Vol. 12, No. 1, 2026, pp. 197-206
DOI: https://doi.org/10.29210/1202627238

Contents lists available at Journal IICET

Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)
ISSN: 2476-9886 (Print) ISSN: 2477-0302 (Electronic)

Journal homepage: https://jurnal.iicet.org/index.php/j-edu

Documentary evaluation of formal obe alignment and operational traceability in english teacher education

Sri Agriyanti Mestari*), Karmila Machmud, Sayama Malabar, Lukman A.R. Laliyo, Mohamad Zubaidi
Universitas Negeri Gorontalo, Indonesia

Article Info

Article history:

Received Aug 15th, 2026
Revised Sep 28th, 2026
Accepted Oct 10th, 2026

Keyword:

English teacher education,
Curriculum alignment,
Operational traceability,
Outcome-based education

ABSTRACT

Outcome-Based Education (OBE) requires coherent links between intended outcomes, learning activities, assessment, and evidence. This evaluative document analysis examined formal alignment and operational traceability in an Indonesian English teacher education curriculum and five purposively selected course plans. Five experts rated 33 indicators across 11 dimensions; a supplementary nine-feature audit assessed documentary links across courses. The Constructive Alignment Index was 79.9% (659/825), with Kendall's W = .662 indicating rater concordance. Outcome specificity (93.3%) and graduate-profile traceability (90.7%) were relatively strong, whereas formative feedback (64.0%) and differentiated learning opportunities (48.0%) were weaker. All five plans contained rubrics, but explicit outcome-specific assessment criteria links appeared in only one; four were partially documented. Feedback-to-revision procedures were partial in all five. The findings distinguish the presence of OBE components from their auditable documentary continuity. They support improved outcome coding, assessment mapping, and feedback design but do not establish classroom implementation or student attainment.

© 2026 The Authors. Published by IICET.
This is an open access article under the CC BY-NC-SA license
(https://creativecommons.org/licenses/by-nc-sa/4.0)

Corresponding Author:

Sri Agriyanti Mestari,
Universitas Negeri Gorontalo,
Email: sri.mestari@ung.ac.id

Introduction

Outcome-Based Education (OBE) has become an influential approach to curriculum design and quality assurance in higher education because it shifts the primary focus of education from what institutions provide and teachers deliver to what students are ultimately able to demonstrate. In an outcome-oriented system, curriculum content, learning experiences, teaching strategies, and assessment practices are expected to contribute coherently to clearly articulated learning outcomes. Spady (1994) conceptualizes this orientation through four interrelated principles: clarity of focus, design down, high expectations, and expanded opportunities. These principles require educational programs to begin with significant intended outcomes and to organize curriculum, instruction, and assessment backwards from those outcomes. More recent formulations similarly emphasize that curriculum, instruction, and assessment should be driven by the learning outcomes students are expected to demonstrate at the end of a course or program (Rao, 2020). Accordingly, the formal specification of learning outcomes alone is insufficient. Effective OBE implementation depends on whether intended outcomes can be consistently traced through curriculum structures, course-level learning experiences, assessment practices, and evidence of student achievement.

At least three contemporary empirical accounts caution against treating declared OBE alignment as evidence of educational impact. Katawazai (2021) documented implementation challenges in public universities despite the adoption of OBE and student-centred learning. Mufanti, Carter, and England (2024) reported gaps between Indonesian tertiary English teachers' reported understanding of OBE and their ability to formulate measurable outcomes and aligned instructional components. Pham and Nguyen (2024) described higher-education changes that could remain primarily formal or documentary. These studies address different institutional settings and research questions; none directly tests whether the particular documents examined here caused improvements in student achievement. They nevertheless justify examining the distance between stated curriculum intentions and the specificity of the relationships documented in course plans. Indonesian English-teacher-education work also reports implementation challenges (Allo, Sudarsi, & Taula'bi', 2024).

The distinction between formal and operational alignment is therefore central to the present study. Formal alignment refers to relationships that are administratively represented through curriculum matrices, course mappings, or semester learning plans. Operationally traceable alignment requires stronger evidence: an intended outcome must be traceable through increasingly specific curriculum components to learning activities, assessment tasks, criteria or rubrics, identifiable evidence of student performance, and, where documented, subsequent feedback and improvement processes. This perspective extends conventional curriculum-mapping analysis by asking not only where an outcome is represented, but also how the intended competence is developed, assessed, evidenced, and supported for further improvement.

Operational traceability is an applied documentary construct grounded in constructive alignment (Biggs, 1996; Biggs, Tang, & Kennedy, 2022), curriculum mapping (Harden, 2001) and alignment evaluation (Martone & Sireci, 2009). It asks whether a reviewer can follow intended outcomes through learning tasks to assessment criteria and planned evidence, with feedback-to-revision treated as an additional link. This is a diagnostic framework, not a new causal theory or evidence of student attainment. Formalization should not become mechanical compliance (Gurukkal, 2020).

