setiap model evaluasi disajikan pada Gambar 5.



Gambar 5. Sebaran Model Evaluasi Program yang Digunakan

Distribusi model evaluasi dianalisis untuk menggambarkan kecenderungan penggunaan berbagai model evaluasi program pendidikan pada artikel yang memenuhi kriteria inklusi. Sebaran penggunaan setiap model evaluasi disajikan pada Gambar 5.

Distribusi penggunaan model evaluasi berdasarkan fokus program dianalisis untuk memberikan gambaran mengenai kecenderungan pemilihan model pada berbagai konteks evaluasi program pendidikan. Hasil analisis tersebut disajikan pada Tabel 4.

Tabel 4. Penggunaan Model Evaluasi Berdasarkan Fokus Program

Fokus Program	Total	CIPP	CIPPO	Kirkpatrick	Lainnya	Model Dominan
Kurikulum & Pembelajaran	6	66,7%	0	0	33,3%	CIPP
Pelatihan & Capacity Building	6	50%	0	33,3%	16,7%	CIPP
Pelatihan Vokasi	2	0	100%	0	0	CIPPO
Program Kewirausahaan	2	100%	0	0	0	CIPP
Pendidikan Nonformal	2	50%	0	0	50%	Mixed
Program Khusus/Kemitraan	1	100%	0	0	0	CIPP

Berdasarkan Tabel 4, model CIPP mendominasi evaluasi pada program kurikulum dan pembelajaran, kewirausahaan, serta program khusus atau kemitraan. Kirkpatrick lebih banyak digunakan pada evaluasi pelatihan dan capacity building, sedangkan CIPPO digunakan pada evaluasi pelatihan vokasi.

3.6 Dimensi Evaluasi

Dimensi evaluasi yang dikaji dalam 19 artikel meliputi context, input, process, product, dan outcome, dengan cakupan yang berbeda pada setiap penelitian. Ringkasan dimensi evaluasi beserta pola temuan utama disajikan pada Tabel 5.

Tabel 5. Temuan Berdasarkan Dimensi dan Cakupan Artikel

Dimensi	Cakupan di artikel	Pola hasil utama
Context (konteks, kebutuhan, keselarasan kebijakan)	Mayoritas artikel yang menggunakan model CIPP/CIPPO dan model evaluasi sejenis, seperti Syam & Ishak (2025), Idrus et al. (2025), Dewi et al. (2025), Hamid et al. (2025), Angelina (2024), Sari et al. (2024), Afriadi et al. (2024), Kaewkumkong & Kaewkamkong (2024), Marisyah et al. (2019), Brooks et al. (2023), dan Eryanto et al. (2019).	Sebagian besar studi menunjukkan bahwa tujuan program telah sesuai dengan kebutuhan serta kebijakan yang menjadi dasar implementasi.
Input (SDM, sarana prasarana, kurikulum, pendanaan)	Mayoritas artikel yang menggunakan model CIPP/CIPPO, terutama Hamid et al. (2025), Idrus et al. (2025), Dewi et al. (2025), Angelina (2024), Padli et al. (2023), dan Eryanto et al. (2019).	Banyak studi melaporkan keterbatasan sarana prasarana, pendanaan, dan sumber daya manusia sebagai aspek yang masih perlu diperkuat.
Process (implementasi program/ pembelajaran/ pelatihan)	Dievaluasi pada sebagian besar artikel yang menggunakan CIPP, CIPPO, Kirkpatrick, maupun Stake's Countenance, mencakup program kurikulum, pelatihan, pendidikan inklusif, dan pendidikan vokasi.	Implementasi program umumnya berjalan baik, meskipun terdapat variasi kualitas pelaksanaan antarprogram.
Product (hasil langsung, capaian belajar, kepuasan)	Dilaporkan pada sebagian besar artikel yang menggunakan CIPP, CIPPO, dan Stake's Countenance.	Mayoritas studi melaporkan tercapainya target atau hasil program sesuai dengan tujuan yang ditetapkan.
Outcome (dampak jangka panjang, <i>employability</i> , perubahan perilaku)	Ditemukan pada sebagian kecil artikel, terutama yang menggunakan CIPPO, Kirkpatrick, dan <i>Theory of Change</i> (Hamid et al., 2025; Padli et al., 2023; Abraham et al., 2024; Lantu et al., 2021; Brooks et al., 2023).	Evaluasi <i>outcome</i> relatif terbatas dan umumnya berfokus pada perubahan perilaku atau dampak jangka pendek.

