System 119 (2023) 103156

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

The role of future L2 selves in L2 speech development: A longitudinal study in an instructional setting

Yang Zhou [a,*], Mostafa Papi [b]

[a] School of Foreign Languages, Guizhou Normal University, China
[b] School of Teacher Education, Florida State University, United States

ARTICLE INFO

Keywords:
Future selves
Speech development
Comprehensibility
Accentedness
Speaking performance

ABSTRACT

The current study examined the role of English-speaking instruction and motivation in learners' development of L2 speech comprehensibility and accentedess, over the course of an academic semester. Eighty-three college students enrolled in English classes in China completed a sentence reading task and a picture description task twice (pre-test and post-test), as well as a questionnaire for measuring their future selves (Papi et al., 2019). The collected speech samples were coded for comprehensibility and accentedness. t-test results indicated that after one semester of English-speaking instruction, the learners made statistically significant gains in both L2 speech measures. Hierarchical multiple regression results showed that Ideal L2 Self/Own positively predicted speech comprehensibility, and Ideal L2 Self/Other negatively predicted L2 speech accentedness. The results suggest that learners' future L2 selves influence how they take advantage of instructional opportunities to improve their L2 speech in qualitatively different manners. Theoretical and practical implications are discussed.

In recent years, a number of longitudinal studies have been carried out to explore language learners' second language (L2) speech development in regular L2 learning classrooms in college. Of those, only a few have shown that college students succeeded in achieving advanced oral L2 proficiency (Derwing & Munro, 2015; Nagle, 2018; Saito et al., 2018). Several studies, however, have found that adult L2 learners generally fail to achieve significant improvements in at least one speech component such as accentedness (e.g., Munoz & Llanes, 2014; Saito et al., 2017), comprehensibility (e.g., Baker-Smemoe & Haslam, 2013), accuracy, or complexity (e.g., Mora & Valls-Ferrer, 2012). Unlike naturalistic L2 learning contexts where most adult L2 learners can make fast and stable progress using frequent opportunities to practice L2 on a daily basis (Sun et al., 2023), classroom language instruction is not as effective due to the lower frequency of interaction and engagement with authentic L2 resources (Larson-Hall, 2008). One of the main reasons for the variability of success in L2 speech development could be individual learner differences, such as their motivation, which could lead to different levels of behavioral engagement in the learning process, and, consequently, variability in L2 speech performance (Li, Hiver, & Papi, 2022). In other words, motivation can affect L2 speech production through promoting language learners' engagement in L2 communication (Hernandez, 2010; Segalowitz, 2010). Moyer (2014) argued that "pronunciation skill development is perhaps the greatest test of one's desire to successfully acquire a new language given the complexity of the task and its inherent relevance for self-concept" (p. 436). Although a few studies have highlighted the association between motivation and learners' L2 speech learning (Saito et al., 2017; Sun et al., 2023), it is still not clear how different motivational factors can lead to individual differences in L2 speech comprehensibility and accentedness.

* Corresponding author. School of Foreign Languages, Guizhou Normal University, Guiyang City, Guizhou Province, 550025, China.

https://doi.org/10.1016/j.system.2023.103156
Received 18 June 2023; Received in revised form 1 October 2023; Accepted 1 October 2023

The second reason for the variability in L2 speech development could be the fact that developing such capacity requires a long-term commitment to the learning process especially within instructional settings where access to L2 input and interaction is limited. Therefore, only learners who persevere in this years-long pursuit can succeed in acquiring advanced levels of proficiency in L2 speech. Given the importance of long-term perseverance in the process, the current study employs Papi et al.'s (2019) 2 × 2 model of future L2 selves to explore the effect of learner motivation on their L2 speech development in a classroom-based instructional setting. The use of this model as our motivational lens is based on the reasoning that L2 selves represent motivational factors that motivate learners to persist throughout the long-term process of language learning (Feng & Papi, 2020). In addition, future L2 selves could lead to qualitative differences in learners' behavioral engagement in language learning (Teimouri, 2017), which in turn can influence the quality of L2 outcomes (e.g., Papi & Khajavy, 2021).

In addition to the motivation model used, a longitudinal design was used in this study to examine changes in L2 speech considering L2 pronunciation is a skill that is developed over a long period of time (Derwing et al., 2008; Nagle, 2021; Saito et al., 2021) and should thus be captured using a design that serves this purpose. Consequently, the present study utilized Papi et al.'s (2019) model and a longitudinal research design to investigate how future L2 selves can explain the development of distinct L2 speech dimensions among L2 learners.

1. L2 speech development

L2 speech development is commonly considered the most challenging aspect of second language acquisition (SLA). There are at least two theoretical views that try to account for the reasons behind the difficulty of L2 speech development by highlighting its difference from L1 speech development. The revised Speech Learning Model (SLM-r, Flege & Bohn, 2021) argued that the varied extent of interaction between sound systems and input conditions in L1 and L2 is the fundamental reason for language learners' varying performance in L2 speech. According to Flege (2002), the sound-processing method used for learning L1 tends to last a lifetime, and it is likely to be used unwittingly during the process of learning L2. The specific elements that denote the L1 and L2 sound inventories are a set of acoustic-phonetic cues that can be further organized into long-lasting and meaningful phonetic categories. Specifically, learners' speech production tends to be directed by such phonetic categories, and incorrect categorization of L2 sounds may further contribute to faulty speech output. In order to generate a new L2 phonetic category in their sound system for L2 speech acquisition, L2 learners must have a clear perception of phonetic differentiation between L1 and L2. After new L2 categories have been established, learners' L2 speech production will eventually improve. Yet, if L2 learners are unable to distinguish the phonetic differences between two sound systems, their phonetic category will be blocked, which may impede their progress in L2 speech production (Flege, 2002).

The viewpoint that L2 learners' lack the capability of mastering native-like pronunciation was also discussed in the Critical Period Hypothesis (DeKeyser & Larson-Hall, 2005), which states that the acquisition of L2 input occurs only in the early stages of L2 immersion and that it is nearly impossible for late L2 learners to gain access to the presumed automatic language learning mechanism that early language learners have (DeKeyser, 2007; Granena & Long, 2013). According to the SLM, age-related effects in L2 speech learning result from a decrease in neurocognitive plasticity, making it increasingly hard for L2 learners to acquire elaborate phonetic-level information about L2 speech sounds (Flege, 1995, p. 266), but this does not necessarily preclude learners from improving their L2 phonetic skills. Flege and Bohn (2021) argues that the quality and quantity of L2 input conditions that learners are exposed to play an important role in influencing the representation and operationalization of their phonetic categories, which is supported by a number of empirical studies conducted by Flege and his colleagues (e.g., Flege, 1995, 2002; Flege et al., 2003). Flege (2002) asserts that, on the surface, the maturational issue may be one of the causes of unequal performance in language learners' L2 speech acquisition, but that the actual cause is the amount of L2 use and the quality of L2 exposure. The present study is based on Flege and Bohn's (2021) assumption that the quality and quantity of such L2 use and exposure are the major determining factors in individual differences in L2 speech quality, and we extend this argument by speculating that such variability in L2 use and exposure is a function of the learners' quantity and quality of motivation (Papi & Khajavy, 2021; Zhang & Papi, 2021), which can, in turn, lead to differences in L2 learning outcomes (e.g., Kermad, 2018; Papi & Khajavy, 2021).

In the past few decades, research on L2 speech learning has developed rapidly, and the inquiry into L2 learners' speech development permeates a variety of disciplines such as education, psychology, and linguistics. Many researchers have employed listeners' intuitive judgments to evaluate the two important dimensions of L2 speech (comprehensibility and accentedness) immediately after listening to L2 speech samples to determine L2 learners' speech performances (e.g., Derwing et al., 1998; Galante & Thomson, 2017; Munoz & Llanes, 2014; Saito et al., 2018; Sun et al., 2023). It has been shown that L2 speech learning is affected by an intricate interplay of individual factors, including age-related developmental variables, quality and quantity of exposure to the target language, and L2 instruction.

