RESEARCH ARTICLE

Disentangling the causal role of motivation, enjoyment, and anxiety in second language speech learning: A final report

Kazuya Saito[1] (iD), Jean-Marc Dewaele[1,2], Yo In'nami[3] and Mariko Abe[4]

[1]University College London, London, UK; [2]Birkbeck, University of London, London, UK; [3]Chuo University, Tokyo, Japan and [4]Okayama University, Tsushima-Naka, Okayama, Japan
Corresponding author: Kazuya Saito; Email: k.saito@ucl.ac.uk

Abstract

While our earlier report focused on the initial four months of the dataset (Saito et al., 2018, *Language Learning*), this study investigates the relationship between individual differences in motivation (Ideal Self and Ought–to Self), emotions (Enjoyment and Anxiety), and L2 speech learning among 121 Japanese English–as–a–Foreign–Language high school students over 1.5–years. Participants' L2 speech proficiency consistently improved at each testing point (6 months, 10 months, and 1.5 years), while their motivation and emotions, measured through questionnaires, remained relatively stable. The results of structural equation modeling suggest that the relationship between motivation, emotions, and acquisition may evolve. Within the first 6-10 months, data indicated a correlational relationship, highlighting a mutual influence among motivation, emotions, and acquisition. However, as the study progressed beyond one year, after students had fully adapted to their educational settings, a clearer causal relationship emerged: Enhanced motivation and more positive emotions were linked to increased classroom practice, leading to significant gains in L2 speech proficiency. The predictive roles of Anxiety remained unclear in this longitudinal dataset.

Introduction

Second language (L2) acquisition researchers have become increasingly interested in the complex relationship between learners' motivations and emotions, and their combined influence on L2 speech learning outcomes in classroom settings. A growing body of cross–sectional evidence suggests that individuals with greater motivation and more positive emotions are more willing to communicate in the L2 and more inclined to engage in extended L2 practice, often resulting in advanced proficiency (Botes,

Dewaele, & Greiff, 2022a; Dörnyei, 2020). Recent developments in this area of study have seen a growing emphasis on longitudinal investigations (Li & Wei, 2023). These studies have shed light on the dynamic interplay between motivation, enjoyment, anxiety, and L2 outcomes over a specific time frame, typically spanning one academic term (3-4 months). We argue that to disentangle the causal relationships between motivation, enjoyment, anxiety, and acquisition, the topic needs to be reexamined in relation to a wider range of potentially important factors over a relatively long period (> 1 semester). This paper presents a longitudinal study focusing on 121 Japanese high school learners of English–as–a–Foreign–Language (EFL). It explores how these EFL learners engaged in L2 English practice in the classroom setting, and how their motivation, emotions, and L2 speech proficiency evolved over a period of 1.5 years. Here, we aim to empirically test the four hypotheses concerning the directionality of this relationship:

1. Changes in motivation and emotions serve as a driving force for L2 speech learning (motivation and emotions → L2 speech learning).
2. Enhanced L2 speech learning fosters increased levels of motivation and enjoyment and lower levels of anxiety (L2 speech learning → motivation and emotions).
3. Motivation, emotions, and L2 speech learning are interwoven and correlated with each other without clear directionality (motivation/emotions ↔ L2 speech learning).
4. The nature of the relationship evolves, shifting from correlational in the short term to causal in the long term.

Background

L2 speech learning in EFL classrooms

To date, extensive research has been dedicated to understanding the factors influencing the success of L2 speech learning in naturalistic settings, demonstrating that adult L2 learners can continue to acquire new sounds as long as they are immersed in sufficient daily input across various contexts (Flege & Bohn, 2021). A growing amount of attention has been directed to exploring the applicability of these findings to *foreign language* learning contexts, where learners typically receive a few hours of language instruction weekly—termed the "minimal input" condition (Larson-Hall, 2008, p. 36). In these studies, L2 speech proficiency outcomes have been assessed using diverse metrics, such as national standardized test scores (e.g., BISTA assessment; Baumert et al., 2020), native speakers' judgments of overall impression (e.g., overall foreign accentedness and comprehensibility; Nagle, 2018), and specific pronunciation tests (e.g., English [r] and [l] for Japanese learners of English; Saito, 2019). Factors traditionally linked to successful outcomes include the age of learning (Larson-Hall, 2008), duration of learning (Jaekel et al., 2017), and participation in extracurricular activities (e.g., Muñoz, 2014 for study abroad; Saito & Hanzawa, 2016 for conversation activities).

While these experience–related factors account for small–to–medium effects (e.g., r = .30-.40; Saito & Hanzawa, 2016), they do not fully explain all variances. This observation has led some researchers to propose that beyond external experiences, intrinsic individual differences significantly impact L2 speech learning success (Trofimovich et al., 2015). Specifically, when learners with similar exposure and practice opportunities are compared, those with certain motivational and emotional profiles may leverage each learning opportunity more effectively, resulting in greater

gains (Moyer, 2014; Nagle, 2018; Saito et al., 2017; Zhou & Papi, 2023). These researchers seem to share an assumed logical sequence: greater motivation and more positive emotions lead to increased L2 practice and experience, which in turn leads to acquisition. This investigation aims to scrutinize the interplay between motivation and emotions in L2 speech learning.

Cross–sectional investigations of motivation, emotions, and L2 learning

Over the last two decades, various conceptual frameworks have been developed to elucidate the complex nature of L2 learners' motivation and emotion profiles. Dörnyei's L2 Motivational Self System has become the default framework in Second Language Acquisition research (Dörnyei, 2005, 2009, 2020). It posits that L2 learners' motivational dispositions can be conceptualized as the Ideal L2 Self (attributes the learner ideally aspires to attain) and the Ought–to L2 Self (attributes the learner believes are necessary to meet external expectations, duties, and obligations). In more recent developments, this self-system has been expanded upon with the introduction of the 2×2 model of future self-guides, delineating the Ideal Self/Own (with promotion focus) versus Others and Ought–to L2 Self/Own (with prevention focus; Papi et al., 2019). L2 learners with Ideal L2 Self/Own are likely to practice the target language and maximize each practice opportunity (eager L2 use) while those with Ought–to L2 Self/Own tend to resort to the minimal use of the target language to avoid making mistakes (vigilant L2 use; Papi & Khajavy, 2021).

Cross–sectional evidence has consistently demonstrated a stronger correlation between the motivation profiles related to Ideal L2 Self and higher levels of L2 proficiency (measured via general proficiency test scores) and achievement (operationalized via in–house exam results), in contrast to the motivation profiles associated with the Ought–to L2 Self (Dörnyei & Chan, 2013; Papi & Khajavy, 2021; Lamb, 2012; Moskovsky et al., 2016; Papi & Teimouri, 2014). This distinction between the roles of the Ideal and Ought–to L2 Self in L2 learning outcomes is further supported by meta–analytic evidence (Al-Hoorie, 2018; $r = .20$ vs. -.05).

The last decade of L2 acquisition research has witnessed a significant transition in scholarly focus. The traditional exclusive focus on negative emotions (i.e., Foreign Language Classroom Anxiety [FLCA]; Horwitz et al., 1986) has been replaced by a more holistic perspective that includes both negative and positive emotions (i.e., Foreign Language Enjoyment [FLE]; Dewaele & MacIntyre, 2014). Researchers have become aware that learner emotions do not emerge in isolation but are part of a highly dynamic system including interacting linguistic, social, and psychological factors (Dewaele et al., 2018).

A meta-analysis demonstrated that FLE and FLCA are independent dimensions that are moderately negatively correlated ($r = -.31$) (Botes et al., 2022a, p. 214). Although FLCA and FLE are linked to several similar learner–internal sociobiographical variables (age, education, proficiency level, multilingualism), they are also associated with distinct learner–external variables. FLE tends to be linked with classroom factors, such as relationships with teachers and peers (Dewaele et al., 2018) and with teaching methods (Dewaele et al., 2024). In contrast, FLCA is more strongly related to a learner's personality traits such as neuroticism (Dewaele & MacIntyre, 2019). Meta–analytic reviews have identified a small–to–medium negative correlation between FLCA and general academic achievement in an L2 ($r = -..36$ in Teimouri et al., 2019; $r = -.39$ in Botes et al., 2020) and a moderate positive correlation between FLE and both self–reported achievement and academic achievement ($r = .27$ and $r = .30$, respectively;

Botes et al., 2022a). A stronger correlation emerged between FLE and Willingness to Communicate (*r* = .48; Botes et al., 2022a).

The relationship between motivation and emotions has been the focus of intense debate in the field of psychology and applied linguistics. They are different in nature but are connected to some extent (MacIntyre et al., 2019; Dörnyei, 2020). In contrast with emotions, motivation is specifically goal-oriented and reflects the effort and investment learners put into the FL learning process and in the development of a new identity (Ushioda & Dörnyei, 2009). Dörnyei (2020) argues that some emotions may have "a certain amount of goal-directed quality" (p. 121) but that they are not directly linked to goal–specific action like motivation. At most they may have vague "action tendencies" (p. 121). He disagrees with MacIntyre et al. (2019) that positive emotions are intrinsically motivating and negative emotions demotivating. He sees emotions as occupying the backseat in the learning process: "they can sustain and amplify existing motivation […] and they can also instigate the generation of new, goal–directed behavioral scripts (i.e., motives proper)" (Dörnyei, 2020, p. 121–122). Dörnyei and Henry (2022) view emotions as the fuel that sustains motivation. This view is not unlike the one proposed in Dewaele et al. (2023) where FLE emerged as a buoy that could help sustain sagging motivation.

The current study adopts a perspective where motivation and emotions are intertwined, analyzing sociopsychological individual differences through both motivational constructs (Ideal and Ought–to L2 Self) and emotional aspects (FLCA and FLE).

Longitudinal investigations of motivation, emotions, and L2 learning

Researchers have been increasingly interested in the dynamic nature of motivation and emotions and how they connect to L2 learning. Longitudinal research designs have been used with multiple data collection points over various periods, ranging from a couple of days (Guedat-Bittighofer & Dewaele, 2023) to one academic semester (Elahi Shirvan & Taherian 2021). Several longitudinal studies have investigated how participants' motivation and emotion profiles at the outset of the project shaped their L2 learning gains. For instance, Saito et al. (2017) explored how 40 Japanese EFL students with diverse motivation profiles, including integrativeness, instrumentality, and international posture, developed their L2 speech proficiency (comprehensibility) over the course of one academic semester (3 months). The results indicated that those with vague, long–term goals for learning L2 English improved their comprehensibility (although not nativelikeness) in L2 speech proficiency.

