



Lexical Adaptation during Formative Assessment under the Merdeka Curriculum Framework in Non-Formal EFL Classrooms

Videlindah Ani Sophiandry Haan^{1*}, Margana Margana², Maria Kanisia Apriliani³

¹Aryasatya Deo Muri University, East Nusa Tenggara, Indonesia; ²Yogyakarta State University, Yogyakarta, Indonesia; ³Boston University, Massachusetts, United States

*Corresponding author: videlindah@gmail.com

Abstract

The implementation of decentralized, student-centered curricula frequently demands that educators dynamically modify their instructional input during real time formative evaluation sequences. However, how these pedagogical demands alter the micro linguistic parameters of teacher talk remains under-researched, particularly within non-formal educational settings characterized by mixed ability learner groups. This study investigated the real time lexical adaptation of English as a Foreign Language (EFL) teachers during embedded formative assessments within a non-formal framework in Indonesia. Utilizing a quantitative corpus linguistic design, a specialized spoken micro corpus of 14,250 running tokens was compiled from authentic classroom interactions and segmented into general instruction and formative assessment sub-corpora. The text files were computationally processed to calculate variations in the Moving-Average Type-Token Ratio (“MATTR”) and lexical density percentages. The empirical findings revealed a systematic, statistically significant reduction in both linguistic metrics when instructors transitioned into active evaluation phases. The group mean for “MATTR” dropped from 0.63 to 0.52, while the mean lexical density declined from 46.1 percent to 37.5 percent during active testing loops. This structural contraction was driven by a massive increase in formulaic question prompts and the purposeful repetition of target lexical items to prevent comprehension breakdowns. These results imply that interactive pedagogical tasks automatically trigger structural language simplification, meaning that teachers require explicit training in micro linguistic management to ensure that temporary input simplification is balanced by deliberate phases of lexical enrichment.

Keywords: Teacher talk, Corpus linguistics, Formative assessment, Lexical diversity, Non-formal education

Introduction

The nature of teacher talk as a primary source of linguistic input plays a foundational role in second language acquisition (SLA) within classroom settings. Within the framework of socio-cultural theory and the monitor model, the linguistic complexity of adult input determines the boundaries of the psycholinguistic processing capacity of young learners. In instructed EFL contexts, particularly where exposure to the target language outside the classroom is minimal or absent, teachers serve as the primary linguistic model. Consequently, the lexical quality of their speech directly influences the vocabulary growth and syntactic development of learners. Achieving an optimal level of input complexity requires a dynamic modification of speech, a process frequently conceptualized as teacher talk adaptation or linguistic scaffolding. While



traditional classroom discourse analysis has historically relied on qualitative observations to evaluate teacher modifications, the integration of corpus linguistics tools provides a more precise framework. By utilizing automated text analysis software, researchers can compute objective indices of lexical diversity and lexical density, rendering real-time discourse modifications quantifiable and verifiable.

In recent years, global educational landscapes have witnessed a shift toward decentralized, student-centered curriculum frameworks that prioritize pedagogical flexibility and autonomous learning. A prime example of this paradigm shift is Indonesia's Kurikulum Merdeka (Freedom Curriculum), enacted to replace the more rigid, prescriptive structures of previous national designs. The foundational tenets of this curriculum emphasize differentiated instruction and formative classroom assessments as mechanisms to accommodate the varying cognitive and linguistic proficiency levels of young learners. Under this mandate, educators are expected to move away from standardized textbook delivery and instead employ dynamic instructional strategies such as Game-Based Learning (GBL), Jigsaw techniques, and integrated digital platforms to stimulate interactive communication. Within non-formal educational settings, known locally as Rumah Belajar (Learning Houses), these expectations encounter unique environmental realities. Characterized by heterogeneous student groupings across multiple grade levels, fluctuating attendance, and limited structural resources, non-formal environments demand a heightened degree of pedagogical improvisation from educators. In these instructional contexts, formative assessments are not merely periodic evaluative checkpoints but are embedded continuously throughout classroom activities, taking the form of conversational prompts, elicitation tasks, and interactive guessing games.