Among the past longitudinal studies that have investigated language learners' oral L2 development (Derwing et al., 1998; Galante & Thomson, 2017; Munoz & Llanes, 2014; Nagle, 2018; Saito, 2011, 2015; Saito et al., 2017, 2018; Segalowitz & Freed, 2004), only a few have examined the effect of motivation on learners' L2 speech performances (Nagle, 2018; Saito, 2015; Saito et al., 2017, 2018; Sun et al., 2023), Additionally, few L2 speech-related studies examined EFL learners' speaking progress in a speaking-specific training program.

This is an important gap because, if we accept the assumptions underlying Flege's (1995; 2002) Speech Learning Model concerning the intentional and malleable nature of speech development and agree that motivation is a key factor in learners' proactive involvement in the learning process and its ultimate success (Papi et al., 2019), exploring the link between motivation and L2 speech development would make intuitive, practical, and theoretical sense. In the present study, therefore, it is assumed that learners with motivational characteristics adaptive to the L2 speech environment are more likely to proactively expose themselves to and take

advantage of quality input and interaction and consequently improve their L2 oral proficiency (Papi et al., 2023). Exploring the connection between learner motivation and L2 speech development in this study can ultimately help better understand the optimal ways that motivational intervention can be employed to improve L2 speech development.

The current study examined the relative impacts of instruction using two dimensions of L2 speech: Comprehensibility and Accentedness. Accentedness is defined as "non-pathological speech that differs in some noticeable respects from native speaker pronunciation norms" (Munro & Derwing, 1995b, p. 289), while comprehensibility is defined as "listeners' perceptions of difficulty in understanding particular utterances" (Munro & Derwing, 1995b, p. 291). These constructs are frequently used in the field as two independent metrics of oral L2 proficiency (Uchihara, 2022) and have been shown to play the primary role in the communicative effectiveness of L2 speech (Derwing et al., 1998; Munro & Derwing, 1995a, 1995b; Saito et al., 2021; Saito & Plonsky, 2019).

2. Motivation

Motivation is a crucial factor in L2 learning and achievement (Papi & Khajavy, 2021). According to Papi and Hiver (2020), various motivation theories account for different dimensions of learner motivation. Some motivation theories deal with the process of learning, some concern learners' curiosity about the target language and community, and others focus on the end-states of the learning pursuit. Considering that developing L2 speech is a long-term and demanding process that requires persistent effort (Moyer, 2014), in the present study we chose a motivation theory that focuses on future L2 selves, which are the long-term end-states of L2 learning, and thus suitable for the study of the link between motivation and L2 speech development.

Dörnyei's (2009) L2 Motivational Self System (L2MSS) was the first model that incorporated the notion of future selves into motivation theory. L2MSS has three components: the *ideal L2 self* represents one's hopes and aspirations for L2 learning; the *ought-to L2 self* represents one's duties and obligations; and the L2 learning experience refers to the motives related to the immediate learning context and experience (e.g., teachers, methods, instructional materials). Whereas the L2 learning experience concerns the process of L2 learning, the two future selves represent the L2 speaker the learner wants to be in the future. According to Dörnyei (2009), the desire to reduce the discrepancy between one's present L2 self and these future selves is assumed to create the necessary motivation for the learner to engage and persist in the learning process.

In response to asymmetric results concerning the two future selves outlined in L2MSS (e.g., Csizér & Kormos, 2009; Papi & Teimouri, 2012, 2014; Taguchi et al., 2009), Papi et al. (2019); (see also Lanvers, 2016; Teimouri, 2017) drew on self-discrepancy theory (Higgins, 1987) as well as the regulatory focus theory (Higgins, 1997) to propose the 2 × 2 Model of L2 Self-Guides, which bifurcated the original ideal and ought-to L2 self by two standpoints (own vs. other). More specifically, the 2 × 2 model consists of four future selves: ideal L2 self/own, ideal L2 self/other, ought L2 self/own, and ought L2 self/other. Ideal L2 self/own represents one's image of the kind of L2 user one aspires to be in the future; ideal L2 self/other refers to the image that one's important others (e.g., parents, family) hope for them to achieve; ought L2 self/own consists of one's own L2 learning obligations; and ought L2 self/other represents the obligations others have imposed on the learner. In Papi et al.'s (2019) 2 × 2 model, ideal selves are argued to have a promotion focus concerned with approaching desirable end-states (e.g., speaking English fluently), whereas the ought selves have a prevention focus concerned with avoiding undesirable end-states (e.g., failure in school). Confirmatory Factor Analysis confirmed the construct validity of the model and Cronbach's alpha value showed that the scales were reliable. The alpha coefficients were 0.81 for ideal L2 self/own, 0.79 for ideal L2 self/other, 0.80 for ought L2 self/own, and 0.72 for ought L2 self/other.

Studies by Papi and his colleagues have shown that both ideal and ought selves could contribute to motivation, emotions, learning behavior, and achievement, but in qualitatively different manners. These studies, which measured the future selves using Papi et al.'s (2019) questionnaires, showed that the ideal selves led to an eager inclination in L2 use concerned with maximal use of the target language (Papi & Khajavy, 2021), learning enjoyment (Tahmouresi & Papi, 2021), more feedback-seeking behavior (Bondarenko, 2020), persistence in L2 learning (Feng & Papi, 2020), and higher levels of L2 achievement (Papi & Khajavy, 2021), whereas ought selves were associated with L2 anxiety (Jiang & Papi, 2021), vigilant and minimal use of the target language (Papi et al., 2019), and lower L2 achievement (Papi & Khajavy, 2021). The results of these studies highlight the theoretical distinctions between ideal and ought L2 selves (Higgins, 1987). The ideal selves have a promotion focus concerned with positive outcomes; this concern with positive outcomes makes learners motivated by ideal selves eager to use the target language, experience elation-related emotions such as enjoyment, seek more feedback, and achieve better learning outcomes. Ought L2 selves, on the other hand, have a prevention focus concerned with avoiding negative outcomes. This concern makes learners motivated by ought selves vigilant in their L2 use, reluctant to seek feedback, and prone to the experience of anxiety, which in turn harms their achievement. It should be noted that English learning in high schools in China is exam-oriented, and English is a course that students are required to take (Apple et al., 2016; Hu & West, 2015). In this context, the primary purpose of English learning could be meeting the requirements of the schools and avoiding low grades due to pressure from parents, teachers, or peers, reasons that are primarily reflected by the ought-to L2 self. However, due to the fact that the current study was conducted in a private institution where learning English was voluntary and not a requirement for the students, it was not surprising that the ideal L2 self emerged as an important factor in stimulating L2 learning. Given that these future selves lead to qualitatively different behavioral, emotional, and learning patterns among learners, it would not be unreasonable to expect such differences to be reflected in their L2 speech development and performance. The purpose of the present study is, thus, to investigate how these future selves could potentially affect L2 speech development.

