Background

Second language (L2) pronunciation learning is a complex cognitive and motor process. On the cognitive side, L2 speakers must learn to perceive the sounds and prosody of the L2, and on the motor side, they must develop and coordinate articulatory gestures. Understanding the relationship between individual differences (IDs) and L2 pronunciation learning therefore requires considering the IDs that affect the perception and production of segmental features (e.g., individual sounds), suprasegmental features (e.g., stress, intonation), and global dimensions of pronunciation (e.g., comprehensibility and foreign accent/accentedness). ID-pronunciation relationships must also be interpreted in light of measurement decisions. L2 perception accuracy can be measured using discrimination and identification tests of varying complexity, and L2 production accuracy can be measured using acoustic analysis, linguistic coding schemes, and listener-based ratings. L2 production accuracy can also be evaluated in controlled and spontaneous speech using different tasks (e.g., word reading vs. picture description; Saito & Plonsky, 2019). Measurement choices are important because IDs may show different relationships to a range of L2 pronunciation measures under different task conditions.

Although some L2 learners master some L2 sounds immediately, most L2 speakers transfer the sounds and prosody of their native language (L1) into the L2, resulting in a foreign accent. Early work on L2 pronunciation focused on understanding the factors that influence the degree of cross-linguistic (i.e., L1-to-L2) interference in bilinguals who had lived in an L2 environment for many years. The underlying goal of this work was to scrutinize the cognitive basis of foreign accents by comparing age-matched bilingual and monolingual speakers. Studies produced three primary findings. First, bilinguals who had learned the L2 earlier in life tended to perceive and produce L2 sounds similarly to monolinguals (Flege et al., 1999a), suggesting that earlier L2 learners have a better chance of becoming nativelike in the L2. Second, some very early learners still had a noticeable foreign accent in the L2, which means that an early age of onset does not guarantee nativelike performance in all areas of L2 phonology (Abrahamsson & Hyltenstam, 2009). Third, L2 learning can affect L1 speech perception and production (Kartushina et al., 2016), underscoring the fact that the L1 and L2 systems influence one another to varying degrees throughout L2 development. Overall, this line of research points to complex cross-linguistic interactions as a result of the age of L2 learning and the amount of L2 experience.

These findings were the basis of the speech learning model (Flege, 1995). The key premise of this model is that L2 sound learning remains possible throughout the lifespan, but the likelihood
of achieving nativelike L2 perception and production diminishes as the age of onset of L2 learning increases. L2 speakers can establish new mental representations in the L2 if they discern subtle phonetic differences between cross-linguistically similar L1 and L2 sounds, but the later the age of onset, the less likely L2 speakers are to attend to such differences. Because accurate perception is a precursor for accurate production, if L2 speakers do not create new mental representations for L2 sounds, they will likely perceive and produce them according to L1 norms. Here, age of onset represents the state of L1 development at the start of L2 learning, or the extent to which perception and production have become optimized for the L1; age does not refer to the presence of a neurobiological critical period for language development. It also bears mentioning that although the speech learning model posits that many production errors have a perceptual basis, not all production errors are perceptual, nor is perception the only factor regulating production accuracy. Thus, the development of accurate perceptual representations can be said to facilitate, but not guarantee, accurate production. The original model focused on ultimate phonological attainment and was developed to account for a specific process of cross-linguistic phonetic interference in experienced bilinguals; it was not designed to address IDs. The revised model (Flege & Bohn, 2021) maintains the same core claims as its predecessor, but it diverges in two important respects: it has a developmental focus, and it includes more information on the (cognitive) IDs that could be important for L2 sound learning. It does not, however, link IDs to the processes that underlie L2 phonetic category formation, nor does it provide specific, testable ID hypotheses.

