



WHERE ALIGNMENT BREAKS: A DOCUMENT-BASED EVALUATION OF CONSTRUCTIVE ALIGNMENT IN A SEQUENCED EFL READING CURRICULUM

La Ode Alfin Haris Munandar¹, I Made Sujana², Lianda Dewi Sartika^{3*}, Darin
Fadhilah⁴, Anak Agung Ngurah Bagus Janitra Dewanta⁵, Ayunda Wulandari⁶

^{1,2,3,4,5,6}Universitas Mataram

E-mail: laode_alfinharis@unram.ac.id¹, car_msujana@unram.ac.id², liandadewi@staff.unram.ac.id³,
darinfadhilah@staff.unram.ac.id⁴, agungjanitradewanta17@staff.unram.ac.id⁵,
ayundawulandari@staff.unram.ac.id⁶

ARTICLE INFORMATION

Keywords: CEFR;
Collaborative Course Design;
Constructive Alignment;
Curriculum Document
Analysis; Reading Instruction

Article history:

Received: 30 August, 2026

Revised: 29 August, 2026

Accepted: 30 August, 2026

ABSTRACT

Quality assurance in Indonesian higher education is enacted largely through course documents, yet the internal coherence of those documents is rarely audited once they have been approved. This study evaluated constructive alignment across four sequenced Reading courses in an undergraduate English Language Education programme at an Indonesian university. Using a qualitative evaluative design based on document analysis, the study examined the programme curriculum, five semester lesson plans covering the four courses, and one draft instructional module. Twenty-eight units of analysis were coded on three component pairings using a three-level alignment rubric, producing 84 coding decisions and a set of descriptive alignment indices. A gap in stated proficiency level was the clearest finding: mapped against CEFR reading descriptors, the four courses indicated levels B1, B1, B2, and B1 against a programme learning outcome targeting C1, and the sequence did not rise, since the first two courses showed no increase in demand and the final course indicated a lower level than the course preceding it. Mean alignment indices fell in the moderate band for all four courses, ranging from 60.0% to 70.4%, and declined consistently from objectives-to-materials, through objectives-to-planned-implementation, to objectives-to-assessment, which was rated low in every course and reached 27.8% in one. Document verification further revealed that the lesson plan for the first course was textually near-identical to that of the second, that two conflicting versions of one lesson plan were in circulation, and that no assessment instruments or scoring rubrics existed for any course. Because these weaknesses originate in lesson plans drafted in isolation by individual convenors, the study recommends cross-course collaborative design, in which the lesson plans of a related course sequence are drafted and revised together in a single scheduled session and read against one another and against the curriculum, supported by content-based rather than checklist-based internal document audit.

This is an open access article under the [CC BY-NC-SA](#) license.



Corresponding Author:

Lianda Dewi Sartika

liandadewi@staff.unram.ac.id

INTRODUCTION

Reading occupies a foundational position in undergraduate English Language Education programmes, since academic literacy in a second language depends on the capacity to construct meaning from written text. Second language reading ability develops unevenly: readers who share a first language can differ substantially from one another across samples, so that a single course-level expectation rarely fits an entire cohort (Kuperman, 2025). Comprehension is further shaped by properties of the text itself, with readability affecting both understanding and reading time (Zainurrahman et al., 2024). Learners in English as a foreign language settings continue to report difficulty with academic reading, particularly with unfamiliar vocabulary and extended argument (Oktavianti et al., 2026), and moving them from literal comprehension toward critical literacy requires deliberate instructional staging rather than exposure alone (Ahmad et al., 2026). For these reasons reading instruction is typically distributed across a sequence of courses expected to rise in cognitive demand.

A sequence of courses, however, is only as coherent as the documents that define it. In outcome-based curricula, each course is described by a semester lesson plan that states the intended learning outcomes, the material to be studied, the learning activities to be conducted, and the assessment to be administered. Constructive alignment, in Biggs's (1996) formulation, holds that learning is supported when these elements point toward the same intended outcomes, and undermined when they diverge. The principle continues to be applied and tested in contemporary course design, in settings ranging from graduate project management (Pereira et al., 2024) to mathematics teacher education (Spangenberg, 2025), and has generated diagnostic tools intended to make the alignment of a course visible to those designing it (Trowsdale & McKay, 2023). Assessment occupies a special position in this account, since from the learner's perspective assessment defines the curriculum: what is assessed is what is learned, regardless of what the stated outcomes claim.