Similarly, by tracking the speech development of 83 Chinese EFL students over one academic semester, Zhou and Papi (2023) discovered that the strength of participants' self-images as their ideal selves could predict the development of L2 comprehensibility (see also Nagle, 2018, involving English speakers learning Spanish). Interestingly, these findings observed in classroom settings were not replicated in naturalistic environments. For example, Sun et al. (2024) did not find the Ideal Self to be a significant predictor for L2 speech development among 50 Chinese learners of English during their four–month study abroad period.

Regarding the relationship between emotions and L2 learning gains, there are short–term intervention studies (1-2 hours) suggesting that individuals experiencing greater negative emotions (anxiety), tend to show less learning gain when receiving instruction (e.g., Sheen, 2008; Miller & Godfroid, 2020). As for longitudinal investigations, Li and Wei (2023) investigated how positive and negative emotions related to L2 achievement,

measured through in–house English tests, among a total of 954 junior secondary English learners at four intervals (T1-T4) over the course of one semester (3 months). Structural equation modeling revealed that these emotions independently showed associations with English achievement at T2 (one week later) and T3 (five weeks later). However, only FLE was significantly associated with achievement at T4 (nine weeks later), while the significant associations between FLCA and acquisition disappeared at the end of the project.

Motivation for the current study

Existing literature has provided both cross–sectional and longitudinal evidence supporting a significant association between motivation, FLE, FLCA, and the rate of L2 speech learning (and general L2 learning proficiency and outcomes). However, only a few studies have explored the directionality of the link between motivation, FLE, and FLCA on the one hand and L2 learning on the other hand. Such an investigation could be directly relevant to the theoretical core of major motivation and emotions frameworks (e.g., L2 Motivational Self System, FLCA, and FLE). These frameworks address the relationship between different types of motivation and emotions and their differential impacts on L2 learning experience and development. Considering the nascent literature on the topic, it seems that three different positions have emerged:

1. **Motivation and emotions as a driving or inhibiting force for L2 learning (see Figure 1 Model 1)**: The first position posits that motivation and emotions are key driving forces for L2 learning. It suggests that learners, when engaged in practice opportunities (i.e., experience) with stronger motivation, strong positive and weak negative emotions can maximize each input opportunity, leading to increased learning gains (e.g., Moskovsky et al., 2016; Papi, 2018; Papi & Teimouri, 2014).
2. **Enhanced motivation and positive emotions through L2 learning (see Figure 1 Model 2)**: The second perspective argues that greater L2 learning experience and development can lead to enhanced motivation and enjoyment and reduce anxiety (Botes et al., 2020, 2022a). For example, incorporating real–life task activities can boost motivation (Heydarnejad et al., 2022), and more meaning–focused activities can increase enjoyment (Dewaele et al., 2024).
3. **Interdependence of motivation, emotions, and L2 learning (see Figure 1 Model 3)**: The third perspective suggests that the relationship between motivation and emotions and L2 learning is intertwined, with multidirectional causality and non-linear trajectories. Within this position, some emphasize strong interdependence (i.e., correlations) between motivation, emotions, and L2 learning without specifying directionality (e.g., Dörnyei, 2005), while others underline the time–specific nature of the relationship. Motivation and emotions may initially drive the learning, but once a certain level of proficiency is attained, the newly developed abilities can strengthen motivation and positive emotions in a positive feedback loop while reducing negative emotions (e.g., Li et al., 2022).

In addition to examining which of these three models best explains data, it is crucial to adopt a long–term longitudinal design to investigate the directionality of the relationship between motivation, emotions, and acquisition. This is because L2 learning is often slow, gradual, and dynamic, especially in countries where learners have limited

A. Causality model 1 (moti vation/emotions → acquisition)
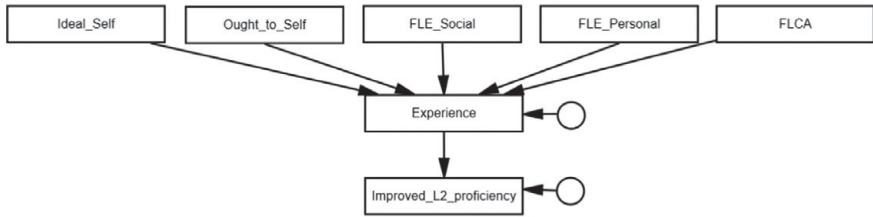
B. Causality model 2 (acquisition → motivation/emotions)
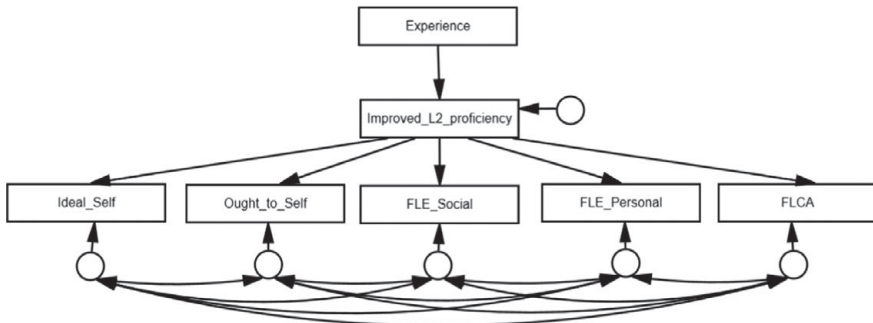
C. Correlational model 3 (motivation/emotions ↔ acquisition)
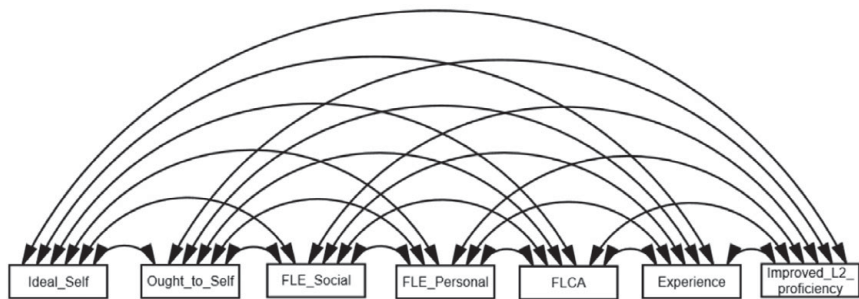

Figure 1. Visual Summary of Model Comparisons. FLE_Social for Foreign Language Enjoyment Social; FLE_Personal for Foreign Language Enjoyment Personal; FLCA for Foreign Language Classroom Anxiety; Experience for percentage of L2 use inside classrooms.

access to L2 input outside the classroom. The investigation of the directionality between motivation, emotions, and acquisition is both theoretically and pedagogically relevant. Educators need to know whether enhancing and maintaining students' motivation and positive emotions while reducing their negative emotions should be a priority, as it can anchor and accelerate L2 learning in the long run (motivation/emotions → learning). Alternatively, helping students perceive their progress first may be the best remedy for boosting motivation and emotions (learning → motivation/emotions). It is also possible that targeting both motivation, emotions, and L2 development as equal pedagogical focuses could lead to an optimal L2 learning loop (motivation, emotions ↔ learning).

Saito et al. (2017) and Zhou and Papi (2023) have tracked L2 learning behaviors at multiple points while measuring motivation and emotions only once. Li and Wei

(2023) took a different approach, examining the links between emotion and learning outcomes at five testing points, revealing limited changes in emotion and L2 proficiency over time. Similarly, other studies conducted longitudinal investigations of changes in motivation and emotions, highlighting the relatively stable nature of these variables (e.g., Dewaele et al., 2023; Elahi Shirvan & Taherian, 2021). The lack of significant change in motivation, emotions, and L2 outcomes may be attributed to the relatively short–term timeframe of these studies (one academic semester: 3-4 months). The current study is designed to address these methodological concerns, looking at the longitudinal relationship between *changes* in motivation, emotions, and L2 speech proficiency for an *extensive* period of classroom experience (i.e., 1.5 years).

Current study

In this research project, we explored the interplay between motivation, emotions, and L2 speech proficiency in a cohort of 121 Japanese high school students over a period of 1.5 years. The study was conducted at a well–regarded public high school located in northern Japan. Data collection was longitudinal, occurring at four intervals. We initially assessed the students' L2 speech proficiency shortly after they commenced high school, providing sufficient time for their adaptation to the new learning environment (T1: Summer 2016). Further assessments of L2 speech proficiency were conducted at intervals of 6 months (T2: Winter 2016), 10 months (T3: Spring 2017), and 18 months (T4: Winter 2018). In addition, surveys to evaluate the students' profiles in terms of motivation, emotions, and language learning experiences were conducted at T2, T3, and T4.

In our initial report based on data from T2 and T3 (Saito et al., 2018), we examined the predictive capacity of motivation and emotions for short–term L2 speech development over a 4–month period. The current final report extends the analysis to the entire duration of 1.5 years, aiming to clarify the directionality of the relationship between *changes* in motivation, emotions, and L2 speech development over a long period. Here, we were particularly interested in three different timings which were assumed to represent three distinct phases of L2 speech learning within the classroom context:

- **T1-T2 (first 6 months):** This short–term period represents how quickly participants can improve their L2 speech while adjusting to the new learning environment.
- **T1-T3 (10 months):** This mid–term period reflects how much proficiency they can achieve after spending sufficient time adapting to the learning environment.
- **T1-T4 (1.5 years):** This long–term period captures the extent to which they continue improving their L2 speech proficiency over the long run.

To explore how changes in motivation, emotions, and L2 speech proficiency evolved at different time points (T1, T2, T3, T4), we employed a two–step approach. The first step involved assessing the magnitude of change in motivation, emotions, and L2 speech proficiency over the research period. This was examined using the mean–based analysis —one–way repeated ANOVAs. Subsequently, we examined the interdependencies between shifts in motivation, emotions, and L2 speech proficiency, utilizing relative gain scores. To this end, the variance-based approach—structural equation modeling (SEM)—was employed to determine which theoretical model in Figure 1 most aptly fits the observed data at each time point.

