



## English tag questions are not confusing, are they? : L1 transfer on L2 among Chinese ESL speakers

Meixing Yi & Christiana Christodoulou\*

**Abstract.** Tag questions frequently occur in both everyday conversations and educational settings (Wei et al. 2023). Due to cross-linguistic differences, Chinese ESL learners often produce confusing responses to English tag questions under the influence of their first language (L1), even after living or studying in English-speaking environments for a lengthy period. Results from 66 participants show that extended time in the U.S. assists Chinese ESL learners in significantly reducing L1 transfer when answering English tag questions. However, analysis of the participants' metalinguistic (self-)awareness reveals that the duration of stay does not markedly diminish the cognitive influence of L1 on L2.

**Keywords.** English tag questions; length of stay; L1 influence; cross-cultural communication

**1. Introduction.** Tag questions, defined as utterances with an interrogative tag (Huddleston & Pullum 2002), are a common occurrence in English. Allerton (2009) defined tag questions as appendages to a sentence, typically a statement, that seek confirmation of the listeners' understanding and, in some cases, agreement with the content of the preceding sentence. They commonly appear not only in daily conversations but also in educational contexts. Chinese and English are typologically distinct languages, belonging to two language families. English belongs to the Indo-European language family, whereas Chinese belongs to the Sino-Tibetan language family. These two languages are different in various aspects, such as those of their phonetic system, word formation, and syntactic rules. The formation of tag questions differs between Chinese and English, as does the way speakers respond to tag questions containing Negation In their Proposition (NIP). For example, for the NIP tag question below:

(1a) You don't need to come today, do you?

an English native speaker would typically respond:

No, (I don't),

given the fact that they do not need to come. In contrast, a native Chinese speaker in the same situation encountering the equivalent question:

(1b) Ni jintian buyong lai, shiba?

*'You don't need to come today, is it?'*

would likely respond:

Shide, wo buyong lai.

*'Yes, I don't need to come.'*

In this Chinese response, the word 'Shide' ('Yes') agrees with the overall meaning of the proposition in the tag question—that the speaker does not need to come—rather than matching the negative grammatical form of the proposition. When Chinese English-as-a-second-language (ESL) speakers respond to English tag questions of this type, they often transfer this strategy of their first language (L1). As a result, a Chinese ESL speaker may answer 'Yes, I don't need to come'

---

\* Authors: Meixing Yi, University of Mississippi ([myi@olemiss.edu](mailto:myi@olemiss.edu)) & Christiana Christodoulou, University of Mississippi ([cchris26@olemiss.edu](mailto:cchris26@olemiss.edu)).

to an English tag question as in (1b). This reflects a deeper cross-linguistic difference in how tag questions are processed and answered across the two languages.

Applying the L1's distinct answering system to the second language (L2) may lead to negative L1 influence (Holmberg 2013). Negative L1 influence in answering English NIP tag questions happens to most Chinese ESL speakers even to those who have studied or lived in an English-speaking country for some time. The current study examines whether the length of stay (LoS) in the U.S. is effective in reducing the negative L1 influence on L2 among Chinese ESL speakers with respect to verbally answering English NIP tag questions.

Our study primarily enhances the understanding of how Chinese ESL speakers respond to English NIP tag questions. The findings also provide a robust foundation for subsequent research into how the LoS in an immersion environment affects the mitigation of L1 transfer with respect to the response to English NIP tag questions.

**2. Background.** Tag questions are a common occurrence in English (Huddleston & Pullum 2002). Tottie & Hoffmann (2006) also argued that tag questions are a conspicuous linguistic phenomenon of spoken language. According to Huddleston & Pullum (2002), tag questions are 'utterances with an interrogative tag', including two components, 'anchor' and 'tag'. The *anchor*, a term used by Huddleston & Pullum (2002), refers to the main clause that provides the propositional content of the sentence. While Huddleston & Pullum (2002) used 'anchor', various other terms have been employed in the literature to describe this component, including *host clause* (Cattell 1973), *main clause* (Biber 1999), *basic clause* (Östman 1981), *matrix clause* (Quirk & Crystal 2010), *stem clause* (McGregor 1995), and *reference clause* (Nässlin 1984). The tag, by contrast, is the interrogative element that follows the anchor and seeks confirmation or disconfirmation. This component is consistently referred to as the 'tag' across most sources. Holmberg (2015) interpreted tag questions as presenting an assertion among two alternatives offered to the listener, with the entire construction falling under the scope of question force, that is to say, the whole sentence of a tag question functions as a *yes/no* question (also referred to as polar questions). So, NIP tag questions, the primary focus of our study, function as negative *yes/no* questions. These definitions focus on the syntactic form of tag questions.

Tottie & Hoffmann (2006) compared (a) the frequency, (b) sentence pattern, (c) the choice of auxiliaries or modal verbs in tags, and (d) the users age group of tag questions in British English and American English and found that there were nine times more tag questions in British English than in American English. Bonsignori (2007) discussed all possible syntactic forms of the tag questions in English and its varieties (e.g. Scottish, Irish, and Geordie) in a study of English tag questions. Wong (2007) investigated the form of tag questions in Hong Kong English based on a corpus study. Cuenca (2011) compared the forms of tag questions across English, German, and some Romance languages, and found that these western languages shared similar tag markers, which confirmed her hypothesis that the form and use of tag questions are interdependent. Later, Barron et al. (2015) conducted a corpus analysis of the form and function of tag questions across Irish English and British English. Westphal (2021) compared the similarities and differences across Philippine and Trinidadian English in terms of tag question forms and functions through a corpus-based study and concluded that cross-variety studies should move beyond surface-form comparisons and recognize that invariant question tags exhibit complex, functionally distinct patterns across text types and contexts.