Other models of L2 sound learning are equally silent with respect to IDs. The L2 perceptual assimilation model (Best & Tyler, 2007) claims that L2 learners compare L1 and L2 contrasts at the phonetic and phonological levels. These comparisons then determine the relative difficulty that L2 listeners will have in the perception of L2 contrasts. If contrastive L2 sounds are assimilated to the same L1 category, but one is a better fit than the other, then L2 learners should not have much trouble discriminating the L2 pair. On the other hand, if both L2 sounds are equally good exemplars of a single L1 category, then L2 learners will likely struggle to perceive the L2 contrast. Although the model discusses experiential factors such as the age of onset of L2 learning and L1 and L2 use, these factors are backgrounded; cross-linguistic patterns of perceptual assimilation are the centerpiece of the model. The L2 linguistic perception model (van Leussen & Escudero, 2015) does not discuss IDs at all, but because learning is meaning-driven, it would be reasonable to surmise that input and interaction are critical to L2 perceptual learning. From this brief review, it is apparent that theoretical models of L2 speech learning were not developed to account for the full range of cognitive, socio-affective, and experiential IDs that are implicated in L2 pronunciation learning. Furthermore, the models deal with L2 sound learning in naturalistic, second language (SL) environments. Thus, it is unclear if they can and should be generalized to instructed, foreign language (FL) contexts and to other aspects of L2 pronunciation learning, such as suprasegmentals. In this chapter, when discussing context of learning, I will use SL to refer to individuals living in an L2 environment and FL to refer to individuals living in an L1 environment and studying the L2.

Fortunately, second language acquisition research can help fill in some of the gaps in relationships between IDs and L2 pronunciation learning. Keeping in mind the cognitive and motor bases of perception and production, it is easy to imagine links between different types of IDs and various aspects of L2 pronunciation. Variation in cognitive IDs such as phonological short-term memory (PSTM), attentional control, and aptitude could affect the accuracy with which L2 learners encode L2 phonetic input, which could in turn influence how quickly they create mental representations for L2 sounds. Given that accurate perception is hypothesized to be one of the primary determinants of accurate production, cognitive IDs might also influence L2 production indirectly, through their connection with L2 perception. Socio-affective and experiential factors are also important. L2 input and use might promote L2 pronunciation development during the early stages of L2 learning, whereas differences in willingness to communicate (MacIntyre, 2007) may explain variations in long-term attainment, especially in an SL environment (e.g., Derwing & Munro, 2013). A similar
argument can be made for motivation. Motivation has been defined, for example, in terms of future self-guides—envisioning oneself as a competent L2 user in the future—that serve to direct motivated behaviors and orient L2 effort (Dörnyei, 2009; see also Papi & Hiver, this volume). Because meaningful L2 interaction is fundamental to all aspects of L2 development, including L2 pronunciation, highly motivated learners may exhibit distinct patterns of L2 use. Frequent and extensive L2 use may ultimately pave the way for more accurate L2 production as L2 speakers practice, adjust, and automatize the articulatory gestures needed to produce L2 sounds (e.g., the Spanish trill /r/). Moreover, different forms of motivation may be associated with qualitatively different learning strategies (Papi et al., 2019), which could also impinge on L2 pronunciation learning. Because SL and FL contexts differ in terms of quantity and quality of L2 exposure, opportunities for L2 use, and the approach to L2 learning (i.e., FL contexts are often characterized by a metalinguistic approach that may include targeted pronunciation instruction), IDs may operate differently in the two contexts.

In this chapter, I use context as an organizational lens, reviewing research on IDs and L2 pronunciation learning in SL and FL contexts. Admittedly, this distinction can be fuzzy because contextual differences in L2 input, use, and so on are not categorical, but rather a matter of degree. I use SL to refer to studies examining individuals living in an L2 environment and FL to refer to individuals living in an L1 environment and studying the L2. For each context, I review research on IDs and L2 perception before research on IDs and L2 production, grouping IDs into cognitive, socio-affective, and experiential categories. I also review research on IDs and pronunciation instruction. Although age is one of the most important IDs for L2 learning (see Singleton & Pfenninger, this volume), I address it only briefly, focusing instead on other IDs that have received attention in the L2 pronunciation literature.