Two developments have made document coherence more consequential in Indonesian higher education. First, outcome-based education has become the organising principle for programme curricula, so that programme learning outcomes

are formally distributed across courses and are expected to be traceable into course-level outcomes; where this has been examined, outcome-based design has been found to depend on constructive alignment for its implementation rather than to guarantee it ([Poolkrajang & Papanai, 2024](#)). Second, proficiency frameworks, most prominently the Common European Framework of Reference for Languages, are increasingly cited in curriculum documents as the reference for stating proficiency targets. Reviews of framework-related research report rapid uptake alongside uneven implementation ([Jeon, 2022](#)), and studies of framework-aligned teaching materials find that alignment claimed at policy level is not always recognisable to the teachers who use those materials ([Govindarajoo et al., 2025](#)). Both developments are documentary in nature: they are enacted primarily through what course documents say ([Council of Europe, 2020](#)).

These developments operate within a specific regulatory environment. Indonesian higher education institutions are required to align programme design with the national qualifications framework and to express curricula in terms of programme learning outcomes, which are then distributed across courses and expressed at course level as course learning outcomes and their constituent sub-outcomes. Each course is documented in a semester lesson plan whose template is fixed at institutional level and whose columns correspond directly to the elements of constructive alignment. The status of these documents is not merely administrative. National quality assurance regulation requires every institution to operate an internal quality assurance system, and the current accreditation provisions name the curriculum and the course lesson plans among the criteria against which a programme is assessed ([Peraturan Menteri Pendidikan Tinggi, Sains, dan Teknologi Nomor 10 Tahun 2026, 2026](#)). A lesson plan is therefore simultaneously a pedagogical instrument and an object of external inspection.

This dual status is consequential, and we take it to be the proximate cause of the misalignment reported below. When a document is produced principally to satisfy inspection, the property that matters to its author is completeness rather than coherence: every column must be filled, and a filled column is indistinguishable from an aligned one at the moment of submission. The lesson plan then drifts from a pedagogical design tool, in which the columns are written against one another, toward an administrative compliance artefact, in which they are written in parallel and never compared. Alignment breakdown at document level is, on this reading, not a failure of individual diligence but a predictable consequence of the function the document is asked to perform. The analysis that follows tests whether the documents in one course sequence show the pattern this account would predict.

Existing research has examined the alignment of reading assessment with proficiency descriptors. Standard-setting work on university academic reading tests has shown that test constructs do not always match the proficiency levels they are assumed to represent (Tangsakul & Poonpon, 2024), and recent guidance on constructing framework-aligned reading assessments treats level alignment as something that must be built deliberately into item development rather than assumed (Shin & Sanonguthai, 2026). Item-level reviews conducted against the framework's can-do descriptors point in the same direction: in a content validity study of a teacher language proficiency test, a number of items were found not to correspond to the descriptor statements they were assumed to reflect, even where the test as a whole was judged valid (Safitry et al., 2023). Similar concerns arise upstream of the items themselves, since the quality of test specifications and item-writer guidance materially affects whether the items produced measure what was intended (Al Lawati, 2023).

What remains comparatively underexamined is the coherence of the course documents themselves, considered as a set rather than one at a time. Work on the consistency between curricula and the materials derived from them has begun to formalise this kind of comparison (Jin & Seo, 2026), and curriculum-level evaluations of reading provision have shown that weaknesses often lie in the documented design rather than in teaching effort (Abejuela et al., 2023). Studies of curriculum design practice similarly locate the problem in how design decisions are made and recorded (McEwan et al., 2023), while recent analyses argue that assessment should be examined as a multilevel architecture rather than as a set of separate instruments (Zaporozhets & Zaporozhets, 2026). Even so, studies of alignment still tend to take a single course as the unit of analysis, which makes it possible to observe internal consistency within a course but not the consistency of a sequence. Yet a sequence is precisely where alignment problems accumulate: a course that duplicates its predecessor, or that targets a lower cognitive level than the course before it, produces a discontinuity that no single-course audit would detect.