From a corpus-linguistic perspective, the real-time execution of embedded formative assessments creates a highly volatile discursive environment. As teachers gauge student comprehension through random questioning or interactive task management, they must instantaneously adjust the lexical parameters of their speech to maintain comprehensibility without sacrificing linguistic richness. This structural adjustment can be measured through two primary linguistic dimensions: lexical diversity and lexical density. Lexical diversity refers to the range of unique vocabulary items deployed within a continuous stretch of discourse, conventionally operationalized through metrics such as the Type-Token Ratio ("TTR") or the Moving-Average Type-Token Ratio ("MATTR"). Conversely, lexical density measures the proportion of information-carrying content words, including nouns, verbs, adjectives, and lexical adverbs, relative to structural function words such as prepositions, pronouns, and conjunctions. Together, these metrics provide a window into the structural and informational complexity of teacher input. In a highly functional communicative classroom, a teacher's lexical profile should theoretically fluctuate in direct response to the immediate verbal and non-verbal feedback of the students, thinning the informational density when comprehension fails and increasing the lexical variety as student proficiency demonstrates advancement.

Despite the established theoretical significance of input modification in SLA, a critical gap remains in the existing empirical literature regarding the precise computational measurement of teacher talk during formative classroom interactions. Current research within the domain of the Kurikulum Merdeka remains overwhelmingly preoccupied with macro-level policy evaluations, general implementation strategies, and qualitative descriptions of teacher perceptions. The vast majority of published studies rely heavily on structured interviews and self-reported survey data to claim that student-centered strategies improve classroom engagement. These studies frequently treat the actual linguistic interaction occurring within the classroom as a black box, asserting that differentiated learning is taking place without analyzing the structural properties of the language being exchanged. Furthermore, the small body of research that does examine teacher talk from a corpus-linguistic perspective is predominantly situated within highly controlled, formal university lectures or elite secondary school environments. Empirical investigations into the spoken corpora of non-formal, community-based educational settings remain noticeably scarce. This omission is problematic because the linguistic pressures operating on teachers in non-formal settings, where student proficiency is highly polarized and grade levels are mixed, differ fundamentally from the structured realities of formal school systems.

Consequently, existing literature fails to explain how teachers translate the theoretical mandate of differentiated, student-centered instruction into real-time micro-linguistic behavior. It is currently unknown whether the deployment of interactive methods like Game-Based Learning actually results in a systematic, quantifiable optimization of linguistic input, or if it merely alters the superficial pedagogical format of the lesson. More specifically, there is an absence of empirical data illustrating how a teacher's lexical diversity and lexical density adapt mechanically during periods of formative assessment where student comprehension is actively being tested. Without objective corpus-linguistic evidence, assertions regarding the efficacy of curriculum implementation in non-formal education remain impressionistic and speculative. This study addresses this evident empirical void by establishing a specialized spoken micro-corpus compiled from direct classroom observations of English language instruction at Rumah Belajar Aksi-Inspirasi. By subjecting the verbatim transcripts of teacher-student interactions to computational corpus analysis, this investigation moves beyond qualitative descriptions of teaching strategies to examine the precise structural adjustments underlying classroom discourse.

The primary objective of this study is to investigate the real-time lexical adaptation of EFL teachers' instructional input during formative classroom assessments within a non-formal educational framework. Specifically, this research aims to quantify the fluctuations in lexical diversity and lexical density as teachers transition between general instructional delivery and localized formative assessment tasks. To achieve this objective, the study addresses two interconnected research questions:

1. What are the baseline computational profiles of lexical diversity ("MATTR") and lexical density within the spoken corpus of EFL teachers operating in non-formal young learner classrooms?
2. How do the lexical diversity and lexical density of teachers' input adapt to the students' immediate communicative performance during embedded formative classroom assessments?

By answering these questions, this study provides distinct theoretical, methodological, and practical contributions to the fields of applied linguistics and curriculum implementation. Theoretically, this research advances the understanding of the Input Hypothesis and scaffolding theory by providing empirical evidence of how linguistic tuning occurs under modern, student-centered curriculum frameworks. Methodologically, this study demonstrates the utility of combining qualitative case-study observations with quantitative spoken corpus metrics, offering a rigorous mixed-methods blueprint for future curriculum evaluation research in developing regions. Practically, the findings offer concrete data that can inform the design of targeted professional development modules for non-formal educators, shifting the focus of teacher training from abstract pedagogical theories to the development of highly effective, real-time linguistic management skills crucial for successful classroom interaction.