Tag questions are commonly seen in daily conversations, and they have been determined to

play a significant role in interactions within an educational context (Wei et al. 2023), in which not only native English speakers but also international ESL speakers might participate. Barbieri's (2015) investigation of discourse features revealed that tag questions were particularly frequent in smaller student classes. Consequently, the analysis of question-and-answer communication formats in educational context interactions has gained increasing attention (Wei et al. 2023). Those analyses examined a variety of dimensions, such as interactional structures, conversation analysis and corpus linguistics, as well as a systemic functional approach (Sinclair & Coulthard 1975; Long et al. 1983; Walsh 2013; Yang 2021; Yang & Yin 2022). However, there are only a handful of studies focusing on the way ESL speakers answer English tag questions, especially the NIP tag questions.

Chinese ESL speakers form and respond to tag questions differently from native English speakers. A comparative study on the structure and function of tag questions in Chinese and English concluded that most English tags are declarative tags, which require strict and complicated, but regular, syntactic formation (Hsin 2016). Specifically, the subject, auxiliary, and verb tense in the tag must match those in the main statement. In contrast, Chinese tags are semantically and syntactically easier than English tags. They can be formed by simply appending particles such as 'Bushi ma' (*'isn't it'*), 'Shi ma/Shi ba' (*'is it'*) or 'Dui ma/Dui ba' (*'right?'*) to the end of any statements, without requiring syntactic alignment or verb agreement between the statement and the tag.

Regarding the manner of answering tag questions, in English, responses to NIP tag questions are aligned with the grammatical form of the proposition part. For instance, when asked:

(2a) You don't like coffee, do you?

a native English speaker would respond:

'No, (I don't)',

if they indeed do not like coffee. In contrast, a native Chinese speaker faced with the equivalent question:

(2b) Ni bu xihuan kafei, shiba?

*'You don't like coffee, is it?'*

in the same situation would likely answer:

Shide, wo bu xihuan.

*'Yes, I don't like (coffee).'*

where 'Shide' (*'Yes'*) reflects agreement with the overall meaning of the proposition part that the speaker indeed does not like coffee—rather than mirroring the negative grammatical structure of the proposition. Answers to negative *yes/no* questions (e.g., *Doesn't she like cats?*) likewise exhibit this sharp distinction (Zhang & Vanek 2021). In English, the *yes/no* part of the answer is typically of the same polarity as the verb in the answer (e.g., *Yes, she does/No, she doesn't*). In Mandarin, however, *shi/shi de* (*yes*) and *bu/bu shi/bu shi de* (*no*) typically oppose the polarity of the verb in the answer (e.g., *Yes, she doesn't/No, she does*) (Holmberg 2015). These two distinct approaches of answering—whether NIP tag questions or negative *yes/no* questions—are commonly referred to as the truth-based system and the polarity-based system. According to Holmberg (2015), Mandarin speakers follow a truth-based system, whereas English speakers mostly use a polarity-based system. Therefore, when Chinese ESL speakers respond to the English NIP tag questions, they often transfer the answering 'system' from their L1 Chinese, which can be confusing or incorrect from an English perspective.

NIP tag questions pose the same answering challenge as negative polarity questions, which

have long been studied. Beyond distinguishing between truth-based and polarity-based answering systems, scholars have proposed different explanations for this contrast. One influential account is syntactic: Holmberg (2013) argued that the distinction arises from the interaction between the syntax of negation and the syntax of polar questions. Other work emphasizes cognitive mechanisms. For example, Zhang & Vanek (2021) compared responses to negative questions among Mandarin monolinguals, English monolinguals, and Chinese ESL bilinguals using a comprehension task measuring reaction times. Their results showed that negative questions impose greater processing demands on speakers of truth-based languages such as Mandarin than on English speakers. Evidence from Chinese ESL participants further suggested cognitive restructuring, indicating a gradual shift from the L1 truth-based system toward the polarity-based system of English. So, existing explanations for the contrast between truth-based and polarity-based answering systems may be syntactic, cognitive, or a combination of both.

Recently, there is a growing body of literature on the negative influence of L1 Chinese on L2 English. Huang (2010) pointed out identified potential L1 Chinese influences on L2 English communication across the word, sentence, and discourse levels. Yuan (2014) conducted a corpus-based study on the influence of L1 Chinese on ESL speakers' use of English prepositions. Wu (2014) investigated the influence of Chinese L1 thinking on use of English L2 in oral expressions. Yuanhua & Xiaoling (2016) and Dipolog-Ubanan (2016) studied the effects of L1 transfer on English writing of Chinese EFL students with different focuses: Yuanhua & Xiaoling (2016) focused on interlanguage use, simplification or overgeneralization of the target language, while Dipolog-Ubanan (2016) focused on grammatical errors and writing proficiency. Kim et al. (2017) conducted an experiment to examine how Chinese L1 influence English L2 reading in brain activity. Bai & Qin (2018) investigated the negative L1 transfer in Chinese college student's English writing and discovered the interference of their L1 shown in idioms, cultural terms and conventions in expression. Their research mentioned that thinking mode played a part in accounting for these interferences. Xu et al. (2024) research also investigated the interference of Chinese (L1) thinking mode on English (L2) focused on writing instead of oral communication. Liu & Yan (2022) analyzed the interference confronted by L1-Chinese speakers from linguistic aspects, such as phonetics, lexical collocation, and word order, as well as cultural perspectives. Zhang (2023) investigated the influence of L1 Chinese on L2 English writing in aspects of grammar, semantics, and syntax, and attributed some interference on L2 writing to the L1 thinking mode and linguistic habits. Although Zhang (2010) briefly discussed Chinese ESL speakers' responses to English tag questions in relation to L1 transfer, and Zhang & Vanek (2021) conducted an experiment on how Chinese ESL speakers answered English negative questions, research specifically addressing L1 Chinese transfer in L2 English responses to tag questions remains scarce.