This study therefore evaluates constructive alignment across a complete sequence of Reading courses in one undergraduate English Language Education programme, treating the programme curriculum and the course documents as the object of analysis. It addresses three questions. First, how well aligned are learning outcomes, materials, planned implementation, and assessment in each Reading course? Second, which components show alignment and which show misalignment? Third, to what extent do the stated outcomes, materials, and assessment plans represent the proficiency level targeted by the programme learning outcome, when

mapped against the reading comprehension descriptors of the Common European Framework of Reference for Languages (Council of Europe, 2020)?

METHOD

Design

The study used a qualitative evaluative design based on document analysis. This design was chosen because the object of evaluation is the coherence of an instructional design as documented, which is accessible through the documents that constitute it. Limited descriptive quantification was used to summarise coding results as alignment indices; these indices support qualitative interpretation and are not intended as measurements of instructional quality or of student achievement.

Setting and Corpus

The setting was an undergraduate English Language Education programme at one of university in Indonesia's western provinces. At the time of the programme's most recent self-evaluation it enrolled approximately 1,100 regular students and was staffed by 38 full-time academic staff whose expertise matched the field of study. Annual intake across the preceding five years ranged from roughly 220 to 275 students against a declared capacity of 220 to 250, drawn from between one and two thousand applicants each year, and the programme held a B accreditation grade from the national accreditation agency. Staff carried a mean total workload of approximately 14 credits per semester, of which around 10 were teaching within the programme itself. Its intake is drawn predominantly from public secondary schools in the surrounding province and neighbouring provinces, with recent cohorts originating from about five provinces and from fourteen provinces across a five-year period. Entering students therefore arrive with a wide range of prior English proficiency and, for many, limited previous exposure to extended academic reading in English. Each of the four Reading courses is convened by a different member of staff.

The programme curriculum was reconstructed in 2022 on outcome-based principles. Students complete a minimum of 146 credits from a total offering of 152, of which 140 are compulsory. The curriculum was developed by a programme-level team and subsequently operationalised course by course, with individual convenors responsible for drafting the semester lesson plan for the course or courses they teach. This division of labour, in which a curriculum is authored collectively and its course documents individually, is common in Indonesian higher education and is directly relevant to the findings reported below.

Within this curriculum, reading instruction is distributed across four sequential two-credit courses placed in the first four semesters, referred to here as Reading 1, Reading 2, Reading 3, and Reading 4. Each course is a prerequisite for the next, and the curriculum assigns the same four programme learning outcomes to all four. Classes are taught face to face with supporting activity on the institutional learning management system.

The corpus comprised the programme curriculum document, five semester lesson plans covering the four courses, and one draft instructional module. The number of lesson plans exceeds the number of courses because two differently dated versions of the Reading 3 lesson plan were in circulation. No assessment instruments and no scoring rubrics were obtainable for any course, and instructional materials in documentary form existed for one course only. Table 1 summarises the corpus.

Table 1. Composition of the Document Corpus

Document type	Number	Note
Programme curriculum	1	Outcome-based revision, 2022
Semester lesson plans	5	Four courses; Reading 3 has two versions
Units of analysis (outcomes)	28	Reading 1 and 2: 5 sub-outcomes each; Reading 3 and 4: 9 weekly blocks each
Instructional module	1	Draft, ten units, Reading 2 only
Assessment instruments	0	Not available for any course
Scoring rubrics	0	Not available for any course

Unit of Analysis and Instrument

The lesson plans followed two incompatible formats. Reading 1 and Reading 2 used an Indonesian outcome-based format in which learning is organised by sub-outcomes, while Reading 3 and Reading 4 used an English-language format organised by weekly blocks with stated expected competencies. To permit comparison, the unit of analysis was defined as the smallest stated outcome in each document: the sub-outcome in the first format and the weekly block competency in the second. This yielded 28 units.

The instrument was an alignment matrix pairing each unit of analysis with three components as stated in the same document: the material, the planned form and method of learning, and the assessment plan. Each pairing was rated using a three-level rubric. A rating of 2, aligned, was assigned when the component could be traced

explicitly to the same stated outcome, occupied the same cognitive level in the revised Bloom taxonomy (Anderson & Krathwohl, 2001), and covered the full scope of the outcome. A rating of 1, partially aligned, was assigned when the component could be traced to the same outcome but differed by one cognitive level or covered only part of its scope. A rating of 0, not aligned, was assigned when the component could not be traced to any stated outcome or differed by two or more cognitive levels. A separate mapping sheet recorded, for each unit, the proficiency scale and level indicated by the wording of the outcome, referenced against the descriptive scales for reading comprehension in the framework companion volume (Council of Europe, 2020).