Berdasarkan Tabel 5, dimensi context, input, process, dan product merupakan dimensi yang paling banyak dievaluasi dalam artikel yang dianalisis. Sebaliknya, dimensi outcome hanya muncul pada sebagian kecil studi, terutama pada penelitian yang menggunakan model CIPPO, Kirkpatrick, dan Theory of Change.

3.7 Karakteristik Model Evaluasi

Selain mengidentifikasi distribusi penggunaan model evaluasi, studi ini juga menyintesis karakteristik masing-masing model yang ditemukan dalam artikel yang dianalisis. Sintesis dilakukan untuk membandingkan pola penggunaan, kekuatan utama, dan keterbatasan setiap model evaluasi berdasarkan temuan dari 19 artikel yang memenuhi kriteria inklusi. Ringkasan hasil sintesis tersebut disajikan pada Tabel 6.

Tabel 6. Karakteristik Model Evaluasi Program Pendidikan Hasil Sintesis Literatur

Model	Pola penggunaan di Indonesia	Kekuatan utama	Keterbatasan yang tampak
CIPP	Model evaluasi yang paling dominan dan digunakan pada berbagai konteks program pendidikan, seperti kurikulum, pendidikan nonformal, pelatihan vokasi, olahraga, dan PPG (Syam & Ishak, 2025; Idrus et al., 2025; Dewi et al., 2025; Rizal et al., 2022; Hamid et al., 2025).	Mengevaluasi aspek context, input, process, dan product secara komprehensif sehingga sesuai untuk berbagai konteks program pendidikan.	Sebagian besar studi masih bersifat deskriptif sehingga analisis hubungan kausal antar-komponen evaluasi relatif terbatas.
Logic Model	Tidak ditemukan sebagai model evaluasi formal dalam artikel yang dianalisis. Namun, sebagian prinsipnya tercermin melalui pendekatan <i>Theory of Change</i> (Brooks et al., 2023).	Memetakan hubungan antara input, aktivitas, output, dan outcome sehingga mendukung analisis jalur perubahan program.	Belum digunakan sebagai model evaluasi formal dalam artikel yang dianalisis.

Kirkpatrick	Digunakan terutama pada evaluasi pelatihan profesional dan <i>capacity building</i> (Abraham et al., 2024; Lantu et al., 2021; Marisya et al., 2019).	Mengevaluasi hasil pelatihan melalui dimensi reaction, learning, behavior, dan results.	Sebagian besar studi lebih banyak mengevaluasi level reaksi dan pembelajaran dibandingkan perubahan perilaku maupun hasil jangka panjang.
CIRO	Tidak ditemukan digunakan secara eksplisit dalam artikel yang dianalisis.	Secara konseptual mengintegrasikan evaluasi context, input, reaction, dan outcome.	Belum ditemukan penggunaannya dalam artikel yang dianalisis sehingga implementasinya dalam konteks pendidikan Indonesia belum dapat disintesis.

Berdasarkan Tabel 6, model CIPP merupakan model evaluasi yang paling dominan digunakan pada berbagai konteks program pendidikan di Indonesia karena mampu mengakomodasi evaluasi secara komprehensif terhadap aspek context, input, process, dan product. Sebaliknya, model Kirkpatrick lebih banyak diterapkan pada evaluasi pelatihan profesional dan capacity building, sedangkan Logic Model tidak ditemukan sebagai model evaluasi formal dan CIRO tidak ditemukan digunakan secara eksplisit dalam artikel yang dianalisis. Temuan ini menunjukkan adanya perbedaan karakteristik dan tingkat adopsi masing-masing model evaluasi dalam praktik evaluasi program pendidikan di Indonesia.

3.8 Faktor Pendukung dan Penghambat

Berdasarkan sintesis terhadap 19 artikel yang dianalisis, teridentifikasi sejumlah faktor yang secara konsisten mendukung maupun menghambat implementasi evaluasi program pendidikan di Indonesia. Faktor pendukung yang paling sering dilaporkan meliputi kejelasan regulasi dan keselarasan program dengan tujuan pendidikan nasional, dukungan kelembagaan, kompetensi evaluator, kolaborasi antara perguruan tinggi, sekolah, dan lembaga terkait, serta penggunaan pendekatan mixed methods dalam proses evaluasi. Sebaliknya, faktor penghambat yang paling banyak dilaporkan mencakup keterbatasan sarana prasarana, pendanaan program yang belum memadai, durasi dan cakupan evaluasi yang relatif terbatas, kesenjangan kapasitas sumber daya manusia antardaerah, serta masih terbatasnya evaluasi outcome jangka panjang, termasuk melalui tracer study, employability, dan keberlanjutan program.