This type of L1 Chinese influence in answering English NIP tag questions has also been observed with Chinese ESL speakers who have studied or lived in an English-speaking country extensively. LoS—referring to the total duration an individual spends living in a foreign language environment—has been proven to be a determining factor in second language development (Juan-Garau 2015). Studies in this category treated the *Study Abroad* (SA) context as a black box or used it as a categorical label, as opposed to at-home curricula (Taguchi 2018). There are also many studies on the influence of the duration of SA programs on second language development. For example, Longcope's (2003) study revealed that SA had an immediate reflection on the learners' fluency, but not necessarily on their grammatical accuracy or syntactic complexity. Avello & Lara (2014) found that the length of SA failed to mitigate the foreign accent of Spanish ESL

speakers. Juan-Garau (2015) concluded that there were several possible main benefits in the oral domain after SA, namely, speech complexity, accuracy, oral fluency, and pronunciation. Her studies also showed that a SA setting was superior to a domestic setting when it came to both oral accuracy and the participants' greater approximation of native speaker norms, although differences between native speakers and non-native speakers were still significant after participating in a SA program.

Long & Solon (2021) studied the impact of a short-term four-week SA program on L2 Spanish syntactic complexity development in narratives by comparing 16 learners who studied abroad with 12 peers who received at-home instruction. They found that the SA group demonstrated a non-significant change over the study period, however, their individual-level analyses revealed that over half of the SA learners increased complexity in narratives in terms of clause length (Long & Solon 2021). McManus et al. (2021) conducted a 21-month longitudinal study to track the linguistic development of 56 advanced university students learning French and Spanish to see how a long-term SA impacted their L2 fluency and vocabulary, accuracy, and syntactic complexity. Their study challenged the narrative that long-term SA program is most beneficial for oral fluency and lexis over accuracy, and confirmed that, for learners who already have a strong linguistic foundation, a nine-month SA also provides the necessary conditions for significant and sustained development in grammatical accuracy. Importantly, this study further established that the length of SA programs is positively correlated with learners' language gains. On the contrary, an extended and prolonged domestic program does not necessarily lead to such gains.

Neff & Apple (2023) investigated the non-linguistic impacts of short-term (one month) versus long-term (one year) SA programs on Japanese learners of English by using pre- and post-SA surveys to measure changes in intercultural communication, L2 speaking confidence, and the sense of an L2 self. The study found that while SA of any duration boosts L2 speaking confidence, a longer stay exerts a stronger effect by fostering an 'ideal L2 self'—the learner's envisioned identity as a fluent L2 user within Dörnyei's (2009) L2 Motivational Self System—and increasing the willingness to engage in intercultural communication. Tseng et al. (2024) reported that learners' age and prior training do not appear to influence the effectiveness of SA language programs. Notably, similar to McManus et al. (2021), their study reinforces the view that immersion in an L2 environment contributes more effectively to language development than prolonged domestic instruction. Despite the fact that there are studies on the influence of LoS in a target language country on L2 development, most of these studies focus on short-term study abroad experiences (typically lasting no more than one year), because, according to Ren (2018), short-term study abroad programs are more popular and the observation period of the long-term program is complex and effort-demanding. Furthermore, only a handful of studies compare different LoS to examine the effect that they may have on similar groups of participants (Borràs & Llanes 2021).

Given what has been reported by previous research (e.g. Zhang (2010)), confusion may frequently emerge when Chinese ESL speakers verbally respond to English tag questions under the influence of their L1 Chinese. However, there is no research available on how their L1 influences their responses, nor has it been explored whether an extended immersion in an English-speaking environment can mitigate this L1 influence. Therefore, with the aim of exploring these potential effects, and specifically the effects of duration on the elimination or decrease of L1 transfer on L2, the current study asks:

*Does an extended stay in an English-speaking environment help Chinese ESL speakers reduce L1 transfer when verbally responding to English NIP tag questions?*

Using a cross-sectional approach, this study compares the verbal responses of participants who have stayed in the U.S. for five years or less ( $\leq 5$ ) with those who have stayed for more than five years ( $> 5$ ) to determine whether LoS significantly influences L2.

Building on prior research in SA contexts, two hypotheses were formulated:

$H_0$ : It is hypothesized that LoS would not exert a significant effect on performance, indicating no statistically significant difference between the  $\leq 5$ -year group and  $> 5$ -year group.

$H_1$ : Conversely, following McManus et al. (2021), it is expected that a longer LoS in an English-speaking environment would significantly enhance the performance of participants in the  $> 5$ -year group, resulting in significant differences between the  $\leq 5$ -year group and  $> 5$ -year group, as well as between the  $\leq 5$ -year group and the native control group.

