International Journal of Applied Linguistics

WILEY

ORIGINAL ARTICLE

Phonetic Transcription Skills Promote L2 Listening Comprehension: Evidence From Longitudinal Cross-Lagged Models of L2 Chinese Learners at Different Proficiency Levels

Dan Mu[1] | Yuqian Li[2] | Caihua Xu[3]

[1]Center for Language and Cognition Groningen, University of Groningen, Groningen, Netherlands | [2]Department of Family Education, China National Children's Center, Beijing, China | [3]School of International Chinese Language Education, Beijing Normal University, Beijing, China

Correspondence: Caihua Xu (xchua@bnu.edu.cn)

Received: 17 June 2024 | **Revised:** 2 March 2025 | **Accepted:** 14 March 2025

Funding: This work was supported by National Social Science Fund of China under Grant number 19FYYB038; International Society for Chinese Language Teaching under Grant number SH23Y29; Academic Support Project for 2024 at School of International Chinese Language Education at Beijing Normal University.

Keywords: cross-lagged model | L2 Chinese | listening comprehension | phonetic transcription system | Pinyin | second language acquisition

关键词: 语音转码 | 汉语拼音 | 听力理解 | 汉语二语 | 交叉滞后模型

ABSTRACT

In deep orthographic languages, a separate phonetic transcription system, such as the Pinyin system in Mandarin Chinese, plays a crucial auxiliary role in helping learners master the pronunciation of written words. While many studies have indicated that learning Pinyin enhances Chinese reading, its relationship with listening comprehension in L2 Chinese remains under-explored. The current study investigated whether Pinyin transcription skills predict short- and long-term listening comprehension performance of L2 Chinese learners at different proficiency levels. A total of 120 non-native Chinese speakers were tested twice (with an interval of one semester, 5 months) on sound-to-Pinyin transcription, character-to-Pinyin transcription, character production, and listening comprehension skills. The results suggest that Pinyin transcription skills had positive effects on listening comprehension. Specifically, sound-to-Pinyin transcription skills showed direct and immediate effects on listening comprehension for both elementary and intermediate learners. Character-to-Pinyin transcription skills directly predicted listening comprehension for elementary learners but with some delay. It also predicted listening comprehension mediated by the character production skill, which occurred without any lag for elementary learners but with a lag for intermediate learners. Our study demonstrated that the benefits of learning Pinyin for listening comprehension vary among L2 Chinese learners at different levels. Some of these effects may not be immediately observable but become apparent after a period of accumulation.

*Dan Mu and Yuqian Li are designated as co-first authors.

摘要

在汉语这样的深层正字法语言中, 独立的语音转码系统 (如汉语拼音系统) 对学习者掌握书面文字发音具有重要辅助作用。尽管已有大量研究表明汉语拼音学习能够促进汉语阅读能力提升, 但其与汉语二语听力理解之间的关系仍然缺乏直接的证据。本研究旨在讨论拼音转码能力是否可以预测不同水平汉语二语者前后两次听力理解的成绩。本研究以120名非汉语母语学习者为被试, 对其语音-拼音转码、汉字-拼音转码、汉字产出及听力理解能力进行了前后两个时间点 (间隔一学期, 5个月) 的测量。结果发现, 拼音转码能力对听力理解具有显著的促进作用。具体而言, 听觉通道的语音-拼音转码能力对初级和中级两个级别被试的听力理解都产生了直接而即时的正向预测作用;视觉通道的汉字-拼音转码能力对初级被试听力理解的直接预测作用则存在一定滞后性, 而且它还可以通过汉字产出能力的中介作用间接预测听力理解, 其中介效应对初级被试是即时的, 对中级被试具有一定滞后性。总体而言, 本研究表明拼音学习对汉语二语学习的影响是深刻而长久的, 不仅会通过听觉通道起直接促进作用, 而且会通过视觉通道稳定而缓慢地产生长期的促进作用。

1 | Introduction

Chinese is widely recognized for the challenges posed by its writing system learning, not only due to the complexity of its orthographic forms but, more importantly, because the Chinese writing system does not directly encode phonological information, making its sound-orthography mappings particularly difficult to master (see McBride 2016, for a review). Given these challenges, from the late 19th century onward, a number of missionaries began experimenting with transcribing Chinese pronunciation using Roman letters (de Francis 1943). In the mid-1950s, building on previous efforts, Mainland China developed the Pinyin system, a phonetic transcription system of Mandarin Chinese using Roman letters to assist with learning and notating Chinese pronunciation (Scheme of the Chinese Phonetic Alphabet 1958). Since its implementation, Pinyin has become an indispensable tool for both native Chinese children and foreign learners of Chinese. However, there has been a lack of systematic research on the cognitive functions of this Chinese phonetic transcription system. That is, beyond its role in transcoding sounds and characters, the use of Pinyin may also have a profound impact on the processes of language production and comprehension. Learning Pinyin has been broadly shown to have a significant positive correlation with Chinese speaking and reading, not only for native Chinese children (Lin et al. 2010; Yin et al. 2011) but also for L2 Chinese learners (Lü 2017; Packard 1990; Xiao et al. 2020). The effect of Pinyin learning on L2 Chinese listening comprehension, however, has not been systematically explored. Therefore, the present longitudinal study aims to investigate the role of phonetic transcription skills (i.e., Pinyin transcription skills) in Chinese L2 acquisition, particularly its impact on listening comprehension skills. We examined two Pinyin transcription skills, sound-to-Pinyin transcription (transcribing sounds into Pinyin) and character-to-Pinyin transcription (transcribing characters into Pinyin) and looked at how these two skills predicted L2 Chinese listening comprehension performance both concurrently and longitudinally. Additionally, we explored whether their effects were mediated by character production skills and whether L2 proficiency modulated these effects. Our study not only aims to offer insights into Chinese-specific Pinyin teaching and L2 Chinese acquisition but also to shed more light on the important role of general phonetic transcription and phonological representation concretization in language development, the mechanisms of interaction between multimodal language forms (i.e., visual and auditory forms) in general language processing, and cross-modal abilities in a broader second language acquisition context.

1.1 | Chinese Phonetic Transcription System: Pinyin

The Chinese writing system (Chinese characters) is well-known for its use of ideograms, where the writing symbols are morphosyllabic-based and are more closely related to the meanings rather than the phonemes. For example, the Chinese character口, meaning mouth, pronounced as [kʰoʊ̯214], does not represent a combination of [kʰ] and [oʊ̯], but rather the shape of the mouth. Since such an ideographic writing system leads to difficulties in character reading for Chinese learners, Chinese Pinyin, as an alphabetic phonemic-based phonetic transcription system, was invented in 1958 in Mainland China to facilitate the learning of Chinese character pronunciation. Taking again the character口 as an example, its Pinyin spelling is /kǒu/, in which /k/, /ou/, and /ˇ/ correspond to [kʰ], [oʊ̯], and its lexical tone, respectively. Thus, the Pinyin system essentially has two functions: (a) provide a visual transcription system that can be mapped onto the Chinese phonological system more transparently than the original characters, and (b) act as a mediator linking Chinese orthographic forms to their phonological forms. Correspondingly, for individuals learning Chinese as a second language (L2) or foreign language (FL), the Pinyin system can aid in memorizing the pronunciation of spoken Chinese and learning how to read the written words.

1.2 | Pinyin Transcription Skills

The skill of Pinyin transcription is the skill of transcribing other forms of the Chinese language into the form of its visual phonetic representations (i.e., Pinyin) according to Pinyin's orthographic rules. It includes two aspects (Xiao et al. 2020): One is the sound-to-Pinyin transcription skill, that is, the skill to transcribe Chinese spoken words into Pinyin, and the other is the character-to-Pinyin transcription skill, that is, the skill to transcribe Chinese written words (i.e., in Chinese characters) into Pinyin. These two aspects reflect learners' abilities in sound-to-Pinyin mapping and character-to-Pinyin mapping, respectively. Compared to Pinyin reading skills (also called Pinyin letter knowledge), which reflect

Pinyin-to-sound mapping, and character writing skills, which reflect Pinyin-to-character mapping, these two Pinyin transcription skills are better suited as a measure for Pinyin proficiency as producing specific Pinyin letters reflects more fine-grained and accurate Pinyin knowledge (Xiao et al. 2020). In addition, Pinyin reading and character writing largely involve the mastery of motion control in articulation and writing movements, which do not directly reflect Pinyin knowledge itself.