2.1. Studies on motivation and L2 speech learning

Few empirical studies have investigated the influence of a varying level of motivation on the improvement of adult learners' L2

speech skills in a language classroom setting. Baker-Smemoe and Haslam (2013), for instance, investigated the effect of motivation on L2 speech achievement in a 10-week speaking session with 31 Chinese EFL learners. The results of the study suggested that the development of fluency and comprehensibility in learners' speech was positively connected with their motivational intensity. In addition, the results indicated that learners with greater motivation were more likely to employ various oral English learning practices that could assist their L2 speech acquisition. In addition, Baker-Smemoe and Haslam (2013) found that adult EFL learners with poor motivation, which was measured using the scale of PLAB (Language Learning Aptitude Battery), could be significantly limited in their long-term development of L2 speech comprehensibility and fluency. In a separate study, Saito et al. (2017) studied the influence of motivation on the L2 speech development of 40 Japanese freshman undergraduates in an EFL classroom setting over a 15-week period. The researchers discovered that motivation, measured using a 13-item tailored questionnaire scale, was a major predictor of the speech comprehensibility gain scores of the students. Similarly, Nagle (2018) discovered that L2 learners improved their L2 (Spanish) accentedness (rated by 18 college students who were native speakers of Spanish) in tandem with their motivation throughout the course of a year-long L2 learning program. Saito et al. (2017), on the other hand, did not find any statistical significance between motivation and speech accentedness (rated by native raters' holistic assessments through a nine-point scale).

Three studies have investigated the association between L2MSS, using Taguchi et al.'s (2009) questionnaire scales, and the development of L2 speech. Saito et al. (2018) discovered, in a longitudinal study involving 108 first-year Japanese high school EFL learners who studied English within a three-month English learning session, that L2 learners' development of comprehensibility (rated by five native speakers of English) was moderately associated with their ideal L2 self, but not with their ought-to L2 self. In another study, Nagle (2018) likewise found no correlation between the two future selves (ideal L2 self and ought-to L2 self) and L2 speech comprehensibility (rated by 18 college students who were native speakers of Spanish) among L1-English Spanish learners. Sun et al. (2023) found ideal L2 self to lead to more eager L2 use with fluent L2 speakers, which in turn predicted gains in L2 comprehensibility. Two studies have employed Papi et al.'s (2019) model of L2 selves in relation to L2 speech. Kermad (2018) found that Ideal L2 Self/Own had a small positive correlation with comprehensibility and accentedness. In addition, multiple regression analyses showed that Ideal L2 Self/Own and Ought L2 Self/Other positively but Ideal L2 Self/Other negatively predicted comprehensibility, (lack of) accentedness, and intelligibility. However, that study included only 20 participants, which makes the validity of the results questionable. In addition, Zhou (2019) found that attitudes toward pronunciation were highly and positively related to Ideal L2 Self/Own and Ought L2 Self/Own, but adversely related to Ought L2 Self/Other among learners of English in China.

In the present study, we have employed Papi et al.'s (2019) 2×2 model because it can enable us to investigate how different future selves might motivate learners to employ qualitatively different learning patterns, which can, in turn, lead to qualitative differences in their L2 speech performance. Given the eagerness and risk-taking tendency of learners motivated by their ideal selves, which can make them more likely to seek target language input and engage in interaction, this group of learners is expected to show better L2 speech development than learners who are motivated by their ought-to L2 selves, which can make them cautious and less willing to engage with L2 input and interaction. The present study, thus, is the first longitudinal study that investigates these L2 speech characteristics through the lens of the 2×2 model of future L2 selves (Papi et al., 2019).

3. Research questions

As the SLM model postulates that the mechanism underpinning language learning remains attainable for L2 learning over the life span and previous research has shown the role of ideal selves in motivating learners to eagerly seek L2 use opportunities (e.g., Papi & Khajavy, 2021), we speculate that learners motivated by these promotion-focused selves are more likely to enhance their L2 speech performance. The following research questions were thus formulated:

1. How do learners' English speech comprehensibility and accentedness develop over one academic semester in a speaking class?
2. How do future L2 selves (ideal L2 self/own, ideal L2 self/other, ought L2 self/own, ought L2 self/other) predict the development of learners' comprehensibility and accentedness of English speech over one academic semester?

4. Methods

4.1. Participants

Eighty-three first-year undergraduate EFL learners (61 females, 22 males) in China participated in this study. These students were from Tianjin University of Finance and Economics (TUFE) and were enrolled in English speaking courses at a private English training school in Tianjin, China. These learners were at an average age of 19.58, with a range of 18–21. The participants reported the length of English learning for an average of over nine years ($m = 9.63$, $SD = 1.06$) within a five-year range (7−12). The mean English score of College Entrance Exam (CEE, also known as Gaokao) these students acquired before they entered university was 114, with a range from 101 to 138 (possible score range: 0 to 150). CEE is required for entry into all undergraduate institutions of higher education in China. Chinese students who obtain an overall score of at least 80% in Gaokao (120 points for all subjects) will be considered for direct entry into first-year study at many universities in the UK.

4.2. Design & procedures

The study followed an observational and longitudinal design in which learners' speech development over an academic semester of

L2 speaking instruction was evaluated in relation to their motivational characteristics. Data related to the participants' motivation and L2 speech performance at the beginning and end of an instructional program were collected following the steps outlined below.

A convenience sampling method was followed to collect data from the TUFE students who had enrolled in the English training program. After obtaining the Institutional Review Board approval for the study and with permission from the principal at the private school, an invitation email with information about the purpose and procedures of the study was sent to students. The students were informed that they had the right to refuse the researchers the right to use their data for research purposes. None of the students denied permission. In addition, a translated consent form in Chinese entailing detailed procedures about the study was sent out to the students who signed up for the study. The first author scheduled mutually convenient times with the participants for data collection. Each participant who completed the survey and oral tasks was rewarded with a 50-CNY gift card as compensation. As described in Table 1, the data collection was completed in three phases. In Phase 1, the participants completed a questionnaire, a sentence reading task and a picture description task; in Phase 2, they received L2 speaking instruction over an academic semester in an instructional program that is described below; and in Phase 3, they completed another sentence reading task and picture description task. Before the students performed the oral tasks, the researcher informed them of the instructions for self-recording via an online meeting and supervised their completion of the oral tasks. The unnecessary portion of each audio file was eliminated, and a file containing the complete speech samples was sent to the raters for further evaluation.

4.3. Instructional program

The students enrolled in the speaking courses attended three 150-min classroom-based lessons per week for 12 weeks at a private English training school in China. The speaking program was intended to improve the communicative abilities of English learners. The students were strictly required to speak English wherever they were in school. The detailed syllabus of the course and the content regarding each of the lessons are demonstrated in Appendix A in the Supplementary Materials. The entire course included 19 major lessons, 14 practice lessons, and 5 activity lessons. In each major lesson, the instructor began with a presentation and explanation of target linguistic features. The presentation process was usually followed by a series of comprehension-oriented (passage reading and listening) and production-oriented (e.g., discussion, oral translation, presentation) activities. Learners were also given multiple opportunities to apply what they have learned to some practical speaking activities (e.g., drama, debate, storytelling, and role-plays). During two class meetings, explicit lessons on pronunciation skills (e.g., articulation rules for American English phonemes, stress and intonation) were provided.

4.4. Instruments

The study used a questionnaire for measuring participants' future L2 selves, a sentence reading task for assessing their L2 speech proficiency in a controlled state, and two picture description tasks for examining participants' L2 spontaneous speech.

4.4.1. Questionnaire

The study adopted Papi et al.'s (2019) L2 self-guides questionnaire for measuring the speakers' L2 future selves, followed by demographic questions (see Appendix B in the Supplementary Materials). The L2 self-guide includes four subscales; each scale consists of four items related to the specific type of self (i.e., ideal L2 self/own, ideal L2 self/other, ought L2 self/own, ought L2 self/other). For each item, learners were asked to rate the degree to which they agreed or disagreed with the statements on a 6-point Likert scale (from 1 = *strongly disagree* to 6 = *strongly agree*). The instrument has been used in multiple major studies on this topic and has shown strong psychometric properties (e.g., Papi et al., 2019; Papi & Khajavy, 2021).