**Research**

**Evidence**

Research on SL (mainly English) learners' perception has concentrated on experiential factors including the age of onset (age of arrival), length of residence, and amount of L1 and L2 use. This body of work has shown that an early age of onset (Flege et al., 1999b), a longer length of residence (Flege et al., 1997), and a lower percentage of L1 use (Mackay et al., 2001) are generally associated with the accurate perception of L2 sounds. Different L1 groups also show different levels of perceptual accuracy for the same L2 contrast (MacKay et al., 2001), reinforcing the view that the phonetic and phonological organization of the L1 relative to the L2 determines the difficulty of L2 perceptual learning. At the same time, L2 learners from the same L1 background may show varying levels of sensitivity to L2 phonetic cues even at the outset of L2 learning, and these differences may be the basis for distinct patterns of L2 perceptual learning (Kim et al., 2018). Apart from experiential factors, at least one study has addressed a cognitive factor. MacKay et al. (2001) cross-sectionally examined the relationship between age of arrival, L1 use, and PSTM and L1 Italian speakers’ perception of L2 English consonants. Between-group comparisons showed that all three age groups (early, mid, and late age of arrival) differed from the native English control group but not from one another, and the low L1 use group made fewer errors than the high L1 use group when identifying L2 consonants. However, when regression was used to clarify the relative importance of each factor, the age of arrival and PSTM explained variances in identification accuracy, whereas L1 use did not. Based on this finding, the authors hypothesized that learners with better PSTM may encode L2 input more effectively, leading to more accurate mental representations for challenging L2 sounds.

Compared to SL perception research, SL production research has demonstrated that a range of experiential, socio-affective, and cognitive factors are important for production accuracy. Early cross-sectional studies showed that the age of arrival was the best predictor of SL English speakers’ segmental accuracy (Flege et al., 1996) and their foreign accent (Flege et al., 1999b). However, age was not the only significant predictor. Rather, participants’ L2 use and integrative motiva-
tion accounted for variance in segmental accuracy, and many different factors, including length of residence, self-estimated sound processing ability, and integrative and instrumental motivation, explained variance in foreign accent. More recent work has highlighted the complex interplay between SL learners’ L1 cultural background, their attitudes toward the L2, and their patterns of L2 use, all of which seem to predict their L2 pronunciation development.

Drawing upon the willingness to communicate framework, Derwing, Munro, and Thomson (2008) and Derwing and Munro (2013) examined L1 Mandarin and L1 Slavic language speakers’ comprehensibility, fluency, and accentedness in English over the participants’ first two years of L2 immersion (2008) and at the seven-year mark (2013). Both studies showed that the L1 Slavic language speakers’ comprehensibility and fluency improved significantly over time, whereas the L1 Mandarin speakers’ pronunciation did not. Based on interview data, the authors suggested that differences in willingness to communicate could account for the between-group differences in learning. Both groups were reluctant to communicate in English and expressed a desire to maintain their native language and culture, but these trends were more pronounced for the L1 Mandarin speakers. In a cross-sectional study on L2 English speakers from various L1 backgrounds, Moyer (2011) highlighted the importance of considering quality of L2 use. She found that quantity of L2 use only became statistically significant when the context of use was considered. L2 use in personal contexts, but not L2 use in professional contexts, was associated with a more nativelike accent, and interactions with native English-speaking friends was the only subdomain of L2 use that accounted for a significant variance in the foreign accent (in addition to the age of arrival and length of residence).

Beyond developing accurate speech patterns, L2 learners must decide how they want to sound and who they want to sound like in the L2. The pronunciation-identity link may be especially acute for SL learners who vary in their desire to assimilate linguistically and culturally to the local L2 community. In fact, SL learners often report complex, time- and context-specific motives for modifying their pronunciation, as Marx (2002) demonstrated in her self-study. Marx was a native speaker of Canadian English who lived in Germany for a number of years. Reflecting upon her L2 experience, she identified multiple stages of identity development, including adopting alternative L2 accents to avoid being perceived as American, using dialectal forms, and channeling native speaker voices to pass as a native speaker (see also Piller, 2002). Marx made a conscious effort to modify her pronunciation and create a new identity in her L2, but identity can also have a subtle, subconscious effect on L2 pronunciation. Gatbonton et al. (2011) found that ethnolinguistic group affiliation (i.e., sense of belonging to a particular ethnic group) was related to Quebec French speakers’ production of English /ð/, a feature known to be especially difficult for non-native speakers. However, this relationship was mediated by the quantity of L2 use. Overall, then, SL studies highlight complex interactions between socio-affective and experiential IDs and remind us that L2 learners are not simply passive vessels within which IDs operate, but rather autonomous agents who can shape their pronunciation usage over time (Moyer, 2017).

In summary, the extensive L2 input that SL learners receive seems to catalyze L2 pronunciation learning at the beginning of L2 immersion. PSTM is related to accurate L2 perception, insofar as SL speakers with better PSTM may encode the L2 more accurately. Willingness to communicate in the L2 is linked to long-term pronunciation learning. SL speakers who display a high willingness to communicate (and a desire to assimilate to the local L2 community) may be more likely to engage in the type of meaningful L2 interaction necessary for improving fluency and comprehensibility over time.