As discussed above, several theoretical accounts have been proposed to explain why L1 Chinese learners experience this difficulty or difference when responding to English NIP tag questions. However, since the underlying causes of this difference—whether semantic, syntactic, or cognitive—fall outside the scope of the present study, they will not be addressed in this work.

**3. Methodology & Data Analysis.** As for the cross-sectional design, this study adopted the methodological design utilized by Simard et al. (1976). Simard et al. (1976) employed a series of experimental conditions where participants were asked to describe a route on a map (stimulus materials), then required participants to complete questionnaires assessing their perceptions of other participants' language abilities and their own efforts in communication. This experiment consisted of a simulated performance session followed by a self-assessment questionnaire, with analyses based on the participants' perceived effort and self-reported language proficiency. Drawing on the framework established by Simard et al. (1976), the current study comprised two sessions: a performance-based questionnaire and a metalinguistic awareness survey. The performance-based questionnaire was designed to examine the participants' actual responses when presented with English NIP tag questions, thereby capturing their real-time performance. Our metalinguistic awareness survey sought to investigate participants' self-reported perceptions of the extent to which their responses to English NIP tag questions were influenced by their L1 Chinese.

3.1. PARTICIPANTS. Two groups of participants were recruited through convenience sampling based on the following criteria: (1) their age (from 18 to 65), (2) being native speakers of Chinese with English as their L2, and (3) their experience in living, studying, or working in the U.S.. Although participants were recruited through convenience sampling, the majority of them were not related to the researchers in any way. **Table 1** shows the participants' demographic information.

3.2. EXPERIMENT DESIGN. To refine the experimental design, a pilot study was conducted prior to the current study. The pilot study involved ten Chinese ESL participants, five of whom had resided in the U.S. for more than two years and five for less than two years. Participants listened to ten short stories in English narrated by the researcher and, following each story, responded orally to an English NIP tag question based on the story content. Insights from the pilot study informed several methodological refinements: an answer sheet was developed to facilitate systematic note-taking during data collection; some story prompts were revised for greater clarity; five additional distracting stories were incorporated to enhance data reliability.

	Age Range				Educational Level			
	18–29	30–39	40–49	50–65	High school	College	Master’s	Ph.D.
Group 1 ( $\leq 5$ )	16 (72.7%)	6 (27.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (27.3%)	11 (50.0%)	5 (22.7%)
Group 2 ( $> 5$ )	4 (18.2%)	6 (27.3%)	4 (18.2%)	8 (36.4%)	0 (0.0%)	1 (4.5%)	8 (36.4%)	13 (59.1%)
Group 3 (Control)	15 (68.2%)	2 (9.1%)	1 (4.5%)	4 (18.2%)	3 (13.6%)	14 (63.6%)	4 (18.2%)	1 (4.5%)

Table 1. Participants’ demographic information with percentages

The experiment of current study consists of two sessions: the performance-based questionnaire and the metalinguistic awareness survey, lasting no more than 15 minutes, collectively. During the performance-based questionnaire, participants listened to 15 short stories in English (10 experimental items and 5 distractors), which were constructed and narrated by the researcher, a native speaker of Chinese. Following each story, participants were verbally asked an English NIP tag question based on the story’s content and were required to provide a verbal response. An example of the experimental items is as follows:

- (1) **Story:** *Lucy and Mike have just completed their lunch, and they are feeling quite full.*

**Tag Question:** *They are not hungry at all, are they?*

The researcher documented participants’ responses using a prepared score sheet that included a list of five anticipated response types. These anticipated response types were informed by the responses collected during the pilot study. In most cases, participants’ responses fell under one of the five predefined types. However, when a response did not fit any of these types, it was transcribed verbatim, examined in detail, and then interpreted accordingly. To minimize potential bias, the anticipated responses were not disclosed to participants under any circumstances, nor was there any direct or indirect indication of what might constitute a correct response. On most occasions, when the participants’ response was originally marked as differing from our five types of anticipated responses, upon further analysis, it was determined that the response did match the criteria of one of the five types. Then any residual responses remained in a Type 6 category and were excluded from statistical analysis. An example of score sheet for the tag question in (1) above is given below (see the [Appendix](#) for more information):

- i. Respond ‘Yes, (they are not...)’;* **(Truth-based Response)**
- ii. A repetition of the proposition ‘They are not...’ without explicit affirmation or negation;* **(Avoidance Response)**
- iii. Initially answer with ‘Yes, (they are not...),’ then later corrected to ‘No, (they are not...)’;* **(Self-correction Response)**
- iv. Contemplate for 3 seconds and reply with ‘No,(they are not...)’;* **(Processing Response)**
- v. Respond with ‘No, (they are not...)’ with no more than 3 seconds;* **(Polarity-based Response)**
- vi. Alternative Responses.*

The measurement of hesitant pauses varies widely across studies. Jefferson (1984) did not specify a temporal threshold, and Campione et al. (2002) cautioned that fixed thresholds may obscure meaningful shorter pauses. In studies of oral fluency, McManus et al. (2021) defined pauses as silences of 0.25 s or longer when calculating the Mean Length of Run. Similarly, Betz et al. (2023) avoided specifying a minimum duration, arguing that no objective detection method exists and relying instead on annotators’ perception of hesitation. By contrast, Coulange et al. (2024) adopted a practical threshold, defining pauses as intervals between words lasting from 180 ms to 2 s. In contrast, the present study focuses on Speech Onset Time (SOT)—the interval between stimulus onset and articulation onset—as a measure of cognitive processing load in Type 4 Processing Responses. For analytical purposes, a cutoff of 3 s (3000 ms) was used to distinguish Type 4 Processing Responses from Type 5 Polarity-based Responses.