1.3 | L2 Listening Comprehension

Some listening comprehension models are widely used in the field of second language acquisition. Field (2008a) has proposed a two-process model of listening comprehension: decoding and meaning-building. Decoding is the lower-level rapid recognition process of language units, while meaning-building is the higher-level process that establishes the literal meaning and beyond-literal meaning within a sociocultural context (Cheng et al. 2023). Similarly, Anderson's (1995, 2009) three-phase model divides listening comprehension into three phases, (a) perceptual processing: recognizing and segmenting language units from the speech stream, (b) parsing: constructing meaning of segmented language units, and (c) utilization: using world knowledge and context information in long-term memory to interpret the intended meaning. Anderson's model has also been regarded as corresponding to the three stages (acoustic-phonetic processor, parser, and conceptualizer) in Levelt's (1995) speech production model (Dörnyei and Kormos 1998).

L2 listeners who are less proficient than native listeners tend to experience significant difficulty with the decoding/perceptual processing phase (Field 2008b; Goh 2002). This is primarily because the spoken language is fast and contains many unclear pronunciations, leading to worse recognition, memorization, and segmentation of phonological information on the part of L2 listeners (Hasan 2000; Goh 2000). Additionally, since the links between phonological representations and concepts in L2 listener's mental lexicon are weaker than those of L1 listeners (Gollan and Acenas 2004; Gollan et al. 2008), it is also harder for them to access the meaning of spoken input in the meaning-building phase.

1.4 | Pinyin Transcription Skills and L2 Chinese Listening Comprehension

The lexical quality hypothesis (Perfetti and Hart 2002) provides a theoretical framework for understanding the potential mechanisms through which Pinyin skills affect language processing. This hypothesis emphasizes that the lexical-level knowledge in comprehension tasks is composed of interconnected orthographic, phonological, and semantic components. All three components must be of high quality to contribute to more coherent and reliable word identification, which in turn benefits the automatization of word processing and, ultimately, fluid comprehension (Perfetti et al. 2005; Perfetti and Hart 2002; Perfetti 2007, 2017). Moreover, lexical quality "allows learning"; that is, "new information about spelling, pronunciation, or meaning to an impoverished presentation" can be added to improve its quality (Perfetti 2017, 59). In the case of Chinese, as Chinese orthography does not directly represent sounds, learners may struggle to memorize its phonological knowledge, which can result in impoverished and lower-quality phonological representations of Chinese words. Therefore, Pinyin can be relied on as a visual-physical aid of phonological encoding and as added "new information about pronunciation" to substantiate and strengthen the lexical phonological component and the relationships between the phonological component and the orthographic and semantic components. This, in turn, enhances the interrelated trinity of the overall lexical quality and promotes more efficient and fluent language comprehension. Correspondingly, numerous empirical studies have demonstrated that Pinyin skills are positively correlated to L1/L2 Chinese phonological perception (Liu and Xu 2023), reading comprehension (e.g., Ma et al. 2020), writing (e.g., Zhang and Roberts 2021), and oral production (e.g., Packard 1990; Xu et al. 2023). Moreover, evidence from several cross-sectional and longitudinal studies has revealed that these Pinyin skills can predict students' L2 Chinese performance even 1–5 years later (Lin et al. 2010; Pan et al. 2011), suggesting that Pinyin learning might have a long-term effect. It has even been found that some benefits of Pinyin on L2 Chinese proficiency can only be long-term and not immediately apparent (Liu et al. 2020).

However, to the best of our knowledge, far fewer studies have focused on the relationship between Pinyin transcription skills and L2 Chinese listening comprehension. Theoretically, based on the lexical quality hypothesis and listening comprehension models, Pinyin transcription skills possibly affect L2 Chinese listening comprehension through two different pathways. One is by influencing the phonological encoding (or perceptual processing) phase, and the other by influencing the meaning-building (or parsing and utilization) phase.

In terms of their effects on the phonological encoding phase of listening comprehension, based on the lexical quality hypothesis, sound-to-Pinyin and character-to-Pinyin transcription skills visually reinforce the formation and memory of L2 Chinese phonological representations. Therefore, these skills may help learners develop higher-quality phonological components in the mental lexicon, thereby benefiting phonological encoding in listening comprehension. Furthermore, according to the perceptual assimilation model (PAM; Best and Tyler 2007), different L2 phonemes can in some cases be auditorily perceptually assimilated into a single phonetic category in the L1, making it difficult for learners to develop distinct phonological representations between these sounds. In this context, the sound-to-Pinyin transcription skill reflects the fine-grained ability to map the auditory phonemes to different written letters, emphasizing the distinctiveness of auditory phonemes in visual perception. It may therefore strengthen more refined and L1-category-independent phonological representations and thereby the abilities of phonological discrimination, categorization, and segmentation (Huang and Hanley 1997; Cheung et al. 2001), which are essential to the phonological encoding in listening comprehension (Vandergrift 2007; Li et al. 2012).

Although Pinyin transcription skills are more closely related to the phonological component in the mental lexicon, they may also influence the meaning-building phase of listening comprehension through the mediation of the character production skill (i.e., transcribing spoken Chinese into written characters, or

character dictation). As Chinese orthography carries less phonological information but is rich in semantic information, Chinese lexical quality has a stronger correlation between its orthographic component and its semantic component as compared to alphabetic languages (Perfetti, 2017). It has also been reported that proficient Chinese learners tend to rely more heavily on the direct connection between orthography and semantics (Lin and Zhang 2025; Zhou et al. 2018). Such a stronger orthography-semantics connection in Chinese not only strengthens its semantic activation through orthography but also aids higher lexical-semantic quality in listening comprehension by enhancing semantic precision and "minimizing confusion" (Perfetti and Hart 2002, 190), as the stricter mappings (i.e., more toward one-to-one mappings) between characters and meanings in Chinese help resolve ambiguities arising from the looser mappings (i.e., more toward one-to-many mappings) between sounds and meanings in Chinese. Thus, by mapping the sounds to orthographies, good character production skills may be able to promote the activation of orthographic information during listening processing, which thereby leads to more efficient and accurate semantic activation and benefits meaning-building during listening comprehension. Given that Pinyin is the bridge connecting sounds and characters, Pinyin skills may play an important role in the abilities of sound-to-character mapping/character production (Shen and Bear 2000; Zhang et al. 2020). Therefore, taken together, it is possible that Pinyin transcription skills favor character production and, in turn, promote meaning-building in listening comprehension.

Few studies have provided empirical evidence for the above theoretical pathways of Pinyin transcription skills' effects on L2 Chinese listening comprehension. Xiao et al. (2020) investigated Pinyin transcription skills' influence on L2 Chinese reading, suggesting that successful performance of sound-to-Pinyin transcription indicates successful categorization and retainment of fine-grained Chinese phonological representations, which facilitates listening comprehension and then reading comprehension. However, they did not find the effects of character-to-Pinyin and character production skills, which might only partially support our theoretical pathways, namely, the sound-to-Pinyin transcription skill's impact on the phonological encoding phase, but not other pathways. On the other hand, Zhang and Roberts (2021) have revealed that the Pinyin-invented spelling (i.e., spelling the Pinyin of pseudo-words) is a strong predictor of character production. Their findings imply the possibility of the mediation of character production in the pathway of sound-to-Pinyin transcription's effect on the meaning-building phase of listening comprehension. More recently, Yuan and Xu (2024) specifically focused on studying the relationship between Pinyin and listening comprehension. They found the sound-to-Pinyin transcription skill had a direct impact on listening comprehension, which occurred at elementary and intermediate levels, whereas the character-to-Pinyin transcription skill impacted listening comprehension in a character-production-mediated way, which only occurred at the intermediate level. No significant effects of Pinyin were found at the advanced level, possibly because advanced learners are rarely exposed to Pinyin at this stage in their language learning. These existing studies provide mixed empirical support for the role of Pinyin in listening comprehension, and they mostly validated only part of the potential pathways and did not systematically explore the full picture. Additionally, these studies considered only the immediate effects of Pinyin without addressing its long-term effects during the learners' L2 learning process.

2 | The Current Study

The current study aims to systematically discuss the relationship between Pinyin transcription skills and L2 Chinese listening comprehension. More importantly, given that the beneficial effects of Pinyin skills have been found to be long-lasting or delayed in Chinese reading (Lin et al. 2010) and oral performance (Liu et al. 2020), we also aim to investigate whether Pinyin transcription skills have longitudinal effects on L2 Chinese listening comprehension, which has never been investigated before.

Furthermore, it is noteworthy that people usually learn L2 Chinese relying on Pinyin more than Chinese characters at the beginning stage (Xu and Liu 2016), which may lead to different concurrent and longitudinal impacts of Pinyin skills on listening comprehension according to the different levels of L2 Chinese proficiency. First, beginners' knowledge of Chinese characters is still not solid, so their character production skills may not be sufficient to help access precise meanings during listening comprehension. Second, some studies have argued that Chinese characters' orthographic information can be more rapidly and strongly activated during spoken word processing in native speakers and highly experienced L2 learners than in beginners (Chen et al. 2019; Qu and Damian 2017). Hence, character production skills are probably more difficult to immediately strengthen online activation of orthographies and in turn help meaning-building processes in listening comprehension for learners with less learning experience. Third, sounds may also be less activated in character-to-Pinyin transcription if one does not have enough experience with character learning, resulting in the effect of character-to-Pinyin transcription skills on improving phonological encoding in listening comprehension being seen probably only in more experienced learners.