Where experiential factors have been the focus of SL perception research, FL perception research has concentrated on cognitive predictors. Research has shown that learners with high PSTM identify L2 vowels more accurately than learners with low PSTM (Cerviño-Povedano & Mora, 2010; Inceoglu, 2019). PSTM has also been linked to phonetic cue use, insofar as high PSTM learners appear to rely less on secondary, or less informative, cues when classifying L2
IDs for L2 Pronunciation

sounds (Cerviño-Povedano & Mora, 2010). Another cognitive factor that has proven to be a significant predictor of FL perception accuracy is inhibitory control, a major component of executive function that broadly refers to the ability to block out task-irrelevant information. Darcy et al. (2016) found a positive link between L1 English and Spanish speakers’ inhibitory control and their accuracy in discriminating L2 vowel contrasts. According to the authors, this finding suggests that inhibitory control promotes L2 perceptual learning by helping learners mitigate L1 interference, which could be especially important in the L1-dominant FL context.

Like SL production research, FL production research has targeted a variety of cognitive, socio-affective, and experiential IDs. Early work in this area assessed relationships between IDs and L2 pronunciation accuracy. For instance, Moyer (1999) found that professional motivation was the best predictor of L1 English speakers’ foreign accent in L2 German, though the length of L2 immersion and pronunciation training were also significant factors. In another study, Elliot (1995b) found that field independence, a cognitive ID that reflects how individuals process information, and concern for nativelike pronunciation, a pronunciation-specific form of motivation, predicted L1 English speakers’ pronunciation accuracy in L2 Spanish.

Contrasting with these cross-sectional, attainment-oriented studies, recent FL production research has increasingly examined IDs longitudinally to understand how they affect the rate of L2 pronunciation learning. Research on cognitive IDs suggests that aptitude and awareness are important for the development of challenging pronunciation features. Saito et al. (2019) studied L1 Japanese speakers over two semesters of FL English instruction. They found that during the first semester, experience and explicit aptitude measures predicted distinct areas of pronunciation development; classroom L2 use was associated with gains in comprehensibility and syllabic accuracy, whereas associative memory and phonemic coding were associated with gains in fluency. During the second semester, none of the experience measures predicted gains. Rather, sound sequence recognition, a measure of incidental learning ability, was associated with gains in three areas: comprehensibility, segmental accuracy, and syllabic accuracy (see also Saito et al., 2020). In a study on L1 Japanese speakers’ production of English /ɹ/, Saito (2019) found that the number of L2 content courses, an L2 experience measure, accounted for variance in F2, an acoustic feature hypothesized to be easy to learn, whereas participants’ awareness of and attention to accurate /ɹ/ production explained variance in duration, a challenging acoustic feature. In contrast, explicit articulatory knowledge was not related to any aspect of /ɹ/ production.

