



Journal of Advanced Research in Social and Behavioural Sciences

Journal homepage:
<https://karyailham.com.my/index.php/jarsbs/index>
ISSN: 2462-1951



Input, Interaction, Feedback, and Classroom Management in EFL Primary English Teaching Behaviors: A Quantitative Classroom Observation Study in Mainland China

Feirong Wang^{1,*}

¹ Institute of International Education, New Era University College, Kajang, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 27 March 2026

Received in revised form 21 May 2026

Accepted 10 June 2026

Available online 14 July 2026

Keywords:

EFL teaching behaviors; classroom observation; primary English education; corrective feedback; teacher questioning

ABSTRACT

For the vast majority of mainland Chinese primary school students, the 40-minute English class is effectively their sole point of access to the target language, thus teachers' observable classroom behavior determines whether there will be any opportunity for language acquisition at all. Current research related to Chinese EFL classrooms is heavily reliant on self-reported information collected from teachers and lacks quantitative findings based on objective classroom observations on large samples. The current study adopted a non-participant structured observation method to analyze teaching behaviors along four axes: language input, teacher questioning, corrective feedback, and classroom management. 54 lessons delivered by 27 primary school teachers working in urban and county-level schools in Hubei Province, China, were coded according to a specially designed instrument (PETBOS). Findings indicated that teachers' speech contained 51% of L1 (Chinese), display questions represented 83% of questions in total with a mean wait time of 1.6 seconds, recasts were the most frequently used (49%) type of feedback but only yielded a 14% repair rate, and instructional time constituted 64% of class time on average. Urban and county teachers demonstrated dichotomy in terms of significant differences between proportions of L2 input usage, referential questions, instructional time, and managerial time ($p = .050$) while exhibiting similar wait times, types of feedback, transitional periods, and off-task periods. These findings offer concrete, quantifiable measures for targeting teacher development and indicate that mere redistribution of resources will not affect entrenched patterns of teachers' habitual classroom behavior, and alternative strategies such as video-stimulated reflective practice must be employed.

1. Introduction

Since 2001, the English as a foreign language (EFL) course has become an obligatory course in China from grade 3 in primary schools. In 2011, the *English Curriculum Standards for Compulsory*

* Corresponding author.

E-mail address: wangfeirong0621@e.newera.edu.my

<https://doi.org/10.37934/arsbs.45.1.3041>

Education" designated "comprehensive language proficiency" as the general objective, while in 2022, the new version proposed a competency-based approach that focuses on language competence, cultural awareness, thinking quality, and learning ability [18]. The evolving policy direction in this field has put the quality of EFL class teaching in primary schools in the spotlight.

In the Chinese EFL context, the 40-minute English class almost represents the only input to students from the target language [3,8]. The language a teacher selects, the questions posed, the feedback provided, and the allocation of class time all constitute observable teaching behaviors that directly shape students' language learning opportunities. All these observable acts of teachers directly influence the language learning environment. Current studies that focus on teaching practices in Chinese EFL contexts are limited along two aspects. Regarding data, the majority of the studies are based on questionnaire and reflective journal data obtained from teachers about their own perception of teaching. However, the gap between actual and perceived teaching practice has been identified consistently through literature [5,23]. As for the educational level, the existing empirical studies on observing ESL teaching in a structured way using COLT, FLINT, and TBOP tools have mostly focused on secondary or higher education contexts. Observation data regarding the primary level in China is extremely rare [26]. The most recent study conducted on primary English teachers in China was still based on survey data, which makes the issue even more prominent [21]. Although clear differences in teacher training, qualifications, and class size between urban and county schools have been well documented [8,21], whether such disparities translate into observable teaching behavior differences has not been quantitatively examined.

Systematic observation is the main method used in the current study to identify and differentiate teaching actions in four aspects in EFL classrooms in China proper. The following research questions drive this study.

RQ1. What is the distribution of teacher language input (L1/L2 ratio, modification strategies), and does it differ between urban and county-level settings?

RQ2. What types and frequencies of teacher questioning are observed, what is the average wait time, and do these differ between settings?

RQ3. What are the distributions of corrective feedback types and learner uptake, and do these differ between settings?

RQ4. How is classroom time allocated, and does this differ between settings?

On an empirical basis, this paper is part of a few big-sample studies which have used a four-dimension framework for observing the behaviors of teachers in Chinese primary schools teaching English as a foreign language. On a comparative basis, the urban-rural comparison highlights those teaching behaviors that differ based on dimensionality and those behaviors which do not. From a practical point of view, the behavioral profiles can be used as quantifiable goals for training programs in order to move away from theory and into practice [4,21].