This threshold is supported by both internal and external benchmarks. Annotation of approximately 11 hours of speech data showed that the highest mean SOT in our dataset was 2,321 ms, placing the 3 s cutoff safely above typical response times. At the same time, it approaches the range associated with higher cognitive load reported by Zhang & Vanek (2021), whose data show mean SOTs of approximately 3,250-4,250 ms for challenging negative yesno questions. Moreover, across all response types in the present dataset (see **Table 2**), only Type 4 Processing Responses exhibited mean SOTs exceeding 3 s. Together, the pilot study and the overall data distribution indicate that the 3 s threshold reliably differentiates hesitant, less automatic responses (Type 4) from faster, more automatic polarity-based responses (Type 5).

	1. Truth-based Response	2. Avoidance Response	3. Self-correction Response	Response Interval	4. Processing Response	5. Polarity- based Response
Group 1 ( $\leq 5$ )	1,562	1,732	779	2,798	4,151	1,096
Group 2 ( $> 5$ )	1,312	1,872	1,925	4,683	3,747	1,177
Group 3 (Control)	1,403	1,073	1,449	3,033	4,412	811

Table 2. Mean Speech Onset Time (ms) for each response type across the three groups

The entire performance-based questionnaire session for each participant was audio-recorded. For confidentiality purposes, anonymity was ensured by using a coding system to both label audio files and record the participants’ names and responses in our database. After the data collection was completed, the five anticipated response types were assigned numerical values for statistical analysis:

$$\begin{aligned} \text{Type 1} &= 1 \text{ point}, & \text{Type 2} &= 2 \text{ points}, & \text{Type 3} &= 3 \text{ points}, \\ \text{Type 4} &= 4 \text{ points}, & \text{Type 5} &= 5 \text{ points}. \end{aligned}$$

Once again, we reiterate that only the researcher had access to the score sheet. The predefined response categories were included solely to facilitate the efficient administration of the task.

After completing their performance-based questionnaire, participants in Group 1 and Group 2 were asked to fill out a metalinguistic awareness survey. In this survey, they self-evaluated the extent to which they were inclined to respond with truth-based answers, such as ‘Yes, they are

### Response Type Distribution across Groups

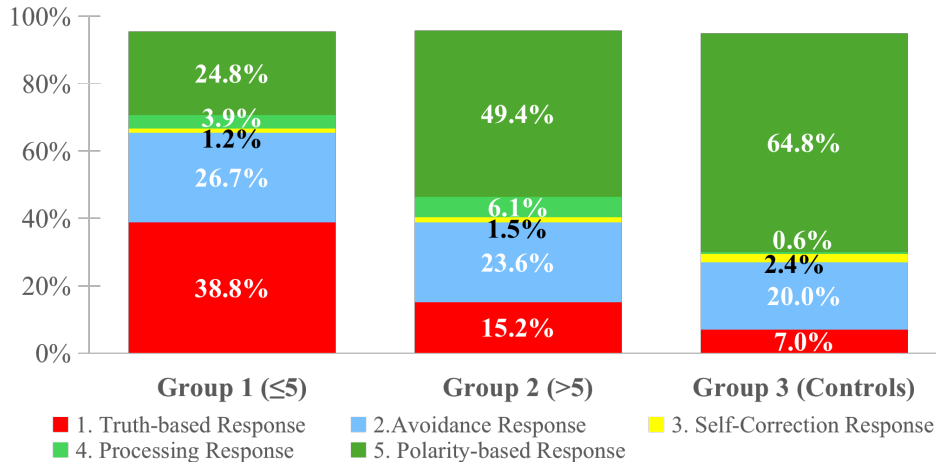


Figure 1. Response type distribution across groups

not’, to the English tag questions (following polarity-based answering system) in the first session, using a four-point Likert scale:

*1 = Strongly inclined; 2 = Moderately inclined;  
3 = Slightly inclined; 4 = Not inclined at all.*

Scores ranging from 1 to 4 were assigned to represent the degree of inclination, with 1 indicating ‘strongly inclined’ and 4 indicating ‘not inclined at all’.

**4. Results.** The one-on-one performance-based questionnaire and the metalinguistic awareness survey were administered in quiet and publicly accessible locations. The data were systematically entered and organized in an Excel database for analysis.

**4.1. PERFORMANCE-BASED QUESTIONNAIRE.** For the performance-based questionnaire, as illustrated in **Figure 1**, 38.8% of participants in Group 1 produced Type 1 Truth-based Responses, compared to 15.2% in Group 2 and 7.0% in Group 3. In contrast, only 24.8% of Group 1 participants produced Type 5 Polarity-based Responses, whereas this response type was observed in 49.4% of Group 2 participants and 64.8% of those in Group 3.

Since higher educational level was proven to correlate with increased language proficiency and metalinguistic awareness (Collier 1987), which may influence response patterns independently of LoS, we first examined whether the demographic factor of educational level might act as a confounding variable. A chi-squared test was conducted to compare the educational composition of Group 1 and Group 2, which included only three levels: college, master’s, and doctoral level. The test confirmed a significant imbalance ( $\chi^2(2)=7.60, p=.022$ ), indicating that the two groups differed in their educational backgrounds.