In summary, our research questions in the current study are (1) whether and how the two Pinyin transcription skills (sound-to-Pinyin and character-to-Pinyin) concurrently and longitudinally predict the performance of L2 Chinese listening comprehension and (2) to what extent these effects differ depending on L2 Chinese proficiency. Based on the previous theoretical and empirical investigations, we expect that both two Pinyin transcription skills would concurrently and longitudinally predict L2 Chinese listening comprehension, but the pathways would be different for different proficiency levels. Specifically, for the sound-to-Pinyin transcription skill, we expect that it would predict listening comprehension performance only directly (i.e., affecting phonological encoding phase) for less proficient learners, but both directly and indirectly mediated by character production (i.e., affecting both phonological encoding and meaning-building phases) for more proficient learners. As for the character-to-Pinyin transcription skill, it would predict listening comprehension performance both directly and as mediated by character production, but these would only appear in the more proficient group.

3 | Methods

3.1 | Participants

159 non-native Chinese speakers studying at Beijing Normal University were recruited for this study to test at two different times (T1 and T2); 39 of them were excluded because they did not complete all tasks at two times. The remaining 120 participants were divided into two groups: (1) the elementary group (i.e., the less proficient group), consisting of 60 participants who studied in the L2 Chinese elementary class (21 males, 39 females; M_{age} = 21.58, SD_{age} = 3.13, *range* = 17–37; number of months of Chinese study: M = 19.57, SD = 14.59, *range* = 1–72; native-language background: 30 from sinographic culture, 30 from non-sinographic culture) and (2) the intermediate group (i.e., the more proficient group), consisting of 60 participants who studied in the L2 Chinese intermediate class (16 males, 44 females; M_{age} = 21.92, SD_{age} = 3.06, *range* = 18–37; number of months of Chinese study: M = 43.71, SD = 35.69, *range* = 6–192; native-language background: 30 from sinographic culture, 30 from non-sinographic culture). The elementary and intermediate classes were divided based on students' L2 Chinese entrance exam scores (a custom-designed exam by Beijing Normal University), Chinese learning duration, and their HSK (Chinese Proficiency Test) levels (if they have one). Therefore, this division can effectively reflect the different Chinese proficiency levels of the participants.

3.2 | Measurements

3.2.1 | The Sound-to-Pinyin Transcription Skill

To measure participants' sound-to-Pinyin transcription skills both at T1 and T2, participants were asked to listen to the pre-recorded spoken materials (16 words and 4 sentences) twice and write down the corresponding Pinyin. 91 syllables were tested, involving 24 consonants (C), 33 vowels (V), and 5 tones, and including 6 types of syllable structures in Chinese (22 CV, 18 CVC, 20 CVV, 10 CVVC, 8 CVVV, and 12 syllables with no consonants). Taking into account the range of vocabulary that students had learned in their Chinese courses and the variety of word syllable structures, 27, 17, 34, 7, 5, and 1 words were selected from HSK Level 1 through Level 6, respectively. For each correct onset, rhyme, or tone, 1 point was given; otherwise, no point was given. The maximum score was 273. The reliabilities of this task in both T1 ($\alpha = 0.97$) and T2 ($\alpha = 0.97$) were strong.

3.2.2 | The Character-to-Pinyin Transcription Skill

To measure participants' character-to-Pinyin transcription skills both at T1 and T2, participants were asked to write down the corresponding Pinyin of 24 printed written words/phases. 54 syllables were tested, involving 22 consonants, 18 vowels, and 5 tones, and including 5 types of syllable structures in Chinese (20 CV, 13 CVC, 11 CVV, 4 CVVC, and 4 syllables with no consonant). 10, 5, 8, 9, 14, and 8 words were selected from HSK Level 1 through Level 6, respectively. To detect more fine-grained character-to-Pinyin skills, homonyms (e.g., 脑 and 恼, both pronounced as /nǎo/), polyphones (e.g., 长 can be pronounced either as /cháng/ or /zhǎng/), and characters with similar sounds (e.g., 请/qǐng/ and 情/qíng/) and orthographies (e.g., 人/rù/ and 人/rén/) were all included. For each correct onset, rhyme, or tone, 1 point was given; otherwise, no point was given. The maximum score was 162. The reliabilities of this task in both T1 ($\alpha = 0.97$) and T2 ($\alpha = 0.98$) were strong.

3.2.3 | The Character Production Skill

To measure participants' character production skills both at T1 and T2, participants were asked to listen to the pre-recorded spoken materials (12 words and 4 sentences) twice and write down the corresponding characters. 50 characters were tested, including homonyms and characters with similar sounds and orthographies. 24, 9, 11, 3, and 3 words were selected from HSK Level 1 through Level 5, respectively. For each correct character, 1 point was given; otherwise, no point was given. The maximum score was 50. The reliabilities of this task in both T1 ($\alpha = 0.88$) and T2 ($\alpha = 0.88$) were very good.

3.2.4 | Listening Comprehension

Listening comprehension performance was assessed by a task where participants were required to listen to some auditory materials (5 sentences and 1 short paragraph) and then answer 10 questions (multiple choice questions or true/false questions) according to the meaning of what they heard (adapted from Xiao et al. 2020). The auditory materials and the stems of multiple-choice questions were pre-recorded (played twice); thus, only the true/false questions and the answer options of multiple-choice questions were printed in Chinese characters on the test paper. The Pinyin of each answer option was also printed along with the corresponding characters (in case participants could not understand the characters). The listening materials and questions were different for elementary and intermediate groups to match their own levels and also different at T1 and T2 to avoid the repetition effect. The maximum score for each listening comprehension task was 10 (1 point for each correct answer). As there were only 10 questions in our listening comprehension task, the reliabilities of this task in T1 ($\alpha = 0.48$) and T2 ($\alpha = 0.50$) were not very strong but acceptable.

3.3 | Procedure

The elementary group and intermediate group were tested separately, for each of which participants were tested together in a quiet classroom. All recordings needed in the tasks were recorded by native Chinese speakers. The interval was 10–15 s between words and 15–20 s between sentences. The four tasks were tested twice. The first time (T1) was at the beginning of one academic semester and the second time (T2) was at the end of this semester (5 months after). Since students' L2 Chinese learning was typically organized on a semester basis and Pinyin skill, as a basic transcoding skill, is acquired quickly (typically 1–2 weeks of systematic teaching; Xiao et al. 2020), one semester was a suitable

interval to measure their learning progress. Each time of the test lasted 35–50 min, and participants were rewarded after each test.

3.4 | Data Analysis

Path models and cross-lagged panel models were constructed in Mplus 8.10 (Muthén and Muthén, 1998–2023) to investigate how Pinyin transcription skills affect L2 Chinese listening comprehension concurrently and longitudinally, respectively. Two Pinyin transcription skills were put in separate models to avoid multicollinearity due to their relatively high correlations (see Results). Each path model was built with three variables (i.e., sound-to-Pinyin/character-to-Pinyin transcription, listening comprehension, and character production) at each test time, with the direct "Pinyin transcription → listening comprehension" and mediated "Pinyin transcription → character production → listening comprehension" pathways tested. In cross-lagged models, based on a previous Pinyin-related cross-lagged modeling study (Zhang et al. 2020), we added sound-to-Pinyin/character-to-Pinyin transcription, listening comprehension, and character production at both T1 and T2 into the models and tested the direct "Pinyin transcription →T2 listening comprehension" and mediated "Pinyin transcription → T1 character production → T2 listening comprehension" pathways while taking into account all possible reciprocal and autoregressive pathways from T1 to T2, as well as all possible pathways that the predictors at one time, through the mediation of the concurrent character production, could affect variables at another time. We only tested the mediation of character production concurrent with the predictors, rather than the non-concurrent one or both, because this approach was the only way to achieve a good model fit. In addition, to avoid biases and ensure the accuracy of the results, age, native-language background, and Chinese learning duration were included in both path models and cross-lagged models as control variables because of their theoretical influences on L2 Chinese proficiency and their correlations with some other variables in our study (see Results). Since gender was not significantly correlated with any other variables in our study (*ps* > 0.05), we did not add it to the models.