2. Literature Review

2.1 Language Input and Classroom Interaction

Input in the form of language is a key prerequisite for second language acquisition. The Input Hypothesis states that learners acquire a second language when they receive comprehensible input that slightly exceeds their current competence level ($i+1$) [11]. In English as a Foreign Language learning environments, teacher talk represents the primary controllable source of language input for learners, making teachers' language choices particularly important [7]. There have been two opposing views regarding this issue; the first advocates maximizing L2 use to ensure sufficient language exposure [22], while the second highlights the potential benefits of L1 use and proposes a

flexible “optimal position” approach [6]. It has been shown that, in EFL classrooms, teachers’ L1 use mainly serves functions such as concept explanation, classroom management, and cultural comparison [24]. However, in Chinese EFL contexts, L1 use has also been associated with teachers’ perceived language limitations and examination-oriented teaching practices [8,9].

The quantity and quality of input directly influence interactional conditions. According to the Interaction Hypothesis, negotiation of meaning promotes learners’ attention to linguistic forms, thereby facilitating second language acquisition [14]. Classroom discourse has commonly been characterized through the IRF (Initiation, Response, Feedback) sequence [20]. Within this structure, the types of questions used by teachers play an important role. Display questions require learners to recall known information, whereas referential questions encourage the expression of opinions or unknown information. Referential questions have been found to generate longer and more meaningful learner responses [2]; however, display questions continue to dominate EFL classroom interactions [15,25]. Another important variable is wait time, which refers to the interval between a teacher’s question and the subsequent response or utterance. Although studies recommend three to five seconds of wait time, the average duration in Chinese EFL classrooms is often less than two seconds [19]. Consequently, insufficient L2 input may restrict opportunities for meaningful interaction, causing teachers to rely on display questions as a safer classroom strategy and creating an input–interaction negative loop.

2.2 Corrective Feedback and Time Allocation

Classroom interaction provides opportunities for CF as no corrective feedback can occur without students generating erroneous output. One of the most widely cited taxonomies distinguishes between explicit correction, recast, clarification request, metalinguistic feedback, elicitation, and repetition [16]. According to the results of the investigation, recast represented 55% of feedback types; however, around 70% of them were unremarked upon, indicating the lowest rate of repair. Subsequently, the meta-analyses found prompts (metalinguistic feedback, elicitation, and clarification request) to have significantly higher effect sizes compared to recasts [17]. This trend can be observed in other EFL environments, with the percentage of recasts being over 50% in multiple EFL settings [13] and 69.14% in Chinese EFL secondary schools [27], where the proportion of unremarked-upon recasts was very high. Currently, research on CF in EFL settings lacks data gathered at primary school level. It appears that young learners with lower metalinguistic development may process recasts differently from adult participants.

In fact, all these dimensions function in the context of the time limitation per each lesson. The ALT framework identifies the portion of time during which students actively engage with appropriate content as a key predictor of learning outcomes [1]. There are different types of time in the classroom: instructional time, management time, transitional time, and off-task time. Out of them, only instructional time helps in acquiring language skills. Considering that each lesson in Chinese primary school is 40 minutes long, it limits instructional time significantly. In addition, young children's short attention span, higher levels of management, and many changes from one activity to another increase management and transitional time.

2.3 Conceptual Framework

From the foregoing review, this study adopts the following Four-Dimension Teaching Behavior Observation Framework. This framework integrates the four dimensions into a progressive order. Input offers the target language materials as the basic condition; Interaction triggers and negotiates

the input through questioning from the teacher; and Feedback corrects any errors made by learners during their output, focusing on form. These make up the basic instructional sequence. Classroom Management takes on two different roles. It acts as the prerequisite condition without which no instructions can take place, as well as a competing dimension where the more it claims in terms of time, the less time left for input–interaction–feedback. The study’s analytical strategy flows from here. Distribution of behavior is observed in all four dimensions, followed by the comparison between urban and county settings on what factors affect which dimensions. Figure 1 outlines this framework.