4. Pembahasan

Dominasi model CIPP menunjukkan bahwa praktik evaluasi program pendidikan di Indonesia masih berorientasi pada pendekatan evaluasi yang komprehensif dan mendukung pengambilan keputusan. Karakteristik CIPP yang mengevaluasi aspek context, input, process, dan product memungkinkan model ini diterapkan pada berbagai konteks pendidikan yang memiliki kompleksitas kebijakan, sumber daya, dan implementasi yang beragam. Kondisi tersebut menjelaskan mengapa CIPP menjadi model yang paling banyak diadopsi dalam publikasi Indonesia. Interpretasi ini konsisten dengan Stufflebeam dan Coryn (2014) yang menempatkan CIPP sebagai decision-oriented evaluation model, serta sejalan dengan Fitzpatrick et al. (2022) yang menekankan keunggulan CIPP sebagai kerangka evaluasi yang komprehensif.

Selain dipengaruhi oleh karakteristik metodologisnya, dominasi CIPP juga diduga berkaitan dengan faktor historis dan kelembagaan. Model ini telah lama diperkenalkan dalam literatur evaluasi pendidikan di Indonesia dan banyak digunakan sebagai rujukan dalam penelitian akademik maupun praktik evaluasi program, sehingga menjadi kerangka yang paling dikenal oleh evaluator. Akibatnya, adopsi model alternatif berkembang lebih lambat meskipun literatur internasional menunjukkan semakin luasnya pemanfaatan Theory of Change maupun Logic Model. Kajian Riawan et al. (2025) juga menunjukkan bahwa CIPP mampu menyediakan informasi evaluatif yang mendukung pengambilan keputusan pada berbagai program pendidikan. Dalam konteks Indonesia yang memiliki keragaman geografis, sosial, dan kapasitas

kelembagaan, kemampuan CIPP mengevaluasi konteks kebijakan, kesiapan sumber daya, kualitas implementasi, dan hasil program secara terpadu menjadikan model ini lebih mudah diadaptasi dibandingkan model evaluasi lainnya (Hamid et al., 2025; Dewi et al., 2025).

Munculnya model CIPPO menunjukkan adanya perkembangan cara pandang evaluator Indonesia terhadap pentingnya dimensi outcome dalam evaluasi program. Penambahan dimensi tersebut merefleksikan kesadaran bahwa keberhasilan program tidak cukup diukur dari capaian produk (product), tetapi juga dari dampak jangka panjang setelah implementasi program selesai (Stufflebeam & Zhang, 2017). Namun demikian, hasil sintesis memperlihatkan bahwa pengukuran outcome masih terbatas pada sebagian kecil studi dan belum didukung oleh mekanisme tracer study, employability, maupun data longitudinal. Kondisi ini menunjukkan adanya kesenjangan antara perkembangan konseptual model evaluasi dengan kapasitas operasional sistem evaluasi pendidikan di Indonesia.

Terbatasnya penggunaan Kirkpatrick mengindikasikan bahwa model ini lebih sesuai diterapkan pada evaluasi pelatihan dibandingkan evaluasi kebijakan pendidikan secara luas. Fokus Kirkpatrick pada perubahan perilaku individu membuat model ini lebih sesuai untuk mengevaluasi pelatihan aparatur, guru, maupun *capacity building* (Abraham et al., 2024). Namun, dibandingkan CIPP, model ini kurang mampu menjelaskan kesiapan konteks maupun kualitas implementasi program. Interpretasi ini mendukung pandangan Kirkpatrick dan Kirkpatrick (2008) serta Fitzpatrick et al. (2022) mengenai karakteristik model evaluasi pelatihan.

Tidak ditemukannya Logic Model sebagai model evaluasi formal maupun belum digunakannya CIRO secara eksplisit menunjukkan bahwa praktik evaluasi program pendidikan di Indonesia masih didominasi oleh model yang telah mapan. Salah satu kemungkinan penyebabnya adalah dominasi penggunaan CIPP dalam lingkungan akademik Indonesia serta belum banyaknya pedoman evaluasi nasional yang merekomendasikan Logic Model maupun CIRO sebagai kerangka evaluasi utama. Meskipun demikian, munculnya pendekatan Theory of Change menunjukkan bahwa sebagian prinsip Logic Model mulai digunakan dalam menjelaskan hubungan kausal antara intervensi dan perubahan yang diharapkan (Mexdadoris & Jama, 2020). Dengan demikian, Logic Model belum diadopsi sebagai kerangka evaluasi formal, tetapi sebagian logikanya telah hadir secara implisit. Sebaliknya, Tidak ditemukannya penggunaan CIRO secara eksplisit dalam 19 artikel yang dianalisis menunjukkan bahwa model ini belum memperoleh ruang dalam publikasi evaluasi program pendidikan Indonesia pada periode kajian. Kondisi tersebut tidak serta-merta menunjukkan bahwa CIRO tidak relevan, tetapi mengindikasikan perlunya penelitian empiris untuk menguji kesesuaian model tersebut pada konteks program pendidikan dan pelatihan di Indonesia.