The central research question probes the dynamic interaction between young EFL students' motivation, emotions, and their L2 speech proficiency, particularly in the context of their language practice in their classroom. We posited three potential patterns of interaction:

1. Enhanced motivation, positive emotions catalyzing increased L2 speech development (motivation/emotions → acquisition; Moskovsky et al., 2016; Papi, 2018)
2. Advancements in L2 speech proficiency bolstering motivation/emotions (acquisition → motivation/emotions; Botes et al., 2020, 2022a; Heydarnejad et al., 2022)
3. A bidirectional relationship between changes in motivation/emotions and L2 speech proficiency (motivation/emotions ↔ acquisition; Dörnyei, 2005; Li et al., 2022)

Given that our longitudinal data highlight three distinct phases of L2 speech learning trajectories (6 months, 10 months, and 1.5 years), we also introduced a fourth possibility:

4. The correlational versus causal nature of the relationship between motivation, emotions, and L2 speech learning may vary depending on the specific time point in the learning process within classroom settings, with motivation and emotions serving as crucial driving forces, especially for long–term L2 speech development (Moyer, 2014).

Method

Participants

A total of 121 first–year high school students (ages 15-16) were recruited. They included 50 male and 71 female participants. Their backgrounds in English as a Foreign Language (EFL) were diverse, as indicated by their varied age of EFL onset (M = 10.0 years, SD = 3.2, *Range* = 3-13 years), total duration of EFL study (M = 874.1 hours, SD = 564.5, Range = 150-4,500 hours), and length of extracurricular EFL instruction, such as study school attendance (M = 492.0 hours, SD = 564.5, *Range* = 0-3,150 hours). Given that all students were admitted to the high school based on a uniform entrance examination, their English proficiency levels were relatively homogeneous. Their general proficiency, as measured by the EIKEN Test in Practical English Proficiency, ranged between Grade Pre-2 and Grade 2. This level of proficiency corresponds to A2 (basic user) and B1 (independent user) on the Common European Framework of Reference for Languages scale (EIKEN Foundation of Japan, 2017).

During the project duration, from Summer 2016 to Winter 2018, the students were taught according to a standard EFL syllabus. They were required to attend seven EFL classes each week, with each session lasting 50 minutes. The classes were delivered by three Japanese teachers, each possessing very advanced proficiency in English. Comprehensive classroom observations were undertaken, and the details of these observations are presented in the Supporting Information-S1. As previously established in EFL studies (e.g., Larson-Hall, 2008; Muñoz, 2014), participants' L2 learning experience profiles were collected using self–report questionnaires at three different time points (T2, T3, and T4).

To obtain ethical approval from both participating students and their parents, a set of steps was carefully taken. First, the teachers invited the students' parents to attend a

meeting where the investigators of the project explained the academic purpose of the project (i.e., tracking high–school students' L2 speaking proficiency development) and the precautionary measures taken for personal information protection. After obtaining the parents' consent, the teachers explained the same details to the students before they took the speaking tests and the experience and motivation/emotion questionnaires. Once the students agreed to participate, they signed the consent form at the end of the questionnaires.

Motivation and emotions questionnaires

To capture participants' motivation and emotions, a composite questionnaire was administered. The questionnaire comprised 8 items for motivation and 18 items for emotion. It was specifically designed to measure the distinct aspects of motivation as conceptualized in Dörnyei's L2 Motivational Self System (Dörnyei, 2005), which includes Ideal Self and Ought–to Self, and FLCA as well as FLE outlined in Dewaele and MacIntyre (2014). Although a more recent model of motivation, the 2×2 model of future self-guides, was proposed by Papi et al. (2019), this could not be incorporated into the questionnaire due to the timing of data collection (2016 to 2018). Similarly, the recent conceptualization of FLE distinguishes one superordinate and three lower–order dimensions, FLE Teacher, FLE Social, and FLE Personal (Botes et al., 2020).

In the current study, FLE was operationalized for Social ($n = 5$ items) and Personal ($n = 5$ items) based on Dewaele and MacIntyre's (2014) Principal Components Analysis of the 21 FLE items. FLE_Social refers to good social relationships with peers and teachers while FLE_Personal refers to "an internal sense of enjoyment in the face of challenges" (p. 231). The decision to focus on the two lower–order dimensions was based on the need for a higher level of granularity and previous findings that FLE_Personal was a better predictor of gain than FLE_Social (Dewaele & MacIntyre). All items were based on a 6-point scale.

FLCA was measured through eight items first used by Dewaele and MacIntyre (2014) and later validated in Botes, Westhuizen et al. (2022b). The items refer to mild–to–strong physical symptoms of anxiety and to manifestations of social anxiety. Participants responded to each statement on a 6-point scale. The reliability of the composite questionnaire was checked via the Cronbach Alpha coefficient. Given that all the reliability coefficients exceeded .70 (see Table 2), they could be considered acceptable in relation to the field–specific guidelines (Larson-Hall, 2015). The composite questionnaire is deposited in L2 Speech Tools and is publicly available for replication (Mora-Plaza et al., 2022: https://sla-speech-tools.com/).

L2 speech proficiency measures

To collect and assess a relatively large number of participants' speech within a short period without any significant delay (within 10 days), all participants took an automated English–speaking test, Telephone Standard Speaking Test (TSST) developed by ALC Press Inc. in Japan. The TSST was modeled after the Oral Proficiency Interview format established by the American Council on the Teaching of Foreign Languages. Scores range from one to nine, with nine being the highest score. Comprehensive validation studies have been conducted to confirm the reliability and validity of the TSST (for a comprehensive summary, see Koizumi, 2017, 2021). For instance, notable

findings regarding the TSST's validation include high test–retest reliability (r = .73), strong agreement with the face–to–face Standard Speaking Test (r = .89), and significant correlations with external proficiency tests (ranging from .6 to .8; e.g., EIKEN, TOEFL). This validation evidence underscores the TSST's effectiveness and reliability in evaluating the speaking proficiency of English language learners in an automated and efficient manner.

Using their home phones or cell phones at their convenience, participants were instructed to take the exam in a quiet room to ensure their speech was clear for accurate ratings, as background noise can negatively impact raters' assessments. The testing agency monitored participants' progress and completion of the test. Their test attendance and scores were factored into their final assessment by their EFL teachers. In cases where technical issues arose (e.g., unstable cell phone signals), participants were permitted to retake the test, but only once. Such instances were exceedingly rare, with only one participant failing and consequently dropping out of the study. To facilitate the tracking and comparison of participants' test scores over time, they were given a deadline to complete the test, typically within 10 days of receiving the phone number for the test.

Each test session lasted 15 minutes, during which participants responded to 10 recorded questions. These questions were randomly selected from a diverse databank, ensuring exposure to varied content at each testing point (T1 to T4). Designed to elicit spontaneous speech, participants had no prior knowledge of the questions and were given 45 seconds to formulate and provide a detailed answer for each. To eliminate any potential confusion regarding task instructions, they were provided in both English and Japanese. The questions varied in response type, including three narrative, three descriptive, and four reasoning questions. Each speech sample submitted through the TSST was evaluated by three trained judges. This tripartite evaluation approach ensured high reliability and agreement among the judges (Koizumi, 2017). The structured and monitored testing procedure, combined with the standardized evaluation by trained judges, contributed to the validity and reliability of the speech proficiency assessments in the study.[1]

---

[1]One way to examine the validity of TSST as a measure of L2 speaking proficiency is to check its relationship with listeners' perceived comprehensibility. Comprehensibility is measured via listeners' intuitive overall impression of ease of understanding without any prescribed descriptors and is considered arguably the most important factor for successful L2 communication in real-life settings (Saito, 2021). As proposed and demonstrated by Isaacs, Trofimovich et al. (2015), comprehensibility is a key predictor for IELTS Speaking Proficiency test scores. From an acquisitional point of view, comprehensibility has been identified as a significant predictor of adult L2 speech learning in various contexts, including both naturalistic and classroom settings (e.g., Huensch & Nagle, 2023). In our prior report, the same participants' speech samples at T2 and T3 were evaluated by our team of experienced L1 English judges. These judges assessed the intuitive perception of comprehensibility using a 9-point scale. The correlation between these comprehensibility ratings and the TSST scores was found to be quite high, i.e., r = .87 and .89 for T2 and T3, respectively. The strong correlation suggests that the judges' evaluations of comprehensibility closely aligned with the TSST scores. Based on this relationship, it is reasonable to assert that the TSST scores can serve as an adequate measure of longitudinal L2 speech development for the purposes of this investigation. The consistently high correlation between the TSST scores and native-speaker judgments of comprehensibility underscores the reliability of TSST as a tool for assessing progress in L2 speech proficiency over time.

EFL experience questionnaire

Participants' experiences of learning L2 English were measured using a questionnaire adapted from the EFL Experience Questionnaire (Saito et al., 2017). While a range of items were surveyed in the questionnaire, the most relevant experience variable for the current study—participants' recent experience of learning L2 English in classroom settings—was included in the analysis. To quantify the degree to which they practiced L2 English, participants were asked to report the percentage of time spent using L2 English during English classes (for similar methods, see Larson-Hall, 2008; Muñoz, 2014).

In the current investigation and analysis, the participants' experience outside classrooms was not included. The participants' experiences outside classrooms varied substantially in many different ways: modes (conversational activities with L1 vs. L2 English speakers; reading vs. listening; presence vs. absence of after–school activities) and quantity (some reporting zero hours, others reporting up to 100 hours per week). As a result, each dataset was neither normally distributed nor comparable. Given that the sample size (n = 121 students) was rather small for SEM analyses, we decided to focus on one key experience variable most relevant to the project (i.e., L2 use inside classrooms) and not to include multiple extracurricular experience variables to avoid weakening statistical power. For open science purposes and for those interested in follow–up analyses, the participants' L2 use outside classrooms is available in the shared dataset. Supplementary materials are available via the Open Science Framework (https://osf.io/tacuy/?view_only=4d07da7d753a4a00915a15cf27db549b).