To formally test the influence of LoS while statistically controlling for this educational imbalance, a Cumulative Link Mixed Model (CLMM) was conducted. This approach was selected as it is appropriate for modeling ordinal outcomes (response rating 1-5) while also accounting for

## Answer Type Distribution in Group 1 & 2

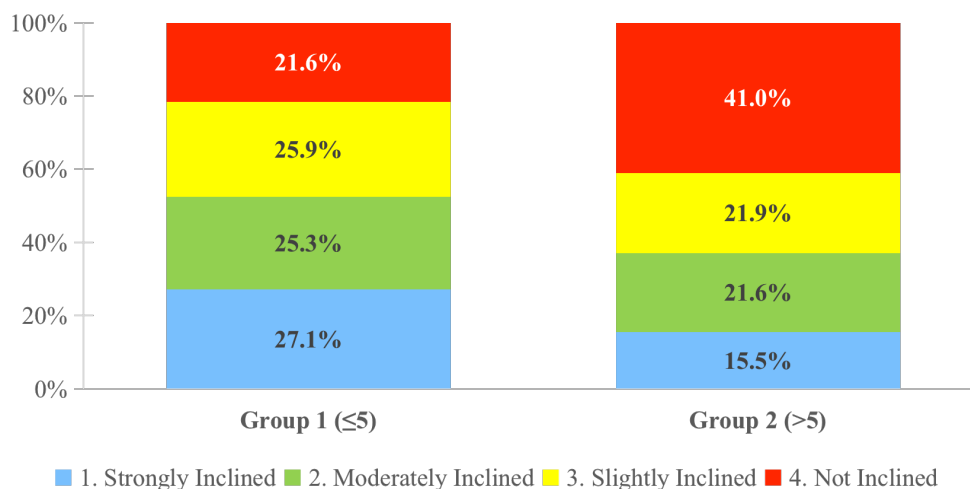


Figure 2. Answer type distribution in Group 1 & 2

the non-independence of repeated measures (multiple response ratings nested within each participant). The CLMM included group (LoS, with Group 1 as the reference) and edu (educational level) as fixed effects. Due to the heterogeneous representation of educational levels across the groups, specifically, the absence of the high school level in Groups 1 and 2, and to ensure model stability and reliable parameter estimation, the lowest two educational levels were aggregated. The high school and college levels were therefore combined into a single reference category, ‘College Level and Below’, for the CLMM analysis. The random intercepts were specified after Likelihood Ratio Tests confirmed that the inclusion of both ParticipantID and QuestionID significantly improved model fit ( $p < .001$  for both components).

The CLMM results indicated that educational level was not a significant predictor of response ratings (all  $p > .30$ ). This suggests that the observed difference in educational composition did not account for the differences in response ratings.

In contrast, LoS remained a highly significant predictor of participants response ratings. Compared to Group 1, the odds of receiving a higher response rating were 5.07 times greater for participants in Group 2 (OR= 5.07,  $\beta=1.62$ ,  $SE=0.54$ ,  $p=.003$ ) and 13.16 times greater for those in Group 3 (OR= 13.16,  $\beta=2.58$ ,  $SE=0.57$ ,  $p < .001$ ), indicating a strong and increasing effect of LoS on participants’ response ratings. A post-hoc analysis with Tukey adjustment was performed to conduct pairwise comparisons among the three LoS groups. The results revealed that Group 1 had significantly lower response ratings than both Group 2 ( $p=.007$ ) and Group 3 ( $p < .0001$ ). The difference between Group 2 and Group 3 was not statistically significant ( $p=.316$ ).

4.2. THE METALINGUISTIC AWARENESS SURVEY. In the metalinguistic awareness survey, participants—excluding those in the control group—self-assessed the extent to which they were inclined to provide a truth-based response caused by their L1 transfer (e.g. respond with ‘Yes, they are not hungry’ to the tag question ‘They are not hungry at all, are they?’ ) to each previously presented tag question. **Figure 2** summarizes the distribution of answer types in Group 1 and Group 2.

As shown, 27.1% of participants in Group 1 reported being strongly inclined to provide Type 1 Truth-based Response to the English tag questions provided in the performance-based section of the study, compared to 15.5% in Group 2. We also see that fewer than half of the participants from both groups (21.6% in Group 1 and 41% in Group 2) reported that they were not inclined to provide a Type 1 Truth-based Response.

To compare self-awareness when answering the experimenting English tag questions between Group 1 and Group 2, the four degrees of inclination were assigned numerical scores from 1 to 4, with 1 representing 'strongly inclined' and 4 representing 'not inclined'. Median scores for each participant in each group were calculated. Normality was assessed for both groups using the Shapiro-Wilk test. The results indicated significant deviations from normality for both Group 1 ( $W=0.880, p=.012$ ) and Group 2 ( $W=0.829, p=.001$ ). Consequently, the non-parametric Mann-Whitney U test ( $\alpha = .05$ ) was employed. The results revealed no statistically significant difference in self-awareness scores between Group 1 and Group 2 ( $W= 184, p= .162$ ).

**5. Discussion.** This study investigated whether extended exposure to an English-speaking environment reduces L1 transfer among Chinese ESL speakers when responding to English NIP tag questions. Results from the performance-based questionnaire revealed that LoS would have no significant effect. Participants who had resided in the U.S. for more than five years produced significantly more target-like responses than those with five years or less of residence, while no significant difference was observed between the longer-stay group and native English speakers. Educational level did not emerge as a significant confounding factor. These findings suggest that increased exposure to the target language environment reduces L1 influence on L2 performance in responses to English NIP tag questions. This pattern is consistent with Zhang & Vanek (2021), who propose that L2 learners initially rely on a truth-based response system but gradually shift toward the target polarity-based system with greater exposure to English.