The CEFR mapping was carried out on the wording of the documents rather than on learner performance, and used four indicators drawn from the structure of the reading comprehension scales. The first was the cognitive operation named by the verb in the stated outcome, distinguishing locating explicit information, inferring unstated meaning, and evaluating a position. The second was the text type and length specified for the task, since the scales differentiate short simple texts from extended texts on unfamiliar topics. The third was the degree of explicitness of the information the reader must recover, that is, whether the required information is stated in the text or must be constructed from it. The fourth was the expected independence of the reader, including whether dictionary or teacher support is assumed. A level was assigned only where at least three of the four indicators pointed to the same band; where they diverged, the lower band was recorded and the divergence noted in the coding sheet.

Analysis

Each of the 28 units was coded on three pairings, producing 84 coding decisions. Every decision was recorded together with the document wording on which it rested, so that the basis of each rating is inspectable. An alignment index was computed for each course and pairing as the sum of obtained scores divided by the maximum possible score, expressed as a percentage, with the maximum defined as twice the number of units. Indices were interpreted using bands agreed by the research team: high at 80% or above, moderate between 60% and 79%, and low below 60%. These bands are a reporting convention rather than a validated scale, and are reported here so that readers may reinterpret the raw sums if they prefer different thresholds.

Trustworthiness was pursued through documentation rather than through independent double coding. The research team comprised the convenors of the courses under analysis together with colleagues from the same programme, so a second coder drawn from within the team would not have been independent of the

documents being judged, while a coder from outside the programme would have lacked the curricular knowledge needed to interpret abbreviated lesson plan entries, several of which consist of a single noun phrase. Rating under those conditions would have produced an agreement figure reflecting the availability of shared background rather than the clarity of the rubric. The study therefore did not carry out independent double coding and reports no inter-rater reliability coefficient.

Ethical Considerations

The study analysed institutional documents rather than human participants. Because several findings concern shortcomings in documents attributable to identifiable individuals, the institution, the course codes, and the names of course convenors are withheld, and courses are referred to by sequence position. Findings are reported as properties of documents rather than as evaluations of the persons who produced them. The identity of the institution can be disclosed confidentially to editors on request.

RESULTS AND DISCUSSION

Results

Document Verification

Verification of the lesson plans against the programme curriculum identified six discrepancies, summarised in Table 2. Three of these bear directly on the interpretation of the alignment indices reported subsequently.

Table 2. Discrepancies Identified through Document Verification

Aspect	Finding	Consequence
Course identity	The Reading 1 lesson plan reproduces the course description, course outcomes, sub-outcomes, content, and weekly plan of the Reading 2 lesson plan; textual overlap between the two documents is approximately 88%. Its course description opens by naming the second course, and it lists itself as its own prerequisite.	The outcomes that the curriculum states for Reading 1, concerning main ideas, supporting detail, scanning, skimming, and reading speed, appear nowhere in its lesson plan.
Document version	Two versions of the Reading 3 lesson plan, dated six months apart,	No document unambiguously governs the

Aspect	Finding	Consequence
	differ in the programme outcomes loaded, the convenor named, and the prerequisite stated.	course; the later version was analysed because its outcome load matches the curriculum.
Outcome load	The curriculum assigns the same four programme outcomes to all four courses; the earlier Reading 3 version and the Reading 4 lesson plan each load a different set of three.	Outcomes stated at programme level are not carried into course-level outcomes.
Semester placement	Semester fields in the lesson plans contradict the curriculum for three of the four courses.	The intended sequence is obscured in the very documents that should express it.
Outcome correlation matrix	The matrix relating course outcomes to sub-outcomes is left blank in two lesson plans.	The relation between the two levels of outcome is undocumented.
Reference list	The Reading 4 reference list is not identical to that of Reading 3 and consists entirely of works on critical thinking and argumentation, whereas the stated content of Reading 4 concerns novels, biographies, letters, comics, and film.	Stated material is not supported by references relevant to the course outcomes.

A seventh finding concerns absence rather than discrepancy. No assessment instruments and no scoring rubrics were available for any of the four courses. Assessment exists in the documents only as indicators, techniques, and weightings recorded in the assessment column of the lesson plans. This absence is reported here as a finding in its own right: it means that no written record exists of what is actually measured, how responses are scored, or whether scoring is consistent across classes and across years.

