Evaluating Curriculum and Materials Implementation in English Language Teaching: A Review of CIPP, Kirkpatrick, and Checklist-Based Evaluation Models

Herlyna Cahya Ningrum, Grace Ester Napitupulu
English Education, Faculty of Language and Arts,
Universitas Negeri Semarang
hlina641@students.unnes.ac.id
DOI: https://doi.org/10.15294/eltlt.v1i1.620
QRCBN 62-6861-2530-756

ABSTRACT
This study reviews the evaluation of curriculum and instructional materials in English Language Teaching (ELT) using three models: CIPP, Kirkpatrick, and checklist-based evaluation. Using a Systematic Literature Review (SLR) guided by PRISMA, studies from 2010 to 2024 were collected from Scopus, ERIC, and Google Scholar, and analyzed thematically. The review demonstrates that the CIPP model is widely utilized to evaluate context, resources, implementation processes, and outcomes comprehensively. Kirkpatrick's model is mainly applied in teacher training and professional development to assess reaction, learning, behavior, and results. Checklist-based evaluation is widely used for textbook and materials selection due to its practicality, but it may be limited if not adapted to local teaching contexts. Common challenges reported across studies include time constraints, limited evaluator training, resource shortages, and misalignment between curriculum goals and available materials. The review recommends integrating these models to support more balanced, context-sensitive, and evidence-based evaluation of ELT curriculum and materials implementation.

Keywords: English Language Teaching (ELT); curriculum evaluation; instructional materials; CIPP model; Kirkpatrick model

INTRODUCTION

Educational systems worldwide have experienced continuous curriculum reforms in response to globalization, technological advancement, and the growing demand for communicative competence in English language teaching (ELT). Curriculum reform is commonly accompanied by the development and adoption of new instructional materials intended to align classroom practices with updated learning objectives, competency standards, and learner-centered pedagogies (Richards, 2017; Nation & Macalister, 2010). However, the effectiveness of such reforms depends not only on curriculum design but also on how the curriculum and materials are implemented in authentic classroom contexts.

Evaluating the implementation of curriculum and instructional materials is therefore crucial in educational settings, particularly in ELT, where instructional quality directly influences language learning outcomes. Curriculum evaluation provides systematic information about the value, effectiveness, and feasibility of educational programs, supporting informed decision-making for improvement (Stufflebeam & Zhang, 2017). Similarly, materials evaluation helps determine whether textbooks and learning resources are appropriate for learners' needs, curricular goals, and contextual constraints (Tomlinson, 2013). Without rigorous evaluation, curriculum reforms risk remaining policy-level intentions rather than meaningful pedagogical change.

Despite well-articulated curriculum frameworks, numerous challenges arise when translating curriculum policy into classroom practice. Teachers may encounter limited resources, inadequate professional development, time constraints, or mismatches between prescribed curriculum goals and available instructional materials (Fullan, 2007; Wedell, 2009). In ELT contexts, these challenges often result in gaps between the intended curriculum, the implemented curriculum, and the actual learning outcomes. Such

discrepancies highlight the importance of evaluation models that can systematically capture contextual conditions, implementation processes, and learning results.

To address these complexities, various evaluation models have been applied in educational research. The Context, Input, Process, and Product (CIPP) model offers a comprehensive framework for evaluating educational programs by examining needs, resources, implementation, and outcomes (Stufflebeam, 2003). Kirkpatrick's four-level evaluation model, initially developed for training programs, has also been adapted to educational contexts to assess reactions, learning, behavior, and results (Kirkpatrick & Kirkpatrick, 2006). In addition, checklist-based evaluation models are frequently used in ELT materials evaluation due to their practicality and focus on specific criteria such as content, language use, methodology, and cultural appropriateness (McGrath, 2016). However, previous studies vary in how these models are applied, combined, or adapted across contexts.