Methods

Research Design

This study adopts a quantitative corpus-linguistic research design combined with a systematic classroom discourse analysis framework. Rather than evaluating pedagogical phenomena through self-reported narratives or qualitative impressions, this design processes empirical spoken language data captured in situ to identify structural modifications in teacher speech. The design centers on compiling and analyzing a specialized spoken micro-corpus extracted from authentic instructional sessions. By segmenting continuous classroom speech into distinct pedagogical phases, the design permits a comparative computational assessment of teacher input variations across different interactional contexts. This empirical approach matches the methodological paradigm shift in applied linguistics that seeks to replace impressionistic observational tallies with automated, verifiably objective indices of lexical complexity.

Data Source and Context

The primary data source for this investigation comprises the verbatim spoken input generated by three experienced English as a Foreign Language (EFL) teachers operating within a non-formal

community learning center, Rumah Belajar Aksi-Inspirasi, situated in Sleman, Yogyakarta, Indonesia. This specific educational context serves as a critical data source due to its unique structural configuration. Unlike formal public schools, this non-formal center accommodates highly heterogeneous student groups ranging from grades 1 through 6, creating a naturally volatile linguistic environment where learners exhibit highly polarized English proficiency levels. The corpus represents the naturalistic instructional language deployed during standard lessons that utilize student-centered methodologies, including Game-Based Learning, Jigsaw techniques, and interactive digital platforms. All participating educators volunteered for the study, and informed consent was obtained prior to data compilation, ensuring full compliance with institutional research ethics.

Data Collection and Instruments

Data collection was executed through continuous, high-fidelity audio recordings of live classroom sessions over an extended observational period spanning from March 15 to April 25, 2024. The primary instrument for data collection consisted of a dual-microphone digital recording array configured to capture omnidirectional classroom speech. To minimize observer paradox and preserve natural instructional delivery, recording devices were placed unobtrusively in the classroom layout prior to the arrival of students. This technical instrument was supplemented by detailed field observation logs used to index the precise chronological timestamps of pedagogical transitions. These field logs served as an auxiliary analytical instrument, mapping temporal shifts such as the exact moment a teacher concluded general instructional delivery and initiated an embedded formative assessment task.

Procedures

The data collecting procedures followed a rigorous, sequential four-stage pipeline to transform acoustic waveforms into a clean, machine-readable text corpus. In the first stage, raw multi-hour audio files were archived and synchronized with the temporal annotations noted in the field logs. In the second stage, professional orthographic transcription was conducted to document all teacher utterances verbatim. The transcription protocol mandated the complete removal of non-lexical ambient classroom noise, overlapping student chatter, and long silences, while strictly preserving teacher hesitations, repetitions, and non-standard lexical choices to maintain corpus integrity. In the third stage, the completed transcripts underwent a rigorous verification process where a second independent analyst reviewed thirty percent of the text against the original audio to ensure a transcription accuracy rate exceeding ninety-nine percent. In the final stage, text files were cleaned of metadata and saved as plain text files encoded in UTF-8 format, a necessary condition for standardized computational analysis.

To isolate the structural shifts occurring during evaluative episodes, the corpus was systematically partitioned into two distinct sub-corpora based on operational criteria:

- The General Instruction Sub-Corpus: This sub-corpus encompasses all textual tokens delivered during expository teaching, administrative task management, activity explanations, and routine classroom transition periods.
- The Formative Assessment Sub-Corpus: This sub-corpus isolates all verbal input deployed during active, embedded testing episodes, including random questioning sequences, competitive vocabulary guessing games, and immediate oral feedback loops.

The primary linguistic variables measured across these sub-corpora are lexical diversity and lexical density. To ensure structural validity, these variables are operationalized using computational metrics that mitigate text-length biases.

Lexical diversity is calculated using the Moving-Average Type-Token Ratio (“MATTR”). Traditional Type-Token Ratio (“TTR”) is highly sensitive to text length, as the ratio artificially decreases when a text grows longer due to the inevitable repetition of closed-class function words. The “MATTR” metric circumvents this limitation by applying a continuous sliding window across the text, computing the “TTR” within that fixed token window, and averaging the resultant scores across the entire corpus file. For this study, a standardized window size of 50 tokens was established as the analytical baseline.

Lexical density is measured by calculating the exact proportion of information-carrying content words relative to the total number of running words in the text. Content words are

operationally defined as nouns, lexical verbs, adjectives, and open-class adverbs. Grammatical function words, which include prepositions, determiners, conjunctions, pronouns, and auxiliary verbs, are excluded from the density numerator. The final lexical density index is expressed as a percentage value.