Results

Measurement invariances

To examine the degree to which constructs were properly measured over time, we investigated the longitudinal measurement invariance of motivation (i.e., Ideal L2 Self and Ought–to L2 Self) and emotions (i.e., FLE [Social, Personal] and FLCA). Specifically, each subconstruct was separately modeled as a latent variable at each time point, and the data–model fit was assessed. For example, to investigate the measurement invariance of Ideal L2 Self, we modeled it as a latent variable loaded on the four observed variables (for these variables or items, see L2 Speech Tools [https://sla-speech-tools.com/]) at each time point. The model was tested for (a) configural invariance (i.e., whether the one-factor model with no equal constraints on parameters across time fit the data well), (b) weak invariance (i.e., whether the model fit the data well, with factor loadings specified to be equal across time), and (c) strong invariance (i.e., whether the model fit the data well, additionally with intercepts specified to be equal across time). Thus, parameters were gradually constrained to be of equal size, with no constraints on configural invariance, constraints on factor loadings for weak invariance, and constraints on factor loadings and intercepts for strong invariance. Satisfying these measurement invariances suggests that changes between time points are related to the latent variable being modeled. Following the procedures in Kline (2023), Nagle (2023), and Vandenberg and Lance (2000), the measurement invariance was analyzed for each construct/sub-construct (Motivation [Ideal, Ought-to], Enjoyment[Social, Personal], and Anxiety) at each time point (T2, T3, and T4). In the current investigation, we considered constructs with levels of configural, weak, and strong invariance as adequate for longitudinal analyses. However, we exercised caution when

Table 1. Descriptive statistics for items measuring motivation and emotions

	T2 (6 months from T1)		T3 (10 months from T1)		T4 (1.5 years from T1)	
	M	SD	M	SD	M	SD
Motivation						
Ideal L2 Self Item 1	3.21	1.55	3.34	1.53	3.19	1.57
Ideal L2 Self Item 2	3.68	1.53	3.62	1.46	3.77	1.43
Ideal L2 Self Item 3	3.64	1.49	3.74	1.49	3.78	1.38
Ideal L2 Self Item 4	3.14	1.66	3.26	1.55	3.16	1.50
Ideal L2 Self Items 1–4[a]	3.41	1.35	3.49	1.33	3.48	1.30
Ought–to L2 Self Item 1	2.88	1.54	2.88	1.49	3.21	1.38
Ought–to L2 Self Item 2	2.85	1.60	3.09	1.61	3.00	1.59
Ought–to L2 Self Item 3	2.88	1.52	3.15	1.52	3.11	1.42
Ought–to L2 Self Item 4	2.70	1.68	2.85	1.61	2.76	1.47
Ought–to L2 Self Items 1–4[a]	2.83	1.17	2.99	1.18	3.02	1.14
Emotions						
FLE Social Item 1	4.43	1.37	4.51	1.26	4.38	1.13
FLE Social Item 2	4.53	1.15	4.45	1.20	4.20	1.10
FLE Social Item 3	5.02	1.10	4.96	1.19	4.81	1.01
FLE Social Item 4	4.80	1.08	4.62	1.16	4.51	0.88
FLE Social Item 5	4.54	1.29	4.45	1.09	4.24	1.01
FLE Social Items 1–5[a]	4.66	.97	4.60	.94	4.43	.83
FLE Personal Item 1	4.50	1.14	4.44	1.14	4.34	1.06
FLE Personal Item 2	4.47	1.18	4.51	1.05	4.21	1.03
FLE Personal Item 3	3.85	1.26	3.96	1.27	4.02	1.22
FLE Personal Item 4	4.13	1.26	4.32	1.13	4.11	1.04
FLE Personal Item 5	5.51	.93	5.35	1.10	5.29	1.01
FLE Personal Items 1–5[a]	4.49	.84	4.52	.81	4.39	.84
FLCA Item 1	3.07	1.45	3.07	1.37	3.24	1.38
FLCA Item 2	4.94	1.22	4.92	1.14	4.70	1.17
FLCA Item 3	3.74	1.49	3.67	1.43	3.53	1.41
FLCA Item 4	3.45	1.43	3.29	1.43	3.39	1.28
FLCA Item 5	4.08	1.29	4.02	1.29	3.87	1.30
FLCA Item 6	3.50	1.43	3.53	1.37	3.43	1.28
FLCA Item 7	3.05	1.41	3.15	1.45	2.91	1.24
FLCA Item 8	2.68	1.37	2.62	1.33	2.82	1.45
FLCA Items 1–8[a]	3.56	1.00	3.53	.94	3.51	1.00

Note: The questionnaires to measure motivation and emotions were administered at Times 2 through 4. They were not administered at Time 1. Each item is on a scale of 1 through 6. Item numbers correspond to those in L2 Speech Tools https://sla-speech-tools.com/ (Composite Motivation & Emotions Questionnaire). For example, Ideal L2 Self Item 1 in the table above corresponds to "I imagine myself as someone who is able to speak English." in the supplementary material. FLE for Foreign Language Enjoyment; FLCA for Foreign Language Classroom Anxiety. [a]Grand means by construct.

constructs failed to reach at least the configural level of invariance, as this indicated that what comprised these constructs may have varied at each time point.

The descriptive statistics for each item in Table 1 show similar means and standard deviations over time, particularly for motivation, suggesting little change longitudinally. For example, for Ideal L2 Self Item 1, these values were 3.21 and 1.55 at Time 2, 3.34 and .53 at Time 3, and 3.19 and 1.57 at Time 4. The results for measurement invariance in Table 2 were mixed depending on the constructs. For motivation, both Ideal L2 Self and Ought–to L2 Self showed strong invariance. For emotions, FLE Social showed strong invariance, whereas FLE Personal showed only configural invariance. The results suggest that the components of Ideal L2 Self, Ought–to L2 Self, and FLE (i.e., how the questionnaire responses factored into these constructs) remained relatively stable and longitudinal changes in Ideal L2 Self, Ought–to L2 Self, and FLE

Table 2. Fit indices for the tests of measurement invariance of motivation and emotions

	χ^2	df	p	CFI	RMSEA (90% CI)	SRMR	Fit?	χ^2 difference test	Retained?
Fit criteria	–	–	> .05	.90	< .08	< .08	–	> .05	–
Motivation									
Ideal L2 Self									
Configural[a]	44.148	39	.263	1.000	.000 (.000, .076)	.037	Yes	–	Yes
Weak[b]	48.256	45	.343	1.000	.025 (.000, .065)	.040	Yes	p = .764	Yes
Strong[c]	54.575	51	.340	1.000	.000 (.000, .064)	.042	Yes	p = .389	Yes
Ought–to L2 Self									
Configural[a]	41.320	39	.370	.997	.018 (.000, .079)	.049	Yes	–	Yes
Weak[b]	45.740	45	.441	1.000	.006 (.000, .072)	.056	Yes	p = .660	Yes
Strong[c]	53.485	51	.379	.997	.017 (.000, .070)	.059	Yes	p = .226	Yes
Emotions									
FLE Social									
Configural[a]	102.403	72	.011	.965	.066 (.030, .095)	.059	Yes	–	Yes
Weak[b]	116.706	80	.005	.958	.067 (.036, .094)	.074	Yes	p = .059	Yes
Strong[c]	124.149	88	.007	.960	.063 (.031, .089)	.075	Yes	p = .557	Yes
FLE Personal									
Configural[a]	134.256	72	.000	.920	.091 (.063, .117)	.086	Yes	–	Yes
Weak[b]	146.270	80	.000	.915	.089 (.063, .114)	.099	No	p = .131	No
Strong[c]	166.563	88	.000	.898	.093 (.068, .116)	.102	No	p = .006	No
FLCA									
Configural[a]	392.375	225	.000	.870	.086 (.071, .101)	.080	No	–	No
Weak[b]	417.812	239	.000	.863	.086 (.071, .100)	.091	No	p = . 029	No
Strong[c]	435.099	251	.000	.859	.085 (.071, .099)	.090	No	p = .151	No

Note: df = degrees of freedom. CFI = comparative fit index. RMSEA = root mean square error of approximation. CI = confidence interval. SRMR = standardized root mean square residual; [a]The configural invariance model specifies the same factor structure with no equal constraints on parameters across time; [b]The weak (i.e., metric) invariance model specifies the same factor structure with equal factor loadings across time; [c]The strong (i.e., scalar) invariance model specifies the same factor structure with equal factor loadings and intercepts across time. FLE for Foreign Language Enjoyment; FLCA for Foreign Language Classroom Anxiety.

(Social, Personal) were considered to reflect changes in these constructs. In contrast, FLCA showed none of these invariances. Any observations related to longitudinal changes in FLCA in the current dataset should be interpreted cautiously since the changes could be due to the measurement nonequivalence of questionnaire items, not actual changes in the constructs.

Motivation, emotions, and L2 speech development

Table 3 summarizes statistics for motivation, emotions, experience, and L2 speech proficiency profiles over time. The initial aim of the statistical analyses was to evaluate the overall development of participants' motivation (Ideal L2 Self and Ought–to L2 Self), emotions (FLE [Social, Personal] and FLCA), and L2 speech proficiency. For motivation, FLE and FLCA, one–way repeated measure ANOVAs were conducted, using Ideal L2 Self, Ought–to L2 Self, FLE [Social, Personal], and FLCA scores as dependent variables and Time as the independent variable (T2-T4). The results indicated no significant development in terms of Ideal L2 Self, $F(2, 202) = .763$, $p = .468$, $\eta^2 = .022$, Ought–to L2 Self, $F(2, 202) = 2.803$, $p = .063$, $\eta^2 = .027$, FLE Social, F

Table 3. Summary statistics for motivation, emotions, experience, and L2 speech proficiency profiles over time

Variables	T1			T2 (6 months from T1)			T3 (10 months from T1)			T4 (1.5 years from T1)		
	M	*SD*	α	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*	α
Motivation (6-point)												
Ideal L2 Self	–	–	–	3.41	1.35	.82	3.49	1.33	.83	3.48	1.30	.85
Ought–to L2 Self	–	–	–	2.99	1.18	.85	2.83	1.17	.88	3.02	1.14	.86
Emotions (6-point)												
FLE Social	–	–	–	4.67	.97	.84	4.59	.93	.83	4.43	.83	.83
FLE Personal	–	–	–	4.49	.84	.82	4.49	.81	.87	4.39	.84	.81
FLCA	–	–	–	3.56	1.00	.86	3.53	.94	.87	3.51	1.00	.86
Experience												
In–class English use (%)	–	–	–	58.48	16.54		56.27	18.87		52.11	19.74	
L2 speech proficiency												
TSST (gain scores)	–	–	–	.00[a]	.51		.00[b]	.51		.00[c]	.61	
TSST (raw scores; 9-point)	3.41	.67	–	3.61	.63		3.79	.65		4.12	.80	

Note: FLE for Foreign Language Enjoyment (6-point); FLCA for Foreign Language Classroom Anxiety (6-point); TSST for Telephone Standard Speaking Test (9-point). [a]This t0_to_t1_relative_gain score is .00000252100840336522. [b]This t0_to_t2_relative_gain score is −.00000181818181818153. [c]This t0_to_t2_relative_gain score is −.00000142857142857808.