4.4.2. Sentence reading task

This study adopted a sentence reading task for examining L2 learners' speech performance at a controlled level, a technique that has been used in previous studies as well. For example, Saito (2011) created four sentences containing 39 loaded words out of 50 words which consisted of targeted problematic phones including/æ, f, v, θ, e, w, l, r/. In the sentence reading task that Saito (2011) developed, at least three loaded phones appeared in different words within the four sentences (e.g.,/w/: when, woman, with). Saito (2011) deliberately created the four sentences in this way to ensure that all targeted phones can be equally assessed. Following Saito's

Table 1
Data collection timeline in three phases.

Timing	Events
Phase 1	Participants completed: • Motivation and background questionnaire (10 min) • Pre-test sentence reading task A and picture description task A (20 min)
Phase 2	• Rater Training Orientation (20 min): The first author walked four raters through both the theoretical definition and the assessment procedure in terms of accentedness and comprehensibility. • Rating Calibration (60 min): The first author and raters synchronously but independently assessed a number of speech samples and compared the results with each other. If there was a disagreement on a specific case, they were asked to negotiate with each other until consensus was reached.
Phase 3	• Participants completed the post-test sentence reading task B and picture description task B (20 min). • The raters started to rate each speech sample.

(2011) task design, in the present study, a total of 20 sentences were composed for the sentence reading task, in which every phoneme of all common American English phonemes was represented in at least three words; this was done to ensure that every English phoneme could be tested sufficiently. For instance, there were at least three words in the sentences containing the phoneme/ɛ/(e.g., get, expensive, pen) or the phoneme/ʊ/(e.g., could, good, book). Likewise, the rest of the phonemes of the standard American phonetic alphabet were used in different words used in these sentences (see Appendix C in the Supplementary Materials for the complete list of sentences).

4.4.3. Picture description task

The current study also adopted a timed picture description task to examine learners' spontaneous speech proficiency. This type of task has been widely used in past studies that examined L2 learners' spontaneous speech proficiency (Derwing et al., 1998; Munro & Mann, 2005; Saito et al., 2017). There were twelve pictures included in the task. The first six pictures were extracted from an ESL textbook used by L2 learners at an intermediate level of English proficiency. Another set of six pictures was created as post-test alternatives. The two sets of pictures in the two tests both describe a similar story about some daily routines occurring in a kindergarten, but the details of the story in each picture vary slightly between the two pairs (see Appendix D in the Supplementary Materials for the picture description task). The two picture description tasks were piloted, and the results indicated that they were comparable in terms of the number of words (task A: $M = 226.90$ words; task B: $M = 235.40$ words) and time (task A: $M = 173.10$s'; task B: $M = 182.80$s') that speakers used to complete them, as well as the oral performance they demonstrated (Comprehensibility/Accentedness rating in task A: $M = 4.75/4.23$ and in task B: $M = 4.82/4.18$).

4.5. Data coding

Four graduate students of applied linguistics who spoke English as their first language and reported normal hearing were recruited to rate the collected speech samples. The raters had little experience with Chinese ESL/EFL learners, given that raters' experience with a specific foreign accent could exert a significant influence on L2 speech assessment (Isaacs & Thomson, 2013).

As shown in Table 1, the raters were also trained in Phase 2 and the data were coded in Phase 3. The speech samples for both the sentence-reading and picture description tasks were coded for comprehensibility and accentedness by all the raters using two different scales. Speech comprehensibility was rated based on a Likert scale (from 1 = *very hard to understand* to 9 = *no effort to understand*), and accentedness was rated based on a similar 9-point scale (from 1 = *native-like* to 9 = *heavily accented*). The decision to adopt a 9-point scale was informed by earlier research (e.g., Derwing & Munro, 1997). The scale is regarded as "practical, useable across contexts, and

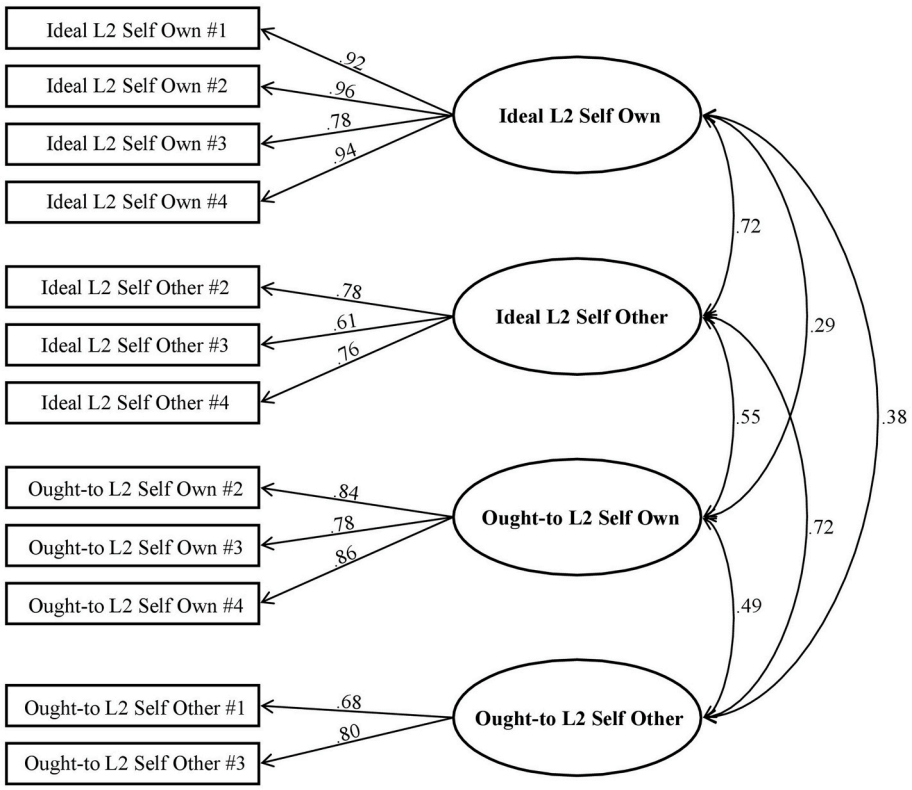

Fig. 1. Cfa on L2 self-guide items.

sufficiently reliable for research purposes" (p. 138), with the advantage of providing adequate choice for the rater (see Isaacs & Thomson, 2013; for a discussion on the 9-point scale).

The operationalization of rating the speech samples of the two tasks slightly varied. For the sentence reading task, the rater listened to the speech sample of every sentence, and the final grade given to each speaker was based on the average score of the speech samples of all sentences. For the picture description task, the rater listened to each speaker's entire description and assigned a total score.

4.5.1. Reliability analysis

Inter-rater reliability analysis was performed on a subset of 83 samples selected and rated at two different time points by four raters. Inter-Class Correlation (ICC) estimates and the 95% confidence intervals were calculated via SPSS version 26 (IBM) based on single ratings ($k = 4$), absolute agreement, and a two-way random effects model. The results of reliability analyses indicated that the four raters demonstrated good inter-rater agreement in rating the speakers' speech performances displayed in both the pretest and posttest. In terms of the speech Comprehensibility or Accentedness ratings, the ICC values ranged from 0.75 to 0.9 for the pretest and posttest.

To measure the intra-rater reliability, 10 of the 83 speech samples were replicated and re-rated by the same raters. ICC estimates between the two sets of samples were higher than 0.75 across all assessment criteria across the four raters, confirming intra-rater agreement. Since both inter- and intra-reliability of the rating were achieved in the study, the scores graded by the four raters were averaged across all tested speech items, and the average score was considered the final score for each speaker.