Pola evaluasi yang ditemukan memperlihatkan bahwa sistem evaluasi program pendidikan di Indonesia lebih berhasil mengidentifikasi kesesuaian kebijakan daripada efektivitas implementasi. Tingginya capaian pada dimensi konteks menunjukkan bahwa sebagian besar program telah dirancang sesuai kebutuhan dan kebijakan nasional (Syam & Ishak, 2025; Idrus et al., 2025; Afriadi et al., 2024). Akan tetapi, kelemahan yang berulang pada dimensi input dan proses mengindikasikan bahwa kualitas implementasi masih dipengaruhi oleh keterbatasan sarana prasarana, pendanaan, dan kapasitas kelembagaan (Hamid et al., 2025). Temuan tersebut sejalan dengan literatur evaluasi di negara berkembang yang menunjukkan bahwa keberhasilan desain kebijakan belum selalu diikuti oleh kesiapan implementasi (implementation gap) (Fitzpatrick et al., 2011; Gough et al., 2017).

Terbatasnya evaluasi outcome jangka panjang menunjukkan bahwa sebagian besar praktik evaluasi masih berorientasi pada hasil jangka pendek. Akibatnya, dampak program terhadap kesiapan kerja lulusan, keberlanjutan program, maupun pembangunan sumber daya manusia belum dapat dijelaskan secara komprehensif. Kondisi ini membatasi penggunaan hasil evaluasi sebagai dasar penyusunan kebijakan pendidikan berbasis bukti dalam jangka panjang.

Keberhasilan maupun keterbatasan praktik evaluasi program pendidikan di Indonesia tampaknya tidak hanya ditentukan oleh model evaluasi yang digunakan, tetapi juga oleh kapasitas kelembagaan yang mendukung pelaksanaan evaluasi. Berbagai studi menunjukkan bahwa kejelasan regulasi, keselarasan program dengan tujuan pendidikan nasional, dukungan institusi, kompetensi evaluator, serta kolaborasi antara perguruan tinggi, sekolah, dan lembaga pelaksana menjadi prasyarat penting bagi terlaksananya evaluasi yang berkualitas (Abraham et al., 2024; Syam & Ishak, 2025). Temuan ini mengindikasikan bahwa efektivitas suatu model evaluasi tidak dapat dilepaskan dari kesiapan organisasi dan kualitas sumber daya manusia yang mengimplementasikannya. Dengan kata lain, keberhasilan evaluasi merupakan hasil interaksi antara kerangka metodologis yang digunakan dan kapasitas institusional yang tersedia.

Sebaliknya, hambatan yang muncul secara konsisten menunjukkan bahwa tantangan utama evaluasi program pendidikan di Indonesia masih bersifat struktural. Keterbatasan sarana prasarana, pendanaan, kapasitas sumber daya manusia, serta belum tersedianya sistem pelacakan outcome jangka panjang menyebabkan sebagian besar evaluasi masih berorientasi pada pencapaian output dan implementasi program, bukan pada dampak yang berkelanjutan (Hamid et al., 2025; Idrus et al., 2025). Kondisi tersebut sejalan dengan temuan Afriadi et al. (2024) yang menunjukkan bahwa ketimpangan kapasitas antardaerah dan karakteristik desentralisasi pendidikan masih menjadi tantangan dalam menghasilkan evaluasi yang konsisten dan berbasis bukti. Oleh karena itu, penguatan sistem evaluasi nasional tidak cukup hanya melalui pemilihan model evaluasi yang tepat, tetapi juga memerlukan investasi pada pengembangan kapasitas evaluator, penguatan sistem data, serta dukungan kelembagaan agar hasil evaluasi dapat dimanfaatkan secara optimal dalam proses pengambilan keputusan pendidikan.