(2, 202) = .463, p = .630, η^2 = .004, FLE Personal, F(2, 202) = 1.922, p = .149, η^2 = .018, and FLCA, F(2, 202) = .050, p = .951, η^2 = .001.

In contrast, a significant increase over time with large effect sizes was observed in participants' L2 speech scores with relatively large effects, F(2, 202) = 41.127, p < .001, η^2 = .272. Multiple comparison analyses revealed significant improvements in participants' TSST scores from T1 to T4 (p < .016; Bonferroni corrected). The gains were statistically significant from T1 to T2 (t = -3.382, p < .001, d = .28), from T2 to T3 (t = -3.107, p = .002, d = .028), and from T3 to T4 (t = -5.00, p < .001., d = .42).

These findings suggest that while there was clear improvement in L2 speech proficiency throughout the project, learners' motivation and emotions remained relatively stable, showing no significant shifts despite the participants' enhanced L2 proficiency. This result did not appear to clearly support any of our hypotheses regarding the link between motivation, emotions, and acquisition.

Model comparisons

The second aim of the statistical analysis was to examine which theoretical model in Figure 1 most aptly fit the observed data at each time point. The three models have different predictions—(1) motivation and enjoyment trigger L2 speech development (while anxiety delays it) (motivation/emotions → acquisition), (2) L2 speech development boosts motivation and enjoyment and it limits anxiety (acquisition → motivation/emotions), and (3) motivation, emotions and L2 speech development are interwoven (motivation/emotions ↔ acquisition). We compared the two causality models (Models 1 and 2) and one correlation model (Model 3) via Structural Equation Model (SEM). To fit models, the Maximum Likelihood with robust standard errors and a scaled test (MLR) was used with the lavaan package (Rosseel, 2012) in R.

In light of our focus on the impact of motivation and emotions on L2 speech *development*, and given the significant improvement in participants' L2 proficiency from T1 to T4, the analyses focused on participants' improved L2 proficiency scores at T2 (six months), T3 (ten months), and T4 (1.5 years) relative to the onset of the project (T1, 0 months). To determine the extent of learning at each data collection point, residual T2, T3, and T4 scores were generated while controlling for T1 scores. T1 scores reflected their L2 English outcomes after years of EFL learning experiences in various settings *before* entering high school. Therefore, the residual T2, T3, and T4 scores indicated what they had learned *after* six months, ten months, and 1.5 years of study at high school, factoring out any influence of their diverse L2 English learning *unrelated* to the current project (i.e., how well they already spoke L2 English prior to the project).[2]

Conversely, due to the lack of significant shifts in participants' motivations and emotions profiles, their raw motivation and emotion scores at T2, T3, and T4 were used, considering these scores as relatively stable traits in the initial ANOVA analyses.

Given that all theories on L2 speech development consider experience as the main catalyst for acquisition (e.g., Flege & Bohn, 2021), the initial two causality models incorporate a hierarchical relationship between experience and enhanced L2 speech proficiency. As mentioned in the Current Study section, the L2 learning experience was operationalized as the extent to which participants practiced L2 English within classrooms, and such experience profiles were measured as participants' self-reports of the percentage of time spent using L2 English in class.

The first causal model posits that motivation and emotions determine the degree of L2 practice within classroom settings, which subsequently influences L2 speech learning gains (Figure 1 Model 1). Meanwhile, the second causal model suggests that while increased practice opportunities bolster L2 speech proficiency, such advancements in L2 speech capabilities can, in turn, bolster motivation and enjoyment while reducing anxiety (Figure 1 Model 2). Conversely, the correlational model suggests a complex interplay among motivation and emotions, experience, and acquisition, without a predefined causal direction. This model integrates two motivation variables, three emotion variables, and one experience variable, each purportedly impacting participants' gain scores (Figure 1 Model 3). The results of SEM are presented at T2 (6 months), T3 (10 months), and T4 (1.5 years), respectively.

To evaluate each model, we used fit indices and criteria recommended in the SEM literature: (a) a Comparative Fit Index (CFI) value of .90 or higher, (b) a Root Mean Square Error of Approximation (RMSEA) value of .08 or lower, and (c) a Standardized Root Mean Square Residual (SRMR) value of .08 or lower (Browne & Cudeck, 1993). To compare models, we used Akaike Information Criterion (AIC) and Bayes information criterion (BIC) values. Models with smaller values were considered more

---

[2]A suggestion was made to fit an SEM that models the continuous effect of time as a latent factor. We agree that this approach could be appropriate when no specific time contrasts are of interest, as it would allow for all possible combinations of multiple comparisons (e.g., "T1-T2, T1-T3, T1-T4," as well as "T2-T3, T2-T4, T3-T4"). However, we did not take this approach because our study was initially designed to focus on three specific time periods—the extent to which participants' L2 proficiency improved at T2 (six months), T3 (ten months), and T4 (1.5 years), relative to the onset of the project at T1 (0 months). This focus led us to calculate residual T2, T3, and T4 scores (controlling for T1 scores) and to model three distinct SEMs, rather than fitting one SEM with time as a latent factor.

Table 4. Comparisons of Model 1 (motivation/emotions → acquisition), Model 2 (acquisition → motivation/emotions), and Model 3 (motivation/emotions ↔ acquisition)

	χ^2 (df)	p	CFI	RMSEA(90% CI)	SRMR	AIC	BIC	Fit?	Best-fitting model
Fit criteria	–	≥ .05	.90	< .08	< .08	The smaller, the better			
A. T2 (6 months)									
Model 1	17.568 (5)	.004	.371	.141 [.072, .217]	.081	659.646	687.437		
Model 2	5.487 (5)	.359	.997	.024 [.000, .133]	.046	1902.773	1980.588	x	x
Model 3	–	–	–	–	–	2248.045	2345.315	x[a]	
B. T3 (10 months)									
Model 1	16.200 (5)	.006	.577	.138 [.063, .219]	.076	636.792	664.750		
Model 2	19.949 (5)	.001	.890	.167 [.092, .249]	.083	1909.855	1988.137		
Model 3	–	–	–	–	–	2244.214	2342.067	x[a]	x[a]
C. T4 (1.5 years)									
Model 1	1.581 (5)	.904	1.000	.000 [.000, .061]	.021	508.963	536.237	x	x
Model 2	15.753 (5)	.008	.929	.156 [.078, .242]	.093	1652.377	1728.743		
Model 3	–	–	–	–	–	1969.708	2065.166	x[a]	

Notes: χ^2 = Chi–square test of model fit; df = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean square Residual. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion. [a]Model 3 is a saturated model—a type of a model that fits the data perfectly, making it impossible to test the model-data fit (e.g., Kline, 2023).

appropriate. Along with these statistical criteria, the substantive interpretability of the models was also considered. Note that Model 3, being a perfect–fit model (i.e., a saturated model; e.g., Kline, 2023), did not allow for the evaluation of its fit using indices such as CFI. However, it was possible to compare it with other models using indices such as AIC.

- **T2-6 Months:** According to the results of the chi–square tests (see Table 4), Model 1 did not fit the data well (CFI = .371, RMSEA = .141 [95% confidence interval = .072, .217], and SRMR = .081). Model 2 fit the data well (CFI = .997, RMSEA = .024 [.000, .133], and SRMR = .046). Model 3, being a saturated model, did not allow for evaluating its fit. Comparatively, Model 2 showed lower AIC and BIC values (1902.773 and 1980.588, respectively) than Model 3 (2248.045 and 2345.315, respectively). Thus, we considered Model 2 to be the best model to explain the data at T2. As visually summarized in Figure 2, the regression analyses revealed significant relationships between experience, enhanced L2 proficiency, and emotions, indicating that the extent of practice in classrooms positively correlated with greater L2 learning gains (β = .188, p = .026). This increase in L2 proficiency was associated with an increase in positive emotions (β = .377, p < .001 for FLE Social; β = .312, p < .001 for FLE Personal) and a decrease in negative emotion (β = -.200, p = .006 for FLCA).
- **T3-10 Months:** Ten months after the project began, neither causal Model 1 nor Model 2 adequately fit the data, as indicated by their fit indices overall (CFI = .577, RMSEA = .138 [.063, .219], and SRMR = .076 for Model 1; CFI = .890, RMSEA = .167 [.092, .249], and SRMR = .083 for Model 2). Consequently, only the correlational Model 3 received support at this time point in the dataset. As visually summarized in Figure 3, the correlational analysis within Model 3 revealed
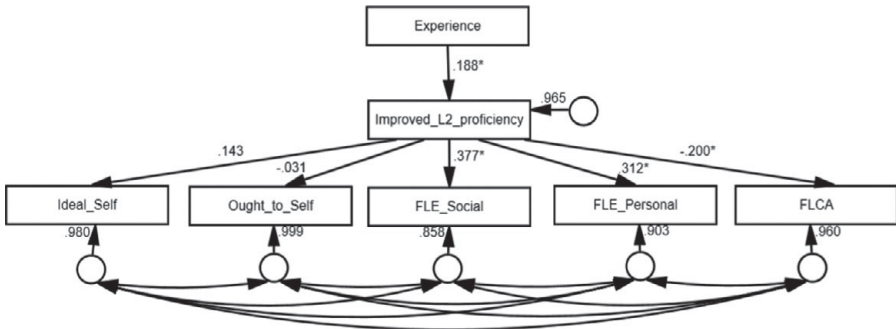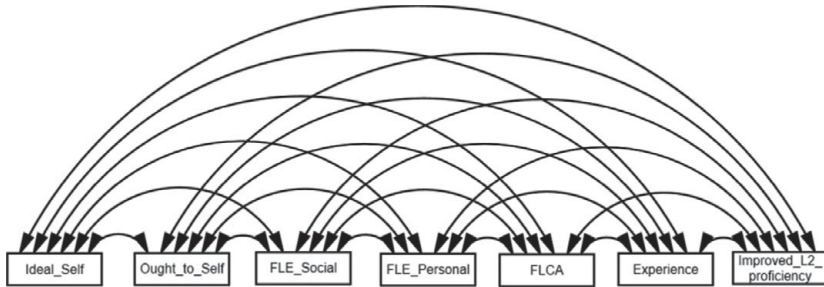
Correlations among Ideal_Self, Ought_to_Self, FLE_Social, FLE_Personl, and FLCA

	Ideal_Self	Ought_to_Self	FLE_Social	FLE_Personal
Ideal_Self	--			
Ought_to_Self	.185	--		
FLE_Social	.169	.028	--	
FLE_Personal	.286*	-.003	.638*	--
FLCA	-.369*	.006	-.344*	-.380*

Note. $p < .05$.