Alignment Indices

Table 3 presents the alignment indices for the three component pairings across the four courses.

Table 3. Alignment Indices by Course and Component Pairing

Component pairing	Reading 1	Reading 2	Reading 3	Reading 4
Outcomes – Materials	70.0%	70.0%	83.3%	72.2%
Outcomes – Planned implementation	60.0%	60.0%	72.2%	88.9%
Outcomes – Assessment	50.0%	50.0%	27.8%	50.0%
Mean index	60.0%	60.0%	61.1%	70.4%

Three patterns are visible. Alignment declines in the same order in every course, from materials through planned implementation to assessment. No course reaches the high band on the mean index. Assessment falls in the low band in all four courses without exception.

Outcomes and Materials

This pairing produced the highest indices, which is unsurprising given that the material column in a lesson plan is typically drafted immediately after the outcome column. Reading 3 scored highest because each weekly block pairs a competency with a topic of the same name. Its three closing blocks depart from this pattern: the stated competency shifts to summary writing while the material continues to be described as the analysis of claims, evidence, and argument.

Reading 1 and Reading 2 obtained identical indices. This is a direct consequence of the duplication reported in Table 2 rather than a coincidence, and it means that the Reading 1 index describes the coherence of the Reading 2 document. When the sub-outcomes in the Reading 1 lesson plan are instead compared with the description of Reading 1 in the programme curriculum, none of the five can be traced to that description, placing the course in the not aligned category at the level of the curriculum.

The clearest instance of misalignment occurred in Reading 4, where one block states an outcome of developing a letter, a productive skill at the creating level, while the corresponding material consists only of general information about letters at the understanding level. A related case arises in a subsequent block in which film, rather than written text, is the object of study in a reading course.

Outcomes and Planned Implementation

The dominant finding for this pairing is the low specificity of planned activities. In Reading 1 and Reading 2, the form and method column is worded identically across three consecutive blocks as lecture, group work exercises and discussion, together with independent reading through the learning management system. This wording does not distinguish an activity that develops prediction from one that develops contextual word inference, although these are separate stated outcomes. A method statement that would serve any outcome equally cannot guarantee that a particular outcome is practised.

Reading 4 recorded the highest index for this pairing at 88.9%, since lecture and discussion are indeed commensurate with outcomes that largely concern recognising text types and their features. A high block-level index, however, does not establish alignment at course level. The course description states that students select their own reading material, read independently and continuously, record vocabulary, and reflect on their reading in an electronic journal. None of this sustained independent reading appears in the weekly plan, which is dominated by lecture and discussion about genres. The planned implementation is thus aligned with the block outcomes while being misaligned with the nature of the course itself.

Coding also revealed repetition of stated outcomes across blocks in Reading 3, where nine weekly blocks contain only six distinct competencies. The final four meetings restate competencies already covered without increasing demand, and these same blocks carry no assessment weighting.

Outcomes and Assessment

Assessment produced the lowest indices in every course and was the only pairing rated low throughout. Three recurring patterns account for this. First, some blocks cannot be traced to any assessment plan at all. In Reading 3, four of nine blocks fall into this category: three closing blocks carry no weighting, and two blocks record weightings that are not legible in the document. Reading 4 shows the same structure, with three unweighted closing blocks alongside a single block carrying half of the course weighting.

Second, group performance is repeatedly used to assess outcomes stated in individual terms. In Reading 3, an outcome concerning the application of note-taking is assessed through group performance, although note-taking is an individual capability. The same occurs in Reading 4 for outcomes concerning the understanding of fiction, non-fiction, and biography. Group assessment does not license inferences about individual attainment, which weakens the validity of the resulting judgements

in the way that studies of reading assessment alignment have cautioned (Al Lawati, 2023; Tangsakul & Poonpon, 2024).

Third, the cognitive level demanded by assessment often diverges from that of the outcome, in both directions. In Reading 1 and Reading 2, a sub-outcome requiring determination of word meaning at the understanding level is assessed through a project-based task demanding the creating level, while a sub-outcome requiring analysis of context and meaning is assessed through a summarising task one level below it. In the same two courses, a mid-semester examination carrying 20% of the final mark is not linked to any sub-outcome in the document, so that one fifth of a student's result has no explicit outcome referent.