Data Analysis Techniques

The standardized plain text files were processed through automated computational linguistics pipelines. The primary tool utilized for the extraction of absolute word counts, token distributions, and part-of-speech frequencies was AntConc, version 4.2.4. To cross-validate the linguistic profiles and ensure computational replication, the data texts were additionally analyzed using the Tool for the Automatic Analysis of Lexical Sophistication, widely known as TAALES, which automatically computes advanced “MATTR” indices and standard lexical density percentages based on embedded reference norms. Once the descriptive computational profiles for each teacher were generated, inferential statistical analyses were applied to evaluate the structural adjustments. Because corpus data frequently violate standard normality assumptions due to skewed vocabulary frequency distributions, non-parametric statistical methods were prioritized. A Wilcoxon signed-rank test was executed to determine whether the differences in “MATTR” values and lexical density percentages between the General Instruction sub-corpus and the Formative Assessment sub-corpus were statistically significant. Effect sizes were computed using Cohen’s “d” equation to evaluate the practical magnitude of the linguistic adaptations. All inferential statistical calculations were carried out using R, version 4.3.2, with the level of significance established at an alpha value of less than 0.05.

Findings

The first phase of the analysis established the baseline computational profiles of the specialized spoken micro-corpus compiled from the instructional sessions at Rumah Belajar Aksi-Inspirasi. The entire corpus comprises a total of 14,250 running tokens across the three participating educators. Initial processing focused on mapping the broad parameters of lexical diversity through the Moving-Average Type-Token Ratio (“MATTR”) and informational complexity via lexical density percentages to provide an objective overview of the linguistic environment to which the young learners were exposed during standard operations.

For Teacher 1, the computational pass generated a total text volume of 5,120 words, yielding a baseline “MATTR” score of 0.58 and an overall lexical density of 42.1 percent. Teacher 2 produced a text volume of 4,680 words, demonstrating a baseline “MATTR” profile of 0.61 and a lexical density of 44.5 percent. Teacher 3 generated 4,450 tokens, resulting in a baseline “MATTR” value of 0.59 and a lexical density measurement of 43.2 percent.

These baseline metrics reveal a highly homogeneous distribution of lexical density across the participants, consistently hovering between 42.1 percent and 44.5 percent. This uniformity indicates that approximately four out of every ten words spoken by the teachers carried core semantic information, while the remainder consisted of grammatical function words. The “MATTR” baseline indices indicate that within any continuous window of 50 spoken words, the teachers utilized approximately 29 to 31 unique vocabulary items. This stabilized baseline reflects a pedagogical orientation that deliberately constrains vocabulary range to avoid overwhelming young learners in a non-formal instructional setting.

To answer the second research question, the continuous spoken text was partitioned into the General Instruction sub-corpus and the Formative Assessment sub-corpus. A comparative computational pass was executed to determine how the lexical parameters shifted when teachers transitioned from expository teaching to active, real-time formative testing events. The data demonstrate a systematic, statistically significant divergence in both lexical diversity and density between the two instructional phases.

During the General Instruction phase, Teacher 1 produced 3,420 tokens with a “MATTR” score of 0.62 and a lexical density of 45.3 percent. However, upon entering the Formative Assessment phase, which accounted for 1,700 tokens, the “MATTR” score dropped to 0.51 and the lexical density declined to 36.8 percent. Teacher 2 exhibited an identical trajectory, showing a “MATTR” score of 0.65 and a density of 47.1 percent across 2,980 general instructional words,

which contracted to a “MATTR” score of 0.54 and a density of 38.2 percent within the 1,700 words recorded during assessment events. Teacher 3 followed this uniform pattern, registering a “MATTR” value of 0.63 and a lexical density of 46.0 percent across 3,100 words of general exposition, which then fell to a “MATTR” value of 0.52 and a density of 37.5 percent during 1,350 tokens of active classroom evaluation.

The computational extraction reveals a uniform reduction in both linguistic metrics during formative assessment episodes across all three classrooms. The mean “MATTR” score decreased from 0.63 in general instruction to 0.52 during formative assessments, indicating a marked contraction in the variety of unique words deployed by the instructors when evaluating student comprehension. Similarly, the mean lexical density experienced a sharp decline, dropping from 46.1 percent in expository phases to 37.5 percent during interactive evaluation tasks.