Another important aptitude-based measure that has been addressed in the literature is an aptitude for oral mimicry, or imitation ability. This measure is grounded in the view that some individuals may be more aware of phonetic variation and more capable of modifying their own speech patterns than others. Flege and Hammond (1982) explored this issue by asking L1 English speakers to read English sentences with a Spanish accent. The authors tallied the number of L2-to-L1 segmental substitutions that participants made and analyzed the speech of the participants who produced the most and fewest substitutions. Acoustic comparisons showed that only the skilled imitators (i.e., the participants who produced the most L2-to-L1 substitutions when imitating an L2 accent in the L1) produced speech that was phonetically different from the speech of a native English control group, which suggests that individual differences in phonemic awareness may underlie individual differences in imitation ability. In a more recent study, Reiterer et al. (2013) sorted L1 German speakers into high and low skill groups based on their direct imitation ability, which they measured by asking participants to imitate Hindi sentences after hearing them. The high and low skill groups were then asked to read sentences in German with an English accent. Behaviorally, skilled imitators used a larger articulation space than unskilled imitators, and neurologically, they showed less extensive activation, indicative of more efficient processing. These findings indicate that skilled imitators exhibit substantial neurocognitive and articulatory flexibility, which could help them learn to produce the L2 more accurately.
Socio-affective factors have been linked to FL learners’ global pronunciation development. In their study on L1 Japanese learners of English, Saito et al. (2017) investigated the link between various types of motivation (e.g., integrativeness, or the desire to learn English to connect with native English speakers and learn about English-speaking culture; instrumentality, or the desire to learn English for professional advancement), metacognitive orientation (e.g., a focus on rich and accurate grammar and vocabulary and a focus on nativelike pronunciation), and gains in comprehensibility and accentedness over one semester. Among the motivation and metacognitive factors, only a general concern for comprehensibility emerged as a significant predictor of comprehensibility. Other longitudinal studies have shown that motivation shapes L2 speech learning in interaction with other factors. Defining motivation in terms of future ideal and ought-to L2 self-guides, Saito et al. (2018) found that a stronger ideal L2 self was associated with stronger positive emotions and weaker negative emotions. These factors were also linked to participants’ previous and current L2 use across a range of activities and to their comprehensibility gains. Nagle (2018) obtained somewhat different findings in his longitudinal study on motivation, effort, and L2 Spanish pronunciation. Like Saito et al. (2018), he defined motivation in terms of participants’ ideal and ought-to L2 selves. Although neither facet of motivation predicted longitudinal gains in comprehensibility and accentedness, self-reported effort was a significant predictor of accentedness. These somewhat divergent findings could be due to differences in learner populations (L2 English vs. L2 Spanish) and speaking tasks (a semi-spontaneous timed picture description task in Saito et al. vs. a semi-controlled sentence formation task in Nagle). Arguably, socio-affective IDs like motivation work by catalyzing motivated behaviors that promote learning. Sardegna et al. (2018) addressed this issue in their study of L1 Korean-speaking learners of English. They found that participants’ self-efficacy beliefs and their attitude toward their own pronunciation had an impact on the amount of effort they invested in pronunciation learning and the strategies they used to improve.

Summarizing FL research, PSTM and inhibitory control are important for FL perception. FL learners with better cognitive skills may extract more accurate information from the limited L2 input they receive, which would support their L2 perceptual learning. Regarding production, L2 coursework can help learners improve their production of easy pronunciation features, especially during the initial stages of L2 learning, but aptitude and awareness may be necessary for improving challenging features. Imitation skill might also account for some variance in the production of L2 sounds. On the other hand, motivation and effort have been linked to the development of global dimensions of pronunciation. The ideal L2 self seems to be related to comprehensibility, whereas motivated effort seems to be related to accentedness.

Two studies comparing SL and FL learners merit special attention. Baker Smemoe and Haslam (2013) examined relationships between aptitude, motivation, and strategy use and global pronunciation outcomes for L1 Chinese learners of English. There were few significant differences in strategy use according to aptitude and the context of learning, but motivation was positively related to different categories of strategy use. When SL and FL learners were combined for the pronunciation analysis, auditory aptitude was correlated with foreign accent, accuracy, and fluency, and motivation was correlated with fluency and comprehensibility. Muñoz and Llanes (2014) focused on L2 input and use factors for L1 Spanish adults and children who were learning L2 English in SL (study abroad) and FL contexts. Patterns of language use were strikingly different across contexts and age groups. Although SL participants reported spending more time speaking English than FL participants did, L2 contact was most consistent for SL children who reported the most L2 use overall and the most L2 use with native English speakers. Both of these variables (i.e., L2 speaking time and speaking time with native speakers) were significantly correlated with foreign accents. Thus, the results of these studies generally align with the SL and FL literature.

One other issue relevant to the present chapter is the role IDs play in shaping the efficacy of pronunciation instruction and pronunciation training. The studies that have examined IDs and explicit pronunciation instruction have shown that cognitive and socio-affective variables play a
IDs for L2 Pronunciation

small but significant role. In Elliott’s (1995a) study, motivation was correlated with post-instruction pronunciation accuracy, but it was not a significant factor in a regression of post-test scores. Kissling (2014) obtained a different result. In her study, motivation was not related to L2 Spanish learners’ overall production accuracy at post-test, but it did predict a small amount of variance in the experimental group’s realization of one challenging L2 sound. In contrast, phonetic aptitude and PSTM (Kissling, 2014) and field dependence (Elliot, 1995a) were not related to production accuracy in either study. However, the fact that these studies did not find a link between cognitive IDs and pronunciation instruction should not be taken as evidence that such a link does not exist. In Bowles et al. (2016), learners with no background in tonal languages participated in tonal word learning training. Each session included a familiarization phase, a training phase with right–wrong feedback, and a testing phase without feedback. The authors examined 22 cognitive IDs representing pitch-perception ability, musicality, general L2 aptitude, and general cognitive ability. All four ID types predicted final accuracy scores, but the domain-specific pitch-perception measures (i.e., tone identification and tone discrimination) were the strongest predictors overall.