This study examined the English Language Education Study Program at Universitas Negeri Gorontalo, Indonesia, asking how formal OBE alignment and operational traceability are documented across the curriculum and five purposively selected RPS. It assessed the 33-indicator CAI, described strengths and weaknesses across 11 dimensions, and compared five course plans. The contribution is a replicable document-audit distinction, not an estimate of the effect of OBE on learning. Contemporary curriculum and alignment work provides related methodological perspectives (Derouich, 2025; Pereira, Nsair, Pereira, & Grant, 2024; Smith & Chittams, 2024).

The study therefore makes a bounded contribution: it operationalizes a formal-versus-traceable distinction within a reproducible document-audit procedure and illustrates why the presence of OBE components must be evaluated separately from their cross-document consistency and assessment-specific links. It does not estimate the causal effect of alignment on attainment. Alternative explanations for incomplete documentation include institutional quality-assurance templates, document revision cycles, academic workload, assessment literacy, and practices that occur but are not recorded; the present design cannot discriminate among these mechanisms.

Method

This evaluative document-analysis study examined intended curriculum alignment in the English Language Education Study Program, Faculty of Letters and Culture, Universitas Negeri Gorontalo, Indonesia, using documents collected during January–December 2025. The analysis combined descriptive document review, structured ratings, and qualitative tracing of links among learning outcomes, learning activities, assessment, and feedback. The documentary unit of analysis was the official program curriculum and five purposively selected course plans; the quantitative scoring unit was each indicator-by-rater judgment, rather than an individual student or course.

The documentary corpus comprised the official program curriculum, its curriculum mapping, and five Semester Learning Plans (Rencana Pembelajaran Semester, RPS): English Intensive Course, English Poetry, Semantics and Pragmatics, Speaking for Professional Context, and Curriculum and Material Development. Purposive selection sought variation in competency domains and available documentary evidence: integrated language skills, literary interpretation, linguistic analysis, professional oral communication, and pedagogical curriculum/materials design. The five RPS formed a purposive documentary sample rather than a probability sample of all courses. The review also considered assessment plans, rubrics or criteria, and supporting documentary evidence when available within the collected corpus. Accordingly, cross-course comparisons are diagnostic and cannot establish the prevalence of a pattern across the entire program.

The analysis proceeded in two interconnected stages. First, the official curriculum was mapped to identify graduate profiles, PLOs, the curriculum structure, course contributions, and stated design-down relationships; the five RPS were then examined for CLOs, sub-CLOs, activities, assessment tasks, criteria, and documented follow-up opportunities. Mapping recorded declared relationships without assuming that their presence established substantive alignment. Curriculum-mapping research similarly emphasizes the visibility of relationships among outcomes, content, and assessment (Okojie, Bastas, & Miralay, 2022; Reniers, Mathany, Farkas, Pollock, & Husband, 2022). Second, the mapped relationships were evaluated using the structured indicator instrument and documentary evidence tracing described below. These stages are sequential in purpose but complementary in interpretation: the first locates declared links, while the second evaluates their specificity and traceability. Figure 1 summarizes this procedure.