Given the expanding body of research on curriculum and materials evaluation, a systematic literature review is necessary to synthesize existing evidence, identify dominant evaluation practices, and reveal methodological and conceptual gaps. Systematic literature reviews enable researchers to analyze previous studies in a transparent, replicable, and comprehensive manner, providing an overview of trends and research directions within a field (Kitchenham & Charters, 2007; Carrera-Rivera et al., 2022). Therefore, this study aims to systematically review research on the evaluation of curriculum and instructional materials implementation in ELT using the CIPP model, Kirkpatrick's model, and checklist-based evaluation approaches. The study aims to address the following research question: *How have curriculum and materials implementation in English language teaching been evaluated in previous studies?* The findings are expected to contribute theoretical insights into evaluation practices and offer practical guidance for educators, curriculum developers, and policymakers in improving ELT implementation.

Curriculum implementation refers to the process of translating designed curriculum plans into actual classroom practices. It involves teachers' interpretation and adaptation of

curriculum documents, materials, and objectives to meet students' needs within the context of specific constraints. According to Afriadi et al. (2024) and Hoang et al. (2020), curriculum implementation is not a static or mechanical process but a dynamic interaction among curriculum intent, teachers' agency, institutional support, and contextual realities. It emphasizes not only achieving predefined learning outcomes but also ensuring alignment between policy intentions and the practical realities of the classroom.

Several models describe how the curriculum is executed in educational settings. Traditional objective-oriented approaches, such as Tyler's model, emphasize measurable behavioral outcomes and accountability. In contrast, Stake's responsive model focuses on understanding stakeholder perceptions and contextual nuances of teaching and learning processes, aligning with constructivist paradigms. The CIPP (Context, Input, Process, Product) model, developed by Stufflebeam, offers a comprehensive and systematic framework for curriculum evaluation and implementation, integrating goal achievement with contextual and process-based reflection. Additionally, contemporary curriculum implementation frameworks emphasize participatory and adaptive approaches, encouraging teachers and learners to co-construct meaning and reflect on outcomes to facilitate continuous improvement.

The effectiveness of curriculum implementation depends on teacher competence, the policy environment, and contextual factors. Teachers' pedagogical skills, attitudes, and evaluation literacy significantly shape how curriculum objectives are operationalized (Yang & Li, 2020). Institutional policies, available resources, and leadership support are also crucial in maintaining curriculum coherence and quality (Uljens, 2018). Moreover, socio-cultural contexts influence how learning goals are localized, emphasizing the need for culturally responsive curricula that reflect learners' linguistic and social realities. Contextual constraints, such as class size, infrastructure, and community expectations, also influence the effectiveness with which curriculum goals are achieved.

Instructional materials are essential mediators between curriculum design and classroom learning. They serve as tools

that translate abstract curriculum objectives into concrete learning experiences. As Saddhono (2018) explains in the context of TISOL (Teaching Indonesian to Speakers of Other Languages), well-designed materials enhance learner motivation, engagement, and comprehension, directly contributing to achieving curriculum goals. In English Language Teaching (ELT), materials facilitate the integration of linguistic input, communicative practice, and cultural context (Kramsch, 2000), ensuring that learning outcomes align with curriculum expectations.

Instructional materials can be categorized into print-based (textbooks, workbooks), digital (multimedia, online platforms), and context-based thematic resources. Saddhono (2018) emphasizes the value of thematic and culturally integrated materials, which connect linguistic learning with cultural content, thereby enhancing relevance and comprehension. Similarly, Alemi and Sadehvandi (2012) highlight that effective ELT materials should reflect learners' needs, local contexts, and institutional goals, ensuring alignment with communicative and task-based methodologies.

The alignment between curriculum objectives and instructional materials is critical for coherence and quality assurance. Checklist-based evaluations of ELT textbooks (e.g., AbdelWahab, 2013; Alemi & Sadehvandi, 2012) demonstrate that materials must correspond to syllabus objectives, teaching methodologies, and learners' proficiency levels. Misalignment may result in gaps between expected outcomes and classroom realities, underscoring the need for ongoing material evaluation and adaptation to ensure alignment.