Other experimental studies have examined whether certain learners benefit more from different types of training depending on their ID profile. Yuan et al. (2018) investigated the relationship between musical ability and L1 Mandarin learners’ production of L2 Spanish intonation contours. Learners were categorized as having a high-, mid-, or low-musical ability and assigned to intonation training conditions with or without visual pitch gestures. Participants with better musical abilities tended to perform better overall, and post-hoc analyses suggested that observing pitch gestures was beneficial for mid-level participants but not for high- and low-musical ability learners. In another study, Perrachione et al. (2011) divided participants into high- and low-aptitude groups and assigned half of the participants in each group to high- and low-variability (i.e., multi-talker and single-talker) training conditions. High-variability training boosted high-aptitude learners’ achievements, but it was detrimental for low-aptitude learners. However, when high-variability training was blocked by talkers to reduce cognitive load while maintaining the same overall level of variability, the low-aptitude group’s performance improved significantly.

Data Elicitation

Discrimination and identification tests are commonly used to assess perceptual accuracy. These tests tap into different aspects (and possible stages) of perceptual learning. Discrimination tests sensitivity to contrasts, whereas identification tests category formation. Thus, discrimination involves comparing auditory stimuli to one another, and identification involves associating a single auditory stimulus with an orthographic or visual representation. For example, to test L2 English speakers’ discrimination of the /i/-/ɪ/ contrast, learners might hear recordings of “sheep” and “ship” and be asked to indicate if the recordings are the same or different. To test their identification, they might see images of a sheep and a ship, hear a recording of “sheep”, and be asked to select the image that matches the word they heard. Even if L2 learners discriminate and identify L2 sounds accurately, they may rely on different cues than native listeners. To examine cue use, a continuum with steps from one phonetic extreme to the other can be used. For instance, to test L2 learners’ reliance on duration and spectrum as cues to the /i/-/ɪ/ contrast, researchers might create a phonetic continuum from one vowel to the other by manipulating duration and spectrum in equal steps (see, e.g., Kim et al., 2018). Each step is presented individually, and learners are asked to associate the target step with the corresponding vowel category (e.g., with an image of a sheep or a ship). Because perceptual learning is a multistage process (Flege & Bohn, 2021), it is important to assess both discrimination and identification accuracy throughout L2 development, but few studies have examined L2 perceptual learning longitudinally using both tests.

Most of the time, researchers aggregate raw data from individual perceptual trials into a composite outcome measure such as percent accuracy or $d$ prime, a sensitivity index that takes response bias
into account. These measures are then used as dependent variables in correlation, regression, and analysis of variance, commonly known as ANOVA. This approach is advantageous because it allows researchers to capture perceptual accuracy straightforwardly with a single score. At the same time, pooling over individual trials entails moving away from the basic task that L2 learners face when processing L2 speech: deciding if an auditory stimulus represents one category or another. This decision can be modeled using a logistic model of trial-level data (e.g., Kartushina & Frauenfelder, 2014). Studies on L2 perceptual learning have focused predominantly on accuracy without considering reaction time data, which could speak to perceptual efficiency. Together, accuracy and reaction time analyses could provide insight into how L2 learners create and refine L2 perceptual categories.