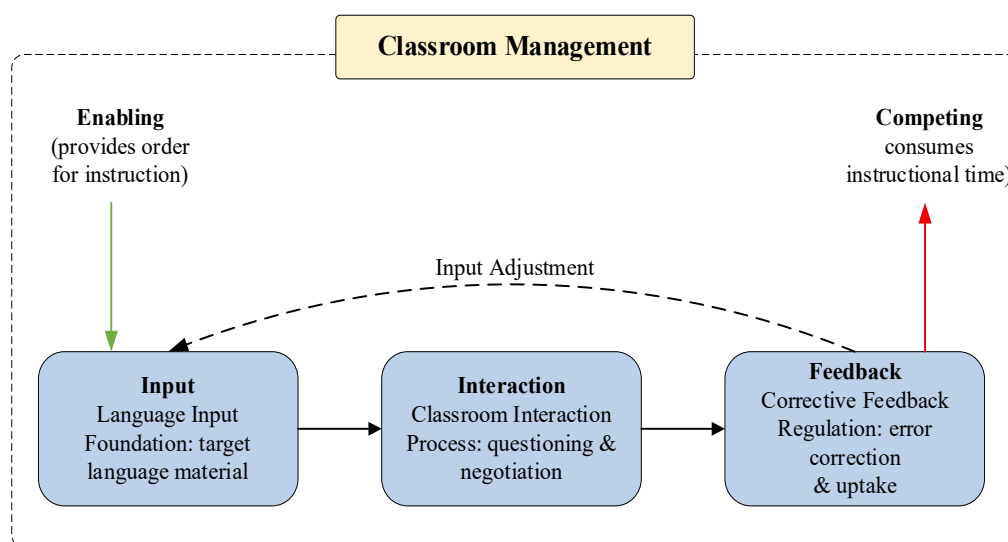


Fig. 1. Four-Dimension Teaching Behavior Observation Framework

Note: Solid arrows indicate sequential progression, dashed arc indicates feedback loop, green/red arrows indicate the dual role of management.

3 Methodology

3.1 Research Design

This research used a quantitative, non-participant observation approach with structured observations in the classroom setting that used a cross-sectional descriptive and comparative method. The purpose was to identify the distribution of teacher behavior within four domains in Chinese primary English language classrooms and compare this distribution between urban and county schools.

3.2 Participants

The study was carried out in Hubei Province, a central Chinese province whose mid-range economic development and mix of urban and county-level schools offer reasonable national representativeness. Sampling was stratified based on purposive sampling where the participants were grouped into urban and county groups. The urban group comprised 15 teachers teaching English while the county had 12 such teachers, making a total of 27 teachers. Every teacher was observed for 2 periods of equal grade level classes that yielded 54 lessons of 40 minutes each, covering grades 3 to 6. All participants were female, consistent with the documented gender composition exceeding 95% female among Chinese primary English teachers [21]. Table 1 shows the

demographics of the participants. The average period of service for urban teachers was high at $M=11.3$ ($SD=5.1$) compared to $M=7.9$ ($SD=5.2$) for county teachers.

Table 1
Participant Demographics

Characteristic	Urban (n = 15)	County (n = 12)	Total (N = 27)
Gender (Female)	15 (100%)	12 (100%)	27 (100%)
Experience (yrs), M (SD)	11.3 (5.1)	7.9 (5.2)	9.8 (5.4)
Bachelor's degree	13 (86.7%)	7 (58.3%)	20 (74.1%)
Associate degree	2 (13.3%)	5 (41.7%)	7 (25.9%)
Grades 3–4 teachers (lessons)	8 (16)	7 (14)	15 (30)
Grades 5–6 teachers (lessons)	7 (14)	5 (10)	12 (24)

3.3 Instrument

The Primary EFL Teaching Behavior Observation Scheme (PETBOS) was created through the incorporation of observation systems into four modules. In Module A (Language Input), the teacher's use of language choice (L1, L2, or Mixed) and input modification strategies (simplification, repetition, elaboration, visual-gestural cueing, or L1 translation) is coded using COLT Part A [23]. In Module B (Classroom Interaction), each utterance is classified into display, referential, and procedural questions, with an estimated wait time in seconds according to Long and Sato [15]. In Module C (Corrective Feedback), each corrective move is coded in relation to types of feedback (explicit correction, recasts, clarification requests, metalinguistic feedback, elicitation, and repetition [16]), as well as the learner's uptake (repair, needs repair, or no uptake). Module D (Classroom Management) uses interval sampling (30 seconds), coding intervals as instructional, managerial, transitional, or off-task times, according to the Academic Learning Time system [1].

The PETBOS protocol underwent piloting in six lessons (three urban and three county), but none of these six lessons made it to the final sample. Piloting informed a few revisions. Specifically, the category "Mixed" was introduced, while "Procedural" was redefined as directives that relate to classroom organization. The coding categories were evaluated by three EFL teacher educators.

3.4 Data Collection and Reliability

The data collection process occurred during September to November 2023. All 54 lessons were audio-video recorded by utilizing two different camera views, with one being specific to teacher-related actions while another included classroom-wide activities. Two independent coders analyzed the entire material after going through 12 hours of training involving three sessions based on pilot-phase videotapes. Both coders reached the predetermined agreement criterion on six validation videos prior to the actual coding procedure.

The inter-rater reliability measure was performed by utilizing Cohen's κ coefficient for categorical data and the intraclass correlation coefficient for continuous measures. The values for Cohen's κ

coefficient were obtained for language use .91, modification technique .84, questioning technique .87, feedback technique .85, learner reaction .82, and time management .89. For wait time, the corresponding value for the two-way mixed ICC was .93. All of the calculated values indicated the almost perfect agreement level [12].