Evaluation serves to determine the effectiveness, relevance, and impact of curriculum and instructional materials. According to Stufflebeam & Shinkfield (2007), the primary purpose of evaluation is not merely to prove effectiveness but to improve educational practice. Evaluation informs decision-making, accountability, and continuous development, encompassing both formative assessment (for improvement during implementation) and summative evaluation (for assessing overall effectiveness and sustainability)

The CIPP Model (Stufflebeam) remains one of the most widely applied frameworks for evaluating both curricula and ELT programs. It examines four dimensions: Context (needs analysis), Input (resources and strategies), Process (implementation monitoring), and Product (outcomes and impact)—providing a comprehensive view of program performance.

The Kirkpatrick Model, originally developed for training evaluation, complements the CIPP model by assessing learning outcomes across four levels: reaction, learning, behavior, and results, making it particularly useful in contexts such as teacher training and professional development.

Finally, Checklist-based evaluation models, such as those developed by Işık & Atmışdört (2010) and AbdelWahab (2013), offer structured instruments for evaluating the quality, practicality, and contextual relevance of ELT materials. These checklists facilitate consistency, validity, and reliability in material selection and adaptation. Common indicators across the CIPP, Kirkpatrick, and checklist-based models include:

- Context indicators include learner needs, institutional policy, and cultural and social context.
- Input indicators include the adequacy of resources, teacher qualifications, and instructional materials.
- Process indicators include fidelity of implementation, teaching methods, and learner engagement.
- Product indicators: achievement of learning outcomes, satisfaction, and sustainability jeehp-16-40.

These indicators support both qualitative and quantitative evaluation approaches, ensuring comprehensive monitoring and evidence-based curriculum refinement.

METHODS

This study employed a qualitative research approach, utilizing a Systematic Literature Review (SLR) design, to examine the evaluation of curriculum and instructional materials in English Language Teaching (ELT) using the CIPP, Kirkpatrick, and checklist-based evaluation models. Qualitative approaches are appropriate for synthesizing conceptual

patterns, methodological trends, and interpretive findings across existing studies rather than measuring statistical effects (Creswell & Poth, 2018).

The SLR design was selected because it enables a transparent, structured, and replicable process for identifying, evaluating, and synthesizing relevant literature within a defined research scope (Kitchenham & Charters, 2007). In educational research, SLRs are particularly valuable for mapping research trends, identifying dominant theoretical frameworks, and revealing research gaps related to curriculum and instructional practices (Petticrew & Roberts, 2006). By systematically reviewing prior studies, this research aims to provide an evidence-based overview of evaluation practices used in the implementation of ELT curriculum and materials.

The review process followed established systematic review guidelines, drawing primarily on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to guide the stages of identification, screening, eligibility, and inclusion (Page et al., 2021). Although PRISMA was initially developed for health research, it has been widely adopted in educational and social science research due to its clarity and methodological rigor (Moher et al., 2009).

The following research questions guided the review:

1. How has curriculum and instructional materials implementation in English Language Teaching been evaluated in previous studies?
2. What evaluation models (CIPP, Kirkpatrick, and checklist-based models) are most frequently used in ELT contexts?
3. What strengths and limitations are reported in the application of these evaluation models?

Relevant studies were retrieved from multiple academic databases, including Scopus, ERIC, and Google Scholar, to ensure broad coverage of peer-reviewed literature in education and applied linguistics. These databases were selected due to their extensive indexing of ELT, curriculum studies, and educational evaluation research (Gusenbauer & Haddaway, 2020).

The review focused on publications from 2010 to 2024 to

capture recent developments in curriculum reform and evaluation practices. The types of publications included peer-reviewed journal articles and conference proceedings that reported empirical or systematic review studies related to the implementation of curriculum and instructional materials in educational contexts. A systematic search strategy was developed using combinations of keywords and Boolean operators. The main search strings included terms such as *curriculum implementation*, *materials implementation*, *English language teaching*, *program evaluation*, the *CIPP model*, the *Kirkpatrick model*, and *materials evaluation checklists*. These keywords were adapted to the syntax requirements of each database.