4.6. Data analysis

4.6.1. Confirmatory Factor Analysis of motivation questionnaire items

Confirmatory factor analysis (CFA) was conducted in the study to test the factor structure of the students' future selves within the framework of the 2 × 2 model (Papi et al., 2019). Before running the analysis, Q–Q plots were examined, which confirmed that the assumption of multivariate normality was met. In addition, our data was factorable because it included close to five cases per parameter (Kline, 2011). CFA was run via AMOS 26 (IBM), with the solution developed based on maximum likelihood estimation. There were 16 questionnaire items (observed variables) incorporated in the original model, and they fell into four L2 self constructs (latent variables). However, the original model underwent modifications based on goodness-of-fit indexes, and four items with factor loading lower than 0.4 (Hair et al., 2010) were removed from the model in order to best fit the model to the dataset. The estimates were recalculated after the removal of the four items, and the final model is shown in Fig. 1. There were 12 observed variables in the final model that loaded strongly onto the proposed four motivation constructs.

Despite the fact that the Chi-square to degrees of freedom was significant ($df = 59$, Chi-square = 96.61, $p < .01$), indicating insufficient fit, the other four widely used goodness-of-fit indexes (Harrington, 2008) demonstrated a good fit of the final model to the current data (CFI: comparative fit index = 0.97, TLI: Tucker-Lewis index = 0.96, NFI: Nordic fit index = 0.90, RMSEA: root mean square error of application = 0.07). In other words, the values extracted from the model summary confirmed that the 2 × 2 model of L2 self-guides was a good fit for our dataset. In addition, Cronbach's alpha coefficients ranged from 0.70 to 0.94, suggesting good internal consistency of the scales (Table 2).

4.6.2. Principal component analysis on L2 speech measures

Principal component analysis (direct oblimin rotation) with Kaiser normalization (eigenvalue is greater than 1; item loading value is greater than 0.4) to identify latent factors underlying the multiple L2 speech measures. Before running the analysis, the assumptions of the analysis were checked and met. Scatterplots showed that the two variables were linearly related and led to the detection of no notable outliers. Kaiser-Meyer-Olkin measure of sampling adequacy was significant, showing the sample was large enough (0.72, $p < .05$). Finally, Bartlett's test of sphericity ($\chi^2 = 146.46$, $p < .01$) was significant, confirming that the data was suitable for this analysis. The analysis led to the emergence of two factors representing comprehensibility and accentedness (Table 3), which accounted for 87% of the total variance in the L2 speech measures (Kaiser-Meyer-Olkin = 0.72; $\chi^2 = 146.46$, $p < .01$). The average scores of the learners' speech comprehensibility and accentedness measures, obtained in sentence reading tasks and picture description tasks, were adopted as the final scores for Comprehensibility ($\alpha = 0.86$) and Accentedness ($\alpha = 0.87$), respectively.

5. Results

The data were analyzed using SPSS version 26 (IBM; see Table 4 for descriptives). In response to the first research question (RQ1: How do learners' English speech comprehensibility and accentedness develop over one academic semester in a speaking class?), paired samples t-tests were carried out to compare the students' L2 speech performance in the pretests and posttests in terms of two different

Table 2
Cronbach's alpha reliability analysis on L2 self-guide scales.

	means	SDs	Cronbach's alpha	N of items
Ideal L2 Self/Own	4.16	1.04	0.94	4
Ideal L2 Self/Other	3.87	.63	0.75	4
Ought L2 Self/Own	2.87	.59	0.74	3
Ought L2 Self/Other	4.28	.54	0.70	2

Table 3
Summary of a two-factor solution for L2 speech measures.

		Factor	
		1	2
Accentedness	Picture description accentedness	.92	
	Sentence reading accentedness	.97	
Comprehensibility	Picture description comprehensibility		.96
	Sentence reading comprehensibility		.92

Note: Factor loadings <0.30 are not shown.

aspects of English speech (i.e., Comprehensibility & Accentedness). The Bonferroni procedure was also used to control for the effects of multiple comparisons, with the significance level set at $p < .025$. Before running the analysis, the assumption of normality of variance in terms of the difference was checked. Most of the cases in the histogram were to a great extent normally distributed in a bell-shaped pattern. In addition, the skewness values for all cases ranged from −0.37 to −0.09, and the kurtosis values ranged from −0.53 to 0.40, indicating an acceptable data range for normal distribution (Mayers, 2013).

The *t*-test results, presented in Table 5, showed that there was a statistically significant difference in the participants' comprehension scores ($t\,(82) = 17.28$, $p < x\,.01$) between the pretest ($Mean = 4.71$, $SD = 1.28$) and the posttest ($Mean = 5.38$, $SD = 1.34$) with a medium effect size ($d = 0.55$). There was also a significant difference in accentedness ($t\,(82) = -15.45$, $p < .01$) between the pretest ($Mean = 4.18$, $SD = 1.20$) and posttest scores ($Mean = 3.62$, $SD = 1.38$), with a medium effect size ($d = 0.61$). These results show that after one semester's English-speaking training in the private English school, these Chinese EFL learners made statistically significant improvements in their English speech performance. Specifically, the learners could speak more comprehensibly in the posttest than in the pretest. Also, the students showed less accented pronunciation in the posttest compared with the pretest.

To answer the main research question of this study (RQ2: How do future L2 selves predict the development of learners' comprehensibility and accentedness of English speech over one academic semester?), two multiple regression analyses were conducted. Regression analysis has been recommended by researchers as a flexible and powerful alternative to mean-comparison methods (Plonsky & Oswald, 2017). In the analyses, the target measure's Pretest scores, Age, and Years of English Learning were used as covariates to control for their effects.

The regression assumptions of normality, linearity, homoscedasticity, and multicollinearity were evaluated. Initially, the histograms of all models displayed a roughly normal distribution of the standard deviations of the residuals. In addition, the dots in the P–P plots were closely aligned with the diagonal line, and the points in the scatterplots were randomly distributed along a horizontal line without a funneling structure. Hence, both linearity and homoscedasticity conditions were met for the regression models (Cohen et al., 2003). Also, the value of collinearity tolerance (VIF) was investigated for each variable in the subsequent models. The findings indicated that multicollinearity was not a problem among the predictors in the study, as the values fell below 5.0 (Hair et al., 2010).

As shown in Table 6, with post-test Comprehensibility as the outcome variable, the model was significant and explained 96% of the variance ($R^2 = 0.96$, $F = 295.20$, $p < .01$). In addition, Ideal L2 Self/Own ($\beta = 0.13$, $p < .05$) emerged as the only significant predictor among the selves, suggesting that the higher one's score in Ideal L2 Self/Own the better the comprehensibility of their speech. With post-test Accentedness as the outcome variable (Table 7), the model was significant and accounted for 96% of the variance in the outcome variable ($R^2 = 0.96$, $F = 284.18$, $p < .01$). In addition, Ideal L2 Self/Other emerged as a significant negative predictor ($\beta = -0.09$, $p < .05$), suggesting that with the increase of the learner's score in Ideal L2 Self/Other, the accentedness of their speech decreases.

6. Discussion

The learners in the study were demonstrated to have achieved significant gains in L2 speech proficiency including Comprehensibility, and Accentedness, after attending one academic semester of formal training in speaking English. The results are generally in line with the findings of the study by Saito et al. (2018), which revealed that Japanese EFL learners improved in overall

Table 4
Descriptive statistics of all the measures.