Proficiency Level Representation

All four lesson plans load a programme outcome specifying proficiency at level C1, while course descriptions state substantially lower targets, describing students as pre-intermediate to intermediate in Reading 1 and lower-intermediate in Reading 2. Mapping the wording of the stated outcomes against the reading comprehension scales produced the pattern in Table 4.

Table 4. Proficiency Level Indicated by Stated Outcomes

Course	Most relevant scale	Level indicated	Level targeted	Gap
Reading 1	Overall reading comprehension	B1	C1	2 levels
Reading 2	Overall reading comprehension	B1	C1	2 levels
Reading 3	Reading for information and argument	B2	C1	1 level
Reading 4	Reading as a leisure activity	B1	C1	2 levels

No course states outcomes at the level its loaded programme outcome targets. More consequentially for a sequence, the progression is not monotonic. The first two courses show no increase in demand because their documents are identical, and the final course in the sequence indicates a lower level than the course preceding it, even though it is placed in a later semester and takes that course as its prerequisite.

Discussion

The ordering of the indices is the most interpretable result of this study. Alignment declines from materials to planned implementation to assessment in every course, which corresponds to the order in which these components appear in the lesson plan template. The further a column sits from the outcome column, the weaker its documented relation to it. This suggests that alignment in these documents is a local property produced during drafting rather than a global property verified afterwards, and that it dissipates as drafting proceeds. Examining assessment as a multilevel architecture rather than as a list of instruments makes the same point from the assessment side ([Zaporozhets & Zaporozhets, 2026](#)).

That assessment is the weakest component in all four courses is consequential rather than incidental. [Biggs \(1996\)](#) argues that assessment defines the curriculum from the student's point of view; on that argument, a curriculum whose assessment is its least aligned component will be experienced by students as something other than what its outcomes describe. The absence of instruments and rubrics compounds this, since it removes the only evidence that could show what is in fact being measured. The situation described here, in which four convenors each documented assessment separately and none produced instruments or rubrics, indicates that assessment design was never treated as a shared object of work.

Work on test specifications and item writer guidance shows that the quality of these intermediate documents determines whether the resulting instruments measure what was intended ([Al Lawati, 2023](#)); where such documents are absent altogether, as here, there is nothing between a stated outcome and a mark. Institutional responses that address only documentation, by requiring that a rubric be attached, will not produce usable rubrics if the capacity to write them has not been built. The finding therefore points toward staff development in assessment design alongside any change in documentation requirements.

The duplication of one lesson plan into another deserves comment beyond its status as an error. A discrepancy of this visibility, in which a document names a different course in its own description and lists itself as its own prerequisite, persisted from the drafting of the documents until this study without being detected. Its persistence is evidence about process rather than about individuals: neither cross-course review among convenors of a related sequence nor document quality checking at programme level was operating. The staffing figures reported earlier make this intelligible without recourse to individual negligence, since convenors carried teaching loads close to the institutional ceiling alongside research and community service obligations, and no scheduled occasion existed in which the four lesson plans would be read side by side. Most of the other discrepancies in Table 2 are explicable

on the same basis. Read against the internal quality assurance system that every Indonesian institution is required to operate, the persistence of this duplication indicates a specific and correctable weakness. An internal audit organised as a compliance checklist asks whether a lesson plan exists for each course, whether each column is filled, and whether the document was submitted on time; a duplicated document satisfies all three tests, because it is present, complete, and punctual. Only a content-based audit, one that reads the substance of a document against the curriculum it is supposed to implement and against the documents of adjacent courses, would detect that the document describes a different course. The discrepancies in Table 2 are, on this reading, not evidence that quality assurance was absent but evidence that it was operating at the level of document existence rather than document substance.

These findings support a qualified reading of framework adoption. The programme has adopted outcome-based curriculum design and cites an international proficiency framework in its programme outcomes, yet neither is doing the work the documents imply. Programme outcomes are loaded onto courses without being carried into course outcomes; a proficiency level is named without any course stating outcomes at that level; and the sequence does not rise. This is consistent with reviews reporting uneven framework implementation despite formal adoption ([Govindarajoo et al., 2025](#); [Jeon, 2022](#)), and extends them by locating the failure specifically at the transition from programme-level statement to course-level document.