The inclusion criteria were:

1. Peer-reviewed publications;
2. Studies focusing on curriculum and/or instructional materials implementation;
3. Research conducted in educational contexts, particularly ELT or EFL settings.
4. Studies employing or discussing evaluation models relevant to CIPP, Kirkpatrick, or checklist-based approaches.

The exclusion criteria were:

1. Opinion papers, editorials, or non-empirical commentaries;
2. Studies conducted outside educational contexts;
3. Publications with incomplete methodological descriptions or inaccessible full texts.

The study selection process consisted of three main stages: identification, screening, and eligibility assessment. During the identification stage, all records retrieved from the databases were compiled, and duplicates were removed. In the screening stage, titles and abstracts were reviewed to assess relevance based on the inclusion and exclusion criteria. Full-text articles were then examined in the eligibility stage to determine their suitability for final inclusion.

Through this process, a total of 12 studies were selected for the final review. The selection process was documented to

ensure transparency and replicability in accordance with SLR standards (Kitchenham & Charters, 2007). Data extraction was conducted using a structured form to ensure consistency across studies. The extracted data included the author(s), year of publication, research context, evaluation model used, research methodology, and key findings related to the implementation of curriculum or materials. The analysis employed thematic analysis to identify recurring patterns and themes across the selected studies (Braun & Clarke, 2006). An initial coding process was conducted to categorize key concepts related to evaluation purposes, implementation challenges, and model effectiveness. These codes were then synthesized into broader themes, allowing for systematic comparison and interpretation of evaluation practices across different ELT contexts.

To enhance the trustworthiness of the review, several strategies were employed. First, the review process was documented transparently, including search strategies, selection criteria, and analysis procedures. Second, consistency and rigor were maintained through the use of predefined inclusion and exclusion criteria. Where applicable, the study selection and coding processes were cross-checked to minimize bias and enhance reliability (Lincoln & Guba, 1985). These measures contribute to the credibility and dependability of the review findings.

RESULTS AND DISCUSSION

Trends in Curriculum Implementation Evaluation

Trends in Curriculum Implementation Evaluation

Curriculum implementation evaluation involves assessing how educational programs are delivered in practice, with a focus on alignment with goals, stakeholder needs, and outcomes. Recent trends emphasize the integration of technology, data-driven decision-making, and adaptive professional development to enhance effectiveness.

Key Trends

High-performing districts prioritize instructional coherence by integrating curriculum, assessments, and data on

unified platforms, such as Kiddom, ensuring that lessons align with prioritized learning goals. AI tools for generating differentiated practice, auto-feedback, and lesson adaptations are gaining traction, reducing teacher workload while personalizing instruction.

Evaluation Models

Formative evaluations during implementation drive continuous improvements, while summative assessments measure final impacts for accountability. Models such as the CIPP (Context, Input, Process, Product) framework and responsive stakeholder approaches provide holistic insights, combining quantitative and qualitative methods.

Professional Learning Shifts

Targeted, job-embedded training tied to specific curricula, such as Illustrative Mathematics or OpenSciEd, enhances teacher confidence and student outcomes more effectively than generic professional development. Districts use data protocols and pacing guides to tailor support based on real-time classroom insights.

Evaluation of Materials Implementation

The evaluation of materials implementation assesses how effectively teaching resources, such as textbooks or digital aids, are used in educational settings to achieve learning objectives. This process identifies strengths, gaps, and areas for improvement during the deployment phase. Common frameworks, such as the CIPP model, provide comprehensive guidance for this evaluation.