3.5 Data Analysis

Frequencies, percentages, mean values, and standard deviations were calculated for each scale. The independent-sample t-test was conducted to compare urban vs. county data on continuous variables with normality checked via Shapiro-Wilk tests; the Mann-Whitney U test replaced the t-test in case of the violation of normality. The χ^2 test was employed for distribution analysis of categorical variables. Effect sizes (Cohen's d for the t-test and Cramér's V for the χ^2 test) were estimated for all inferential tests conducted.

4. Findings

4.1 Language Input (RQ1)

Table 2 shows the proportion of utterance language use and the frequency of use of input modifications in the 54 lessons observed. In the total 8,463 utterances that were analyzed, there was a mean of about 157 utterances per lesson; with regard to language use in utterances, 50.9% were in L1 (Chinese), 37.8% in L2 (English), and 11.3% were mixed-language utterances. The L2 proportion was significantly higher among urban teachers (42.8%) than among county-level teachers (31.5%), $t(52) = 3.14$, $p = .003$, $d = 0.87$, large effect size.

In the strategies involving changes to inputs, there were a total of 2,475 occurrences. The most common of these were L1 translation, which occurred 34.2% of the time, followed by repetition, with 28.0%, and visual-gestural clues, with 18.2%. Simplification made up 13.2% of strategies, whereas elaboration made up the smallest percentage, of 6.4%. It is clear from this that teachers tended to utilize L1 input change instead of TL strategies for comprehension.

Table 2

Teacher Language Input

Dimension	Urban	County	Total	Comparison
Language choice				
L2 (English)	42.8%	31.5%	37.8%	$t(52) = 3.14$, $p = .003$, $d = 0.87$
L1 (Chinese)	44.6%	58.7%	50.9%	
Mixed	12.6%	9.8%	11.3%	
Top 3 modification strategies				
L1 translation	—	—	34.2%	—
Repetition	—	—	28.0%	
Visual/gestural	—	—	18.2%	

Note: Based on 8,463 utterances across 54 lessons. Modification strategy total N = 2,475.

4.2 Teacher Questioning (RQ2)

Table 3 provides the classification of the types of questions posed by teachers together with the wait time. A total of 1,339 questions were observed among which the average was 24.8 per class

period. Display questions dominated with 82.9%, whereas referential questions accounted for 10.2% and procedural questions amounted to 6.9%. The distribution of question types varied significantly between the urban and county samples, $\chi^2(2) = 13.67$, $p = .001$, Cramér's $V = .10$. Teachers from urban areas used more referential questions approximately twice as often as those from the county area (12.7% versus 6.7%) but the latter utilized more display questions (87.7% versus 79.4%). Overall, county classes featured an average of only 1.6 referential questions per class period.

With respect to wait time, the mean wait time after asking a display question was 1.4 seconds ($SD = 0.8$) while that for a referential question was 2.9 seconds ($SD = 1.6$), $t(1244) = 9.82$, $p < .001$, $d = 1.19$, indicating differential engagement in terms of cognitive processing by the two question types. Overall, mean wait time for all questions asked in the study was 1.6 seconds ($SD = 1.1$) – a significantly shorter duration compared to the ideal wait time of three to five seconds (Rowe, 1986). Notably, the urban–county difference in overall wait time did not reach significance, $t(1337) = 1.52$, $p = .129$, $d = 0.08$.

Table 3
Teacher Questioning

Dimension	Urban	County	Total	Comparison
Question type				
Display	612 (79.4%)	498 (87.7%)	1,110 (82.9%)	$\chi^2(2) = 13.67, p = .001, V = .10$
Referential	98 (12.7%)	38 (6.7%)	136 (10.2%)	
Procedural	61 (7.9%)	32 (5.6%)	93 (6.9%)	
Wait time (seconds)				
Display, M(SD)	—	—	1.4 (0.8)	Ref. vs. Disp.: $t(1244) = 9.82, p < .001, d = 1.19$
Referential, M(SD)	—	—	2.9 (1.6)	
Overall, M(SD)	—	—	1.6 (1.1)	Urban–County: $t(1337) = 1.52, p = .129 (n.s.)$

4.3 Corrective Feedback and Learner Uptake (RQ3)

A total of 444 feedback occasions occurred, an average of 8.2 per lesson, which is not a very high number due to the lack of oral production in teacher-dominated primary classes. These six types of feedback along with their corresponding repair rates and percentages are shown in Table 4. Recasts were used most often (49.1%, 218 times), but only had a 14.2% repair rate, of which 75.7% did not result in any uptake. On the other hand, explicit correction, elicitation, and metalinguistic feedback had repair rates of 66.7%, 61.9%, and 55.0%, respectively, even though they were rarely used. Clarifications and repetitions had a repair rate of 45.2% and 27.6%, respectively.