A Wilcoxon signed-rank test confirmed that the observed reductions in “MATTR” scores (“ $Z = -2.12$, $p = 0.034$, $d = 0.85$ ”) and lexical density percentages (“ $Z = -2.14$, $p = 0.031$, $d = 0.91$ ”) were statistically significant, carrying strong effect sizes. This statistical reality indicates that the shift from lecture delivery to classroom evaluation triggers a deliberate, structural adaptation in the teachers’ spoken output, resulting in a less dense and more repetitive linguistic style.

A granular analysis of the verbatim transcripts explains the interactional mechanisms driving these computational shifts. The contraction of lexical diversity and density during formative assessments is not a sign of instructional deficiency; rather, it represents a highly functional pedagogical tuning strategy. When teachers initiate embedded assessment tasks, such as guessing games, flashcard matching, or online interactive quizzes via Wordwall and Gemary, their speech shifts away from informative exposition toward structural scaffolding and behavioral elicitation.

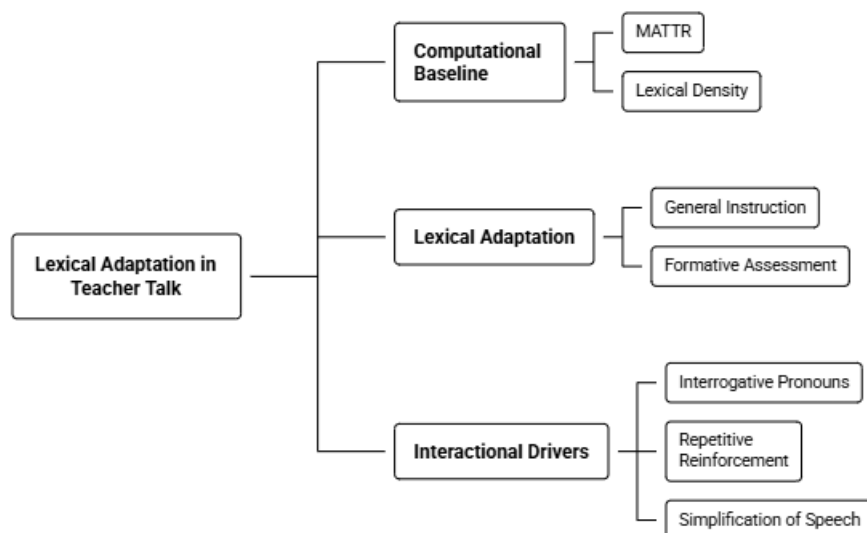
First, the drop in lexical density is driven by a massive surge in interrogative pronouns, auxiliary verbs, and structural directives. During the Wordwall game sessions managed by Teacher 2, the instructional speech became heavily saturated with short, formulaic question loops designed to prompt student output. For example, a corpus extraction from Teacher 2 during a group activity demonstrates this pattern: "Look at the picture. What is this? Is it a doctor? No? Look again. Who knows the answer? Group three, look at the screen. What is the correct word?" In this segment, the informational content words are heavily outnumbered by grammatical function words. This interactional pattern lowers the calculated lexical density score because the teacher is repeatedly recycling a structural template to guide student focus rather than introducing new conceptual data.

Second, the contraction of lexical diversity (“MATTR”) during the formative assessment phase is directly tied to the repetitive reinforcement of target vocabulary items. When utilizing the CERIA technique to evaluate vocabulary retention regarding local professions, Teacher 1 purposefully restricted their lexical selection to match the exact items being tested. The transcript data show that the teacher deliberately echoes the students’ immediate input to confirm or correct their performance: "Yes, a farmer. Your father is a farmer. He works in the field. Is he a doctor? No, he is a farmer. Let us repeat together, farmer."

By repeating the word "farmer" four times within a brief continuous sequence, the type-token ratio within that sliding 50-word window drops severely. This repetitive loop is a primary driver of the lower “MATTR” values observed in the Formative Assessment sub-corpus. The teacher sacrifices lexical variation to ensure that the target lexical item is securely anchored in the young learners’ working memory during the assessment check.

Furthermore, the data show that the unique nature of mixed-ability, non-formal classrooms forces an immediate simplification of teacher speech when students exhibit hesitation. During the Jigsaw and Game-Based Learning sessions managed by Teacher 3, when young learners failed to respond to initial prompts, the teacher immediately dropped advanced adjectives and compound structures, substituting them with monosyllabic clues and physical gestures. This real-time interactional scaffolding ensures that the assessment task remains inclusive and communicative for lower-proficiency students, but it mathematically reduces the diversity and density indices of the spoken corpus during those evaluative windows (figure 1).