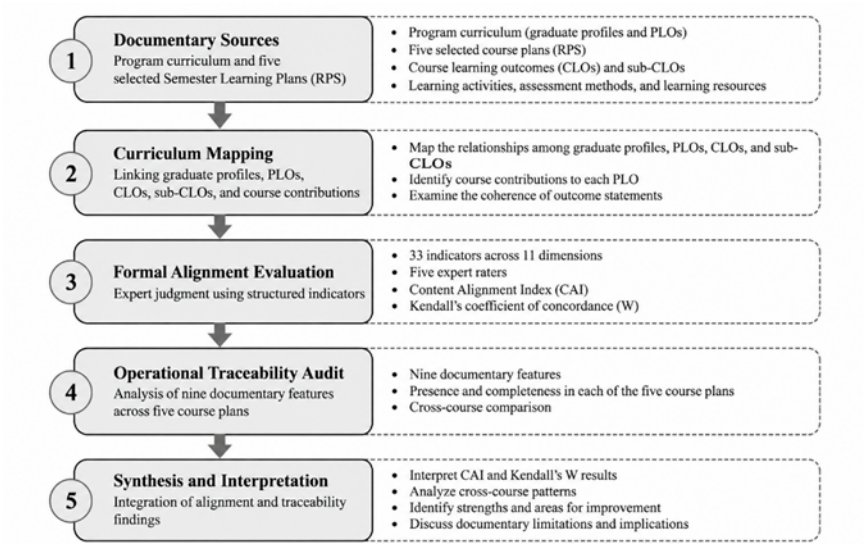


Figure 1. Documentary Curriculum Mapping and Alignment Evaluation Workflow

The 33-indicator instrument covered 11 dimensions (three indicators each): graduate-profile/PLO traceability; PLO specificity; cognitive/performance consistency; PLO–CLO mapping; curriculum progression; content; activities; authentic assessment; criterion-based rubrics; formative feedback; and differentiated opportunities. It drew on OBE, constructive alignment, curriculum mapping and assessment coherence (Biggs, 1996; Harden, 2001; Martone & Sireci, 2009; Porter, 2002; Spady, 1994). Experts reviewed indicator relevance, wording and documentary applicability; no numerical content-validity coefficient is claimed.

Five evaluators (three doctoral and two master's degree holders; 18–32 years' teaching experience) independently rated each of 33 indicators on a five-point ordinal scale, yielding 165 judgments. A score of 1 indicated inadequately documented alignment and 5 clearly traceable alignment supported by documentary evidence. Evidence notes and recommendations supported interpretation but did not enter the score. The records do not document formal rater calibration or consensus adjudication, and neither is claimed. Quantitative alignment was summarized using the Constructive Alignment Index (CAI). The aggregate CAI was calculated from all valid ratings using the following formula:

$$CAI = (Total\ actual\ score\ across\ five\ raters\ /\ [33\ indicators\ x\ 5\ raters\ x\ maximum\ score\ of\ 5])\ x\ 100$$

The maximum possible aggregate score was therefore 825. Dimension-level scores were also calculated because each of the 11 dimensions comprised three indicators; thus, the maximum possible score for each dimension was 75 (3 indicators x 5 raters x 5). Dimension-level analysis was retained because a single aggregate percentage could conceal variation among different parts of the alignment chain. The CAI was used as a descriptive summary of the ratings generated by the study-specific instrument and was not treated as a standardized psychometric scale.

For interpretive purposes, CAI percentages were grouped into five evaluative bands: 0-20 = not/poorly aligned; >20-40 = emerging alignment; >40-60 = moderately aligned; >60-80 = aligned; and >80-100 = highly aligned. These bands were used as descriptive communication categories rather than as universal or externally validated psychometric cut-offs. Consequently, interpretation did not depend solely on whether a percentage fell immediately above or below a categorical threshold. Greater interpretive weight was given to the pattern across the 11 dimensions, indicator-level results, and the documentary evidence explaining why particular links were strong, partial, absent, or difficult to trace.

Inter-rater concordance was examined using Kendall's coefficient of concordance (W) to determine the extent to which the five raters showed similar patterns of judgment across the 33 indicators. Kendall's W was treated strictly as evidence concerning consistency in the pattern of judgments among raters. It was not incorporated into the CAI calculation, interpreted as a coefficient of curriculum alignment, or used as a stand-alone validity coefficient for the instrument. Variation among raters was retained rather than statistically removed, while documentary evidence and rater notes were used to clarify the basis of their judgments. Table 3 presents a qualitative matrix of the course-level evidence.

A supplementary nine-feature audit coded outcome statements, lesson-level outcomes, identifier consistency, relevant activities, assessment tasks, weight consistency, rubrics, outcome-specific criteria links, and feedback-to-revision provisions once per RPS. P required explicit sufficient evidence; S denoted incomplete or conflicting evidence; A meant no evidence after checking relevant sections; and NV meant evidence was unavailable or unverified. Weight consistency required components totaling 100% without contradiction; feedback-to-revision P required an explicit procedural link, whereas generic feedback provisions were S. Frequencies count documents, not students or tasks. The supplementary source check was not independently recoded; no new intercoder coefficient is claimed. Four RPS bear 2024 revision dates and the Semantics and Pragmatics RPS is dated 2023.

Quality checks distinguished expert content review, documentary evidence trails, rater concordance (Kendall's W = .662), and exploratory Rasch–Partial Credit Model diagnostics. These sources are complementary, not interchangeable evidence of instrument validity; the five-rater design precludes strong psychometric claims. Kendall's W describes concordance on the 33 ordinal indicators, not reliability of the supplementary P/S/A matrix. No classroom observation, student work, or attainment data were collected.

Results and Discussions

The document analysis showed that the English Language Education Study Program had established a relatively complete formal architecture for an outcome-based curriculum. At the program level, the curriculum contained four graduate profiles, 12 PLOs, an eight-semester curriculum structure, curriculum mapping, learning-content domains, and course-level learning plans. The five RPS examined in this study also contained the principal components expected in OBE-oriented course design, including PLO references, CLOs, sub-CLOs, learning content, teaching and learning activities, assessment tasks, and, in several courses, assessment criteria or rubrics. The formal availability of these components indicates that OBE had been incorporated into the intended curriculum rather than remaining solely at the level of institutional policy or terminology.