Key Evaluation Models

The CIPP model (Context, Input, Process, Product) assesses material implementation by examining the needs and environment (context), available resources (input), delivery methods (process), and outcomes (product). In ELT contexts, it reveals issues such as theory-practice gaps or inadequate preparation in programs like school internships. The

Discrepancy Evaluation Model identifies gaps between expected and actual results, prompting corrective actions for teaching methods or resources.

ELT-specific checklists

A three-stage system is commonly used for evaluating ELT materials: screening for quick disqualification, detailed content and usability analysis, and practical testing. Checklists assess practicality, reliability, and alignment with learner needs, and pilots have reported high satisfaction rates (up to 94% reliability). Clear goals, democratic conception, and continuous evolution based on stakeholder input are some of the criteria.

Best Practices.

- Conduct a context analysis to ensure that materials are appropriate for students' needs and societal demands.
- Before the full rollout, pilot materials will be used to test engagement and resource adequacy.
- For ongoing feedback, use a combination of methods, including surveys, observations, and data analytics, to gather insights.
- Measure outcomes based on student performance and teacher attitudes toward implementation fidelity.

Discussion

Teachers are consistently identified as key actors in curriculum and materials evaluation across the reviewed studies; however, their roles are frequently limited by institutional structures and workload. According to studies that employ the CIPP model, teachers play a crucial role in providing contextual and process-related information, as they possess firsthand knowledge of classroom realities and learner needs (Karim, 2025). Teachers who participate in CIPP-based evaluations view the model as comprehensive and improvement-oriented, particularly in identifying mismatches between curriculum goals and actual classroom practices.

Nonetheless, teachers frequently perceive CIPP as time-consuming, complex, and data-collection intensive, which limits its widespread use. In contrast, checklist-based evaluation studies show that teachers generally hold positive attitudes toward checklists due to their practicality, clarity, and ease of use (Işik, 2018). Checklists are regarded as particularly useful for screening and selecting ELT materials, allowing teachers to make more timely and systematic decisions.

However, the literature suggests that when teachers rely solely on checklists, evaluations can become mechanical and superficial. Teachers emphasize the importance of professional judgment and contextual adaptation when using checklist-based models, implying that evaluative competence is just as important as the instrument itself. Overall, the studies' teacher perceptions indicate a preference for evaluation tools that strike a balance between theoretical rigor and practical feasibility.

Several recurring challenges emerge from the four studies reviewed. Limited evaluation time, insufficient model training, and disconnects among evaluation frameworks and local ELT contexts are all common issues. Teachers frequently lack formal preparation for using comprehensive models like CIPP, resulting in partial or inconsistent implementation.

Another issue is overreliance on summative evaluation results, particularly in Kirkpatrick- and checklist-based approaches, which may overlook classroom processes and contextual variables. The literature also discusses the difficulties in maintaining objectivity and validity when evaluation instruments are not tailored to specific teaching contexts.

Given the challenges, several best practices have been identified. The studies recommend combining evaluation models to capitalize on their respective strengths, for example, integrating the holistic perspective of CIPP with the efficiency of checklist-based tools (Işik, 2018).

The findings confirm previous views that teachers are critical to the success of curriculum and materials implementation. However, the literature reveals an ongoing tension between teachers' evaluative potential and their

limited participation in decision-making processes. While models like CIPP theoretically emphasize stakeholder involvement, in practice, teachers are frequently seen as data providers rather than evaluative decision-makers. This suggests a mismatch between the conceptual design of evaluation models and their implementation in institutional settings.

Furthermore, the results show that teachers' preferences for evaluation models are heavily influenced by practicality and feasibility. Although comprehensive models like CIPP are valued for their depth and diagnostic capability, their complexity and time requirements frequently make them unsuitable for long-term use.

In contrast, checklist-based evaluations are preferred for their efficiency, but they risk compromising instructional quality when used without local adaptation. This provides support to the argument that evaluation effectiveness is determined not only by the model itself, but also by teachers' evaluation literacy and institutional support.