In contrast, the metalinguistic awareness survey revealed no significant group difference in participants' perceived L1 influence. Both groups reported similar levels of awareness regarding their tendency toward truth-based responses, even when their actual performance differed. Qualitative feedback further indicated that some participants experienced discomfort and a strong inclination toward truth-based responses despite recognizing that such responses were not appropriate in English.

Taken together, we conclude that while the extended stay in the U.S. does not necessarily increase the self-awareness of reduced L1 interference, it does help to mitigate the negative L1 influence in practical L2 performance when responding verbally to English NIP tag questions.

This study contributes to L2 acquisition research by clarifying how Chinese ESL speakers respond to English NIP tag questions. It highlights the role of study-abroad experiences, showing that extended exposure to an English-speaking environment can promote more target-like responses to conflicting L1-L2 structures. The study also examines the relationship between LoS and L1 influence on L2 in responses to NIP tag questions, an area with limited empirical attention. The findings provide a basis for future research and offer implications for improving cross-cultural communication.

## References

- Allerton, DJ. 2009. 16 tag questions. *One Language, Two Grammars?* 306.  
Avello, Pilar & Ann Rebecca Lara. 2014. Phonological development in l2 speech

- production during study abroad programmes differing in length of stay. *Language acquisition in study abroad and formal instruction contexts* 13. 137–166. <https://doi.org/http://digital.casalini.it/9789027270238>.
- Bai, Lin & Jie Qin. 2018. A study of negative language transfer in college students' writing from cultural perspective. *Theory and Practice in Language Studies* 8(3). 306–313.
- Barbieri, Federica. 2015. Involvement in university classroom discourse: Register variation and interactivity. *Applied linguistics* 36(2). 151–173. <https://doi.org/https://doi.org/10.1093/applin/amt030>.
- Barron, Anne, Irina Pandarova & Karoline Muderack. 2015. Tag questions across irish english and british english: A corpus analysis of form and function. *Multilingua* 34(4). 495–525.
- Betz, Simon, Nataliya Bryhadyr, Olcay Türk & Petra Wagner. 2023. Cognitive load increases spoken and gestural hesitation frequency. *Languages* 8(1). 71.
- Biber, Douglas. 1999. A register perspective on grammar and discourse: Variability in the form and use of english complement clauses. *Discourse studies* 1(2). 131–150.
- Bonsignori, Veronica. 2007. Tag questions in english. a syntactic, pragmatic and prosodic account .
- Borràs, Judith & Àngels Llanes. 2021. Re-examining the impact of study abroad on L2 development: A critical overview. *The Language Learning Journal* 49(5). 527–540. <https://doi.org/https://doi.org/10.1080/09571736.2019.1642941>.
- Campione, Estelle, Jean Véronis et al. 2002. A large-scale multilingual study of silent pause duration. In *Speech prosody*, vol. 2002, 199–202.
- Cattell, Ray. 1973. Negative transportation and tag questions. *Language* 612–639. <https://doi.org/https://doi.org/10.2307/412354>.
- Collier, Virginia P. 1987. Age and rate of acquisition of second language for academic purposes. *TESOL quarterly* 21(4). 617–641.
- Coulangue, Sylvain, Tsuneo Kato, Solange Rossato & Monica Masperi. 2024. Enhancing language learners comprehensibility through automated analysis of pause positions and syllable prominence. *Languages* 9(3). 78.
- Cuenca, Maria Josep. 2011. Form-use mappings for tag questions. In *Discourse and perspective in cognitive linguistics*, 3–20. John Benjamins Publishing Company.
- Dipolog-Ubanan, GF. 2016. L1 influence on writing in L2 among ucsi chinese students: A case study. *Pertanika Journal of Social Sciences & Humanities* 24(4).
- Dörnyei, Zoltán et al. 2009. The L2 motivational self system. *Motivation, language identity and the L2 self* 36(3). 9–11.
- Holmberg, Anders. 2013. The syntax of answers to polar questions in english and swedish. *Lingua* 128. 31–50. <https://doi.org/https://doi.org/10.1016/j.lingua.2012.10.018>.
- Holmberg, Anders. 2015. *The syntax of yes and no*. Oxford University Press.
- Hsin, Ai-li C. 2016. An analysis of chinese tag questions with a cross-linguistic comparison to english tags. *Taiwan Journal of Linguistics* 14(1).
- Huang, Li-Shih. 2010. The potential influence of L1 (chinese) on L2 (english) communication. *ELT journal* 64(2). 155–164.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The cambridge grammar of the english language*. Cambridge University Press. <https://doi.org/https://doi.org/10.1017/9781316423530>.
- Jefferson, Gail. 1984. Notes on a systematic deployment of the acknowledgement tokens yeah; and mm hm .