The aggregate Constructive Alignment Index (CAI) was 79.9%, based on an actual score of 659 out of a maximum possible 825. According to the descriptive evaluative bands adopted in this study, the intended curriculum therefore fell within the aligned category. Individual rater scores ranged from 70.3% to 95.2%, indicating differences in rating stringency, while Kendall's coefficient of concordance was W = .662. The latter was interpreted as evidence of a meaningful degree of concordance in the pattern of judgments across the five raters rather than as a measure of curriculum quality itself.

Table 1 reports the 11 dimension scores. Strongest were PLO specificity (93.3%), graduate-profile traceability (90.7%), content (88.0%), and PLO–CLO mapping (85.3%). Curriculum coherence was 81.3%; activities, authentic assessment, and rubrics were each 84.0%; cognitive/performance consistency was 76.0%. Formative feedback was 64.0%, and expanded learning opportunities and differentiated support 48.0%. Thus, the aggregate CAI concealed weaker downstream documentation. Indicator-level results were strongest for graduate-profile/PLO mapping (100%) and weaker for feedback mechanisms (56%) and differentiated support (48%).

Table 1. Constructive Alignment Profile across Eleven Dimensions

Dimension	Score	Maximum	Percentage	Interpretation
Graduate profile clarity and traceability to PLOs	68	75	90.7%	Highly aligned
PLO specificity, measurability, and competence orientation	70	75	93.3%	Highly aligned
Cognitive/performance-level consistency of PLOs and CLOs	57	75	76.0%	Aligned
PLO-CLO design-down and explicit mapping	64	75	85.3%	Highly aligned
Curriculum coherence and progression	61	75	81.3%	Highly aligned
Content alignment with CLOs and PLOs	66	75	88.0%	Highly aligned
Alignment of teaching and learning activities	63	75	84.0%	Highly aligned
Authenticity and alignment of assessment tasks	63	75	84.0%	Highly aligned

Dimension	Score	Maximum	Percentage	Interpretation
Criterion-referenced assessment and rubrics	63	75	84.0%	Highly aligned
Formative feedback and opportunities for improvement	48	75	64.0%	Aligned
Expanded learning opportunities and differentiated support	36	75	48.0%	Moderately aligned

In Table 1, all five RPS documented outcomes, lesson-level outcomes, relevant activities, tasks, and rubrics (5/5 present each). Identifier consistency was partial in all five. Outcome-specific criteria links were present in one and partial in four; feedback-to-revision was partial in all five. Weighting was consistent in four; Curriculum and Material Development conflicted on Assignment 3 (25% in the summary, 30% in the assignment narrative). The four previously NV cells were resolved, leaving none unverified. These counts concern documentary features, not enacted teaching.

Among the five courses, Speaking for Professional Context showed comparatively direct micro-operational traceability. Its performative communication outcomes were supported by video modelling, case-based learning, rehearsal, project-based performance, speaking rubrics, case assignments, and mid- and final-video projects. The relationship between what students were expected to communicate or produce and what they were required to demonstrate in assessment was therefore comparatively direct. However, a discrepancy between PLO codes used in the program-level mapping and those used in the RPS weakened traceability across levels.

The 2023 Semantics and Pragmatics RPS lists six CLOs and ten LLOs with a competency map and assessments weighted 10%, 30%, 10%, and 50%. Detailed task and project rubrics are present; outcome-to-criterion crosswalks remain incomplete and feedback provisions generic. Its authorization sheet also bears a different course name and code. Rubric presence is therefore not equivalent to fully traceable assessment.

By contrast, English Poetry showed more fragmented traceability. The course contained relevant analytical and creative learning activities, including analysis, comparison, group work, poem creation, and an observation project. However, duplicated sub-CLOs and relatively generic assessment evidence weakened the relationship between higher-order outcomes and the tasks used to demonstrate them. Where outcomes were expressed at higher cognitive levels such as analysis, evaluation, or creation, the available evidence did not always demonstrate an equivalent level of cognitive or performative demand.

A similar pattern was observed in the English Intensive Course. The course provided extensive integrated language practice, contextual activities, discussion, project-based learning, performance tasks, progress checks, and a final magazine project. However, some analytical outcomes and CLO-sub-CLO relationships were not matched by equally explicit direct evidence. Thus, the course demonstrated considerable activity-level richness, but activity richness did not always translate into equally strong outcome-level traceability. The cross-course evidence is consolidated in Table 3.

Curriculum and Material Development showed a visible progression from curriculum analysis toward material design, indicating an appropriate product-oriented learning trajectory. The use of assignments and a final material-development project provided relevant evidence for several intended competencies. However, some rubrics remained generic, and the technology-related outcome was not matched by sufficiently direct assessment evidence. In this case, the main issue was not the absence of an authentic task, but the limited specificity of the criteria used to judge whether the intended competencies had actually been demonstrated.

