Journal of Language
Volume 7, Number 2, pp: 400-406, November 2025
e-ISSN: 2685-8878 | p-ISSN: 2655-9080

PATTERNS AND AWARENESS OF CODE-MIXING AMONG INDONESIAN YOUTHS IN AN ENGLISH-SPEAKING COMMUNITY: A SOCIOLINGUISTIC ANALYSIS

Riziq Prio Alfarras, Purwanto Siwi
Universitas Islam Sumatera Utara, Medan, Indonesia
E-mail: riziqalfarras2004@gmail.com

Received: 2025-09-09 Accepted: 2025-11-04 Published: 2025-11-15

Abstract

This qualitative study investigates the patterns and underlying factors of code-mixing (CM) among Indonesian youths in Aaron's English-Speaking Club, a community dedicated to practicing English in Medan, Indonesia. Employing a grounded theory approach by Charmaz (2016), the research involved in-depth interviews with ten club members, selected via probability sampling. The conversations were analyzed to quantify the frequency of CM (the insertion of Indonesian words into English discourse) and to identify the sociolinguistic factors influencing this phenomenon using the framework of Holmes (1992). Findings reveal that all participants engaged in CM, with percentages ranging from 0.52% to 24.34% of their total words. The analysis indicates that CM is not an indicator of low proficiency but a strategic, communicative resource. Key factors influencing CM included occupation, cultural attitude, and regional background. Notably, participants were initially unaware of the linguistic concepts of CM and code-switching (CS), but upon explanation, all identified more strongly with CM than CS in their typical language use. The study concludes that CM is a natural and prevalent feature of bilingual communication among these youths, reflecting their dynamic linguistic identity and the specific social context of their English practice community. It recommends increased metalinguistic awareness to empower speakers in their bilingual journey.

Keywords: *Bilingualism; code-mixing; English speaking community; Indonesian youth; language awareness; sociolinguistics*

1. Introduction

The linguistic landscape of Indonesia, particularly among its youth, has undergone a significant transformation in the digital age. The pervasive influence of social media and globalized content has made English more accessible than ever, fostering a generation of young Indonesians who are not just passive learners but active users of the language in their daily digital interactions (Thara & Poornachandran, 2018). This engagement often moves beyond pure English usage into a fluid blend of languages, a phenomenon prevalent in multilingual societies.

This blending is systematically studied in sociolinguistics as code-mixing (CM) and code-switching (CS). Code-mixing refers to the embedding of linguistic units—such as words, phrases, or clauses—from one language into an utterance of another language within a single sentence or discourse (Bhatia & Ritchie, 2004; Kim, 2006). Code-switching, while related, describes the alternation between two or more languages, or varieties of a language, at the level of discourse, often triggered by a change in context, topic, or participant (Aurer, 1998). In the Indonesian context, public figures like the celebrity Cinta Laura, known for mixing German, Indonesian, and English, and educational influencers like Aaron O'Brien, an Australian who integrates Indonesian into his English teaching, have popularized and normalized this linguistic behavior (Astari & Marantika, 2023; Kasanah & Hidayati, 2023).

Aaron's English Speaking Club (AESC) provides a fertile ground for observing these phenomena. It is a community where Indonesian youths consciously gather to practice English, creating a unique social context where the expectation is to use English, yet their underlying linguistic repertoire is firmly rooted in Indonesian. This tension between intention and habit makes the club an ideal microcosm to study how CM manifests among motivated learners in a relaxed, peer-driven environment.

While much research has focused on CM in educational or formal settings, there is a gap in understanding its patterns within informal, self-motivated learning communities like speaking clubs. This study aims to fill that gap by analyzing the sociolinguistic dynamics of CM within AESC. It seeks to move beyond mere quantification to understand the why behind the mix—exploring the factors that trigger CM and the speakers' awareness of their own linguistic choices. Therefore, this study is guided by the following research questions:

1. What percentage of CM do respondents use in an English conversation?
2. What are the factors causing each respondent to use CM?
3. Are the respondents aware that they engage in CM or CS?