- Juan-Garau, María. 2015. Speaking an additional language: Can study abroad do the trick? *Language Value* 7. 45–60.
- Kim, Say Young, Li Liu & Fan Cao. 2017. How does first language (L1) influence second language (L2) reading in the brain? evidence from Korean-English and Chinese-English bilinguals. *Brain and Language* 171. 1–13.
- Liu, Xinyi & Runfei Yan. 2022. The negative transfer confronted by Chinese English learners: from the perspective of linguistics and culture. In *2022 5th international conference on humanities education and social sciences (ichess 2022)*, 3151–3159. Atlantis Press.
- Long, Avizia Y & Megan Solon. 2021. The impact of a short-term stay abroad on L2 Spanish syntactic complexity development in narratives. *Study Abroad Research in Second Language Acquisition and International Education* 6(1). 163–188.
- Long, Michael H, Charlene J Sato et al. 1983. Classroom foreigner talk discourse: Forms and functions of teachers' questions. *Classroom oriented research in second language acquisition* 268–285.
- Longcope, Peter Duncan. 2003. *What is the impact of study abroad on L2 learning? a descriptive study of contexts, conditions, and outcomes*. University of Pennsylvania.
- McGregor, B. 1995. The 'English tag question': A new analysis, is (n't) it? *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES* 4 91–122. <https://doi.org/https://doi.org/10.1075/cilt.118.04mcg>.
- McManus, Kevin, Rosamond Mitchell & Nicole Tracy-Ventura. 2021. A longitudinal study of advanced learners' linguistic development before, during, and after study abroad. *Applied Linguistics* 42(1). 136–163. <https://doi.org/https://doi.org/10.1093/applin/amaa003>.
- Nässlin, Siv. 1984. *The English tag question: a study of sentences containing tags of the type 'isn't it?', 'is it?'*: Almquist & Wiksell international dissertation.
- Neff, Peter & Matthew Apple. 2023. Short-term and long-term study abroad: The impact on language learners' intercultural communication, L2 confidence, and sense of L2 self. *Journal of Multilingual and Multicultural Development* 44(7). 572–588. <https://doi.org/https://doi.org/10.1080/01434632.2020.1847125>.
- Östman, Jan-Ola. 1981. You know .
- Quirk, Randolph & David Crystal. 2010. *A comprehensive grammar of the English language*. Pearson Education India.
- Ren, Wei. 2018. Developing L2 pragmatic competence in study abroad contexts. In *The Routledge handbook of study abroad research and practice*, 119–133. Routledge.
- Simard, Lise M, Donald M Taylor & Howard Giles. 1976. Attribution processes and interpersonal accommodation in a bilingual setting. *Language and Speech* 19(4). 374–387.
- Sinclair, John & Malcolm Coulthard. 1975. Towards an analysis of discourse: The English used by teachers and pupils. (*No Title*) .
- Taguchi, Naoko. 2018. Contexts and pragmatics learning: Problems and opportunities of the study abroad research. *Language Teaching* 51(1). 124–137. <https://doi.org/10.1017/S0261444815000440>.
- Tottie, Gunnel & Sebastian Hoffmann. 2006. Tag questions in British and American English. *Journal of English Linguistics* 34(4). 283–311. <https://doi.org/https://doi.org/10.1177/0075424206294369>.

- Tseng, Wen-Ta, Yeu-Ting Liu, Yi-Ting Hsu & Hsi-Chin Chu. 2024. Revisiting the effectiveness of study abroad language programs: A multi-level meta-analysis. *Language Teaching Research* 28(1). 156–200. <https://doi.org/https://doi.org/10.1177/136216882098>.
- Walsh, Steve. 2013. *Classroom discourse and teacher development*. <https://doi.org/https://doi.org/10.1515/9780748645190>.
- Wei, Lifang, Alex Ho-Cheong Leung & Yufeng Sun. 2023. we dont normally require that in other contexts, do we: Interpersonal meanings of tag questions in british university seminars based on the base corpus. *Frontiers in Psychology* Volume 13 - 2022. <https://doi.org/10.3389/fpsyg.2022.1070937>.
- Westphal, Michael. 2021. Question tags across new englishes. *World Englishes* 40(4). 519–533.
- Wong, May Ly. 2007. Tag questions in hong kong english: A corpus-based study. *Asian Englishes* 10(1). 44–61.
- Wu, Shu-Ling. 2014. Influence of l1 thinking for speaking on use of an l2: The case of path expressions by english-speaking learners of chinese. *Studies in second language acquisition of Chinese* 1–29.
- Xu, Wandong et al. 2024. Metacognitive strategy use in chinese (l1) and english (l2) writing: A within-subject comparison .
- Yang, Xueyan. 2021. How can efl teachers make their questions more interactive with students? interpersonal patterns of teacher questions. *System* 99. 102509. <https://doi.org/https://doi.org/10.1016/j.system.2021.102509>.
- Yang, Xueyan & Shunli Yin. 2022. Interpersonal projection as efl teachers' discourse strategy to enhance students' willingness to communicate: A systemic-functional perspective. *System* 104. 102687. <https://doi.org/https://doi.org/10.1016/j.system.2021.102687>.
- Yuan, Hu. 2014. A corpus-based study on the influence of l1 on efl learners' use of prepositions. *Theory and Practice in Language Studies* 4. 2513–2521. <https://doi.org/https://api.semanticscholar.org/CorpusID:17094572>.
- Yuanhua, CHEN & LI Xiaoling. 2016. Effects of l1 transfer on english writing of chinese efl students. *Cross-Cultural Communication* 12(4). 50–53.
- Zhang, Haoruo & Norbert Vanek. 2021. From no, she does to yes, she does: Negation processing in negative yes–no questions by mandarin speakers of english. *Applied Psycholinguistics* 42(4). 937–967.
- Zhang, Qianqian. 2023. The influence of l1(chinese) on writing in l2(english): A case study. *Communications in Humanities Research* <https://doi.org/https://doi.org/10.54254/2753-7064/3/20220511>.
- Zhang, Qiyun. 2010. A study of chinese learning of english tag questions. *Journal of Language Teaching & Research* 1(5).