Table 2. Cross-Course Synthesis of Operational Traceability

Course	Main documented activities	Main assessment evidence	Traceability diagnosis
English Intensive Course	Integrated-skills practice, contextual practice, discussion, project-based learning	Progress checks, speaking/writing performance, project, final magazine	Partial traceability; performance evidence is available, but some analytical outcomes and coding relationships are not explicitly evidenced
English Poetry	Analysis, comparison, group work, poem creation, observation project	Participation, tests, project, generic project rubric	Broken/partial links; higher-order outcomes are present, but duplicated sub-CLOs and uneven C4-C6 evidence weaken traceability
Semantics and Pragmatics	Lexical analysis, discussion,	Lexical-relations assignment (10%), mid-term (30%),	Relatively strong course-level analytical evidence; incomplete

Course	Main documented activities	Main assessment evidence	Traceability diagnosis
	contextual conversation analysis, scientific-paper project	implicature/presupposition assignment (10%), project (50%); detailed rubrics	CLO-to-rubric links, generic feedback, and inconsistent course identification on the authorization sheet
Speaking for Professional Context	Video modelling, case-based learning, rehearsal, project-based performance	Speaking rubric, case assignments, mid/final video projects	Strongest micro-operational traceability; performative outcomes are directly assessed, although program-level coding is inconsistent
Curriculum and Material Development	Discussion, inquiry, mapping, reading, material-design project	Assignments and final material project	Product-oriented alignment is evident, but rubrics are relatively generic and technology-related outcomes lack sufficiently direct evidence

Table 3. Evidence-Based Coding Matrix of Operational Traceability Across Five Course Plans

Course	Outcome / coding links	Activity and assessment evidence	Rubric / criteria evidence	Documented limitation
English Intensive Course	CLO–sub-CLO links documented; CLO 09 reused for distinct statements	Integrated-skills tasks, performance checks and final magazine project	Assessment criteria documented, with variable task specificity	Repeated CLO 09 coding; some analytical outcomes lack explicit assessment links
English Poetry	Lesson outcomes documented; LLO 7 and LLO 12 duplicate the same statement	Poetry analysis, comparison, composition and observation/project activities	Project rubric is present	Duplicated LLOs; uneven explicit C4–C6 outcome-to-criteria links
Semantics and Pragmatics	Six CLOs and ten LLOs; competency map links LLOs to CLOs and meetings	Lexical-relation task (10%), mid-term test (30%), implicature/presupposition task (10%), scientific-paper project (50%)	Task-specific rubrics, test rubric, participation and project rubrics are present	Not all CLOs map explicitly to rubric criteria; feedback described generally; cover identifies Semantics Pragmatics/CAC61142 while authorization sheet identifies Listening for Academic Purposes/CAC60333
Speaking for Professional Context	Course outcome mapping documented; program and RPS outcome codes differ	Speaking practice, case tasks, mid-term and final video projects	Speaking performance rubric and assessment requirements present	Cross-document PLO/CLO code mismatch reduces program-to-course traceability
Curriculum and Material Development	Curriculum-analysis and material-development outcomes and tasks documented	Mapping and materials-design assignments, final development project	Task/project criteria and feedback rubric present	Assignment 3 weighting differs between sections (25% vs 30%); technology-related outcome not directly assessed

Note. This qualitative evidence matrix records the availability and specificity of course-plan artifacts, not classroom implementation or measured student learning. It does not constitute a binary frequency analysis; no additional intercoder-reliability statistic is inferred.

Table 4. Supplementary Cross-Course Documentary Coding and Feature Frequencies

Feature	IC	Poetry	Sem.	Speak.	Curr.	P	S	A	NV
Outcome statements	P	P	P	P	P	5	0	0	0
Lesson-level outcomes	P	P	P	P	P	5	0	0	0
Consistent outcome identifiers	S	S	S	S	S	0	5	0	0
Relevant learning activities	P	P	P	P	P	5	0	0	0
Identifiable assessment tasks	P	P	P	P	P	5	0	0	0
Consistent assessment weighting	P	P	P	P	S	4	1	0	0

Feature	IC	Poetry	Sem.	Speak.	Curr.	P	S	A	NV
Assessment criteria or rubrics documented	P	P	P	P	P	5	0	0	0
Outcome-specific criteria linkage	S	S	S	P	S	1	4	0	0
Feedback-to-revision documentation	S	S	S	S	S	0	5	0	0

Note. P = present; S = partial; A = absent; NV = not verified. Frequencies count course plans, not rating judgments. All four previously unverified cells were resolved through examination of the source plans. The Semantics and Pragmatics source is the 2023 plan, not the 2026 replacement. The other four inspected course plans carry December 2024 revision dates; these document versions must not be described collectively as five 2023 plans.