Figure 2. Best-fitting model at Time 2 (6 months from Time 1): Model 2.
Note: Values in the figure are standardized estimates. FLE for Foreign Language Enjoyment; FLCA for Foreign Language Classroom Anxiety. Improved L2 proficiency refers to TSST gain scores from Time 1 to Time 2.

Correlations among the variables

	Ideal Self	Ought-to Self	FLE Social	FLE Personal	FLCA	Experience
Ideal Self	--					
Ought-to Self	.275*	--				
FLE Social	.214*	.124	--			
FLE Personal	.251*	.099	.724*	--		
FLCA	-.376*	-.009	-.302*	-.307*	--	
Experience	.143	-.075	.371*	.285*	-.092	--
Improved L2 proficiency[a]	.332*	.046	.177	.185	-.187	.026

Note. $p < .05$.

Figure 3. Best(-fitting) model Time 3 (10 months from Time 1): Model 3.
Note: Values in the figure are standardized estimates. FLE for Foreign Language Enjoyment; FLCA for Foreign Language Classroom Anxiety. [a]TSST gain scores from Time 0 to Time 3.

significant associations between FLE and in–classroom experience ($r =.371$, $p < .001$ for FLE Social enjoyment; $r = .285$, $p < .001$ for FLE Personal enjoyment), as well as the one between motivation and improved L2 proficiency ($r = .332$, $p =.001$ for Ideal L2 Self).
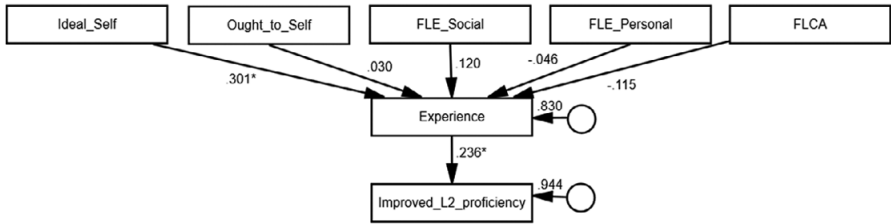
Figure 4. Best-fitting model at Time 4 (1.5 years from Time 1): Model 1.
Note: Values in the figure are standardized estimates. FLE for Foreign Language Enjoyment; FLCA for Foreign Language Classroom Anxiety. Improved L2 proficiency refers to TSST gain scores from Time 1 to Time 4.

- **T4-1.5 Years:** After an extensive period since the outset of the project (1.5 years), Model 1 fit the data well (CFI = 1.000, RMSEA = .000 [.000, .061], and SRMR = .021), whereas Model 2 showed poor fit (CFI = .929, RMSEA = .156 [.078, .242], and SRMR = .093). Model 3, as a saturated model, was not suitable for fit evaluation. Comparatively, Model 1 showed lower AIC and BIC values (1652.377 and 1728.743, respectively) than Model 3 (1969.708 and 2065.166, respectively). The results of the regression analysis indicated the following relationship between motivation, emotions and acquisition: Those with stronger motivation (Ideal L2 Self) showed more input and engagement within their classroom experience (β = .301, p = .017); such increase in in–class L2 experience led to more L2 speech learning gains (β = .236, p = .018). For a visual summary, see Figure 4.

Discussion

This study investigated the relationship between individual differences in motivation, emotions (FLE and FLCA), and L2 speech learning among 121 Japanese EFL high school students over a 1.5–year period. We tested four hypotheses: two causality models positing that enhanced motivation and emotions drive L2 speech learning (Model 1: motivation/emotions → acquisition) and that improved L2 speech proficiency boosts motivation and emotions (Model 2: acquisition → motivation/emotions), alongside a correlation model suggesting a bidirectional relationship between motivation, FLE, FLCA and acquisition without a predefined direction (Model 3: motivation/emotions ↔ acquisition). Finally, we explored a fourth hypothesis suggesting that the nature of the relationship between motivation, emotions, and L2 speech learning may evolve over different phases of EFL speech learning (6 months, 10 months, 1.5 years), with the predictive power of motivation and emotions becoming stronger over time (Models 2 and 3 → Model 1).

The findings revealed three main insights: First, the motivation and emotions measured through questionnaires remained stable throughout the project. Second, participants' L2 speech proficiency consistently improved at each testing point (6 months, 10 months, and 1.5 years). Third, while the causal roles of motivation and emotions in L2 speech learning remained unclear within the first one year of the study, the motivation/emotions → acquisition model (Model 1) emerged as the best–fitting model for explaining long–term outcomes (1.5 years).

Overall, our findings corroborate previous longitudinal evidence suggesting that L2 learners' motivation and emotion profiles exhibit only mild fluctuations in classroom

settings (e.g., Dewaele et al., 2023). This in turn suggests that motivation and emotions could be a statistic construct when profiles are operationalized via questionnaire instruments and averaged on a group level (Jiang & Dewaele, 2015). One reason for this could be the fact that the instruments are too blunt to measure what happens below the surface (i.e., group averages). In contrast, the participants' L2 speech proficiency significantly improved in classroom environments through consistent participation in English education (Jaekel et al., 2017) and through increased in–class practice within such settings (Muñoz, 2014). Our findings align with the major theoretical paradigm demonstrating that L2 learners possess the capacity to acquire new sounds and languages regardless of age and contexts, given sufficient input opportunities (Flege & Bohn, 2021).

The results of the ANOVA analyses presented here—highlighting the stability of motivation/emotions alongside the dynamism of L2 speech learning—do not straight-forwardly align with any of the postulated hypotheses (motivation/emotions → acquisition, acquisition → motivation/emotions, and motivation/emotions ↔ acquisition). Longitudinal analysis elucidated that motivation, emotions, and acquisition are not directly comparable due to the contrasting nature of these elements (i.e., stable motivation/emotions vs. dynamic L2 speech learning). Instead, the findings suggest a possible pathway: motivation and emotions might initially amplify opportunities for practicing the target language in classroom settings; this positive link between motivation, emotions, and practice could, in turn, lead to increased autonomy and optimism which could lead to more L2 use in classrooms, which could—in turn—enhance L2 speech proficiency (stable motivation and emotions → increased practice opportunities → acquisition; Dewaele & Meftah, 2024; Dörnyei, 2020).

However, the assertion about the stability of motivation and emotions needs to be interpreted with caution and thus further examined with more balanced research designs. Scholars have repeatedly argued that individuals' motivation and emotions are fluctuating at every point of L2 use (Waninge, Dörnyei, & De Bot, 2014). The practice of measuring and averaging learners' motivation and emotions may not accurately reflect its dynamic, time–sensitive nature (Zhang, Dai, & Wang, 2020; Talebzadeh et al., 2020). To more effectively capture this phenomenon, it is crucial to employ diverse instruments beyond questionnaires, such as motometers and interviews (Elahi Shirvan & Taherian, 2021 for emotion; Liu & Wang, 2017 for motivation), or classroom observations and interviews to understand the specific causes of fluctuations (Dewaele & Pavelescu, 2021).

Turning to the results of the SEM, our findings support an additional possibility that the causal role of motivation and emotions in L2 speech learning (stable motivation/emotions → acquisition [Model 1]) could be clearer when the relationship was examined for a longer period of time whereas this model was not the best fit within the first year of the project.

In the initial six months (T1-T2), for example, the best fit model suggested that an increase in classroom L2 practice could be associated with enhanced L2 speech proficiency, which in turn influenced motivation and emotions profiles, particularly emotional aspects—FLE and FLCA. This in turn supports the acquisition → motivation/emotions model (Model 2) that posits L2 proficiency gains as determinants of motivation and emotions (Botes et al., 2020, 2022a). The influence of L2 speech proficiency on motivation and emotion is plausible, especially given that substantial learning tends to occur during the first few months in new L2 learning environments (Flege & Bohn, 2021). During this early phase, learners' speech proficiency may improve regardless of their initial motivation and emotional state; rather, any

measurable gains in L2 proficiency during this period may further enhance their motivation and emotions toward learning the L2.

Around the time of 10 months (T1-T3), the causal relationship became less clear, suggesting a mutual influence between motivation, emotions, experience, and acquisition (motivation/emotions ↔ acquisition [Model 3]). This suggests that not only do motivation and emotions have strong associations with acquisition, but these relationships are also equally connected to the extent of target language practice in classroom settings.

Over the longer term (1.5 years; T1-T4), a clear directionality finally emerged between greater motivation and emotions, specifically Ideal L2 Self, and more classroom practice opportunities, which then positively affected L2 speech learning gains, lending empirical support to the motivation/emotions → acquisition model (Model 1). The results here support the views that Ideal Self can lead learners to engage in more practice (Dörnyei, 2020) and that L2 speech proficiency improves as a function of increased input and output opportunities (Flege & Bohn, 2021).

Taken together, the findings support our fourth hypothesis that the relationship between motivation, emotions, and L2 speech learning evolves at different stages of EFL learning (short–, mid–, and long–term), with the predictive power of motivation and emotions becoming more pronounced in the long term. This observation corroborates the feedback loop model proposed by Li, Dewaele, et al. (2022), which posits that elevated motivation and enjoyment, combined with lower anxiety, foster language acquisition, which in turn, impacts motivation and emotions. Here, our findings introduce the notion of a *reverse* feedback mechanism that applies across both short–, mid– and long–term contexts of classroom L2 speech learning:

- **Short– and Mid–Term** (< 1 year): It appears that learners are driven to enhance their L2 speech proficiency, irrespective of their initial profiles of motivation and emotions, particularly within the setting of students newly enrolled in a prestigious school. This motivational phase seems to extend up to a year of EFL learning, during which motivation, emotions and acquisition begin to mutually influence one another as has been found in other contexts (Dewaele & Meftah, 2024).
- **Long-Term** (> 1 year): Beyond the one–year mark within the same educational context, learners who have attained a certain level of L2 speech proficiency might hit a developmental plateau. It is at this juncture that a stable motivation and an optimal emotional profile become crucial, motivating learners to persist in their extensive practice and extracting joy in doing so while keeping anxiety in check. The ability to nurture motivation and enjoyment allowing them to intertwine is the basis for further progress and achievement in the L2 speech. The framework proposed here relates to cross–sectional evidence indicating that among highly experienced L2 learners in classroom settings, those with highly advanced L2 speech proficiency often exhibit strong professional motivation (Moyer, 2014).