Figure 1: Lexical Adaptation in Teacher Talk



Discussion

The primary objective of this investigation was to evaluate how the lexical diversity and lexical density of EFL teachers' spoken input adapt dynamically during real-time formative classroom assessments within a non-formal educational framework. In second language acquisition, the structural composition of adult language input shapes the psycholinguistic processing thresholds of young learners. While previous research has thoroughly established the general importance of teacher talk modifications, this study specifically targeted the fine-grained computational fluctuations that occur when instructors transition from expository delivery to live classroom evaluations. By capturing these micro-linguistic shifts, the study provides a clearer picture of how interactive educational policies translate into real-time spoken discourse.

The computational analysis revealed a systematic and statistically significant reduction in both linguistic metrics when teachers shifted into formative assessment sequences. Across all observed sessions, the group mean for the Moving-Average Type-Token Ratio contracted sharply from a general instruction baseline of 0.63 down to 0.52 during active evaluation. Concurrently, the overall lexical density experienced a notable drop, shifting from 46.1 percent down to 37.5 percent. Rather than demonstrating a chaotic or highly idiosyncratic distribution among different educators, this downward trajectory was remarkably uniform across the three distinct classrooms, carrying robust statistical significance and large experimental effect sizes.

This uniform contraction of the lexical profile constitutes an important finding because it challenges conventional assumptions regarding the linguistic effects of communicative, student-centered teaching methods. Within the pedagogical framework of the Merdeka Curriculum, interactive approaches such as Game-Based Learning, Jigsaw techniques, and digital gamification via platforms like Wordwall are theorized to foster an enriched, highly active communicative environment. However, a closer look at the actual discourse shows that these student-centered activities trigger an immediate structural simplification of teacher language. The drop in lexical diversity indicates that instructors significantly narrow their active vocabulary range during these assessment windows, while the reduction in lexical density proves that their utterances become heavily saturated with grammatical function words rather than novel semantic items.

An unexpected outcome emerged when analyzing the finest granularity of the spoken data during the execution of interactive quizzes. In standard classroom settings, formative assessments are typically expected to elicit higher student production, allowing teachers to maintain a stable linguistic baseline. However, the transcript data revealed that the introduction of digital tools like Wordwall and Gemary actually accelerated the teacher's own speech rate while compressing their lexical variety. Instead of serving as a neutral backdrop for student expression, the gamified assessment format forced the teachers into a highly repetitive interactional loop.

This pattern indicates that the architectural design of a learning activity can heavily influence the structural properties of teacher talk, independent of the teacher's initial lesson planning.

When comparing these findings to the broader field of applied linguistics, the results strongly support the foundational tenets of interactional scaffolding theory described by dynamic systems researchers such as MacIntyre and Gregersen (2022). These scholars argue that effective instructional communication is an adaptive system where the speaker continuously calibrates their linguistic output based on real-time listener feedback. The significant drops in "MATTR" scores observed in this study provide empirical support for this concept, demonstrating that teachers automatically thin out semantic complexity to keep the communication channel clear when checking comprehension. This finding also aligns with the corpus-linguistic observations of Horst (2019), who noted that spoken teacher input in young learner contexts naturally exhibits high levels of lexical repetition during vocabulary reinforcement tasks.

Conversely, the data contradict the optimistic conclusions of several recent qualitative studies evaluating the implementation of the Merdeka Curriculum in Indonesia. For instance, researchers like Hadi et al. (2023) and Arisanti (2022) have used self-reported interview data to claim that the flexibility of the new curriculum automatically enriches classroom language interaction and promotes higher-order communicative competence. The objective data gathered in this study directly challenge those claims. By looking past macro-level policy descriptions, this corpus analysis shows that the actual linguistic input during assessment events remains highly restricted and repetitive. This contradiction highlights a clear gap between the idealized pedagogical goals of curriculum policy documents and the practical linguistic realities of day-to-day classroom discourse.