Perkembangan praktik evaluasi program pendidikan di Indonesia menunjukkan adanya kecenderungan menuju penggunaan pendekatan evaluasi yang lebih integratif. Kombinasi model CIPP dan Kirkpatrick pada beberapa studi mencerminkan upaya evaluator untuk mengatasi keterbatasan model tunggal dengan memanfaatkan keunggulan masing-masing pendekatan. CIPP memberikan kerangka yang komprehensif untuk mengevaluasi konteks organisasi, kesiapan sumber daya, proses implementasi, dan hasil program, sedangkan Kirkpatrick melengkapi evaluasi melalui pengukuran perubahan kompetensi, perilaku, dan dampak pelatihan pada tingkat individu. Pendekatan ini menunjukkan bahwa pemilihan model evaluasi mulai bergeser dari penggunaan satu model secara eksklusif menuju kombinasi model yang lebih adaptif sesuai karakteristik program yang dievaluasi (Bates, 2004; Engriyani, 2022). Kecenderungan tersebut sejalan dengan pandangan Fitzpatrick et al. (2022) yang menekankan bahwa tidak ada satu model evaluasi yang mampu menjawab seluruh kebutuhan evaluasi program, sehingga integrasi berbagai pendekatan sering kali menghasilkan informasi evaluatif yang lebih komprehensif dan mendukung pengambilan keputusan.

Temuan tersebut juga didukung oleh hasil analisis jaringan co-authorship yang memperlihatkan adanya kolaborasi internasional antara peneliti Indonesia dan peneliti dari institusi luar negeri, khususnya Inggris. Pola kolaborasi ini mengindikasikan bahwa perkembangan metodologi evaluasi program di Indonesia tidak berlangsung secara terisolasi, tetapi dipengaruhi oleh proses knowledge transfer dan capacity building melalui jejaring penelitian internasional. Kolaborasi semacam ini berpotensi mempercepat adopsi pendekatan evaluasi yang lebih mutakhir, meningkatkan kualitas metodologis penelitian, serta memperluas pemanfaatan model evaluasi yang lebih beragam dalam konteks pendidikan Indonesia.

Temuan penelitian ini memberikan implikasi teoretis bahwa praktik evaluasi program pendidikan di Indonesia masih didominasi oleh paradigma evaluasi yang berorientasi pada pengambilan keputusan melalui model CIPP, sementara pemanfaatan model evaluasi alternatif masih relatif terbatas. Kondisi tersebut menunjukkan bahwa diversifikasi pendekatan evaluasi masih menjadi tantangan sekaligus peluang untuk memperkaya perspektif evaluasi program

pendidikan. Integrasi model CIPP dengan pendekatan lain, seperti Kirkpatrick maupun kerangka berbasis Theory of Change, berpotensi menghasilkan evaluasi yang lebih komprehensif karena tidak hanya menilai kualitas implementasi program, tetapi juga mampu menjelaskan mekanisme perubahan serta dampak jangka panjang program. Dengan demikian, penelitian ini memperluas pemahaman mengenai posisi berbagai model evaluasi dalam konteks pendidikan Indonesia dan memberikan dasar konseptual bagi pengembangan pendekatan evaluasi yang lebih adaptif.

Dari sisi praktis, hasil sintesis menunjukkan bahwa penguatan sistem evaluasi program pendidikan tidak cukup hanya melalui pemilihan model evaluasi yang tepat, tetapi juga memerlukan dukungan kapasitas kelembagaan, kompetensi evaluator, serta ketersediaan sistem data yang mampu mendokumentasikan outcome jangka panjang secara berkelanjutan. Oleh karena itu, pengembangan pedoman evaluasi nasional, peningkatan kapasitas evaluator melalui kolaborasi akademik dan profesional, pemanfaatan teknologi digital dalam pengelolaan data evaluasi, serta pengembangan sistem tracer study nasional menjadi langkah strategis untuk meningkatkan kualitas pengambilan keputusan berbasis bukti. Selain itu, penelitian selanjutnya perlu menguji penerapan model evaluasi yang belum banyak diadopsi, seperti Logic Model dan CIRO, maupun mengembangkan pendekatan hibrida yang mengombinasikan keunggulan berbagai model sesuai dengan karakteristik program pendidikan yang dievaluasi.

Temuan penelitian ini memberikan implikasi praktis bahwa penguatan sistem evaluasi program pendidikan di Indonesia tidak hanya bergantung pada pemilihan model evaluasi yang tepat, tetapi juga memerlukan dukungan kelembagaan yang memadai. Pengembangan sistem data outcome jangka panjang, termasuk tracer study yang terintegrasi dengan sistem informasi pendidikan, menjadi penting agar dampak program dapat diukur secara berkelanjutan. Di samping itu, peningkatan kapasitas evaluator melalui pelatihan, kolaborasi akademik, dan knowledge transfer internasional perlu terus diperkuat agar kualitas metodologis evaluasi semakin meningkat. Pengembangan pedoman evaluasi nasional yang adaptif, pemanfaatan teknologi digital dalam pengelolaan data evaluasi, serta eksplorasi penggunaan model evaluasi alternatif seperti Logic Model dan CIRO maupun pendekatan hibrida yang mengombinasikan keunggulan berbagai model juga menjadi langkah strategis untuk menghasilkan evaluasi yang lebih komprehensif dan mendukung pengambilan keputusan berbasis bukti.