The analogy that initial motivation and enjoyment act as a nurturing nest from which learners, akin to fledgling birds, eventually venture out, underscores the importance of these affective factors. Once motivation and enjoyment are fully grown, they encourage learners to persist in high–quality language practice, thereby enhancing and achieving advanced L2 proficiency (Dewaele & Meftah, 2024).

Finally, unlike FLE, the current dataset did not identify FLCA as a significant predictor of the amount of L2 practice or development at any point. The only suggested pathway is that improved L2 speech proficiency leads to reduced FLCA, particularly

within the first six months (T1-T2). These findings align with recent research, which shows that while cross–sectional evidence supports a correlation between FLCA and L2 learning outcomes (e.g., Teimouri et al., 2019), longitudinal data have not found FLCA to play a significant role in L2 learning outcomes (Li & Wei, 2023).

Our findings echo Dewaele and MacIntyre's (2014) original hypothesis that, although FLE and FLCA may overlap, they represent fundamentally different phenomena. Together with recent research, our tentative conclusion is that FLE is more closely related to the factors currently driving L2 learning in classroom settings, such as ongoing classroom practices (e.g., peer engagement and cooperation; Li, Huang et al., 2021), teaching methods (e.g., more communicative vs. traditional/language–focused methods; Dewaele et al., 2024), teacher characteristics (e.g., the use of humor; Dewaele et al., 2022), and long–term development behaviors (Li & Wei, 2023; Saito et al., 2018). Consequently, FLE appears to predict the extent to which L2 learning occurs (Dewaele et al., 2024; Li & Wei, 2023; Saito et al., 2018). In contrast, FLCA seems to be a more stable, trait–like emotion that can have a disruptive, narrowing effect in the moment (Dewaele & MacIntyre, 2014) and/or a consequence (rather than predictor) of L2 learning and use (as suggested by the current dataset).

Limitations

Looking forward, we acknowledge limitations and propose directions for further research as follows:

1. Only a limited number of motivation dimensions and learner emotions were considered. The effect of more dimensions from within the self–system framework could be explored (e.g., Papi et al., 2019) as well as a wider range of positive and negative emotions (e.g., Boredom and Peace of Mind; Dewaele & Meftah, 2024; Li & Wei, 2023).
2. Second, participants' L2 use was operationalized through self-reports in the current study. To narrow down the scope of the experience directly relevant to the main focus of the study (how much participants had practiced L2 English within classroom settings), the method itself has been problematized by many scholars. For example, a majority of previous L2 motivation studies have used criterion measures wherein participants rated their L2 use using a 6-point scale for a range of broad statements (e.g., "*I am prepared to expend a lot of effort in learning English*"). As Al-Hoorie et al. (2018) pointed out, such practice (described as the "questionnaire curse") does not clearly capture the multifaceted nature of participants' EFL experience. Notably, certain scholars have attempted to capture the dynamic nature of L2 learning experiences in various settings by asking participants to track precisely how much and how they practiced L2 English daily using a range of online instruments (e.g., The Lang-Track-App; Arndt, Granfeldt, & Gullberg, 2023). Other scholars have proposed more sophisticated measures to capture the depth and intensity of the EFL experience (e.g., social network analysis; Strawbridge, 2023). However, we also caution that simply including as many experience variables as possible may not be a realistic option for future L2 speech research. To our knowledge, there is little research that can successfully link such highly complex experience profiles to any aspects of L2 learning. Additionally, including numerous experience variables as predictors would necessitate a large sample size, which would otherwise result in very low statistical power. Finally, some scholars have

argued that experience variables alone may not fully explain the outcomes of L2 speech learning, as certain learners with greater aptitude may be able to make better use of every input and output opportunity than others (Mora, 2022). More relevant discussion and future studies are needed to further address these methodological and conceptual concerns.

3. Relatedly, Papi and Hiver (2024) proposed a new framework of qualitatively different L2 learning experiences. Under this framework, L2 learning experiences can be further categorized according to learners' agentic behaviors, such as input-seeking, interaction-seeking, information-seeking, and feedback-seeking. Future studies should disentangle the dynamic interplay between different types of motivation, qualitatively different learning experiences and behaviors, and acquisition (cf., Papi & Khajavy, 2021; Papi et al., 2019).

4. It is important to remember that the degree of measurement invariance for FLCA was below the configural level, suggesting that the way the questionnaire items factored into the construct of FLCA varied at each time point. That is, any relevant findings related to FLCA may not be comparable across these different time points (T2-T4). However, there is still much ongoing discussion regarding how many constructs underlie the complex phenomenon of L2 emotions (e.g., Dewaele & MacIntyre, 2014, for Enjoyment and Anxiety; Li & Wei, 2023 for Enjoyment, Boredom, and Anxiety) and what kinds and how many questionnaire items are needed to reliably tap into these supposedly different constructs (e.g., Botes et al., 2022a, for the Short–Form Foreign Language Classroom Anxiety Scale). To our knowledge, no studies have ever explored how these constructs remain the same or change when data is collected at different time points. Relatedly, there is a consensus that the key emotion constructs may be difficult to fully validate. The emotions are not "pure" and they do not operate in a social vacuum. Dewaele and MacIntyre (2014) pointed out that Enjoyment is a multifaceted emotion, including learners' feelings of pride, creativity, pride, interest, fun, excitement in class combined with social factors that may change over time such as the relationship with peers and the perception of support and friendliness of the teacher. Learner emotions are part of a complex and dynamic system which means that isolated incidents such as mocking by peers, a harsh remark by a teacher, or a disappointing test result could all undermine the learners' positive feelings and dent their enjoyment. MacIntyre (2017) underlined the existence of fluctuations in anxiety but he could have been talking about any other learner emotion: "Anxiety is continuously interacting with several other learner, situational and other factors including linguistic abilities, physiological reactions, self–related appraisals, pragmatics, interpersonal relationships, specific topics being discussed, type of setting in which people are interacting and so on" (p. 23). The decision about what items to include in instruments to tap into different but interacting emotions is thus by definition fraught. In the current investigation, we would like to acknowledge (a) that we simply adopted well–researched questionnaire instruments for motivation (Dörnyei, 2005) and emotion (Dewaele & MacIntyre, 2014) and followed the analysis protocols for comparability with the ample previous literature using the same methods; and (b) that the development and validation of context–specific motivation and emotion questionnaires were beyond the main focus of the current study. Thus, we now call for future studies to further examine whether, to what degree, and how these oft–used questionnaires (typically used in cross–sectional datasets) can be adapted when applied to longitudinal datasets.

5. Lastly, L2 speech proficiency was measured using a proficiency test (TSST). Research shows diverse progress and results across L2 speech proficiency subdimensions; while learners may quickly advance in fluency and lexicogrammar, segmental accuracy tends to show greater individual variance and is influenced by learners' aptitude (e.g., Saito, Suzukida, & Sun, 2019 for phonemic coding; Kachlicka et al., 2019 for auditory processing; Darcy, Mora & Daidone, 2016 for inhibitory control). Future studies should investigate if the relationship between motivation, emotions, and acquisition varies between relatively easy aspects (e.g., fluency) and more difficult aspects (e.g., segmental details; see Saito, Suzuki et al., 2021).

Conclusion

Tracking the development of 121 Japanese EFL students over 1.5 years, this study explored the dynamic relationships between motivation, positive and negative emotions, and L2 speech proficiency. The mean–based analyses (ANOVAs) suggest that motivation and emotions remain relatively stable for these beginner EFL learners, fueling greater engagement with in–class practice and subsequent gains in L2 speech. Notably, the variance-based analyses (SEM) indicate the nature of the relationship between motivation, emotions, and L2 speech proficiency—whether causal or correlational—may evolve (Li et al., 2022). At the outset of their language learning journey, many L2 learners actively engage with the target language, enhancing their proficiency. This initial phase may exhibit a reciprocal influence between learners' motivation, emotions, and their language learning, particularly within the first year of EFL learning. As learners adapt and progress, the significance of robust motivation, growing enjoyment and limited anxiety becomes evident. It fuels consistent language practice and facilitates ongoing improvement in L2 speech proficiency.

References

Arndt, H. L., Granfeldt, J., & Gullberg, M. (2023). The Lang-Track-App: Open-source tools for implementing the experience sampling method in second language acquisition research. *Language Learning*, *73*(3), 869–903. https://doi.org/10.1111/lang.12555

Al-Hoorie, A. H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*(4), 721–754. https://doi.org/10.14746/ssllt.2018.8.4.2

Baumert, J., Fleckenstein, J., Leucht, M., Köller, O., & Möller, J. (2020). The long-term proficiency of early, middle, and late starters learning English as a foreign language at school: A narrative review and empirical study. *Language Learning*, *70*(4), 1091–1135. https://doi.org/10.1111/lang.12414

Botes, E., Dewaele, J.-M., & Greiff, S. (2020). The Foreign Language Classroom Anxiety Scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *The Journal for the Psychology of Language Learning*, 2, 26–56.

Botes, E., Dewaele, J.-M., & Greiff, S. (2022a). Taking stock: A meta-analysis of the effects of foreign language enjoyment. *Studies in Second Language Learning and Teaching*, 12(2), 205–232. https://doi.org/10.14746/ssllt.2022.12.2.3

Botes, E., van der Westhuizen, L., Dewaele, J.-M., MacIntyre, P.D., & Greiff, S. (2022b) Validating the Short-Form Foreign Language Classroom Anxiety Scale (S-FLCAS). *Applied Linguistics*, 43(5), 1006–1033. https://doi.org/10.1093/applin/amac018

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, 66(4), 741–773. https://doi.org/10.1111/lang.12161

Dewaele, J. M., Guedat-Bittighoffer, D., & Dat, M. A. (2024). Foreign language enjoyment overcomes anxiety and boredom to boost oral proficiency in the first year of English foreign language learning. *International Journal of Applied Linguistics*, 1–16. https://doi.org/10.1111/ijal.12607

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, 4(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., & MacIntyre, P.D. (2019). The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of Instructed Second Language Acquisition studies* (pp. 263–286). Routledge.