The city-county difference regarding the distribution of feedback types was not significant, $\chi^2(5) = 8.93$, $p = .112$, Cramér's $V = .14$. Teachers in both locations demonstrated the same distribution characterized by recasts being predominant. The fact testifies to stability of such behavior patterns regardless of the school environment. Figure 2 presents a visual demonstration of such distribution in form of horizontally divided bars with a predominance of “no uptake” in the recast bar.

Table 4
Corrective Feedback and Learner Uptake

CF Type	n (%)	Repair	Needs-repair	No Uptake
Recast	218 (49.1%)	31 (14.2%)	22 (10.1%)	165 (75.7%)
Explicit correction	72 (16.2%)	48 (66.7%)	9 (12.5%)	15 (20.8%)
Elicitation	63 (14.2%)	39 (61.9%)	14 (22.2%)	10 (15.9%)
Clarification request	42 (9.5%)	19 (45.2%)	12 (28.6%)	11 (26.2%)
Repetition	29 (6.5%)	8 (27.6%)	9 (31.0%)	12 (41.4%)
Metalinguistic	20 (4.5%)	11 (55.0%)	5 (25.0%)	4 (20.0%)
Total	444 (100%)	156 (35.1%)	71 (16.0%)	217 (48.9%)

Note: Urban–county CF distribution: $\chi^2(5) = 8.93$, $p = .112$, $V = .14$ (n.s.).

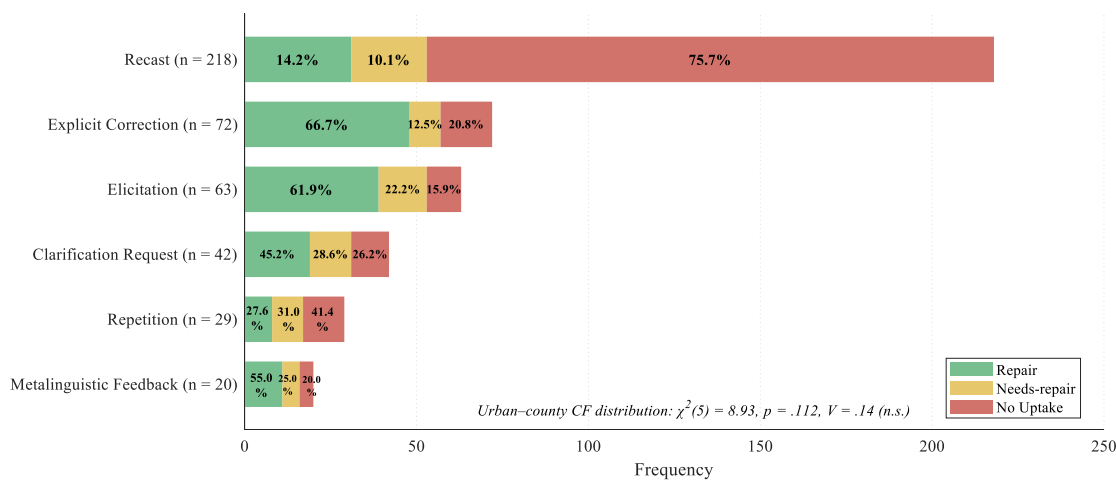


Fig. 2. Distribution of Corrective Feedback Types with Learner Uptake Breakdown

4.4 Time Allocation (RQ4)

Table 5 shows classroom time allocation statistics. Instructional time constituted, on average, 63.6% (SD = 9.1) of all the time, which implies that about 25.4 minutes out of total 40 minutes were spent on teaching activities. The rest of 36.4% was distributed as follows: 17.1% went to management activities, 13.4% to transitional periods, and 5.9% to off-task time. The amount of time that urban classrooms spent on instructions was significantly higher compared to that observed for county classrooms: 66.1% versus 60.5%, $t(52) = 2.05$, $p = .045$, $d = 0.62$. Conversely, county classrooms spent slightly more time on management (19.2% versus 15.4%), $t(52) = 2.01$, $p = .050$, $d = 0.63$. Differences between transitional and off-task time percentages were not statistically significant ($p = .357$ and $p = .662$).

Table 5
Classroom Time Allocation and Urban–County Comparison

Category	Urban M(SD)	County M(SD)	Total M(SD)	t(52)	p	d
Instructional	66.1 (8.2)	60.5 (9.7)	63.6 (9.1)	2.05	.045	0.62
Managerial	15.4 (5.1)	19.2 (6.8)	17.1 (6.1)	2.01	.050	0.63
Transitional	12.8 (4.3)	14.1 (4.9)	13.4 (4.5)	0.93	.357	0.28
Off-task	5.7 (3.4)	6.2 (3.9)	5.9 (3.6)	0.44	.662	0.14

4.5 Summary of Urban–County Comparisons

Combined, the eight urban vs. county comparisons revealed an obvious dichotomy of findings. While three were significantly different (the L2 utterance proportion, the type of questions, and the instructional time), another one was borderline significant (managing time, $p = .050$), and the rest four showed no difference (total waiting time, the type of CFs, transitional time, and off-task time). All the significant findings were found for the factors related to the English proficiency of teachers, availability of professional training programs, and other objective factors, such as the size of the class. The remaining findings were related to deeply ingrained behavioral habits developed during preservice training and practice.