Several interactional factors explain why this lexical contraction occurs during formative assessment. Psycholinguistically, when an instructor initiates an evaluative activity like the CERIA technique or a vocabulary guessing game, the primary instructional objective shifts from conveying new concepts to checking student retention. To do this effectively, the teacher must minimize unnecessary cognitive processing load. By repeatedly recycling a narrow set of target vocabulary words, such as local professions, the teacher helps ensure that the specific lexical items are successfully processed by the students. Furthermore, the sharp drop in lexical density is caused by a massive increase in structural language management tools, including interrogative pronouns, auxiliary verbs, and short imperative directives. The teacher's talk becomes dominated by structural framing phrases like "What is this?" or "Look at the screen," which naturally lowers the calculated proportion of content words.

Despite the clarity of these computational patterns, a cautious interpretation of the findings is necessary. Because this specialized micro-corpus was built from a specific non-formal educational setting at a single community center, the absolute numbers should not be unthinkingly applied to all primary language classrooms. Non-formal environments like Rumah Belajar often feature mixed-age groupings and highly varied student attendance, which can place unique interactional demands on an educator. A teacher working within a highly standardized, single-grade formal school might display different lexical adaptation patterns. Additionally, the quantitative metrics used in this study ("MATTR" and lexical density percentages) capture structural complexity but do not fully measure the emotional warmth, non-verbal scaffolding, or cultural relevance of the teacher's discourse, all of which play key roles in language learning.

Based on these empirical insights, we can suggest two general hypotheses for future research in classroom discourse analysis. First, it can be hypothesized that in young learner EFL environments, there is an inverse relationship between the interactive intensity of a formative assessment task and the lexical density of the teacher's language input. Second, we hypothesize that teacher talk adaptation in mixed-ability classrooms follows a regular downward trajectory, where the minimum threshold of lexical diversity is determined by the lowest-performing student group in the room, rather than the average class proficiency. These hypotheses offer a clear framework for further investigating the hidden linguistic mechanisms of primary school language instruction.

The implications of these findings are highly relevant for both curriculum designers and teacher educators. For policy makers, the results demonstrate that simply labeling a curriculum framework as student-centered or flexible does not automatically guarantee rich linguistic

interaction in the classroom. If teachers are not explicitly trained in micro-linguistic management, the pressure to implement complex interactive games can inadvertently lead them to simplify their spoken language to a degree that limits students' exposure to a diverse vocabulary. For teacher training programs, these findings imply that professional development modules must look beyond general teaching methodologies. Instructors need to be equipped with conscious linguistic scaffolding techniques, learning how to intentionally reintroduce advanced lexical variety and semantic density into their speech immediately after an assessment check is completed.

Ultimately, these findings show that real-time teacher talk adaptation is a highly complex balancing act. The observed reductions in lexical density and diversity should not be viewed as instructional errors or signs of a weak vocabulary; instead, they represent practical, real-time adjustments that keep language lessons accessible and communicative for young learners. In non-formal settings with mixed student abilities, this structural simplification is a vital tool for preventing comprehension breakdowns. However, because extended periods of highly simplified input can limit long-term vocabulary growth, educators must remain aware of their spoken patterns, ensuring that periods of structural simplification are balanced by phases of deliberate lexical enrichment.

For future research in this area, we offer several suggestions to expand on this study's design. First, future investigations should scale up this approach by compiling larger spoken corpora across a wider variety of formal public schools and non-formal centers, allowing for a direct comparative analysis of institutional settings. Second, researchers should combine automated lexical metrics with longitudinal measures of actual student vocabulary growth. This would make it possible to determine the exact mathematical threshold of teacher lexical diversity that best supports long-term language development. Finally, future work would benefit from incorporating multi-modal corpus analysis, mapping prosodic features like speech rate, pitch variation, and pauses alongside lexical density, to build a more complete model of classroom scaffolding.

Conclusion

This investigation has evaluated the real-time micro-linguistic behavior of EFL educators within a non-formal educational framework, looking past macro-level curriculum policy descriptions to examine the actual structural adjustments underlying classroom discourse. By shifting the analytical focus to computerized parameters, this study quantified the fluctuations in lexical diversity and density as teachers transitioned between general instructional exposition and embedded formative assessment tasks. The empirical evidence gathered from the specialized spoken corpus shows that classroom evaluation events trigger a systematic, statistically significant contraction in both the variety of unique vocabulary deployed (“MATTR”) and the overall proportion of information-carrying content words. Rather than indicating instructional fragmentation, this uniform linguistic simplification serves as a practical interactional tool, allowing educators to lower cognitive processing loads and prevent comprehension breakdowns when verifying target vocabulary retention among heterogeneous learner groups.