The robust maximum likelihood (MLR) estimator was used, which allows adjusted standard errors and chi-square statistics (with Huber–White Sandwich estimator) and is robust to the non-normal distributed observations (Little and Rubin 2019). Following Hu and Bentler (1999) and Hair et al. (2010), the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR) were the indices to evaluate the fit of our models. CFI and TLI above 0.95, RMSEA below 0.06, and SRMR below 0.08 are considered a good fit. The two-tailed significance level was set at 0.05 for all tests of direct effects. To test the significance of mediated effects, 95% biased-corrected bootstrapped confidence intervals were recommended (Preacher and Hayes 2008; Shrout and Bolger 2002) and thus were estimated in our study with 1000 iterations using the maximum likelihood (ML) estimator. A 0-excluded 95% confidence interval (CI) indexes a significant mediated effect.

4 | Results

4.1 | Descriptive Statistics and Correlations

Table 1 shows the mean, standard deviation, minimum value, maximum value, skewness, and kurtosis for each task in the elementary and intermediate groups. Tables 2 and 3 show the correlations of all variables for these two groups separately. Pearson *r* was reported as the correlation coefficient except for the variables largely violating normal distribution (i.e., age, Chinese learning duration, and T2 listening comprehension in the intermediate group; Spearman's ρ was used for them) and the categorical variables (i.e., gender and native-language background; Kendall's τ was used for them).

From Tables 2 and 3, the scores of almost all tasks were significantly (*ps* < 0.05) or marginally significantly (*ps* < 0.1) correlated with each other. The exceptions were the correlations between T1 listening comprehension and T2 character-to-Pinyin transcription and T2 character production in the elementary group, as well as the correlations between T2 listening comprehension and T1 and T2 character-to-Pinyin transcription in the intermediate group.

4.2 | Path Models

Eight path models were established first for two Pinyin transcription tasks, elementary and intermediate groups, and at T1 and T2 separately, to examine the concurrent impacts of two Pinyin transcription skills on listening comprehension and whether character production played a role between them.

Figure 1a,b shows the path models of sound-to-Pinyin transcription skill's predicting effect on listening comprehension for the elementary and intermediate groups at T1 and T2, which had excellent fits. For the elementary group, at T1, MLR$\chi 2$ (12) = 27.19, $p = 0.007$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00; at T2, MLR$\chi 2$ (12) = 45.55, $p < 0.001$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00. We can see that the sound-to-Pinyin transcription score significantly and directly predicted listening comprehension at T1 ($\beta = 0.24$, $p = 0.02$). At T2, there was a marginally significant direct pathway from sound-to-Pinyin transcription to listening comprehension ($\beta = 0.22$, $p = 0.051$) and a significant pathway from sound-to-Pinyin transcription to character production ($\beta = 0.39$, $p = 0.004$) and from character production to listening comprehension ($\beta = 0.28$, $p = 0.04$). However, the results of bootstrapping showed that this character-production-mediated pathway was not significant ($\beta = 0.11$, 95%CI [0.00, 0.03]). For the intermediate group, at T1, MLR$\chi 2$ (12) = 53.44, $p < 0.001$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00; at T2, MLR$\chi 2$ (12) = 29.48, $p = 0.003$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00. The results only found sound-to-Pinyin transcription's prediction on listening comprehension directly ($\beta = 0.31$, $p = 0.01$) at T1. No other significant pathways toward listening comprehension were found.

Figure 1c,d shows the path models of character-to-Pinyin transcription skill's predicting effect on listening comprehension for the elementary and intermediate groups at T1 and T2, which also

TABLE 1 | Descriptive statistics of all tasks (accuracy percentage).

Group	Time	Task	Mean	SD	Min	Max	Skew	Kurt
Elementary	T1	Sound-to-Pinyin transcription	0.78	0.10	0.54	0.99	−0.2	−0.38
		Character-to-Pinyin transcription	0.62	0.13	0.33	0.88	−0.27	−0.91
		Character production	0.69	0.14	0.36	0.96	−0.45	−0.04
		Listening comprehension	0.67	0.17	0.30	1.00	−0.08	−0.49
	T2	Sound-to-Pinyin transcription	0.81	0.09	0.58	0.99	−0.38	−0.28
		Character-to-Pinyin transcription	0.72	0.13	0.42	0.91	−0.67	−0.65
		Character production	0.78	0.14	0.32	1.00	−0.98	1.25
		Listening comprehension	0.75	0.22	0.20	1.00	−0.62	−0.54
Intermediate	T1	Sound-to-Pinyin transcription	0.89	0.08	0.69	1.00	−0.71	−0.5
		Character-to-Pinyin transcription	0.82	0.10	0.56	0.99	−0.3	−0.53
		Character production	0.86	0.09	0.60	1.00	−0.73	0.09
		Listening comprehension	0.85	0.13	0.50	1.00	−0.59	−0.47
	T2	Sound-to-Pinyin transcription	0.91	0.07	0.72	1.00	−0.71	−0.1
		Character-to-Pinyin transcription	0.86	0.08	0.63	0.99	−0.6	0.16
		Character production	0.91	0.07	0.72	1.00	−0.89	0.13
		Listening comprehension	0.93	0.13	0.40	1.00	−1.86	3.24

TABLE 2 | Correlation coefficients between all variables in the elementary group.

	1	2	3	4	5	6	7	8	9	10	11
1 Age	—										
2 Gender	−0.06	—									
3 Native-language background	0.09	−0.04	—								
4 Chinese learning duration	0.01	0.05	0.34**	—							
5 T1 sound-to-Pinyin transcription	0.004	−0.11	−0.04	0.06	—						
6 T1 character-to-Pinyin transcription	−0.05	0.13	−0.37***	0.11	0.32*	—					
7 T1 character production	0.08	0.21†	−0.12	0.05	0.25†	0.64***	—				
8 T1 listening comprehension	0.13	−0.15	−0.03	0.11	0.33*	0.26*	0.35**	—			
9 T2 sound-to-Pinyin transcription	−0.16	−0.004	−0.08	−0.06	0.64***	0.50***	0.44***	0.31*	—		
10 T2 character-to-Pinyin transcription	0.01	0.05	−0.36***	−0.001	0.26*	0.79***	0.66***	0.20	.45***	—	
11 T2 character production	−0.12	0.09	−0.17	0.08	0.29*	0.57***	0.74***	0.17	0.43***	0.71***	—
12 T2 listening comprehension	0.02	0.16	−0.20†	0.05	0.26*	0.46***	0.45***	0.27*	0.34**	0.48***	0.37**

†$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

had excellent fits. For the elementary group, at T1, MLR$\chi 2$ (12) = 68.25, $p < 0.001$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00; at T2, MLR$\chi 2$ (12) = 100.14, $p < 0.001$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00. The model reflected that character-to-Pinyin transcription significantly predicted listening comprehension via character production at T1 (from character-to-Pinyin transcription to character production: $\beta = 0.72$, $p < 0.001$; from character production to listening comprehension: $\beta = 0.33$, $p = 0.03$), whereas predicted listening comprehension directly at T2 ($\beta = 0.35$, $p = 0.049$). The bootstrapping test demonstrated that the mediation of character production at T1 was significant ($\beta = 0.24$, 95% CI [0.001, 0.05]). For the intermediate group, at T1, MLR$\chi 2$ (12) = 67.53, $p < 0.001$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00; at T2, MLR$\chi 2$ (12) = 29.32, $p = 0.004$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00, but no significant pathways toward listening comprehension were found.

4.3 | Cross-lagged Models

Four cross-lagged models were next established to see how the two Pinyin transcription skills predict listening comprehension longitudinally and whether character production still mediates their effects. Figure 2a shows the cross-lagged model of the sound-to-Pinyin transcription skill in the elementary group. The model fit was excellent, MLR$\chi 2$ (33) = 165.75, $p < 0.001$, CFI = 1.00,

TABLE 3 | Correlation coefficients between all variables in the intermediate group.