Measures	Range	Mean	SD
Comprehensibility (Pretest)	2.13–6.88	4.71	1.28
Comprehensibility (Posttest)	2.50–7.75	5.38	1.34
Accentedness (Pretest)	2.37–6.88	4.18	1.20
Accentedness (Posttest)	1.63–6.68	3.62	1.38
F1: Ideal L2 Self/Own	2.00–5.75	4.16	1.04
F2: Ideal L2 Self/Other	2.33–5.00	3.87	.63
F3: Ought L2 Self/Own	2.00–4.67	2.87	.59
F4: Ought L2 Self/Other	2.00–5.50	4.28	.54
Age	18.00–21.00	19.58	.83
Years of Learning English	7.00–12.00	9.63	1.06

Table 5
Paired samples T-test.

		Paired Differences				Cohen's d	p
		Mean	SD	95% Confidence Interval of the Difference			
				Lower	Upper		
Pair 1	Comprehensibility	.67	.35	.59	.75	.55	<.01
Pair 2	Accentedness	−.56	.33	−.64	−.49	.61	<.01

Table 6
Regression results with posttest comprehensibility as the outcome variable.

Predictor Variable	*B*	Std. Error	*β*	*t*	*p*	*VIF*
Pretest Comprehensibility	.96	.05	.80	20.65	<.01	3.21
Age	.02	.04	.01	.42	.68	1.10
Years of English Learning	.11	.04	.09	2.82	<.05	2.04
Ideal L2 Self/Own	.17	.07	.13	2.35	<.05	4.15
Ideal L2 Self/Other	.05	.07	.03	.77	.44	2.34
Ought L2 Self/Own	−.01	.11	−.01	−.11	.91	4.25
Ought L2 Self/Other	.01	.08	.01	.12	.90	1.42

Model Summary: $R^2 = 0.96$, F = 295.20, $p < .01$.

Table 7
Regression results with posttest accentedness as the outcome variable.

Predictor Variable	*B*	Std. Error	*β*	*t*	*p*	*VIF*
Pretest Accentedness	.98	.04	.86	23.53	<.01	2.76
Age	−.05	.04	−.03	−1.26	.21	1.08
Years of English Learning	−.04	.04	−.03	−.89	.38	2.13
Ideal L2 Self/Own	−.09	.07	−.06	−1.16	.25	4.40
Ideal L2 Self/Other	−.20	.07	−.09	−2.77	<.05	2.41
Ought L2 Self/Own	.02	.11	.01	.21	.84	4.37
Ought L2 Self/Other	.02	.08	.01	.26	.80	1.40

Model Summary: $R^2 = 0.96$, F = 284.18, $p < .01$.

comprehensibility over one academic year of EFL classroom learning. The present study, however, contrasts with the findings of Derwing et al. (1998), who found that individuals did not make significant progress in accentedness in narrative tasks. The variance may be related to the different L2 instruction methods employed in each study. In the study conducted by Derwing et al. (1998), students participated in comprehensive language courses that placed little emphasis on pronunciation-focused instruction. However, the courses that the students in the current study took were specifically designed for English pronunciation training that was explicitly implemented by providing auditory and articulatory exercises about L2 segmental and suprasegmental characteristics, which could have a substantial effect on the pronunciation proficiency of adult L2 learners (e.g., Derwing et al., 2014; Kissling, 2013; Trofimovich et al., 2017). Similarly, Saito's (2011) study found no significant progress among a group of Japanese ESL learners regarding their speech accentedness and comprehensibility in both sentence-reading and picture description tasks. The difference could be due to the length and intensity of the training. The learners in Saito's (2011) study merely engaged in a total of 4 hours of English instruction, which might have been inadequate to lead to statistically significant improvements in those speech characteristics.

The findings indicated that L2 pronunciation acquisition is attainable for adult EFL learners, even though they are less likely to be as successful as younger learners (Munoz & Llanes, 2014). Even among adult language learners, maybe some benefit more from studying in an ESL classroom as it is a more naturalistic setting than those in EFL contexts (Levis, 2005; Sun et al., 2023). Nevertheless, the results suggest that it is possible to help EFL learners improve their English-speaking skills in an EFL environment if they are trained to engage in frequent communication in an immersive context, which is critical to the development of L2 speech skills (Galante & Thomson, 2017; Lazaraton, 2013).

Regression analyses showed that Ideal L2 Self/Own in the study emerged as a statistically significant predictor of comprehensibility. Previous studies have shown that the Ideal L2 Self/Own is a strong predictor of L2 motivated behaviors (Taguchi et al., 2009), persistence in L2 learning (Feng & Papi, 2020), enjoyment (Teimouri, 2017), willingness to communicate in a second language (e.g., Khajavy & Ghonsooly, 2017), eager use of the target language (Papi et al., 2019) and finally L2 achievement (Papi & Khajavy, 2021; Tahmouresi & Papi, 2021). The results also echo the findings of studies on the role of Ideal L2 Self/Own in relation to L2 pronunciation. For instance, Zhou (2019) found that Ideal L2 Self/Own strongly and positively correlated with English learners' attitudes towards the importance of pronunciation. Other studies found Ideal L2 Self (Own) to be a more consistent predictor of accurate articulation in a language setting where L2 use was optional (Nagle, 2013), and positively associated with L2 speech comprehensibility (Saito et al., 2018; Sun et al., 2023), fewer segmental deviations, and better pronunciation fluency (Kermad, 2018). Given the results of these

studies, it was not surprising that Ideal L2 Self/Own positively predicted L2 speech comprehensibility.

Ideal L2 Self/Own has probably led to improvements in the Comprehensibility of the participants in at least two important ways. The first way that Ideal L2 Self/Own has led to these positive outcomes is probably through increases in the perceived value of L2 pronunciation, which has in turn contributed to the participants' motivated learning behavior and eventually their L2 pronunciation. Learners with a strong Ideal L2 Self/Own may see good pronunciation as a critical part of their vision of the kind of L2 speaker they desire to be in the future, as represented in the items measuring this construct (e.g., *I can imagine a day when I speak English fluently with international friends/colleagues*). The importance and value that these learners attach to speaking the target language in a fluent and comprehensible way (Zhou, 2019) seems to have led to their stronger motivation to improve their L2 pronunciation, which in turn has been found to contribute to their L2 pronunciation development (Baker-Smemoe & Haslam, 2013), and comprehensibility (Saito et al., 2017).

A second and more direct way that Ideal L2 Self/Own could have led to such improvements in L2 pronunciation concerns the regulatory focus of this motivational construct and its effects on learner behavior. The Ideal L2 Self/Own is a future goal with a promotion regulatory focus concerned with accomplishment, growth, and advancement (Dörnyei, 2009), which also leads to an eager strategic inclination involving the maximal use of L2 speaking opportunities (Papi & Khajavy, 2021). Such an eager strategic inclination has probably led students motivated by the Ideal L2 Self/Own to take maximal advantage of L2 speaking opportunities, take more risks in engaging in trial and error in speaking the language, potentially receive more corrective feedback on their L2 pronunciation, and eventually improve their L2 pronunciation skills. Such characteristics can be highly instrumental in learning how to speak a target language. Improving one's L2 speaking skills is a challenging task that can only be accomplished if a student is willing to take risks to eagerly communicate in the target language, engage in the process of learning through trial and error, and finally, seek and use corrective feedback on their L2 speaking performance (Papi, 2018; Zhang & Papi, 2021). These characteristics seem to be associated with learners motivated by promotion-focused motives such as their Ideal L2 Self/Own.

Ideal L2 Self/Other significantly predicted Accentedness negatively, which indicates that the stronger learners' desire to make their important others proud, the less their degree of accentedness in L2 speech. It is a common goal among Chinese students to make their parents or teachers academically proud by following their parents' hopes and aspirations for them (Wang & Rao, 2022). Learners driven by ideal L2 self/other may focus more on establishing a native-like accent to make their important others (e.g., parents, family, teachers, etc.) proud. These important others may not be in a position to assess the comprehensibility of the students' speech but can be impressed by the accent that sounds like that of native English speakers.