Alternative explanations for inconsistent identifiers, weights, and generic feedback include template or version-control errors, institutional compliance practices, implementation lag, staff workload, and unrecorded teaching routines. The documentary design cannot test these mechanisms or infer impaired student learning. It identifies opportunities to reconcile identifiers and weights and document criterion-specific feedback and revision.

Three interpretive lenses clarify these limits: curriculum mapping examines declared coverage (Harden, 2001; Lawton, Pacey, Jones, & Dean, 2025); assessment validity examines task–criterion correspondence (Martone & Sireci, 2009; McArthur, 2023); and feedback literacy concerns learners' use of judgments (Boud & Dawson, 2023; Dawson, Carless, & Lee, 2021). The RPS support a documentary diagnosis, not conclusions about classroom enactment. Curriculum-mapping and quality-assurance applications offer related examples (Khailova, 2021; Xu, 2026). Programmatic assessment and student-generated feedback further contextualize this gap (Baartman & Quinlan, 2024; Nicol & Kushwah, 2024). Related studies address assessment rubrics, authenticity, feedback and curriculum mapping (Gyamfi, Hanna, & Khosravi, 2022; Hattie & Timperley, 2007; Morris, Perry, & Wardle, 2021; Okojie et al., 2022; Sokhanvar, Salehi, & Sokhanvar, 2021; Vlachopoulos & Makri, 2024). Outcome taxonomy and mapping research also address alignment depth (Karanja & Malone, 2021; Sarkar & Kurup, 2024; Spangenberg, 2025).

The distinction between formal alignment and operational traceability also warrants critical scrutiny. Havnes and Prøitz (2016) questioned whether predetermined learning outcomes can capture unexpected learning, while Erikson and Erikson (2019) identified tensions between specified outcomes and complex capacities such as critical thinking. Loughlin, Lygo-Baker, and Lindberg-Sand (2021) cautioned against reducing constructive alignment to a procedural quality-assurance exercise. These critiques suggest that the presence of mapped outcomes, activities, and assessment components should not be equated with their educational coherence or with meaningful learning.

Earlier critiques reinforce this distinction. Hussey and Smith (2002) challenged the assumption that predefined outcomes adequately describe complex educational processes, and Torrance (2007) warned that explicit assessment criteria can encourage compliance rather than substantive learning. In the present audit, all five course plans contained rubrics, but only one documented sufficiently explicit outcome-specific criterion links; outcome identifiers were partly inconsistent across all five, and feedback-to-revision procedures were partial throughout. These documentary gaps support more transparent mapping and feedback design, not more rigid specification of learning. Because classroom practice and student work were not examined, the findings cannot establish whether these weaknesses affected teaching or attainment.

The documentary evidence suggests that the formal portion of this chain was more visible than the later evidence-and-improvement portion. Graduate profiles, PLOs, course mapping, learning activities, assessment formats, and rubrics were frequently documented; the supplementary audit found assessment criteria in all five course plans, not an absence of rubrics. The weaker links concerned the specificity of outcome-to-rubric cross-references, stable coding across documents, and explicit feedback-to-revision procedures. Thus, an aggregate CAI of 79.9% does not establish uniformly complete traceability or successful classroom implementation. Because four examined RPS bear 2024 revision dates whereas the Semantics and Pragmatics source is dated 2023, the comparison is also a documentary snapshot of available versions rather than a strictly synchronized 2023 cohort. Explicit attainment anchors and learner feedforward are relevant to these gaps (Mahalingam, 2026; Sadler, Reimann, & Sambell, 2023; Winstone, Balloo, & Carless, 2022).

Conclusions

This evaluative document analysis found substantial formal OBE alignment in the curriculum of an Indonesian English teacher education program, with a Constructive Alignment Index of 79.9% (659/825). The strongest dimensions concerned documented outcome specification and program-level relationships, while formative feedback (64.0%) and differentiated learning opportunities (48.0%) remained less fully specified. Agreement among the five evaluators was moderate to substantial (Kendall's W = .662); it does not constitute evidence of

construct validity or classroom effectiveness. The study contributes an applied distinction between formal alignment and operational traceability. Formal alignment describes the documented presence and declared relationships of OBE components; operational traceability describes whether a reviewer can follow a consistent documentary chain from intended outcomes through learning activities, assessments, criteria, and planned evidence and improvement opportunities. In the supplementary five-course audit, all plans contained assessment criteria, but only one provided sufficiently explicit outcome-specific criterion links. All five showed partial documentation of feedback-to-revision processes. Thus, the main issue was the continuity of documented links, not a wholesale absence of assessment rubrics.