	1	2	3	4	5	6	7	8	9	10	11
1 Age	—										
2 Gender	−0.22†	—									
3 Native-language background	0.29*	−0.08	—								
4 Chinese learning duration	0.24†	0.04	0.08	—							
5 T1 sound-to-Pinyin transcription	−0.10	0.19†	0.004	0.09	—						
6 T1 character-to-Pinyin transcription	−0.21	0.04	−0.14	−0.11	0.50***	—					
7 T1 character production	−0.13	0.05	−0.07	−0.03	0.50***	0.63***	—				
8 T1 listening comprehension	−0.33*	0.06	−0.34**	0.05	0.39**	0.31*	0.35**	—			
9 T2 sound-to-Pinyin transcription	−0.16	0.14	0.01	0.06	0.72***	0.50***	0.57***	0.45***	—		
10 T2 character-to-Pinyin transcription	−0.23†	−0.02	−0.25*	−0.15	0.40**	0.76***	0.58***	0.29*	0.52***	—	
11 T2 character production	−0.14	0.18	−0.21†	0.07	0.37**	0.42***	0.65***	0.32*	0.36**	0.47***	—
12 T2 listening comprehension	−0.10	0.06	−0.10	0.01	0.30**	0.09	0.25†	0.23†	0.23†	0.13	0.22†

†$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

TLI = 1.00, RMSEA = 0.00, SRMR = 0.01. However, T1 sound-to-Pinyin transcription did not predict T2 listening comprehension in any way. Figure 2b shows the cross-lagged model of the sound-to-Pinyin transcription skill in the intermediate group. The model fit was excellent, MLR$\chi 2$ (33) = 145.70, $p < 0.001$, CFI = 1.00, TLI = 1.27, RMSEA = 0.00, SRMR = 0.003, where T2 listening comprehension was not found to be predictable by T1 sound-to-Pinyin transcription either. Figure 2c shows the cross-lagged model of the character-to-Pinyin transcription skill in the elementary group. The model fit was excellent; MLR$\chi 2$ (33) = 247.84, $p < 0.001$, CFI = 1.00, TLI = 1.00, RMSEA = 0.02, SRMR = 0.01. The delayed direct prediction of T1 character-to-Pinyin transcription on T2 listening comprehension was significant ($\beta = 0.24$, $p = 0.047$). Although the path from T1 character-to-Pinyin transcription to T1 character production was significant ($\beta = 0.67$, $p < 0.001$) and the path from T1 character production to T2 listening comprehension was marginally significant ($\beta = 0.24$, $p = 0.08$), no mediation effect was confirmed by bootstrapping tests ($\beta = 0.16$, 95% CI [−0.01, 0.04]). Figure 2d shows the cross-lagged model of the character-to-Pinyin transcription skill in the intermediate group. The model fit was also excellent, MLR$\chi 2$ (33) = 146.08, $p < 0.001$, CFI = 1.00, TLI = 1.26, RMSEA = 0.00, SRMR = 0.004. In this model, the pathways from T1 character-to-Pinyin transcription to T1 character production ($\beta = 0.59$, $p < 0.001$) and from T1 character production to T2 listening comprehension ($\beta = 0.41$, $p = 0.03$) were significant, in which the mediating effect of T1 character production was also manifested to be significant by the bootstrapping test ($\beta = 0.24$, 95% CI [0.002, 0.05]).

5 | Discussion

The current study investigated whether Pinyin transcription skills are predictors of L2 Chinese listening comprehension concurrently and longitudinally and whether their predictions differ according to L2 Chinese proficiency. Our results revealed that both Pinyin transcription skills were able to predict listening comprehension performance in general, supporting the lexical quality hypothesis (Perfetti and Hart 2002), in which Pinyin, as a reinforcement of the phonological component of the mental lexicon, benefits the interrelated trinity of lexical quality and the improvement of language comprehension.

5.1 | Two Pinyin Transcription Skills

The core cognitive function of Pinyin learning is to facilitate the processing of Pinyin transcription (Xu et al. 2023). This current study measured two aspects of Pinyin transcription skills of L2 Chinese learners, sound-to-Pinyin transcription and character-to-Pinyin transcription, which correspond to the abilities to transcode spoken and written forms of language into visual phonetic symbols, respectively. Overall, our study confirmed the positive effects of Pinyin, as the visual enhancement of phonological information, on auditory information processing. These effects, combined with previously found effects of Pinyin on auditory phonological perception (Liu and Xu 2023), visual reading (Xiao et al. 2020), and oral production (Xu et al. 2023), highlight the role of phonetic transcription skills in the cross-modal interaction and integration mechanisms of language processing and emphasize its significance as an intrinsic factor in language development. Moreover, our results showed that two Pinyin transcription skills had different patterns of effects on L2 Chinese comprehension, which was consistent with the findings in previous research (Yuan and Xu 2024; Xiao et al. 2020). This suggests that two Pinyin transcription skills reflect different auditory/visual information decoding processes and possibly different activation of relevant language representations (Ziegler and Muneaux 2007), which also supports the rationality of dividing Pinyin transcription skills into these two sub-skills in our study. At the same time, the significant correlations between two Pinyin transcription skills indicated that both of them involve the encoding of visual phonetic representations and reflect the ability of Pinyin transcription.

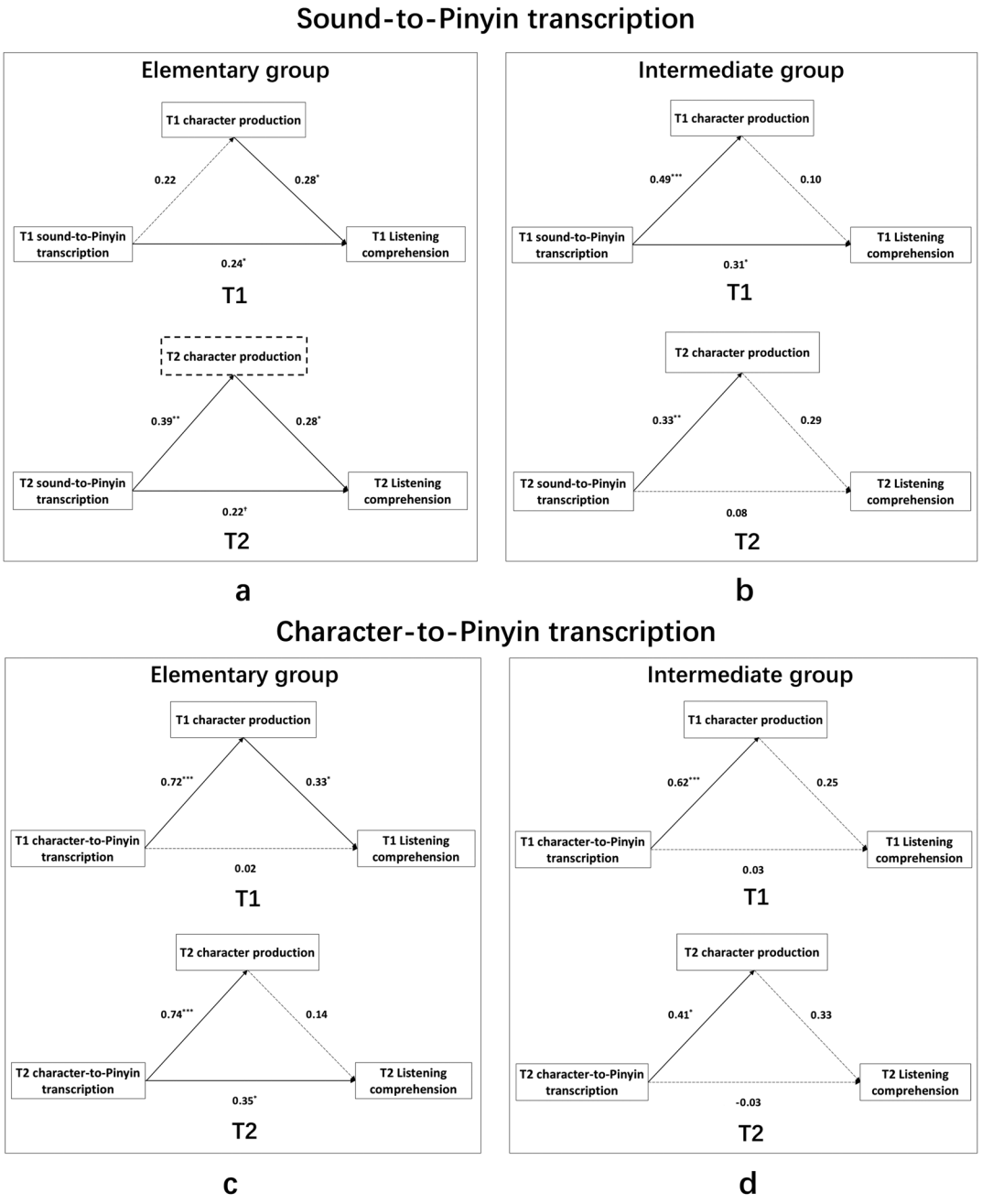

FIGURE 1 | Path models of the effects of Pinyin transcription skills on listening comprehension.

5.2 | The Sound-to-Pinyin Transcription Skill and Listening Comprehension

The sound-to-Pinyin transcription skill reflects the skill of mapping auditory pronunciations to the alphabetic Pinyin system. We hypothesized that this skill would predict L2 Chinese listening comprehension directly (i.e., aid phonological encoding in listening) and as mediated by the character production skill (i.e., aid meaning-building in listening). However, considering that elementary learners may be not familiar with the character system, we hypothesized this mediated effect would only exist in our intermediate group.