6.1. Pedagogical implications

The results of the present study highlight the importance of speaking-specific instruction in L2 speech development. In addition, they showed that Ideal L2 Self/Own contributed to L2 speech Comprehensibility whereas Ideal L2 Self/Other was associated with less Accentedness. Considering the limited chances for L2 practice outside of the classroom, especially in EFL conditions (Derwing et al., 2008), explicit speech training is critical. The findings presented here demonstrate that pedagogical practices derived from an explicit and intensive speaking learning program may improve L2 learners' speech capacity and control over the global functions of language (Galante & Thomson, 2017), such as vocal projection, stress, rhythm, pace, and the ability to be easily understood. Explicit instruction can include immediate and consistent corrections of learners' pronunciation errors (Elliott, 1997), explicit teaching of phonetics (Derwing et al., 1998), adequate exposure, practice, and feedback given to learners regarding the target phones (Kissling, 2013). Such an instructional approach can provide students with sufficient time and opportunity to practice their L2 oral performances, and is more beneficial than discontinuous instruction sparsely scheduled over the course of a semester.

Encouraging learners to follow their ideal selves and even using different techniques to enhance such future self-images could also provide a highly instrumental tool that L2 instructors can employ in different instructional practices that could potentially lead to improvements in learners' motivation and L2 speech development. For this purpose, instructors are recommended to integrate vision-enhancing exercises and activities into their lessons (Dörnyei & Kubanyiova, 2014). For instance, visualization techniques can be integrated into class activities to enhance the learners' ideal L2 self, which in turn can benefit their speech development through boosting their motivation, willingness to communicate, and eager L2 use (Magid & Chan, 2012; Safdari, 2021; Sampson, 2012). There are several crucial steps included in visualization exercises, such as asking learners to visualize their ideal English selves, elaborating on it, writing down their feasible goals, and coming up with concrete action plans and strategies to achieve those goals (for a review, see Papi, 2022).

7. Conclusion

The results of this study indicated that after one academic semester of training in speaking English at a private English school in China, the speech performance of the EFL undergraduate students improved significantly. Adult EFL learners are likely to progress in L2 speech learning if they receive adequate and systematic phonetic training as well as sufficient exposure to comprehensible L2 input (Derwing & Munro, 1997, 2015; Derwing et al., 2008; Levis, 2005; Saito, 2011, 2017). Although the participants in the study were not physically surrounded by native speakers during their daily academic activities, the private English training school provided a setting in which English was required for everyday communication, thereby practicing their oral English skills. The results of this study support the usefulness of L2 immersion for the development of L2 speech skills; such an L2 immersion context provides exceptional possibilities for pronunciation practice, which is likely to polish and automate students' articulation manners (DeKeyser, 2007; Sun et al., 2023).

The study revealed that learners' motivation, and more specifically, their ideal L2 selves, played significant roles in facilitating their L2 speech gains over the long-term English-speaking training sessions. More specifically, Ideal L2 Self/Own was found to improve the participants' L2 speech Comprehensibility and Ideal L2 Self/Other contributed to the reduction of Accentedness in their speaking. These results confirm the prominence of promotion-oriented motives and future goals in the L2 learning pursuit in general and L2 speech development in particular and provide more support for the utility of a promotion-oriented approach in second language learning and teaching (Papi & Khajavy, 2021; Papi & Teimouri, 2014). Ideal L2 selves, as promotion-focused motives, were found to contribute to the learners' L2 speech development, presumably through their tendency to enhance an eager and maximal tendency to engage in L2 communication (Papi & Khajavy, 2021; Sun et al., 2023).

These results support the proposal that learners motivated by various future selves show qualitative differences in their learning behavior and experience, which in turn contribute to qualitative differences in L2 learning outcomes. The findings, however, cannot be generalized to other populations until the study has been replicated in other contexts. In the present study, to examine the effects of motivation on L2 speech, multiple regression analyses were used, which have more statistical power and utility over mean-comparison methods (Plonsky & Oswald, 2017). The current design, however, does not allow for the elimination of the effects of intervening variables such as learning behaviors outside of the classroom or maturational effects. Due to logistical reasons, we could not have a control group in this study. The inclusion of a control group in future studies could thus enhance the rigor of the design and rule out the possibility of hidden variables affecting the results. The present study explored the motivational effects of L2 self-guides on L2 speech development; future studies can explore the role of the motivational constructs such as L2 learning experience (Taguchi et al., 2009) that focus on the quality of the learners' language learning experience. Ideal L2 Self/Other included only two items in the final CFA model. Even though the scale was still reliable (Eisinga et al., 2013), adding more items could strengthen the psychometric properties of the scale and more confidence in its construct validity. With a larger sample, more sophisticated statistical tools such as structural equation modelling could be used in future studies.

Author statement

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

Declaration of competing interest

We have no conflicts of interests to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2023.103156.

References

Apple, M. T., Da Silva, D., & Fellner (Eds.). (2016). *L2 selves and motivations in Asian contexts*. Multilingual Matters.

Baker-Smemoe, W., & Haslam, N. (2013). The effect of language learning aptitude, strategy use and learning context on L2 pronunciation learning. *Applied Linguistics, 34*, 435–456. https://doi.org/10.1093/applin/ams066

Bondarenko, A. V. (2020). *Self-efficacy as a generative mechanism for future self-guides and feedback-seeking behavior in language learning*. Doctoral dissertation. The Florida State University.

Cohen, J., Cohen, P., West, S., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Csizér, K., & Kormos, J. (2009). Modelling the role of inter-cultural contact in the motivation of learning English as a foreign language. *Applied Linguistics, 30*(2), 166–185. https://doi.org/10.1093/applin/amn025

DeKeyser, R. (2007). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). Cambridge University Press.

DeKeyser, R., & Larson-Hall, J. (2005). What does the Critical Period really mean? In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). Oxford University Press.

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning, 64*(3), 526–548. https://doi.org/10.1111/lang.12053

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29*, 359–380. https://doi.org/10.1093/applin/amm041

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). https://doi.org/10.21832/9781847691293-003. Multilingual Matters.

Dörnyei, Z., & Kubanyiova, M. (2014). *Motivating learners, motivating teachers: Building vision in the language classroom*. Cambridge University Press.

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Elliott, R. A. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania, 80*, 95–108. https://doi.org/10.2307/345983