Curriculum quality assurance should therefore focus on stabilizing outcome identifiers across program and course documents, reconciling assessment weights, making links between outcomes and rubric criteria explicit, and specifying how assessment judgments inform feedback and subsequent revision. These improvements are practical design recommendations; the document evidence does not show whether lecturers already implement these practices informally. The findings are limited to an intentionally varied sample of five course plans and associated curriculum documents. The plans were not statistically representative; four inspected plans carried 2024 revision dates, while the Semantics and Pragmatics plan was dated 2023. The study did not observe teaching, analyse actual student work, or measure student attainment. Future research should compare successive plan versions with classroom practices, student assessment artifacts, and longitudinal feedback cycles to test whether stronger documentary traceability corresponds to improved learning evidence.

Acknowledgments

Ac The authors gratefully acknowledge the English Language Education Study Program, Faculty of Letters and Culture, Universitas Negeri Gorontalo, which provided the institutional context and documentary basis for this study. The authors also extend their appreciation to the five raters who contributed to the structured evaluation of curriculum alignment and provided evidence-based judgments that strengthened the analytical process. Appreciation is further extended to colleagues who provided academic feedback during the development and refinement of the study. Generative artificial intelligence was used to assist with language refinement, manuscript organization, and drafting support. All AI-assisted content was reviewed, verified against the original data and cited sources, and revised by the authors, who take full responsibility for the final manuscript. AI was not used to generate research data or make final analytical decisions.

References

Allo, M. D. G., Sudarsi, E. T., & Taula'bi', N. (2024). Implementation of outcome-based education (OBE) principles in the curriculum of the English Education Study Program at a higher education in Toraja. *Jurnal Onoma: Pendidikan, Bahasa, Dan Sastra*, *10*(2), 1379–1391. https://doi.org/10.30605/onoma.v10i2.3505

Baartman, L. K. J., & Quinlan, K. M. (2024). Assessment and feedback in higher education reimagined: Using programmatic assessment to transform higher education. *Perspectives: Policy and Practice in Higher Education*, *28*(2), 57–67. https://doi.org/10.1080/13603108.2023.2283118

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364. https://doi.org/10.1007/BF00138871

Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for quality learning at university* (5th ed.). Maidenhead, England: Open University Press.

Boud, D., & Dawson, P. (2023). What feedback literate teachers do: An empirically-derived competency framework. *Assessment & Evaluation in Higher Education*, *48*(2), 158–171. https://doi.org/10.1080/02602938.2021.1910928

Dawson, P., Carless, D., & Lee, P. P. W. (2021). Authentic feedback: Supporting learners to engage in disciplinary feedback practices. *Assessment & Evaluation in Higher Education*, *46*(2), 286–296. https://doi.org/10.1080/02602938.2020.1769022

Derouich, M. (2025). Ensuring outcome-based curriculum coherence through systematic CLO–PLO alignment and feedback loops. *Discover Education*, *4*. https://doi.org/10.1007/s44217-025-00915-7

Erikson, M. G., & Erikson, M. (2019). Learning outcomes and critical thinking – good intentions in conflict. *Studies in Higher Education*, *44*(12), 2293–2303. https://doi.org/10.1080/03075079.2018.1486813

Gurukkal, R. (2020). Outcome-based education: An open framework. *Higher Education for the Future*, *7*(1), 1–4. https://doi.org/10.1177/2347631119886402

Gyamfi, G., Hanna, B. E., & Khosravi, H. (2022). The effects of rubrics on evaluative judgement: A randomised controlled experiment. *Assessment & Evaluation in Higher Education*, *47*(1), 126–143.