Partially in line with our hypothesis, in the path models, the direct predicting effects were found in both elementary and intermediate groups, but no indirect mediated effects were found in any of them. These direct effects confirm the positive impacts of Pinyin on phonological encoding in listening processing, which is in line with Yuan and Xu's (2024) and Xiao et al.'s (2020) findings. Indeed, the acoustic differences between phonemes are sometimes difficult for L2 learners to discern. However, if we transcode the phonemes into alphabets that they are more familiar with, the differences in the letter's shape will be more recognizable for such learners and may remind them of the corresponding phonological categories. Furthermore, some

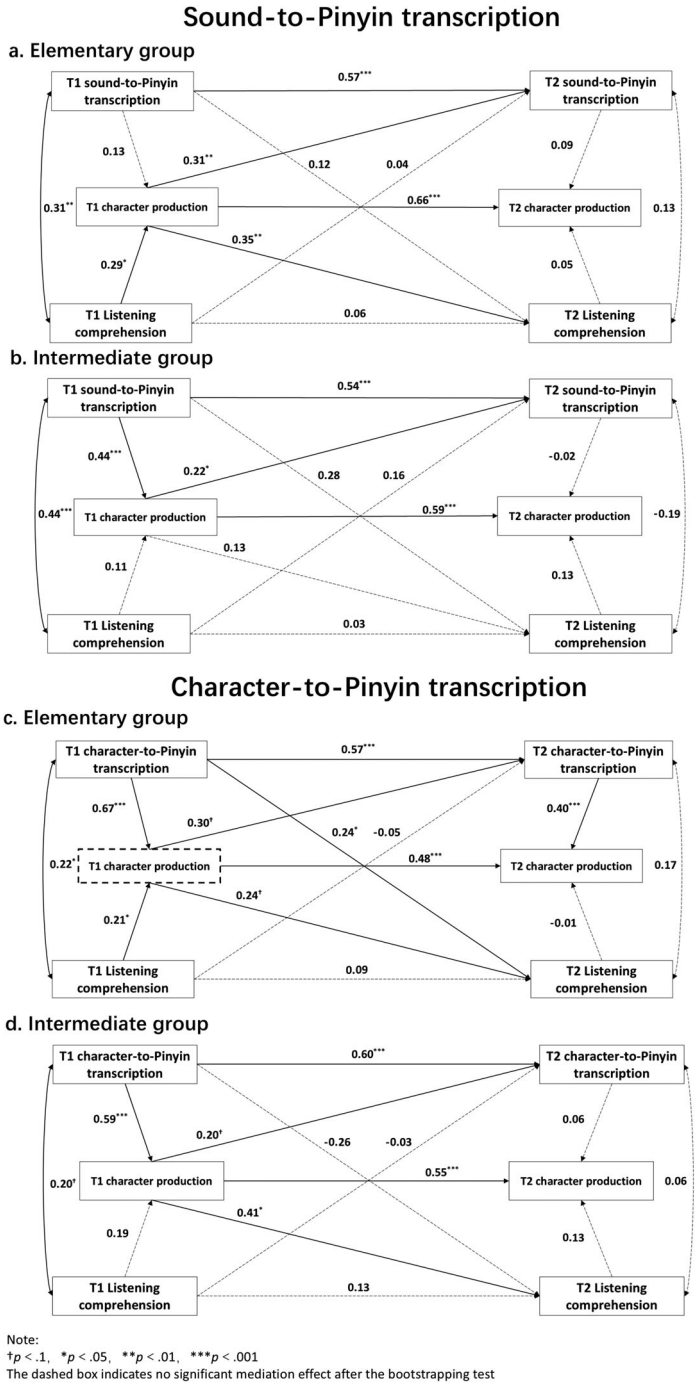

FIGURE 2 | Cross-lagged models of the effects of Pinyin transcription skills on listening comprehension.

psychological evidence also suggests that auditory information is more difficult to reserve, review, and retain in long-term memory than visual information (Cohen et al. 2009; Lindner et al. 2009). Thus, "writing the sound down" can be an effective strategy for remembering the individual sounds and their differences, which is useful to the phonological encoding phase during listening comprehension.

The nonexistence of indirect mediated effects suggested that the sound-to-Pinyin transcription skill may not influence the meaning-building of listening comprehension via the mediation of character production skills. From the concurrent path models, we can see that sound-to-Pinyin transcription significantly predicted character production for the elementary group at T2 and for the intermediate group at both T1 and T2, suggesting that sound-to-Pinyin transcription can consistently enhance character production after learners have accumulated a period of learning experience. However, the relationships between character production and listening comprehension were neither strong nor stable. This is probably due to the difficulty of making good use of character skills in listening comprehension. This may also be due to the similarity between the sound-to-Pinyin transcription task and the character production task, both of which are auditory processing and character production may

also involve the sound-to-Pinyin process. Thus, the effects from these two predictors may be hard to distinguish clearly, and some effects from character production may be diluted by sound-to-Pinyin transcription. As a result, no reliable and robust character-production-mediated paths were readily visible here.

Considering both path models and cross-lagged models, it is worth mentioning that the direct effects of sound-to-Pinyin transcription all appeared immediately. No longitudinal predictions were found. Such straightforward effects may be because the phonological processing in listening comprehension is foundational, so better phonological abilities resulting from better sound-to-Pinyin transcription skills could be immediately used to improve listening. Note that these concurrent effects were significantly or marginally significantly detected in almost all tests except for the intermediate group at T2. This may be due to the non-normal distribution of their T2 listening comprehension scores (although we already used the analysis methods robust to non-normal data). Upon inspecting the data, we also found many participants scored full marks, once again indicating its skewness and lack of sufficient differentiation. We would expect that the effect could be found if the listening comprehension task had higher reliability and a more reasonable distribution of scores.

Comparing two proficiency levels, the results did not differ greatly; that is, sound-to-Pinyin transcription skill predicted listening comprehension directly and immediately for both elementary and intermediate learners at T1 or/and T2. This pattern of results reveals that the facilitation of sound-to-Pinyin transcription skills to the phonological processing in listening comprehension exists across different stages of L2 Chinese learning.

5.3 | The Character-to-Pinyin Transcription Skill and Listening Comprehension

The character-to-Pinyin transcription skill involves more character knowledge. Thus, we hypothesized that such a skill may only have an impact in the intermediate group, both directly on phonological processing and as mediated by character production on meaning-building processing of listening comprehension. The results, however, were not exactly the same as what we expected.

First, the direct effects of character-to-Pinyin transcription skills on listening comprehension were shown in the elementary group, which was delayed at T1 but was immediate at T2. No direct effects were found in the intermediate group. There are two possible explanations. One is that since elementary learners were not familiar with character reading, to help them better spell Pinyin, they may tend to explicitly activate the phonological representations when they read characters in the character-to-Pinyin transcription task. Such activation provides learners with more "training" on practicing phonological knowledge, which can be useful for processing phonological information during listening comprehension. However, due to their unfamiliarity with Chinese characters at the beginning of their Chinese study (i.e., T1), they could not activate phonological representation efficiently at T1, and as a result, the effect of character-to-Pinyin transcription skills had a period of delay. This is probably also why no direct effects of character-to-Pinyin transcription skills for elementary learners were found in Yuan and Xu's (2024) study, where only immediate effects were tested. On the contrary, intermediate learners had a richer experience and more training in character recognition and Chinese reading comprehension. Given that advanced Chinese reading comprehension requires the ability to quickly map characters' orthography to their meanings, the phonological activation may be weakened when they see the characters, and thus the connection between their character-to-Pinyin transcription skills and phonological processing in listening comprehension may not be strong either. This explanation could be supported by the results of some studies showing that more experienced L2 Chinese learners' phonological activation during their Chinese reading processing is weaker than that of beginners (Chen and Shu 2001; Zhou et al. 2018).

The other possible explanation is that such direct effects of character-to-Pinyin transcription skills may not be actually relevant to the phonological encoding involved in listening comprehension. Given that one Chinese character may have different pronunciations according to its different meanings, knowing characters' meanings helps to determine their pronunciations. Thus, the difference in character-to-Pinyin transcription scores might be related to participants' different levels of knowledge about what these characters mean, which could actually reflect their vocabulary knowledge levels. As some previous studies have claimed that vocabulary knowledge significantly predicts or is related to L2 listening comprehension performance (Matthews 2018; Xiao et al. 2020), it is likely that vocabulary knowledge plays a role between character-to-Pinyin transcription skills and listening comprehension in our study. Considering that vocabulary usually grows more rapidly at the beginning stage of L2 learning and it might take time from acquiring vocabulary knowledge to applying vocabulary knowledge to listening comprehension, this effect could be more obvious in the elementary group but have a period of delay. We did not take vocabulary knowledge into account in our current study, so we cannot decide which explanation is more plausible. Future research may add vocabulary knowledge as a mediating variable into the models to better understand the effect of the character-to-Pinyin transcription skill.