2. Literature Review

2.1 Code-Mixing and Code-Switching in Sociolinguistics

Sociolinguistics, the study of the relationship between language and society, provides the theoretical foundation for understanding CM and CS. It posits that language use is not merely a grammatical exercise but a social act, deeply intertwined with identity, context, and group membership (Holmes, 1992). Code-mixing is a natural consequence of language contact in multilingual societies. Bhatia and Ritchie (2004) define it as the mixing of various linguistic units from two different grammatical systems within a single sentence. This differs from code-switching, which Aurer (1998) describes as a communicative system where speakers alternately switch from one language to another, often at sentence boundaries or in response to a change in the social situation.

The study of CM challenges the deficit perspective that views it as a sign of linguistic incompetence. Instead, scholars like Milroy and Muysken (1995) conceptualize language mixing as a "resource of linguistic wealthiness" through which various social meanings are produced. It can serve numerous functions, such as expressing solidarity, emphasizing a point, compensating for a lexical gap, or signaling a shared cultural identity (Pfaff, 1979; Jacobson, 1998).

2.2 Diglossia and the Indonesian Context

The Indonesian linguistic situation can be partially understood through the lens of diglossia, a concept where two varieties of a language exist side-by-side throughout a community, each serving a different social function (Alaiyed & Abdullah, 2018). In a broader sense, Bahasa Indonesia functions as the high (H) variety for formal, educational, and national purposes, while various regional languages (e.g., Javanese, Batak, Sundanese) and colloquial Indonesian dialects act as low (L) varieties for daily, informal communication. English enters this landscape as a global H variety, associated with education, modernity, and prestige. Bassiouney (2006) argues that CM and CS can be studied within the diglossic framework, as speakers navigate between these different varieties. The members of AESC operate in this complex diglossic environment, constantly negotiating between their L1 (Indonesian/regional languages) and the target H variety (English).

2.3 Factors Influencing Code-Mixing and Code-Switching

Janet Holmes (1992) provides a comprehensive framework of eight social factors that influence language choice and, by extension, CM and CS:

a. Regional: Geography and accent can influence language use, as moving to a new region often necessitates learning and mixing with the local language.
b. Ethnic: Ethnic pride and heritage can motivate speakers to use their ethnic language or accent, even in a predominantly different linguistic environment.
c. Social Status: Language can be a social marker. Speakers may use certain lexical items associated with a higher social class or switch to align with a desired social image.
d. Gender: Sociolinguistic studies often find differences in the linguistic behavior of women and men, including politeness strategies and speech functions, which can influence mixing patterns.
e. Age: Younger speakers are often more open to linguistic innovation and borrowing, while older speakers may adhere more closely to traditional language norms.
f. Education: Higher levels of education often correlate with proficiency in a standard or formal language variety, affecting how and when one mixes codes.
g. Occupation: Professional jargon and the need to communicate with specific audiences can be a significant trigger for CS and CM.
h. Cultural Attitude: Norms of politeness and cultural values can dictate language choice. In some cultures, mixing is a sign of identity, while in others, it may be frowned upon.

This study employs Holmes's framework to categorize and understand the motivations behind the CM observed in the AESC communit.

3. Research Method

3.1 Research Design

This study employed a qualitative research design using the Grounded Theory Method (GTM) as developed by Kathy Charmaz (2016). This approach was chosen for its disciplined, systematic, and iterative process for analyzing data to produce inductive findings. Rather than imposing a rigid pre-existing theory, GTM allows patterns and concepts to emerge organically from the data itself, which is ideal for exploring the nuanced and personal nature of language use. The research design relied on observation

techniques through semi-structured interviews and documentation (voice recordings) (Ciesielska, Bostrom, & Ohlander, 2018).