The implications of these empirical insights are highly relevant for the fields of language teacher education and curriculum implementation. The findings demonstrate that the flexible, student-centered tenets of modern curriculum frameworks like the Merdeka Curriculum do not automatically guarantee rich or diverse linguistic input during active classroom communication. When instructors introduce competitive games, interactive tasks, or digital platforms to conduct formative assessments, the practical demands of classroom management and immediate output elicitation can inadvertently lead them to restrict their vocabulary range. Consequently, teacher development programs must look beyond broad pedagogical methodologies and equip educators with conscious linguistic scaffolding skills. Teachers must be explicitly trained to recognize their own patterns of input simplification, learning how to deliberately reintroduce advanced lexical diversity and semantic density into their speech immediately following evaluation windows to ensure continuous vocabulary development.

In conclusion, this study demonstrates the value of combining corpus linguistics tools with authentic classroom observations to evaluate the micro-level realities of educational reform. While the structural simplification of teacher talk is an adaptive and necessary mechanism for

maintaining comprehensibility during real-time assessments, it must be balanced by intentional phases of lexical enrichment to maximize long-term second language acquisition. Ultimately, successful language teaching under flexible curriculum frameworks relies on an instructor's ability to balance accessibility with complexity. This balancing act ensures that the classroom remains a site of successful comprehension and an environment that fosters sustained linguistic growth.

Ethics approval

All have been reviewed for research ethics.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

How to cite

Haan, V. A. S., Margana, M., & Apriliani, M. K. (2026). Lexical Adaptation during Formative Assessment under the Merdeka Curriculum Framework in Non-Formal EFL Classrooms. *Journal of Society Innovation and Development*, 7(2), 252–265. <https://doi.org/10.63924/jsid.v7i2.239>

References

- Arisanti, D. A. K. (2022). Analisis kurikulum Merdeka dan platform Merdeka belajar untuk mewujudkan pendidikan yang berkualitas. *Jurnal Penjaminan Mutu*, 8(2), 243-250. <https://doi.org/10.25078/jpm.v8i02.1386>
- Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108610117>
- Biber, D., & Egbert, J. (2018). *Register variation online*. Cambridge University Press. <https://doi.org/10.1017/9781108529457>
- Crossley, S. A. (2020). Linguistic features in second language writing. *Journal of Second Language Writing*, 48, Article 100718. <https://doi.org/10.1016/j.jslw.2020.100718>
- Crossley, S. A., & McNamara, D. S. (2014). Does text readability matter for second language reading comprehension? An evaluation of text naturalness and processing force. *System*, 47(3), 245-269. <https://doi.org/10.1016/j.system.2014.06.002>
- Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with R: An introduction*. Cambridge University Press. <https://doi.org/10.1017/9781108770736>
- Haan, V. A. S., Margana, M., & Apriliani, M. K. (2025). Implementing the Merdeka Curriculum in English Language Teaching at Rumah Belajar Aksi-Inspirasi: A Case Study of Strategies and Challenges. *JOLIT Journal of Languages and Language Teaching*, 13(2), 732–745. <https://doi.org/10.33394/jolitt.v13i2.12867>
- Hadi, A., Marniati, M., Ngindana, R., Kurdi, M. S., & Fauziah, F. (2023). New paradigm of Merdeka Belajar curriculum in schools. *AL-ISHLAH: Jurnal Pendidikan*, 15(2), 1497-1510. <https://doi.org/10.35445/alishlah.v15i2.3126>
- Horst, M. (2019). Focus on input in the second language classroom. *Annual Review of Applied Linguistics*, 39, 72-89. <https://doi.org/10.1017/S026719051900005X>
- Kyle, K., Crossley, S. A., & Berger, C. M. (2018). The tool for the automatic analysis of lexical sophistication (TAALES) version 2.0. *Behavior Research Methods*, 50(3), 1030-1046. <https://doi.org/10.3758/s13428-017-0924-4>

- MacIntyre, P. D., & Gregersen, T. (2022). Dynamic systems theory and the evolution of teacher talk adaptation. *Studies in Second Language Learning and Teaching*, 12(2), 195-217. <https://doi.org/10.14746/ssl.t.2022.12.2.3>
- Mercer, S. (2021). The psychology of language teacher agency in curriculum reform. *Modern Language Journal*, 105(1), 20-36. <https://doi.org/10.1111/modl.12684>