We now focus on the mediating pathways from character-to-Pinyin transcription skills to listening comprehension, which were observed immediately in the elementary group at T1 and with a delay in the intermediate group. The existence of these character-production–mediated effects supports our above hypothesis that Pinyin transcription skills enhance character production and, in turn, help learners to better access precise meanings during listening comprehension. However, it is necessary to explain why the mediation effect appeared both at the beginning stage of L2 Chinese learning and at the later stage with a lag. As we proposed theoretically, we believe that taking better advantage of the benefits of character production skills in listening comprehension is an advanced skill that requires relatively high proficiency in Chinese characters and rapid activation of character representation in the fleeting sound stream. Therefore, learners need to accumulate sufficient character knowledge, but even so, it will still take some time before

such knowledge can be truly applied in the listening processing and contribute to listening comprehension. That is possibly why we saw the lagged effect from the more proficient intermediate group. The reason that such an effect can be immediately shown at T1 in the elementary group is probably their sensibility to characters at the beginning of character learning. Additionally, as the learning curve of character learning is relatively steep, it is also possible that the learners with high character production scores at the beginning stage were generally those who were diligent in studying, so their listening comprehension scores were correspondingly higher. To have a more convincing conclusion, this mediation effect is worthy of investigation in the future to see whether the pattern of our results can be replicated.

6 | Implications and limitations

From the theoretical perspective, our findings support the lexical quality hypothesis, providing the evidence that Pinyin transcription skills enhanced L2 Chinese comprehension both in the short and the long term through enhancing the quality of phonological representation and the quality of the interactions between phonological, orthographic, and semantic representations. We also shed more light on the mechanism underlying L2 development, highlighting the broader significance of the phonetic transcription, that is, the concretization and visual symbolization of phonological representations in L2 development. While previous research has primarily focused on the facilitative role of phonetic transcription in visual reading processes, our study extended this understanding by demonstrating its impact on phonological encoding and meaning-building in auditory listening comprehension. This also underscores the interaction between different modalities of language abilities. From the practical perspective, our results indicate the necessity of learning and teaching Pinyin or phonetic spelling in educational practice in L2 Chinese or other languages. In practice, some people may claim that Pinyin learning is only necessary at the beginning of L2 Chinese teaching and should be reduced at the higher levels. However, our results underline the importance of Pinyin learning for even intermediate learners from the perspective of its positive impact on listening comprehension performance. Our results also showed that the benefits of Pinyin learning may not be immediately apparent, requiring sustained practice and cumulative effort over time. Therefore, teachers should not be eager for quick results. Laying a solid foundation for students in Pinyin knowledge will help improve their language ability in the long run.

The current study has several limitations. First, to prevent an excessively lengthy testing session, the listening comprehension test used in the study contained a relatively limited number of items, which may have affected its reliability. In addition, while we identified a predictive role of Pinyin transcription skills on listening comprehension, the deeper mechanisms underlying this relationship are worth further clarifying. For instance, it remains unclear whether Pinyin's influence on listening comprehension works through its impact on more fundamental metalinguistic skills in listening comprehension, such as phonological awareness and vocabulary knowledge. Further research could explore this aspect to provide a more comprehensive understanding.

7 | Conclusion

The current study investigated whether two Chinese Pinyin transcription skills have concurrent or longitudinal predictions on L2 Chinese listening comprehension performance and whether their effects vary based on learners' Chinese proficiency levels. Our findings showed direct and concurrent predictive effects of the sound-to-Pinyin transcription skill on the phonological-encoding stage of listening comprehension for both elementary and intermediate learners. We also found the predictive effects of the character-to-Pinyin transcription skill, which directly acted on listening comprehension (the phonological-encoding stage) with some delay at the begining stage for elementary learners, while indirectly acting on listening comprehension (the meaning-building stage) with the mediation of character production skills without delay for elementary learners at the beginning stage, but with a delay for intermediate learners. Our study highlighted the benefits of learning phonetic transcription systems for L2 development. Some of these benefits are immediate, while others may lay the foundation for long-term improvement in L2 proficiency.

Acknowledgments

This work was supported by National Social Science Fund of China under Grant number 19FYYB038; International Society for Chinese Language Teaching under Grant number SH23Y29; Academic Support Project for 2024 at School of International Chinese Language Education at Beijing Normal University. We would like to thank all the participants for their involvement in this study.

Ethics Statement

This study was approved by the relevant research ethics committee at Beijing Normal University. All participants signed the consent form before participating in the study.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/ijal.12741

References

Anderson, J. R. 1995. *Cognitive Psychology and Its Implications*. 4th ed. Freeman.

Anderson, J. R. 2009. *Cognitive Psychology and Its Implications*. 7th ed. Worth Publishers.

Best, C. T., and M. D. Tyler. 2007. "Nonnative and Second-Language Speech Perception: Commonalities and Complementarities." In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro, 13–34. John Benjamins. https://doi.org/10.1075/lllt.17.07bes.

Chen, H. C., and H. Shu. 2001. "Lexical Activation During the Recognition of Chinese Characters: Evidence Against Early Phonological Activation."

Psychonomic Bulletin & Review 8, no. 3: 511–518. https://doi.org/10.3758/bf03196186.

Chen, L., C. Perfetti, and Y. Leng. 2019. "Reading Pinyin Activates Character Orthography for Highly Experienced Learners of Chinese." *Bilingualism: Language and Cognition* 22, no. 1: 103–111. https://doi.org/10.1017/S136672891700058X.

Cheng, J., J. Matthews, K. Lange, and S. McLean. 2023. "Aural Single-Word and Aural Phrasal Verb Knowledge and Their Relationships to L2 Listening Comprehension." *TESOL Quarterly* 57, no. 1: 213–241. https://doi.org/10.1002/tesq.3137.

Cheung, H., H.-C. Chen, C. Y. Lai, O. C. Wong, and M. Hills. 2001. "The Development of Phonological Awareness: Effects of Spoken Language Experience and Orthography." *Cognition* 81, no. 3: 227–241. https://doi.org/10.1016/S0010-0277(01)00136-6.

Cohen, M. A., T. S. Horowitz, and J. M. Wolfe. 2009. "Auditory Recognition Memory Is Inferior to Visual Recognition Memory." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 14: 6008–6010. https://doi.org/10.1073/pnas.0811884106.

de Francis, J. 1943. "The Alphabetization of Chinese." *Journal of the American Oriental Society* 63, no. 4: 225. https://doi.org/10.2307/594356.

Dörnyei, Z., and J. Kormos. 1998. "Problem-Solving Mechanisms in L2 Communication: A Psycholinguistic Perspective." *Studies in Second Language Acquisition* 20, no. 3: 349–385. https://doi.org/10.1017/S0272263198003039.

Field, J. 2008a. *Listening in the Language Classroom*. Cambridge University Press. http://doi.org/10.1017/CBO9780511575945.

Field, J. 2008b. "Revising Segmentation Hypotheses in First and Second Language Listening." *System* 36: 35–51. http://doi.org/10.1016/j.system.2007.10.003.

Goh, C. C. 2000. "A Cognitive Perspective on Language Learners' Listening Comprehension Problems." *System* 28, no. 1: 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3.

Goh, C. C. 2002. "Exploring Listening Comprehension Tactics and Their Interaction Patterns." *System, 30* 185–206. https://doi.org/10.1016/S0346-251X(02)00004-0.

Gollan, T. H., and L. A. Acenas. 2004. "What Is a TOT? Cognate and Translation Effects on Tip-of-the-Tongue States in Spanish-English and Tagalog-English Bilinguals." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 30, no. 1: 246–269. https://doi.org/10.1037/0278-7393.30.1.246.

Gollan, T. H., R. I. Montoya, C. Cera, and T. C. Sandoval. 2008. "More Use Almost Always Means a Smaller Frequency Effect: Aging, Bilingualism, and the Weaker Links Hypothesis." *Journal of Memory and Language* 58, no. 3: 787–814. https://doi.org/10.1016/j.jml.2007.07.001.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. 2010. *Multivariate Data Analysis*. 7th ed. Pearson Prentice Hall.

Hasan, A. S. 2000. "Learners' Perceptions of Listening Comprehension Problems." *Language Culture and Curriculum* 13, no. 2: 137–153. https://doi.org/10.1080/07908310008666595.

Hu, L., and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling* 6, no. 1: 1–55. https://doi.org/10.1080/10705519909540118.

Huang, H. S., and J. R. Hanley. 1997. "A Longitudinal Study of Phonological Awareness, Visual Skills, and Chinese Reading Acquisition Among First-Graders in Taiwan." *International Journal of Behavioral Development* 20, no. 2: 249–268. https://doi.org/10.1080/016502597385324.

Levelt, W. 1995. "The Ability to Speak: From Intentions to Spoken Words." *European Review* 3, no. 1: 13–23. https://doi.org/10.1017/S1062798700001290.