5. Kesimpulan

Studi literatur sistematis ini menganalisis 19 publikasi mengenai evaluasi program pendidikan di Indonesia periode 2015–2025. Hasil penelitian menunjukkan bahwa model CIPP merupakan model evaluasi yang paling dominan digunakan pada berbagai konteks program pendidikan di Indonesia. Model Kirkpatrick digunakan secara terbatas, terutama pada evaluasi pelatihan dan capacity building, sedangkan CIPPO, Theory of Change, Bradley, Stake's Countenance, dan DEM hanya muncul pada sebagian kecil studi. Logic Model tidak ditemukan sebagai model evaluasi formal dalam artikel yang dianalisis, meskipun sebagian logika evaluasinya tercermin melalui pendekatan Theory of Change. Selain itu, model CIRO tidak ditemukan digunakan secara eksplisit dalam studi yang dianalisis.

Temuan ini menunjukkan bahwa praktik evaluasi program pendidikan di Indonesia masih sangat didominasi oleh pendekatan CIPP, yang dalam berbagai studi digunakan karena kemampuannya mengevaluasi konteks, input, proses, dan produk secara komprehensif. Sebaliknya, penggunaan model evaluasi alternatif masih relatif terbatas sehingga keragaman pendekatan evaluasi belum berkembang secara optimal. Selain itu, hasil sintesis menunjukkan bahwa meskipun aspek konteks umumnya memperoleh hasil yang baik, aspek input dan proses masih menghadapi berbagai kendala, terutama terkait sarana prasarana, pendanaan, dan kapasitas sumber daya manusia.

Evaluasi outcome jangka panjang, termasuk tracer study dan pengukuran employability lulusan, masih sangat terbatas sehingga dampak jangka panjang program pendidikan belum dapat diukur secara memadai melalui bukti empiris. Oleh karena itu, diperlukan penguatan sistem evaluasi melalui pembangunan infrastruktur data outcome, peningkatan kapasitas evaluator, alokasi pendanaan khusus, pengembangan pedoman evaluasi yang adaptif, serta pemanfaatan teknologi digital. Penelitian selanjutnya disarankan mengeksplorasi pendekatan hibrid yang mengintegrasikan CIPP dengan model evaluasi alternatif sesuai karakteristik program dan tujuan evaluasi. Dengan demikian, sistem evaluasi yang lebih komprehensif dan berorientasi pada outcome diharapkan mampu menghasilkan informasi yang lebih akurat sebagai dasar pengambilan keputusan serta mendorong perbaikan berkelanjutan terhadap program-program pendidikan di Indonesia.