CONCLUSION

Evaluation Models: The CIPP model is highly effective at providing a comprehensive overview of the context, resources, processes, and outcomes of ELT programs. The Kirkpatrick model focuses on evaluating teacher training and professional development. Meanwhile, checklist-based evaluation is the preferred method for textbook selection due to its utility, though it runs the risk of becoming superficial if not tailored to the local context. Time constraints, a lack of evaluator (teacher) training, insufficient resources, and a misalignment between curriculum objectives and available materials frequently impede the effective implementation of evaluation.

To achieve a more balanced and accurate evaluation, it is recommended to integrate these evaluation models (for example, combining the depth of CIPP with the efficiency of checklists) and enhance educators' evaluation literacy, allowing them to make contextual adjustments.

REFERENCES

Agustina, N. Q., & Mukhtaruddin, F. (2019). The Cipp Model-Based Evaluation on Integrated English Learning (IEL) Program at Language Center. *English Language Teaching Educational Journal*, *2*(1), 22. https://doi.org/10.12928/eltej.v2i1.1043

Carrera-rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). MethodsX How-to conduct a systematic literature review : A quick guide for computer science research. *MethodsX*, *9*, 101895. https://doi.org/10.1016/j.mex.2022.101895

Gifari, M. K., Arifin, Z., Indonesia, U. P., & Bandung, K. (2024). *Inovasi Kurikulum*. *21*(3), 1747–1762.

Işik, A. (2018). ELT Materials Evaluation: A System and Criteria. *Theory and Practice in Language Studies*, *8*(7), 797. https://doi.org/10.17507/tpls.0807.11

Jispendiora, J., No, V., Karakter, P., Sekolah, D. I., Norlita, D., Nageta, P. W., & Faradhila, S. A. (2023). *SYSTEMATIC LITERATURE REVIEW (SLR) : Peran Pendidikan memiliki dampak yang sangat penting dalam perkembangan dan pengembangan diri individu.*

Karim, Md. Z. (2025). Applying CIPP Model for Syllabus Evaluation: Assessing the MA in ELT Syllabus of a Private University in Bangladesh. *International Journal of Research and Innovation in Social Science*, *IX*(II), 1406–1414. https://doi.org/10.47772/ijriss.2025.9020114

Lee, S. Young, Shin, J.-S., & Lee, S.-H. (2019). How to execute Context, Input, Process, and Product evaluation model in medical health education. *Journal of Educational Evaluation for Health Professions*, *16*(40), 40.

Made, L., Wedayanthi, D., Pradnyan, P. B., & Adiwijaya, P. A. (2024). *EDUHUMANIORA : Jurnal Pendidikan Dasar The Implementation of CIPP as Evaluation Model on Teaching Assistance in SD Bali Bilingual School as The Practice of Kurikulum Merdeka*. *16*(1), 37–44.

Model, C., & Evaluasi, D. (2021). *Mixed Model Cipp Dan Kickpatrick Sebagai Pendekatan Evaluasi Program Pengembangan Kompetensi Guru Berbasis Kebutuhan Peningkatan Kemampuan Asessement Literasi - Numerasi*. *7*(1), 203–223.

Saleh, Y., Mahat, H., Hashim, M., Nayan, N., Suhaili, S., Khairul, M., Ghazali, A., & Hayati, R. (2021). *A Systematic Literature Review (SLR) on the Development of Sustainable Heritage Cities in Malaysia*. *32*(3), 290–310. https://doi.org/10.5614/jpwk.2021.32.3.6

Thesis, A., & Zainuddin, A. H. A. (2025). *EVALUATING THE IMPLEMENTATION OF THE MERDEKA CURRICULUM IN ENGLISH TEACHING AT MAN PALOPO ENGLISH EDUCATION STUDY PROGRAM POSTGRADUATE EVALUATING THE IMPLEMENTATION OF THE MERDEKA CURRICULUM IN ENGLISH TEACHING AT MAN PALOPO POSTGRADUATE*.