3.2 Population and Sample

The population for this study was the members of Aaron's English Speaking Club (AESC) in Medan. Following Arikunto's (2012) guidance that if a population is less than 100, the entire population can be taken, a sample of ten (10) members was selected using probability (random) sampling to ensure every member had an equal chance of being selected (Etikan & Bala, 2017). The sample consisted of five males and five females, aged between 20-25 years. To protect their privacy, respondents are referred to by their initials throughout this article.

3.3 Data Collection

Data was collected through one-on-one, audio-recorded interviews conducted in various relaxed settings at different times to suit the respondents' schedules. Each respondent was asked to choose one of four topics—Mental Health, Personal Development, Travelling, or Arts—and then answered five predetermined questions related to that topic. The interviews were conducted in English to elicit natural speech, but respondents were free to respond at length without time constraints. This method was designed to create a comfortable environment where CM could occur naturally. After the main interview, a short debriefing session was held where the concepts of CM and CS were explained, and respondents were asked about their awareness of these phenomena in their own speech.

3.4 Data Analysis

The recorded interviews were transcribed verbatim. The primary data analysis involved identifying and counting every instance where an Indonesian word or phrase was inserted into an English sentence (code-mixing). The analysis did not focus on grammatical errors but strictly on the conscious or unconscious insertion of L1 lexical items into L2 discourse.

The percentage of CM for each respondent was calculated using the formula by Sihaloho and Siregar (2020):

Percentage of CM = (Y / X) * 100

Where:

Y = Total number of Indonesian words (CM) found

X = Total number of words in the conversation

Following the quantitative analysis, a qualitative analysis was conducted. The transcripts were reviewed to identify the potential factors influencing each respondent's CM, using the framework of Holmes (1992). This involved considering the respondent's background, the content of their speech, and the context of the conversation. Finally, the responses from the awareness debrief were analyzed to gauge the participants' metalinguistic knowledge.

4. Discussion

4.1 Prevalence and Percentage of Code-Mixing

The analysis of the ten interviews revealed that code-mixing was a universal feature among the participants, though the extent varied significantly. The total words

spoken and the instances of CM for each respondent, along with the calculated percentage, are presented in Table 1.

No.	Respondent	Total Words (X)	CM Words (Y)	Percentage of CM
1.	SAP	544	76	13.97%
2.	SW	515	48	9.32%
3.	WR	412	90	21.84%
4.	NM	533	95	17.82%
5.	IAP	577	50	8.66%
6.	DH	399	18	4.51%
7.	FHZ	390	32	8.20%
8.	SA	190	1	0.52%
9.	ANP	423	103	24.34%
10.	SS	545	59	10.82%

Table 1. Percentage of Code-Mixing in Respondents' Speech

The results show a wide range, from a very low 0.52% (SA) to a high of 24.34% (ANP). It is crucial to interpret these numbers carefully. A low percentage does not necessarily indicate higher English proficiency, nor does a high percentage indicate a lack thereof. For instance, SA's very low CM rate (0.52%) was characterized by short, direct answers, while ANP's high rate (24.34%) reflected deeply personal and emotional narratives about mental health, where resorting to L1 might provide comfort and precision for expressing complex feelings. This suggests that the topic of conversation and the speaker's communicative goals are significant determinants of CM frequency.

4.2 Factors Influencing Code-Mixing

Applying Holmes's (1992) framework, the study identified a primary influencing factor for each respondent, as shown in Table 2. The analysis was based on the content of their conversations and their personal backgrounds as shared during the interviews.

No.	Respondent	Primary Influencing Factor
1.	SAP	Cultural Attitude
2.	SW	Regional

No.	Respondent	Primary Influencing Factor
3.	WR	Occupation
4.	NM	Occupation
5.	IAP	Regional
6.	DH	Cultural Attitude
7.	FHZ	Gender
8.	SA	Age
9.	ANP	Social Status
10.	SS	Education

Table 2. Primary Sociolinguistic Factor for Each Respondent's Code-Mixing

5. Conclusion

This study set out to investigate the patterns and awareness of code-mixing among Indonesian youths in the specific social context of an English speaking club. The findings clearly demonstrate that code-mixing is not a peripheral error but a central, strategic feature of their bilingual communication. The varying percentages of CM reflect individual differences in speaking style, topic emotionality, and communicative intent rather than a hierarchy of proficiency.