5. Discussion

5.1 The Input–Interaction Chain

The result that L1 occurred in 50.9% of teacher utterances, whereas L2 was just 37.8% is much lower than the minimum threshold set for L2 use by the literature [22]. The high frequency of L1 translation as the dominant input modification strategy (34.2%), followed by elaboration at 6.4% demonstrates the existence of an automatic response tendency. In responding to student failure to comprehend, teachers would resort directly to the native language instead of continuing exposure to L2 by using target language internal strategies, such as simplification or paraphrasing. This is best described as a compensatory fallback response due to poor English competence, rather than being a form of "strategic L1 use" described by Ha [6]. The significantly lower L2 proportion in county schools (31.5% versus 42.8%) provides additional evidence of the impact of teacher English proficiency on language selection during instruction [8]. It must be noted here that the dependence on L1 is a self-perpetuating one, where extensive L1 use denies students the opportunity to develop their L2 processing abilities, which, in turn, encourages teachers to believe that "students are unable to understand English."

The negative effects of inadequate L2 input spill over into the interaction aspect, because of lack of input, the space for referential questioning is restricted; questions posed in Chinese cannot serve as a basis for meaning negotiation of target language. Indeed, the data revealed that display questions were 82.9% of teacher questions, and referential questions were 10.2%; urban teachers asked referential questions approximately twice as many as county teachers. Combined with the very short average wait time, which is well below the ideal three to five seconds [19], classroom interaction became confined to surface operation of IRF sequence, such as teacher questioning, student response and confirmation from teacher, and lacking real meaning negotiation. It is

noteworthy that there was no urban/county difference in wait time. It implies that fast pace was not the result of limited resources but rather systemic rhythm routine in teaching. As for practical interventions, it should be noted that increasing input quantity and promoting referential questioning share the same goal. The former offers linguistic content, while the latter stimulates cognitive process on that content.

5.2 The Feedback–Time Nexus

Recasts represented 49.1% of all corrective feedback, but were associated with the lowest repair rate (14.2%), with 75.7% yielding no learner uptake. Again, the results follow the typical findings by Lyster and Ranta [16] and confirm the latest observations of secondary-level Chinese classrooms (69.14%), which yielded very low learner uptake [27]. Given the developmental characteristics of young learners with underdeveloped metalinguistic awareness, it becomes even more challenging for them to identify recasts from natural repetitions, reducing even further the likelihood of their effectiveness. Explicit correction (66.7%) and elicitation (61.9%) provided the highest repair rates, further confirming the meta-analytic conclusions on the superiority of prompting strategies in relation to recasts [17]. The lack of an urban-county difference in the frequency distribution of types of corrective feedback may be seen as evidence that teachers' behavior in terms of correction is deeply entrenched as a result of preservice education, remaining unchanged despite different conditions at the school level.

The low efficiency of the predominant feedback strategies exacerbates the issue of insufficient class time. Average instructional time made up merely 63.6% of total class time, amounting to 25.4 minutes per lesson, while 36.4% was spent on managerial and transitional activities. County classrooms used significantly more time for management (19.2% vs. 15.4%), possibly owing to higher class numbers and less developed classroom management routines. Comparable levels of off-task behavior (5.7% vs. 6.2%) imply that wasted time was due to teacher management and transition activities rather than student disengagement. With this tight schedule, the excessive number of inefficient recasts further constrains the operational space for the input–interaction–feedback process. Improving feedback efficiency through increased use of elicitation and explicit correction, and improving time efficiency through development of standard classroom routines to reduce management time are therefore two complementary intervention avenues.

5.3 Systemic Habits vs. Resource Gaps

In the categorization of the eight urban-county contrasts according to significance, there appears to be an obvious pattern emerging. It is observed that all the four differences which were statistically significant (proportion of L2 inputs, proportion of referential questions, instructional time, and managerial time) relate directly to aspects that depend upon teacher proficiency in English language, professional development, and school characteristics including class size. However, all four non-significant differences (wait time, preference for feedback types, transitional time, and off task time) represent deeply entrenched behavioral trends due to prior education and teaching culture.