Three practical implications follow, each stated so that it can be acted on. The first concerns who writes the documents. Lesson plans for a sequence of related courses should be drafted and revised in a single scheduled cross-course session rather than separately by each convenor, with the four documents read side by side against the curriculum and against one another. This is the smallest change that would have prevented every discrepancy in Table 2, and reported applications of constructive alignment suggest that it is workable when alignment is treated as an explicit design step with a shared representation ([Pereira et al., 2024](#); [Spangenberg, 2025](#); [Zhou et al., 2026](#)). The second concerns what internal quality assurance checks. Programme-level audit should include a content-based verification of course identity, outcome load, semester placement, and the traceability of every assessment weighting to a stated outcome, since the last of these can be checked arithmetically by a reader who is not a subject specialist. The third concerns how proficiency levels are stated. A two-credit second-semester course cannot deliver C1, and stating that it targets C1 is not ambition but incoherence; each course should state the band it contributes within a rising sequence, so that the relation between course level and graduate level becomes

explicit and checkable. Beyond the programme studied, the coding framework and the three checks above are offered as a reusable instrument for programmes that wish to audit their own course sequences.

Several limitations bound these claims. The study covers one programme, so the indices describe that programme and are not offered as norms. Because no instruments or rubrics existed and no classroom observation was conducted, findings are confined to the design level and say nothing directly about teaching or learning as enacted. Independent double coding was not carried out, for the reasons set out in the method, so the indices rest on the recorded evidential basis of each decision rather than on a reliability coefficient. Where an independent coder with the necessary curricular knowledge is available, applications of this coding framework should include at minimum a partial reliability check: a stratified random subsample of twenty to thirty per cent of the units of analysis, which for a corpus of this size would be six to eight units spread across the four courses and the three component pairings, re-coded independently against the same rubric, with agreement reported as both simple percentage agreement and a chance-corrected coefficient such as Cohen's kappa. Adopting that procedure would give the indices a reliability warrant that the present study cannot provide. Finally, the band thresholds used to interpret the indices are a convention, and readers who prefer other thresholds can recompute from the reported sums.

CONCLUSION

This study evaluated constructive alignment across four sequenced Reading courses through document analysis of a programme curriculum, five lesson plans, and one instructional module. Mean alignment indices fell in the moderate band for all four courses, between 60.0% and 70.4%, with no course reaching the high band. Alignment declined in the same order in every course, and assessment was rated low in all four, falling to 27.8% in one. Document verification revealed duplication of one lesson plan into another, two conflicting versions of a third, deviations in outcome load and semester placement, and the complete absence of assessment instruments and rubrics. Stated outcomes indicated levels B1, B1, B2, and B1 against a targeted C1, a progression that does not rise across the sequence.

Taken together, these results indicate that misalignment in this programme is systemic in one component rather than idiosyncratic to one course, and that it originates in documents produced individually, never reviewed as a set, and audited for existence rather than for substance. The study accordingly proposes three measures: cross-course collaborative drafting of lesson plans for related sequences,

content-based rather than checklist-based internal document audit, and the statement of course-level proficiency targets as stages within a rising sequence. These measures require no new documentation beyond the lesson plan template that institutions already mandate. Further research might apply the coding framework used here to other course sequences and disciplines, extend it to the level of assessment items where instruments exist, and test whether joint authoring of sequenced lesson plans measurably improves documented alignment.

CONFLICTS OF INTEREST: The authors declare no conflict of interest.

REFERENCES

- Abejuela, H. J., Akut, K., Del Rosario, A. S., & Balane, C. (2023). Assessment of the reading curriculum in basic education in the Philippines context. *International Journal of Language Education*, 7(1), 26–45. <https://doi.org/10.26858/ijole.v1i1.23641>
- Ahmad, D. R. P., Widiati, U., Laksmi, E. D., Muniroh, S., & Istiqomah, D. M. (2026). From reading comprehension to critical literacy: Adapting Huh's model in online Indonesian EFL classrooms. *Online Learning in Educational Research*, 6(2), 39–54. <https://doi.org/10.58524/oler.v6i2.1087>
- Al Lawati, Z. A. (2023). Investigating the characteristics of language test specifications and item writer guidelines, and their effect on item development: A mixed-method case study. *Language Testing in Asia*, 13(1), 21. <https://doi.org/10.1186/s40468-023-00233-5>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
- Govindarajoo, M. V. V., Shorouk, A., Benlahcene, A., & Taddese, E. T. (2025). Exploring perceptions of Malaysian primary school teachers towards the implementation of Common European Framework of Reference for Languages-aligned English textbooks. *Humanities and Social Sciences Letters*, 13(2), 767–780. <https://doi.org/10.18488/73.v13i2.4300>
- Jeon, J. (2022). A systematic review of CEFR-related research of English education in South Korea. *Journal of Curriculum and Teaching*, 11(8), 363–374. <https://doi.org/10.5430/jct.v11n8p363>