Feng, L., & Papi, M. (2020). Persistence in language learning: The role of grit and future self-guides. *Learning and Individual Differences, 81*, Article 10194. https://doi.org/10.1016/j.lindif.2020.101904
Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of henning wode* (pp. 217–243). Wissenschaftlicher Verlag.
Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.
Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467–491. https://doi.org/10.1002/tesq.290
Galante, A., & Thomson, R. I. (2017). The effectiveness of drama as an instructional approach for the development of second language oral fluency, comprehensibility, and accentedness. *Tesol Quarterly, 51*, 115–142. https://doi.org/10.1002/tesq.290
Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. doi: 10.1177/0267658312461497.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Prentice Hall.
Harrington, D. (2008). *Confirmatory factor analysis*. Oxford University Press.
Hernandez, T. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal, 94*(4), 600–617. https://doi.org/10.1111/j.1540-4781.2010.01053.x
Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319–340. https://doi.org/10.1037/0003-066x.52.12.1280
Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280. https://doi.org/10.1037/0003-066x.52.12.1280
Hu, B., & West, A. (2015). Exam-oriented education and implementation of education policy for migrant children in urban China. *Educational Studies, 41*(3), 249–267.
Isaacs, T., & Thomson, R. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*(2), 135–159. https://doi.org/10.1080/15434303.2013.769545
Jiang, C., & Papi, M. (2021). The motivation-anxiety interface in language learning: Aregulatory focus perspective. *International Journal of Applied Linguistics, 32*(1), 25–40. https://doi.org/10.1111/ijal.12375
Kermad, A. A. D. (2018). *Speaker and listener variability in the perception of non-native speech.* Doctoral dissertation. Northern Arizona University.
Khajavy, G. H., & Ghonsooly, B. (2017). Predictors of willingness to read in English: Testing a model based on possible selves and self-confidence. *Journal of Multilingual and Multicultural Development, 38*(10), 871–885. https://doi.org/10.1080/01434632.2017.1284853
Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal, 97*, 720–744. https://doi.org/10.1111/j.1540-4781.2013.12029.x
Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
Lanvers, U. (2016). Lots of selves, some rebellious: Developing the self discrepancy model for language learners. *System, 60*, 79–92. https://doi.org/10.1016/j.system.2016.05.012
Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), 35–63. f10.1177/0267658307082981.
Lazaraton, A. (2013). Teaching oral skills. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 106–120). Heinle & Heinle.
Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly, 39*(3), 369–377. https://doi.org/10.2307/3588485
Li, S., Hiver, P., & Papi, M. (2022). Individual differences in second language acquisition: Theory, research, and practice. In L. Shaofeng, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 3–34). Routledge.
Magid, M., & Chan, L. (2012). Motivating English learners by helping them visualise their ideal L2 self: Lessons from two motivational programmes. *Innovation in Language Learning and Teaching, 6*(2), 113–125. https://doi.org/10.1080/17501229.2011.614693
Mayers, A. (2013). *Introduction to Statistics and SPSS in psychology*. Pearson.
Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *Tesol Quarterly, 46*(4), 610–641. https://doi.org/10.1002/tesq.34
Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics, 35*(4), 418–440. https://doi.org/10.1093/applin/amu012
Munoz, C., & Llanes, A. (2014). Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal, 98*, 432–449. https://doi.org/10.1111/j.1540-4781.2014.12059.x
Munro, M., & Mann, V. (2005). Age of Immersion as a predictor of foreign accent. *Applied PsychoLinguistics, 26*(3), 311–341. https://doi.org/10.1017/s0142716405050198
Nagle, C. (2013). A reexamination of ultimate attainment in L2 phonology: Length of immersion, motivation, and phonological short-term memory. In E. Voss, S.-J. D. Tai, & Z. Li (Eds.), *Selected proceedings of the 2011 second language research forum* (pp. 148–161). Cascadilla Proceedings Project.
Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development. *The Modern Language Journal, 102*, 199–217. https://doi.org/10.1111/modl.12461
Nagle, C. (2021). Assessing the state of the art in longitudinal L2 pronunciation research. *Journal of Second Language Pronunciation, 7*(2), 154–182. https://doi.org/10.1075/jslp.20059
Papi, M. (2018). Motivation as quality: Regulatory fit effects on incidental vocabulary learning. *Studies in Second Language Acquisition, 40*(4), 707–730. https://doi.org/10.1017/s027226311700033x
Papi, M. (2022). The L2 motivational self system: Using the selves in the classroom (pp. 91–98). In A. H. Al-Hoorie, & F. Szabó (Eds.), *Researching Language learning motivation: A concise guide. Bloomsbury*.
Papi, M., Bondarenko, A. V., Mansouri, S., Feng, L., & Jiang, C. (2019). Rethinking L2 motivation research: The 2 × 2 model of L2 self guides. *Studies in Second Language Acquisition, 41*(2), 337–361. https://doi.org/10.1017/S0272263118000153
Papi, M., Eom, M., Zhang, Y., Zhou, Y., & Whiteside, Z. (2023). Motivational dispositions predict qualitative differences in oral task performance. *Studies in Second Language Acquisition*, 1–26.
Papi, M., & Hiver, P. (2020). Language learning motivation as a complex dynamic system: A global perspective of truth, control, and value. *The Modern Language Journal, 104*(1), 209–232. https://doi.org/10.1111/modl.12624
Papi, M., & Khajavy, G. H. (2021). Motivational mechanisms underlying second language achievement: A regulatory focus perspective. *Language Learning, 71*(2), 537–572. https://doi.org/10.1111/lang.12443
Papi, M., & Teimouri, Y. (2012). Dynamics of selves and motivation: A cross-sectional study in the EFL context of Iran. *International Journal of Applied Linguistics, 22*(3), 287–309. https://doi.org/10.1111/j.1473-4192.2012.00312.x
Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning, 64*(3), 493–525. https://doi.org/10.1111/lang.12065
Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition, 39*(3), 579–592. https://doi.org/10.1017/s0272263116000231
Safdari, S. (2021). Operationalizing L2 motivational self system: Improving EFL learners' motivation through a vision enhancement program. *Language Teaching Research, 25*(2), 282–305. https://doi.org/10.1177/1362168819846597
Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness, 20*(1), 45–59. https://doi.org/10.1080/09658416.2010.540326

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*, 563–595. https://doi.org/10.1111/lang.12120

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning, 67*, 665–693. https://doi.org/10.1111/lang.12244

Saito, K., Dewaele, J.-M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning, 68*, 709–743. https://doi.org/10.1111/lang.12297

Saito, K., Dewaele, J.-M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech, 60*, 614–632. https://doi.org/10.1177/0023830916687793

Saito, K., Suzuki, S., Oyama, T., & Akiyama, Y. (2021). How does longitudinal interaction promote second language speech learning? Roles of learner experience and proficiency levels. *Second Language Research, 37*(4), 547–571. https://doi.org/10.1177/0267658319884981

Sampson, R. (2012). The language-learning self, self-enhancement activities, and self perceptual change. *Language Teaching Research, 16*(3), 317–335. https://doi.org/10.1177/1362168812436898

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*, 173–199. https://doi.org/10.1017/s0272263104262027

Sun, H., Saito, K., & Dewaele, J.-M. (2023). Cognitive and sociopsychological individual differences, experience, and naturalistic second language speech learning: A longitudinal study. *Language Learning*. https://doi.org/10.1111/lang.12561

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dornyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66–97). Multilingual Matters.

Tahmouresi, S., & Papi, M. (2021). Future selves, enjoyment and anxiety as predictors of L2 writing achievement. *Journal of Second Language Writing, 53*, Article 100837. https://doi.org/10.1016/j.jslw.2021.100837

Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. Studies in Second. *Language Acquisition, 39*, 691–709. https://doi.org/10.1017/s0272263116000243

Trofimovich, P., Kennedy, S., & Blanchet, J. (2017). Development of second language French oral skills in an instructed setting: A focus on speech ratings. *Canadian Journal of Applied Linguistics, 20*, 32–50. https://doi.org/10.7202/1042675ar

Uchihara, T. (2022). Is it possible to measure word-level comprehensibility and accentedness as independent constructs of pronunciation knowledge? *Research Methods in Applied Linguistics, 1*(2), Article 100011. https://doi.org/10.1016/j.rmal.2022.100011

Wang, J., & Rao, N. (2022). What do Chinese students say about their academic motivational goals - reasons underlying academic strivings? *Asia Pacific Journal of Education, 42*(2), 245–259. https://doi.org/10.1080/02188791.2020.1812513

Zhang, Y., & Papi, M. (2021). Motivation and second language pragmatics: A regulatory focus perspective. *Frontiers in Psychology, 12*, Article 753605. https://doi.org/10.3389/fpsyg.2021.753605

Zhou, Y. (2019). How motivation as quality influences ESL learners' attitudes toward English pronunciation learning. *International Journal of English Language Teaching, 6*(2), 33–51. https://doi.org/10.5430/ijelt.v6n2p33