The theoretical importance of this distinction is that it illustrates that there are no "all or nothing" issues when it comes to teaching behaviors. Various behavioral dimensions have different drivers and therefore call for different kinds of solutions. Structural interventions, such as increasing teachers' competence in English and providing more resources for professional development and class size reductions, may help with resource capacity sensitive dimensions like provision of L2 input and asking questions for meaning checking. With habit-stable dimensions, such as wait time and

preference for specific feedback strategies, simply providing the same amount of resources will be ineffective since these behavior patterns are deeply ingrained into the daily classroom practice. Such behavioral dimensions must be tackled by training techniques that focus specifically on altering teachers' behavior through video stimulated reflection [10], microteaching and feedback strategy training. This result echoes prior descriptions of how "structural constraints" and "teacher behavioral inertia" coexist in Chinese EFL education context [8].

6. Conclusion

The analysis of 54 primary EFL classes conducted by 27 teachers at urban and county-level schools in Hubei Province, China, indicated four dimensions of behavior: L1 domination (51% of teacher talk), display question lock-in (83%), recast-based error feedback with low repair (49% frequency, 14% repair rate), and substantial time erosion (only 64% of class time spent on teaching). This dimension pattern constitutes a sequential chain whereby inadequate L2 exposure decreases meaningful teacher-student interaction, display questions impede student talk, ineffective error feedback does not direct students to correct their errors, and extra time spent on management diminishes the teaching/learning experience. In comparing urban and county schools, two contrasting patterns emerged: resource-capacity related behaviors (L2 input, referential questions, instructional time, and managerial time) were significantly different between urban and county contexts, while habit-based behaviors (wait time, error feedback type, transitional time, and off-task time) were similar across both contexts. These results indicate four behavioral foci for teacher training, including increased L2 use, more referential questioning, wait time of no less than 3 seconds, and recasts supplemented with elicitation and explicit error correction. Because habit-stable dimensions resist change via only resource equalization, video-based reflective practice and micro-teaching offer potential routes of intervention; with respect to schools at the county level, setting up routine classroom practices that will reduce management time is crucial. Some of the limitations of the study include: the limitation of a single province in sample collection, the influence of the Hawthorne effect on L2 levels, the failure to consider teaching experience as a covariate, no application of multilevel modeling to handle the hierarchical nature of data, and finally, the lack of student outcome measures.

Acknowledgement

The authors would like to express their highest appreciation to New Era University College for the institutional support and facilities provided throughout this study.

References

- [1] Berliner, David C. "What's All the Fuss about Instructional Time?" In *The Nature of Time in Schools*, edited by Miriam Ben-Peretz and Rainer Bromme, 3–35. New York: Teachers College Press, 1990.
- [2] Brock, Cynthia A. "The Effects of Referential Questions on ESL Classroom Discourse." *TESOL Quarterly* 20, no. 1 (1986): 47–59. <https://doi.org/10.2307/3586388>.
- [3] Butler, Yuko Goto. "English Language Education among Young Learners in East Asia: A Review of Current Research (2004–2014)." *Language Teaching* 48, no. 3 (2015): 303–342. <https://doi.org/10.1017/S0261444815000105>.
- [4] Guna, S. "Corrective Feedback: Pragmatic Exposures in EFL Classroom Interactions." *International Journal of Education and Learning* 5, no. 1 (2023). <https://doi.org/10.31763/ijele.v5i1.820>.
- [5] Gu, Meng. "A Myth in Language Teacher Learning: Lesson Observation." *Frontiers in Psychology* 13 (2022): 1026833. <https://doi.org/10.3389/fpsyg.2022.1026833>.
- [6] Ha, Xuan V. "High School EFL Teachers' Oral Corrective Feedback Beliefs and Practices, and the Effects of Lesson Focus." *International Review of Applied Linguistics in Language Teaching* 61, no. 4 (2023): 1799–1826. <https://doi.org/10.1515/iral-2021-0213>.
- [7] Ha, Xuan V., and James C. Murray. "Corrective Feedback: Beliefs and Practices of Vietnamese Primary EFL Teachers." *Language Teaching Research* 27, no. 1 (2023): 137–167. <https://doi.org/10.1177/1362168820931897>.