The application of Holmes's (1992) sociolinguistic factors proved effective in explaining the motivations behind CM, revealing that these young Indonesians mix codes due to a complex interplay of their occupational, regional, and cultural identities. Their language is a true reflection of their hybrid identity: globally oriented yet locally rooted.

Most significantly, the study highlights a gap in metalinguistic awareness. While these youths are adept practitioners of bilingual communication, they lack the formal vocabulary to describe their own skillful linguistic behavior. Recognizing CM as a natural and resourceful strategy, rather than a deficit, is crucial for building linguistic confidence.

In conclusion, communities like Aaron's English Speaking Club are not merely spaces for practicing a foreign language; they are dynamic contact zones where a new, hybrid language practice is being forged. This practice is systematic, meaningful, and deeply tied to the speakers' social identities.

References

Alaiyed, M., & Abdullah, S. (2018). *Diglossic code switching between standard Arabic and Najdi Arabic in religious discourse*. (Doctoral dissertation, Durham University, Durham, England).

Arikunto, S. (2012). *Prosedur penelitian suatu pendekatan praktik (Research procedures a practical approach)*. Jakarta, IDN: Rineka Cipta.

Astari, N. K. A. D., & Marantika, I. M. Y. (2023). Code mixing used in Cinta Laura YouTube Channel's "Bicara Cinta". *Wanastra: Jurnal Bahasa dan Sastra*, 15(1), 22-30.

Aurer, P. (1998). *Code-switching in conversation: Language, interaction, and identity*. London, UK: Routledge.

Bassiouney, R. (2006). *Functions of code switching in Egypt*. Leiden, NL: Brill.

Bhatia, T. K., & Ritchie, W. C. (2004). Social and psychological factors in language mixing. In T. K. Bhatia & W. C. Ritchie (Eds.), The handbook of bilingualism (pp. 336-352). New Jersey, USA: Blackwell Publishing.

Charmaz, K. (2016). *Constructing grounded theory (2nd ed.)*. Los Angeles, USA: Sage.

Ciesielska, M., Bostrom, K. W., & Ohlander, M. (2018). Observation methods. In *Qualitative methodologies in organization studies* (pp. 33-52). London, UK: Palgrave Macmillan.

Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, 5(6), 00149.

Holmes, J. (1992). *An introduction to Sociolinguistics*. London, UK: Longman.

Jacobson, R. (1998). *Code switching worldwide*. Berlin, DE: Mouton de Gruyter.

Kasanah, M. U., & Hidayati, F. N. (2023). *An analysis of code switching in Aaron's English YouTube channel* (Doctoral dissertation, UIN Surakarta, Surakarta, Indonesia).

Kim, E. (2006). Reasons and motivations for code mixing and code switching. *Issues in EFL*, 4(1), 43-61.

Milroy, L., & Muysken, P. (Eds.). (1995). *One speaker, two languages: Cross-disciplinary perspectives on code-switching.* Cambridge, UK: Cambridge University Press.

Pfaff, C. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, 55(2), 291–318.

Sihaloho, R. D., & Siregar, H. (2020). Pengaruh lingkungan kerja terhadap kinerja karyawan pada PT. Super Setia Sagita Medan (The influence of the work environment on employee performance at PT. Super Setia Sagita Medan). *Jurnal Ilmiah Socio Secretum*, 9(2), 275-285.

Thara, S., & Poornachandran, P. (2018). Code mixing: A brief survey. *International Conference on Advances in Computing, Communications, and Informatics (ICACCI)*, 2382-2385.