https://doi.org/10.1080/02602938.2021.1887081
Harden, R. M. (2001). AMEE Guide No. 21: Curriculum mapping: A tool for transparent and authentic teaching and learning. *Medical Teacher*, *23*(2), 123–137. https://doi.org/10.1080/01421590120036547
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487
Havnes, A., & Prøitz, T. S. (2016). Why use learning outcomes in higher education? Exploring the grounds for academic resistance and reclaiming the value of unexpected learning. *Educational Assessment, Evaluation and Accountability*, *28*(3), 205–223. https://doi.org/10.1007/s11092-016-9243-z
Hussey, T., & Smith, P. (2002). The trouble with learning outcomes. *Active Learning in Higher Education*, *3*(3), 220–233. https://doi.org/10.1177/1469787402003003003
Karanja, E., & Malone, L. C. (2021). Improving project management curriculum by aligning course learning outcomes with Bloom's taxonomy framework. *Journal of International Education in Business*, *14*(2), 197–218. https://doi.org/10.1108/JIEB-05-2020-0038
Katawazai, R. (2021). Implementing outcome-based education and student-centered learning in Afghan public universities: The current practices and challenges. *Heliyon*, *7*(5). https://doi.org/10.1016/j.heliyon.2021.e07076
Khailova, L. (2021). Using curriculum mapping to scaffold and equitably distribute information literacy instruction for graduate professional studies programs. *The Journal of Academic Librarianship*, *47*(1). https://doi.org/10.1016/j.acalib.2020.102281
Lawton, V., Pacey, V., Jones, T. M., & Dean, C. M. (2025). Developing work readiness in university graduates: A case study in mapping a university curriculum to work readiness domains in an Australian physiotherapy degree. *Physiotherapy Theory and Practice*, *41*(5), 1070–1082. https://doi.org/10.1080/09593985.2024.2384654
Loughlin, C., Lygo-Baker, S., & Lindberg-Sand, Å. (2021). Reclaiming constructive alignment. *European Journal of Higher Education*, *11*(2), 119–136. https://doi.org/10.1080/21568235.2020.1816197
Mahalingam, T. (2026). Program learning outcome assessment: A scalable framework for quality assurance in outcome-based education. *Quality Assurance in Education*, *34*(2), 360–377. https://doi.org/10.1108/QAE-07-2025-0215
Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, *79*(4), 1332–1361. https://doi.org/10.3102/0034654309341375
McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, *85*(1), 85–101. https://doi.org/10.1007/s10734-022-00822-y
Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3). https://doi.org/10.1002/rev3.3292
Mufanti, R., Carter, D., & England, N. (2024). Outcomes-based education in Indonesian higher education: Reporting on the understanding, challenges, and support available to teachers. *Social Sciences & Humanities Open*, *9*. https://doi.org/10.1016/j.ssaho.2024.100873
Nicol, D., & Kushwah, L. (2024). Shifting feedback agency to students by having them write their own feedback comments. *Assessment & Evaluation in Higher Education*, *49*(3), 419–439. https://doi.org/10.1080/02602938.2023.2265080
Okojie, M. U., Bastas, M., & Miralay, F. (2022). Using curriculum mapping as a tool to match student learning outcomes and social studies curricula. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.850264
Pereira, E., Nsair, S., Pereira, L. R., & Grant, K. (2024). Constructive alignment in a graduate-level project management course: An innovative framework using large language models. *International Journal of Educational Technology in Higher Education*, *21*. https://doi.org/10.1186/s41239-024-00457-2
Pham, H. T., & Nguyen, P. V. (2024). ASEAN quality assurance scheme and Vietnamese higher education: A shift to outcomes-based education? *Quality in Higher Education*, *30*(2), 285–312. https://doi.org/10.1080/13538322.2023.2261974
Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3–14. https://doi.org/10.3102/0013189X031007003
Rao, N. J. (2020). Outcome-based education: An outline. *Higher Education for the Future*, *7*(1), 5–21. https://doi.org/10.1177/2347631119886418
Reniers, J., Mathany, C., Farkas, M., Pollock, H., & Husband, B. C. (2022). Uncharted territory: Curriculum mapping multiple majors simultaneously. *The Canadian Journal for the Scholarship of Teaching and Learning*, *13*(1). https://doi.org/10.5206/cjsotlrcacea.2022.1.8553
Sadler, I., Reimann, N., & Sambell, K. (2023). Feedforward practices: A systematic review of the literature. *Assessment & Evaluation in Higher Education*, *48*(3), 305–320. https://doi.org/10.1080/02602938.2022.2073434

Sarkar, D., & Kurup, A. (2024). Scale of outcome-based education: Beyond the knowledge-skill dichotomy. *Higher Education for the Future*, *11*(2), 204–219. https://doi.org/10.1177/23476311241263380

Smith, H. R., & Chittams, J. (2024). Defining best practices and validation for curriculum mapping. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2342662

Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation*, *70*. https://doi.org/10.1016/j.stueduc.2021.101030

Spady, W. G. (1994). *Outcome-based education: critical issues and answers*. American Association of School Administrators.

Spangenberg, E. D. (2025). Constructive alignment of a mathematics methodology module. *Instructional Science*, *53*(4), 729–760. https://doi.org/10.1007/s11251-025-09712-z

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, *14*(3), 281–294. https://doi.org/10.1080/09695940701591867

Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, *83*. https://doi.org/10.1016/j.stueduc.2024.101425

Winstone, N. E., Balloo, K., & Carless, D. (2022). Discipline-specific feedback literacies: A framework for curriculum design. *Higher Education*, *83*, 57–77. https://doi.org/10.1007/s10734-020-00632-0

Xu, Y. (2026). Curriculum mapping for the AACSB Assurance of Learning (AOL) process: A case study. *The International Journal of Management Education*, *24*(1). https://doi.org/10.1016/j.ijme.2025.101315