Dewaele, J.-M., & Meftah, R. (2024). How motivation and enjoyment spread their wings during English Foreign Language learning: A pseudo-longitudinal investigation into Moroccan learners. *Learning and Motivation*, 85. https://doi.org/10.1016/j.lmot.2024.101962

Dewaele, J.-M., & Pavelescu, L. (2021). The relationship between incommensurable emotions and willingness to communicate in English as a foreign language: A multiple case study. *Innovation in Language Learning and Teaching*, 15(1), 66–80. https://doi.org/10.1080/17501229.2019.1675667

Dewaele, J.-M., Saito, K., & Halimi, F. (2023). How foreign language enjoyment acts as a buoy for sagging motivation: A longitudinal investigation. *Applied Linguistics*, 44(1), 22–45. https://doi.org/10.1093/applin/amac033

Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, 22(6), 676–697. https://doi.org/10.1177/1362168817692161

Dewaele, J.-M., Saito, K.& Halim, F. (2022). How teacher behaviour shapes Foreign Language learners' Enjoyment, Anxiety and Attitudes/Motivation: A mixed modelling longitudinal investigation. Language Teaching Research. https://doi.org/10.1177/13621688221089

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.

Dörnyei, Z. (2020). *Innovations and challenges in language learning motivation*. Routledge.

Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63(3), 437–462. https://doi.org/10.1111/lang.12005.

Dörnyei, Z., & A. Henry (2022). Accounting for long-term motivation and sustained motivated learning: Motivational currents, self-concordant vision, and persistence in language learning. In A. Elliot (Ed.), *Advances in motivation science* (pp. 1–46). Elsevier.

EIKEN Foundation of Japan. (2017). Investigating the relationship of the EIKEN tests with the CEFR. https://www.eiken.or.jp/eiken/en/research/#anc04

Elahi Shirvan, M. & Taherian, T. (2021). Longitudinal examination of university students' foreign language enjoyment and foreign language classroom anxiety in the course of general English: Latent growth curve modelling. *International Journal of Bilingual Education and Bilingualism*, 24(1), 31–49. https://doi.org/10.1080/13670050.2018.1441804

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.

Guedat-Bittighoffer, D. & Dewaele, J.-M. (2023) Fluctuations des émotions éprouvées par des apprenants débutants dans cinq cours de Français Langue Etrangère: Une étude de cas multiples. *Language, Interaction, and Acquisition*, *14*(2), 279–305. https://doi.org/10.1075/lia.22025.gue

Heydarnejad, T., Tagavipour, F., Patra, I., & Farid Khafaga, A. (2022). The impacts of performance-based assessment on reading comprehension achievement, academic motivation, foreign language anxiety, and students' self-efficacy. *Language Testing in Asia*, *12*(1), 51. https://doi.org/10.1186/s40468-022-00202-4

Horwitz, E. K., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132. https://doi.org/10.1111/j.1540–4781.1986.tb05256.x

Huensch, A., & Nagle, C. (2023). Revisiting the moderating effect of speaker proficiency on the relationships among intelligibility, comprehensibility, and accentedness in L2 Spanish. *Studies in Second Language Acquisition*, *45*(2), 571–585. https://doi.org/10.1017/S0272263122000213

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*, *67*(3), 631–664. https://doi.org/10.1111/lang.12242

Jiang, Y., & Dewaele, J. M. (2015). What lies bubbling beneath the surface? A longitudinal perspective on fluctuations of ideal and ought-to L2 self among Chinese learners of English. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 331–354. https://doi.org/10.1515/iral-2015-0015

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, *192*, 15–24. https://doi.org/10.1016/j.bandl.2019.02.004

Kline, R. B. (2023). *Principles and practice of structural equation Modeling* (5th ed.). Guilford.

Koizumi, R. (2017). *Telephone Standard Speaking Test no datousei kenshou* [A validation study of the Telephone Standard Speaking Test]. *ALC Report, 10.* https://cdn2.alc.co.jp/sa/www/report/alc_report_20171114.pdf

Koizumi, R. (2021). The Telephone Standard Speaking Test: An outside evaluator's investigation of a rebuttal to the generalization inference. In C. A. Chapelle & E. Voss (eds.), *Validity argument in language testing: Case studies of validation research* (pp. 154–175). Cambridge University Press. https://doi.org/10.1017/9781108669849.010

Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, *62*(4), 997–1023. https://doi.org/10.1111/j.1467-9922.2012.00719.x

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. https://doi.org/10.1177/0267658307082981

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). New York: Routledge. https://doi.org/10.4324/9781315775661

Li, C., Dewaele, J.-M., Pawlak, M., & Kruk, M. (2022). Classroom environment and willingness to communicate in English: The mediating role of emotions experienced by university students in China. *Language Teaching Research.* https://doi.org/10.1177/13621688221111623

Li, C., Huang, J., & Li, B. (2021). The predictive effects of classroom environment and trait emotional intelligence on foreign language enjoyment and anxiety. *System*, *96*, 102393. https://doi.org/10.1016/j.system.2020.102393

Li, C., & Wei, L. (2023). Anxiety, enjoyment, and boredom in language learning amongst junior secondary students in rural China: How do they contribute to L2 achievement? *Studies in Second Language Acquisition*, *45*(1), 93–108. https://doi.org/10.1017/S0272263123000031

Liu, X., & Wang, L. C. C. (2017). Motivation, learning strategies, and language competency in a technology facilitated Chinese as a second language classroom. *Chinese Language Teaching Methodology and Technology*, *1*(2), 1.

MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development. In C. Gkonou, M. Daubney, & J.-M. Dewaele (eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 11–30). Bristol: Multilingual Matters.

MacIntyre, P. D., Ross, J., & Clément, R. (2019). Emotions are motivating. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 183–202). Palgrave Macmillan.

Miller, Z. F., & Godfroid, A. (2020). Emotions in incidental language learning: An individual differences approach. *Studies in Second Language Acquisition*, *42*(1), 115–141. https://doi.org/10.1017/S027226311900041X

Mora, J. C. (2022) Aptitude and individual differences. In Tracey M. Derwing, Murray J. Munro & Ron I. Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking*, 68–82. Routledge.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J-M., & Tierney, A. (2022). Tools for second language speech research and teaching. http://sla-speech-tools.com. http://doi.org/10.17616/R31NJNAX

Moskovsky, C., Assulaimani, T., Racheva, S., & Harkins, J. (2016). The L2 motivational self system and L2 achievement: A study of Saudi EFL learners. *The Modern Language Journal*, *100*(3), 641–654. https://doi.org/10.1111/modl.12340

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

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*(4), 463–482. https://doi.org/10.1093/applin/amu024

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*(1), 199–217. https://doi.org/10.1111/modl.12461

Nagle, C. L. (2023). A design framework for longitudinal individual difference research: Conceptual, methodological, and analytical considerations. *Research Methods in Applied Linguistics*, *2*(1), 100033. https://doi.org/10.1016/j.rmal.2022.100033

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., 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., & 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., & Hiver, P. (2024). Proactive language learning theory. *Language Learning*. https://doi.org/10.1111/lang.12644

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

Rosseel, Y (2012). Lavaan: An R package for structural equation modelling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saito, K. (2019). Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English/ɹ/pronunciation. *Second Language Research*, *35*(2), 149–172. https://doi.org/10.1177/0267658318768342

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*(3), 866–900. https://doi.org/10.1002/tesq.3027

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*(4), 614–632. https://doi.org/10.1177/0023830916687793

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*(3), 709–743. https://doi.org/10.1111/lang.12297

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*, 201–225. https://doi.org/10.1017/S0272263117000432

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, *37*(4), 813–840. https://doi.org/10.1017/S0142716415000259

Saito, K., Suzuki, S., Oyama, T., & Akiyama, Y. (2021). How does longitudinal interaction differentially promote experienced vs. inexperienced learners' L2 speech learning? *Second Language Research*, *37*, 547–571. https://doi.org/10.1177/0267658319884981

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning*, *58*(4), 835–874. https://doi.org/10.1111/j.1467-9922.2008.00480.x

Strawbridge, T. (2023). The relationship between social network typology, L2 proficiency growth, and curriculum design in university study abroad. *Studies in Second Language Acquisition*, *45*(5), 1131–1161. https://doi.org/10.1017/S0272263123000049

Sun, H., Saito, K., & Dewaele, J.-M. (2024). Cognitive and sociopsychological individual differences, experience, and naturalistic second language speech learning: A longitudinal study. *Language Learning*, *74*(1), 5–40. https://doi.org/10.1111/lang.12561

Talebzadeh, N., Elahi Shirvan, M., & Khajavy, G. H. (2020). Dynamics and mechanisms of foreign language enjoyment contagion. *Innovation in Language Learning and Teaching*, *14*(5), 399–420. https://doi.org/10.1080/17501229.2019.1614184

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), 363–387. https://doi.org/10.1017/S0272263118000311

Trofimovich, P., Kennedy, S., & Foote, J. A. (2015). Variables affecting L2 pronunciation development. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 353–373). Wiley.

Ushioda, E., & Dörnyei, Z. (2009). Motivation, language identities and the L2 Self: A theoretical overview. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 1–8). Multilingual Matters.

Waninge, F., Dörnyei, Z., & De Bot, K. (2014). Motivational dynamics in language learning: Change, stability, and context. *The Modern Language Journal*, *98*(3), 704–723. https://doi.org/10.1111/modl.12118

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Zhang H., Dai Y., Wang Y. (2020). Motivation and second foreign language proficiency: The mediating role of foreign language enjoyment. *Sustainability*, *12*, 1302. https://doi.org/10.3390/su12041302

Zhou, Y., & Papi, M. (2023). The role of future L2 selves in L2 speech development: A longitudinal study in an instructional setting. *System*, *119*, 103156. https://doi.org/10.1016/j.system.2023.103156