- [8] Hu, Guangwei. "English Language Education in China: Policies, Progress, and Problems." *Language Policy* 4, no. 1 (2005): 5–24. <https://doi.org/10.1007/s10993-004-6561-7>.
- [9] Hu, Yiming, Muhammad Afzaal, and Hanan Alfadda. "The Perceptions of International Learners toward Teacher Code-Switching in the Elementary and Intermediate Chinese Foreign Language Classrooms." *Frontiers in Psychology* 13 (2022): 860567. <https://doi.org/10.3389/fpsyg.2022.860567>.
- [10] Hußner, Isabel, Rebecca Lazarides, William Symes, Ulrich Trautwein, and Jürgen Möller. "Reflect on Your Teaching Experience: Systematic Reflection of Teaching Behaviour and Changes in Student Teachers' Self-Efficacy for Reflection." *Zeitschrift für Erziehungswissenschaft* 26 (2023): 1301–1320. <https://doi.org/10.1007/s11618-023-01190-8>.
- [11] Krashen, Stephen D. *The Input Hypothesis: Issues and Implications*. London: Longman, 1985.
- [12] Landis, J. Richard, and Gary G. Koch. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33, no. 1 (1977): 159–174. <https://doi.org/10.2307/2529310>.
- [13] Le, Thi Thanh, Le Quang Nguyen, and Nguyen Hoang Son Phan. "EFL Students' and Teachers' Beliefs about Oral Corrective Feedback and the Role of Students' Proficiency Levels." *SAGE Open* 15, no. 4 (2025). <https://doi.org/10.1177/21582440251392806>.
- [14] Long, Michael H. "The Role of the Linguistic Environment in Second Language Acquisition." In *Handbook of Second Language Acquisition*, edited by William C. Ritchie and Tej K. Bhatia, 413–468. San Diego: Academic Press, 1996.
- [15] Long, Michael H., and Catherine J. Sato. "Classroom Foreigner Talk Discourse: Forms and Functions of Teachers' Questions." In *Classroom Oriented Research in Second Language Acquisition*, edited by Herbert W. Seliger and Michael H. Long, 268–285. Rowley, MA: Newbury House, 1983.
- [16] Lyster, Roy, and Leila Ranta. "Corrective Feedback and Learner Uptake: Negotiation of Form in Communicative Classrooms." *Studies in Second Language Acquisition* 19, no. 1 (1997): 37–66. <https://doi.org/10.1017/S0272263197001034>.
- [17] Lyster, Roy, and Kazuya Saito. "Oral Feedback in Classroom SLA: A Meta-Analysis." *Studies in Second Language Acquisition* 32, no. 2 (2010): 265–302. <https://doi.org/10.1017/S0272263109990520>.
- [18] Ministry of Education of the People's Republic of China. *English Curriculum Standards for Compulsory Education (2022 Edition)*. Beijing: Beijing Normal University Press, 2022.
- [19] Rowe, Mary Budd. "Wait Time: Slowing Down May Be a Way of Speeding Up!" *Journal of Teacher Education* 37, no. 1 (1986): 43–50. <https://doi.org/10.1177/002248718603700110>.
- [20] Sinclair, John McHardy, and Malcolm Coulthard. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: Oxford University Press, 1975.
- [21] Song, Jie, Jyoti Mullick, and Tian Jiang. "English Language Education in Chinese Primary Schools: Exploring EFL Teachers' Attitudes, Self-Efficacy, and Perceived School Environment." *Asian-Pacific Journal of Second and Foreign Language Education* 10 (2025): 18. <https://doi.org/10.1186/s40862-025-00320-5>.
- [22] Soruc, Adem, Derya Yuksel, Jim McKinley, and Thomas Grimshaw. "Factors Influencing EFL Teachers' Provision of Oral Corrective Feedback: The Role of Teaching Experience." *The Language Learning Journal* 53, no. 1 (2025): 98–113. <https://doi.org/10.1080/09571736.2024.2338346>.
- [23] Spada, Nina, and Maria Fröhlich. *COLT Observation Scheme: Communicative Orientation of Language Teaching Observation Scheme—Coding Conventions and Applications*. Sydney: National Centre for English Language Teaching and Research, 1995.
- [24] Temesgen, Alemu, and Endalew Hailu. "Teachers' Codeswitching in EFL Classrooms: Functions and Motivations." *Cogent Education* 9, no. 1 (2022): 2124039. <https://doi.org/10.1080/2331186X.2022.2124039>.
- [25] Tong, Zhi, Fang An, and Yuhui Cui. "Exploring Teacher Discourse Patterns: Comparative Insights from Novice and Expert Teachers in Junior High School EFL Contexts." *Heliyon* 10, no. 16 (2024): e36435. <https://doi.org/10.1016/j.heliyon.2024.e36435>.
- [26] Wang, Hui, Ting Xu, and Min Zhang. "A Mixed Method Study on Chinese Primary School EFL Teachers' Preparation, Affecting Factors and Support Needed to Implement Intercultural Foreign Language Teaching." *PLOS ONE* 18, no. 4 (2023): e0284146. <https://doi.org/10.1371/journal.pone.0284146>.
- [27] Zhang, Chen, Nicholas E. J. A. Bowen, and Nicola Thomas. "Oral Corrective Feedback and Learner Uptake: An Analysis of Chinese High-School Teachers' Practices in a National Teaching Competition." *Classroom Discourse* (2025). <https://doi.org/10.1080/19463014.2025.2474233>.