6. Referensi

- Abraham, I., Ridwan, A., & Triana, D. D. (2024). Evaluating integrated training for juvenile criminal justice system at the national police education and training center. *International Journal of Evaluation and Research in Education*, 13(5), 3056-3067. <https://doi.org/10.11591/ijere.v13i5.28604>
- Afriadi, B., Komarudin, K., & Sutisna, A. (2024). Evaluation of the effectiveness of the civic education curriculum in Indonesia: A case study of the KTSP curriculum, the 2013 curriculum, and the independent curriculum. *Journal of Curriculum and Teaching*, 13(5), 331-343. <https://doi.org/10.5430/jct.v13n5p331>
- Angelina, P. (2024). Evaluating English pre-service teacher professional education program implementation. *LLT Journal: Journal on Language and Language Teaching*, 27(2), 1078–1089. <https://doi.org/10.24071/llt.v27i2.7172>
- Arikunto, S., & Cepi, S. A. J. (2018). *Evaluasi program pendidikan: Panduan praktis bagi pendidik dan calon pendidik* (3rd ed.). Bumi Aksara.
- Bandu, D. J., Abdulhak, I., Wahyudin, D., Rusman, R., & Indah, R. N. (2021). Context evaluation on implementation of English for Islamic Studies program in IAIN Palu, Indonesia. *Kasetsart Journal of Social Sciences*, 42(2), 307–312. <https://doi.org/10.34044/j.kjss.2021.42.2.14>
- Bates, R. (2004). Analisis kritis praktik evaluasi: Model Kirkpatrick dan prinsip kebijakan. *Evaluasi dan Perencanaan Program*, 27(3), 341–347. <https://doi.org/10.1016/j.evalprogplan.2004.04.011>
- Brooks, H. L., Irmansyah, I., Syarif, A. K., Pedley, R., Renwick, L. J., Rahayu, A. P., Manik, C., Prawira, B., Hann, M., Brierley, H., Lovell, K., & Bee, P. (2023). Evaluating a prototype digital mental health literacy intervention for children and young people aged 11–15 in Java, Indonesia: A mixed methods, multi-site case study evaluation. *Child and Adolescent Psychiatry and Mental Health*, 17(1), 79. <https://doi.org/10.1186/s13034-023-00608-9>
- Dewi, K. S., Padmadewi, N. N., Artini, L. P., Adnyani, N. L. P. S., Santosa, M. H., & Yeo, Y. S. (2025). Inclusive education for students with autism spectrum disorder in an Indonesian bilingual elementary school: A program evaluation study using CIPP model. *International Journal of Learning, Teaching and Educational Research*, 24(4), 199-220. <https://doi.org/10.26803/ijlter.24.4.10>
- Elmizar, E., Ambiyar, A., Verawardina, U., Rizal, F., & Tee, T. K. (2022). Evaluation of educator's functional training program of teaching and learning studio (SKB) at Learning Activity Development Center (BPKB) of West Sumatra. *Educational Administration: Theory and Practice*, 28(2), 39–48. <https://doi.org/10.17762/kuey.v28i02.404>
- Engriyani, E., & Rugaiyah, R. (2022). Evaluasi Program Pendidikan dan Pelatihan Menggunakan Model Kirkpatrick Pada Institusi Pemerintah di Indonesia. *Jurnal Ilmiah Wahana Pendidikan*, 8(23), 683-692. <https://doi.org/10.5281/zenodo.7451683>

- Eryanto, H., Swaramarinda, D.R., & Nurmallasari, D.R. (2019). Effectiveness of Entrepreneurship Practice Program: Using CIPP Program Evaluation. *Journal of Entrepreneurship Education*, 22, 1.
- Fitzpatrick, J. L., Sanders, J. R., Worthen, B. R., & Wingate, L. A. (2022). *Program evaluation: Alternative approaches and practical guidelines* (5th ed.). Pearson.
- Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. SAGE Publications.
- Hamid, M. A., Siswantoyo, Gunadi, Setiawan, D., Mutiara, I., Martias, & Ismail, A. B. (2025). Lesson learned from vocational training center as a school-to-work transition program in Indonesia. *International Journal of Evaluation and Research in Education*, 14(2), 1310-1322. <https://doi.org/10.11591/ijere.v14i2.31142>
- Idrus, A., Setiyadi, B., Syarif, M. I., Wahyu, A., & Suharto. (2025). Evaluation of education program implementation at Pusat Kegiatan Belajar Masyarakat (PKBM), Jambi City. *Evaluation and Program Planning*, 111, 102535. <https://doi.org/10.1016/j.evalprogplan.2025.102535>
- Istiyani, D., Wibowo, A., Taruna, M. M., Rahmawati, T., & Atmanto, N. E. (2024). Challenges and Opportunities in Early Childhood Religious and Moral Education: A Perspective from the Evaluation of Logical Models. *Nazhruna: Jurnal Pendidikan Islam*, 7(2), 233-249. <https://doi.org/10.31538/nzh.v7i2.4843>
- Kaewkumkong, A., & Kaewkamkong, U. (2024). Enhancing ASEAN Community educational cooperation in Thailand through a school partnership program with Indonesia: Policy performance, choices, and recommendations. *Kasetsart Journal of Social Sciences*, 45(1), 1-10. <https://doi.org/10.34044/j.kjss.2024.45.1.01>
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2008). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.
- Lantu, D. C., Labdhagati, H., Razanaufal, M. W., & Sumarli, F. D. (2021). Was the training effective? Evaluation of managers' behavior after a leader development program in Indonesia's best corporate university. *International Journal of Training Research*, 19(1), 77-92. <https://doi.org/10.1080/14480220.2020.1864446>
- LeCroy, Craig Winston. (2021). 'Logic Models' in Cynthia Franklin (ed.), *Encyclopedia of Social Work*. <https://doi.org/10.1093/acrefore/9780199975839.013.1273>
- Marisya, R., Akbar, M., & Asmawi, M. (2019). Education and training evaluation for apparatus in regional public hospital. *International Journal of Innovative Technology and Exploring Engineering*, 8(6C2), 126-130.
- Mexdadoris, M., & Jama, J. (2020). The Effectiveness of the Logic Evaluation Model in K-13 Curriculum Analysis. *Journal of Education Research and Evaluation*, 4(3), 312-318. <https://doi.org/10.23887/jere.v4i3.29004>
- Padli, P., Kiram, Y., Sin, T. H., Iyakrus, I., Raja Azidin, R. M. F., Denay, N., & Rudyanto, R. (2023). Evaluation of the sepak takraw training program in the Student Sports Education and Training Center (SSETC). *Journal of Physical Education and Sport*, 23(Art 381), 3332-3340. <https://doi.org/10.7752/jpes.2023.12381>
- Page, MJ, McKenzie, JE, Bossuyt, PM, Boutron, I., Hoffmann, TC, Mulrow, CD, Shamseer, L., Tetzlaff, JM, Akl, EA, Brennan, SE, Chou, R., Glanville, J., Grimshaw, JM, Hróbjartsson, A., Lalu, M. M., Li, T., Loder, EW, Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). Pernyataan PRISMA 2020: Pedoman yang diperbarui untuk melaporkan tinjauan sistematis. *Kedokteran PLOS*, 18(3), e1003583. <https://doi.org/10.1371/journal.pmed.1003583>
- Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using a logic model. *Prevention Forum*, 20(2), 3-14.