- Jin, H., & Seo, J. (2026). Curriculum textbook consistency framework for evaluating Chinese conversation textbooks in Korean secondary education. *Discover Education*, 5(1), 144. <https://doi.org/10.1007/s44217-026-01117-5>
- Kuperman, V. (2025). Intersample variance of second-language readers should not be overlooked. *Bilingualism: Language and Cognition*, 28(3), 757–763. <https://doi.org/10.1017/S1366728924000932>
- McEwan, M., Pate, A., & Wilder-Davis, K. (2023). Academic development: Leading by example with an authentic and practical approach to curriculum design. *Journal of University Teaching and Learning Practice*, 20(4). <https://doi.org/10.53761/1.20.4.12>
- Oktavianti, I. N., Budiwati, T. R., Prayudha, Prayogi, I., & Hidayah, I. N. (2026). Students' attitudes, challenges, and needs in English reading comprehension: Foundations for AI-integrated corpus-based language pedagogy. *Discover Education*, 5(1), 34. <https://doi.org/10.1007/s44217-026-01115-7>
- Peraturan Menteri Pendidikan Tinggi, Sains, dan Teknologi Nomor 10 Tahun 2026 tentang Perubahan atas Peraturan Menteri Pendidikan Tinggi, Sains, dan Teknologi Nomor 39 Tahun 2025 tentang Penjaminan Mutu Pendidikan Tinggi. (2026). *Berita Negara Republik Indonesia Tahun 2026 Nomor 477*. <https://peraturan.bpk.go.id/Details/352678>
- 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(1), 25. <https://doi.org/10.1186/s41239-024-00457-2>
- Poolkrajang, A., & Papanai, R. (2024). Reflective thinking on enhancing student competencies in learning management through outcome-based learning according to constructive alignment. *International Journal of Engineering Pedagogy*, 14(6), 4–22. <https://doi.org/10.3991/ijep.v14i6.48409>
- Safitry, T. S., Yuwono, U., & Halimi, S. S. (2023). A review of the test items of the LPATE from the perspective of the CEFR's 'can-do' descriptors. *Journal of Higher Education Theory and Practice*, 23(11), 182–203. <https://doi.org/10.33423/jhetp.v23i11.6229>
- Shin, S.-Y., & Sanonguthai, S. (2026). Constructing CEFR-aligned English reading assessments with AI: A practical guide for Thai English teachers in higher education. *LEARN Journal: Language Education and Acquisition Research Network*, 19(2), 71. <https://doi.org/10.70730/FDFL1823>
- 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>
- Tangsakul, S., & Poonpon, K. (2024). Aligning academic reading tests to the Common European Framework of Reference for Languages (CEFR). *rEFLections*, 31(2), 614–638. <https://doi.org/10.61508/refl.v31i2.275057>

- Trowsdale, D., & McKay, A. (2023). EdVee: A visual diagnostic and course design tool for constructive alignment. *Teaching and Learning Inquiry*, 11. <https://doi.org/10.20343/teachlearninqu.11.3>
- Zainurrahman, Z., Yusuf, F. N., & Sukyadi, D. (2024). Text readability: Its impact on reading comprehension and reading time. *Journal of Education and Learning (EduLearn)*, 18(4), 1422–1432. <https://doi.org/10.11591/edulearn.v18i4.21724>
- Zaporozhets, A., & Zaporozhets, T. (2026). Architecture of assessment systems in higher education: Multilevel model and diagnostic criteria. *Journal of Vasyl Stefanyk Precarpathian National University*, 13(1), 145–160. <https://doi.org/10.15330/jpnu.13.1.145-160>
- Zhou, X., Chai, Q., Chilukuri, B., & Quach, J. J. Y. (2026). From experimentation to integration: Embedding generative artificial intelligence in business higher education through the lens of constructive alignment. *Journal of University Teaching and Learning Practice*, 23(2). <https://doi.org/10.53761/pc04tp05>