Li, M., L. Cheng, and J. R. Kirby. 2012. "Phonological Awareness and Listening Comprehension Among Chinese English-Immersion Students." *International Education* 41, no. 2: 4. https://trace.tennessee.edu/cgi/viewcontent.cgi?article=1225&context=internationaleducation.

Lin, D., C. McBride-Chang, H. Shu, et al. 2010. "Small Wins Big: Analytic Pinyin Skills Promote Chinese Word Reading." *Psychological Science* 21, no. 8: 1117–1122. https://doi.org/10.1177/0956797610375447.

Lin, J., and H. Zhang. 2025. "How Phonological and Orthographic Decoding Complicates the Simple View of Reading in Chinese: Examining Mediation Through Listening comprehension." *Cognitive Processing* 26: 133–148. https://doi.org/10.1007/s10339-023-01143-3.

Lindner, K., G. Blosser, and K. Cunigan. 2009. "Visual Versus Auditory Learning and Memory Recall Performance on Short-Term Versus Long-Term Tests." *Modern Psychological Studies* 15, no. 1: 6. https://scholar.utc.edu/mps/vol15/iss1/6.

Little, R. J., and D. B. Rubin. 2019. *Statistical Analysis With Missing Data*. Vol. 793. John Wiley & Sons.

Liu, G. Z., C. H. Xu, J. Q. Zhang, and H. Shu. 2020. "A Study on the Influence of Chinese Pinyin Learning on American Elementary School Students' English Phonological Awareness." *TCSOL Studies* 4: 49–57. https://doi.org/10.3969/j.issn.1674-8174.2020.04.006.

Liu, L., and C. Xu. 2023. "The Role of Pinyin Letter Knowledge in Arabic Speakers' Perception and Production of Chinese Consonants." *Chinese Language Learning* 06: 91–102. https://oversea.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&filename=HYXX202306010&dbname=CJFDLAST2023.

Lü, C. 2017. "The Roles of Pinyin Skill in English-Chinese Biliteracy Learning: Evidence From Chinese Immersion Learners." *Foreign Language Annals* 50: 306–322. https://doi.org/10.1111/flan.12269.

Ma, S., X. Zhang, H. Hatfield, and W. H. Wei. 2020. "Pinyin Is an Effective Proxy for Early Screening for Mandarin-Speaking Children at Risk of Reading Disorders." *Frontiers in Psychology* 11: 327. https://doi.org/10.3389/fpsyg.2020.00327.

Matthews, J. 2018. "Vocabulary for Listening: Emerging Evidence for High and Mid-Frequency Vocabulary Knowledge." *System* 72: 23–36. https://doi.org/10.1016/j.system.2017.10.005.

McBride, C. A. 2016. "Is Chinese Special? Four Aspects of Chinese Literacy Acquisition That Might Distinguish Learning Chinese From Learning Alphabetic Orthographies." *Educational Psychology Review* 28, no. 3: 523–549. https://doi.org/10.1007/s10648-015-9318-2.

Muthén, L. K., and B. O. Muthén. 1998-2023. *Mplus User's Guide*. 8th ed. Muthén & Muthén.

Packard, J. L. 1990. "Effects of Time Lag in the Introduction of Characters Into the Chinese Language Curriculum." *Modern Language Journal* 74, no. 2: 167–175. https://doi.org/10.2307/328120.

Pan, J., C. McBride-Chang, H. Shu, H. Liu, Y. Zhang, and H. Li. 2011. "What Is in the Naming? A 5-Year Longitudinal Study of Early Rapid Naming and Phonological Sensitivity in Relation to Subsequent Reading Skills in both Native Chinese and English as a Second Language." *Journal of Educational Psychology* 103, no. 4: 897–908. https://doi.org/10.1037/a0024344.

Perfetti, C. A. 2017. "Lexical Quality Revisited. Developmental Perspectives in Written Language and Literacy: In Honor of Ludo Verhoeven." 51–67. https://doi.org/10.1075/z.206.04per.

Perfetti, C. A., and L. Hart. 2002. "The Lexical Quality Hypothesis." In *Precursors of Functional Literacy*, edited by L. Vehoeven, C. Elbro, and P. Reitsma, 189–213. John Benjamins. https://doi.org/10.1075/swll.11.14per.

Perfetti, C. A., Y. Liu, and L. H. Tan. 2005. "The Lexical Constituency Model: Some Implications of Research on Chinese for General Theories of Reading." *Psychological Review* 112, no. 1: 43–59. https://doi.org/10.1037/0033-295X.112.1.43.

Perfetti, C. 2007. "Reading Ability: Lexical Quality to Comprehension." *Scientific Studies of Reading* 11, no. 4: 357–383. https://doi.org/10.1080/10888430701530730.

Preacher, K. J., and A. F. Hayes. 2008. "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models." *Behavior Research Methods* 40, no. 3: 879–891. https://doi.org/10.3758/brm.40.3.879.

Qu, Q., and M. F. Damian. 2017. "Orthographic Effects in Spoken Word Recognition: Evidence From Chinese." *Psychonomic Bulletin & Review* 24, no. 3: 901–906. https://doi.org/10.3758/s13423-016-1164-9.

Scheme of the Chinese Phonetic Alphabet. 1958. Ministry of Education of the People's Republic of China. http://www.moe.gov.cn/jyb_sjzl/ziliao/A19/195802/t19580201_186000.html.

Shen, H. H., and D. R. Bear. 2000. "Development of Orthographic Skills in Chinese Children." *Reading and Writing* 13: 197–236. https://doi.org/10.1023/A:1026484207650.

Shrout, P. E., and N. Bolger. 2002. "Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations." *Psychological Methods* 7, no. 4: 422–445. https://doi.org/10.1037//1082-989X.7.4.422.

Vandergrift, L. 2007. "Recent Developments in Second and Foreign Language Listening Comprehension Research." *Language Teaching* 40, no. 3: 191–210. https://doi.org/10.1017/S0261444807004338.

Xiao, H., C. Xu, and H. Rusamy. 2020. "Pinyin Spelling Promotes Reading Abilities of Adolescents Learning Chinese as a Foreign Language: Evidence From Mediation Models." *Frontiers in Psychology* 11: 596680. https://doi.org/10.3389/fpsyg.2020.596680.

Xu, C., and J. Liu. 2016. "A Brief Discussion on Chinese Pinyin and Chinese Teaching for Overseas Children." *Journal of Yunnan Normal University (Teach Res CFL)* 14, no. 05: 1–6. https://doi.org/10.16802/j.cnki.ynsddw.2016.05.002.

Xu, C., Y. Shu, D. Hong, and Q. Li. 2023. "Effects of Pinyin Learning on Chinese Oral Word Acquisition in Spanish Middle School Students." *Frontiers of International Chinese Language Education* no. 01: 27–40+208–209. https://oversea.cnki.net/kcms/detail/detail.aspx?dbcode=CCJD&filename=DWHJ202301004&dbname=CCJDLAST2.

Yin, L., W. Li, X. Chen, et al. 2011. "The Role of Tone Awareness and Pinyin Knowledge in Chinese Reading." *Writing Systems Research* 3, no. 1: 59–68. https://doi.org/10.1093/wsr/wsr010.

Yuan, L., and C. Xu. 2024. "The Influence of Pinyin Transcoding on L2 Chinese Listening Comprehension:Evidence From Path Analysis." *TCSOL Studies* 01: 34–44. 10.16131/j.cnki.cn44-1669/g4.2024.01.005.

Zhang, C., G. E. Bingham, X. Zhang, S. A. Schmitt, D. J. Purpura, and F. Yang. 2020. "Untangling Chinese Preschoolers' Early Writing Development: Associations Among Early Reading, Executive Functioning, and Early Writing Skills." *Reading and Writing: An Interdisciplinary Journal* 33, no. 5: 1263–1294. https://doi.org/10.1007/s11145-019-10006-3.

Zhang, H., and L. Roberts. 2021. "A Comparison of Pinyin Invented Spelling and Oddity Test in Measuring Phonological Awareness in L2 Learners of Chinese." *Journal of Psycholinguistic Research* 50, no. 2: 375–396. https://doi.org/10.1007/s10936-020-09700-z.

Zhou, W., H. Shu, K. Miller, and M. Yan. 2018. "Reliance on Orthography and Phonology in Reading of Chinese: A Developmental Study." *Journal of Research in Reading* 41, no. 2: 370–391. https://doi.org/10.1111/1467-9817.12111.

Ziegler, J. C., and M. Muneaux. 2007. "Orthographic Facilitation and Phonological Inhibition in Spoken Word Recognition: A Developmental Study." *Psychonomic Bulletin & Review* 14, no. 1: 75–80. https://doi.org/10.3758/bf03194031.