- Rahmat, Z., & Ambiyar. (2025). Model evaluasi CIPP dalam program sekolah: Systematic literature review. *Jurnal Kajian Ilmu Pendidikan (JKIP)*, 5(4). <https://doi.org/10.11591/ijere.v13i4.28014>
- Riawan, I. M. O., Citariani, N. M., & Lokasanti, I. K. (2025). Efektivitas Model Evaluasi CIPP dalam Program Pendidikan: Kajian Literatur Sistematis. *Jurnal Evaluasi Pendidikan*, 16(2), 108–115. <https://doi.org/10.21009/jep.v16i2.58457>
- Rogers, P (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. 14(1). <https://doi.org/10.1177/1356389007084674>
- Sari, L. I., Rukmini, D., Faridi, A., & Mujiyanto, J. (2024). English for specific purposes curriculum evaluation from the social semiotic perspective. *International Journal of Evaluation and Research in Education*, 13(4), 2719-2730. <http://doi.org/10.11591/ijere.v13i4.28014>
- Sholeh, M., & Trihantoyo, S. (2019). Evaluating teachers' assignment implementation as school principals preparation program in Indonesia. In *Proceedings of the 3rd International Conference on Education Innovation (ICEI 2019)* (pp. 134–137). Atlantis Press. <https://doi.org/10.2991/icei-19.2019.36>
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education*, 5(1), 19-25.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). Jossey-Bass.
- Stufflebeam, D. L., & Zhang, G. (2017). *The CIPP evaluation model: How to evaluate for improvement and accountability*. Guilford Press.
- Syam, A. S. M., & Ishak. (2025). Bridging policy and practice: Evaluating the curriculum effectiveness in fostering religious moderation and 21st-century competencies. *Nazhruna Jurnal Pendidikan Islam*, 11(2), 403-414. <https://doi.org/10.15575/jpi.v11i2.49596>
- Turmuzi, M., Ratnaya, I. G., Al Idrus, S. W., Paraniti, A. A. I., & Nugraha, I. N. B. S. (2022). Literature Review: Evaluasi Keterlaksanaan Kurikulum 2013 Menggunakan Model Evaluasi CIPP (Context, Input, Process, dan Product). *Jurnal Basicedu*, 6(4), 7220–7232. <https://doi.org/10.31004/basicedu.v6i4.3428>
- Wandi Wandu, Lona Mardiaty, Akma Khairun Nisa, Ahmad Sabri, & Yusran Lubis. (2024). Evaluasi dan Akuntabilitas dalam Manajemen Pendidikan : Strategi untuk Meningkatkan Kualitas Pembelajaran. *Jurnal Manajemen Dan Pendidikan Agama Islam*, 3(1), 42–51. <https://doi.org/10.61132/jmpai.v3i1.820>
- Widoatmodjo, S., Hamzah, D., Sudirman, I., Mardiana, R. (2021). Assessment of The Competency-Based Economies Through Formation of Entrepreneurs (Cefe) Training Evaluation Model. *International Journal of eBusiness and eGovernment Studies*, 13 (2), 85-102.
- Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training: A practical framework, with cases, for evaluating training needs and results*. Gower Press.
- Yuliansari, Jaedun, A., Aminah, S., & Rizaldi (2019). Evaluation of English Subject Learning of Senior High School in Mataram City, Indonesia. *International Journal of Scientific & Technology Research*, 8, 370-373.