Saito and Plonsky (2019) have proposed three primary dimensions along which L2 production studies may differ: construct, which refers to the specific or global nature of the outcome measure (e.g., the production of specific structures such as L2 sounds vs. global comprehensibility and accentedness scores); scoring method, which represents how the data were coded (e.g., acoustic analysis of phonetic features vs. listener-based ratings of comprehensibility); and task type, where controlled and spontaneous speaking tasks (e.g., word reading vs. picture description) are associated with different types of pronunciation knowledge. Although L2 production research has addressed the production of specific and global features using both controlled and spontaneous speaking tasks, few studies have done so within a single design (but see, e.g., Saito, 2019; Saito et al., 2019). Thus, the field is still in the early stages of understanding how controlled and spontaneous pronunciation knowledge develops throughout L2 pronunciation learning. The same analytical trends observed for L2 perception research are also evident in the L2 production literature. Namely, researchers tend to aggregate individual acoustic measurements or listener ratings into speaker averages that are subsequently submitted to analysis. However, more sophisticated quantitative techniques, such as mixed-effects modeling, are becoming common (see, e.g., Linck & Cunnings, 2015). These techniques offer researchers the opportunity to quantify variance in speakers, listeners, and items through random effects and thus may be particularly appropriate for the hierarchical data sets typical in pronunciation research.

Although different types of IDs have been studied to varying degrees in the SL and FL literature, the way they have been operationalized and measured does not vary much by research context. Experiential factors have typically been quantified using surveys that ask L2 learners to provide information on their language learning history. Such self-report measures present some challenges. For instance, researchers often use closed-ended items to assess L2 use across a variety of domains. However, in doing so, they may omit potentially meaningful categories, and respondents may under or overestimate their L2 use depending on when the survey is administered. These pitfalls can be avoided by triangulating experiential IDs through the use of survey and interview data. Ideally, data should be collected on multiple occasions to capture variations in L2 use over time (for discussion, see Ranta & Meckelborg, 2013).

Whereas experiential IDs can be challenging to quantify comprehensively and accurately, cognitive IDs are multidimensional constructs that lend themselves to a range of direct and indirect measurement approaches. On the direct end of the continuum, researchers have used both domain-general measures, such as the LLAMA tests (Meara & Rogers, 2019), and domain-specific measures, such as pitch-perception tests. Researchers have examined links between the subcomponents of aptitude and different dimensions of L2 pronunciation learning (Baker Smemoe & Haslam, 2013; Saito et al., 2019) and training (Bowles et al., 2016; Perrachione et al., 2011). Notably, when domain-general and domain-specific measures are considered simultaneously, domain-specific measures seem to show a stronger relationship to pronunciation outcomes (Bowles et al., 2016). Another prominent direct cognitive measure is PSTM, which is typically assessed using nonword repetition and nonword recognition tasks to capture the storage and rehearsal components of the phonological loop. Recent research has also begun to tease apart explicit and implicit forms of pronunciation learning aptitude (Saito, Sun, & Tierney, 2019). Although less common, researchers
have sometimes used indirect (i.e., self-report) measures. For instance, Saito (2019) used Likert-type survey items to assess participants’ explicit pronunciation awareness. Most of the time, cognitive factors are treated as continuous variables in correlation and regression analyses, but some researchers have created categorical grouping variables using median splits. While this approach can be advantageous, it may prevent a finer-grained examination of gradient cognitive effects.

Among socio-affective IDs, motivation has been the most extensively studied. Motivation encompasses both general language learning motivation, commonly operationalized via the L2 motivational self-system (Dörnyei, 2009), and specific concern for pronunciation (Elliot, 1995a, 199512b). In either case, Likert-type surveys are used, and responses to individual items are averaged into a single score that represents motivational intensity. One issue that comes to the forefront in measuring motivation is the fact that motivation inevitably changes over time. Thus, longitudinal studies that aim to examine links between motivation and L2 pronunciation learning must take care to sample motivation longitudinally (Nagle, 2018). What’s more, methodological practices in L2 motivation research are rapidly evolving (e.g., Papi et al., 2019), so it is important that L2 pronunciation researchers use the most up-to-date, validated instruments.

**Practical Applications**

The pronunciation-ID links that have been documented in the SL and FL literature suggest ways to support pronunciation development in both learning contexts. First, it is important to acknowledge that the basic goal of perception training is to help L2 learners perceive L2 speech accurately, and the basic goal of production training is to help them develop intelligible and comprehensible pronunciation (Levis, 2005). Although SL learners’ pronunciation skills improve with experience, they can still benefit from targeted perception and production training. High-variability phonetic training can help them improve their perception of L2 sounds and can be accessed online using programs such as English Accent Coach (Thomson, 2012). Teachers working in an SL context can also simulate high-variability training by exposing learners to different talkers and can maximize benefits for all learners by presenting stimuli from one talker at a time (Perrachione et al., 2011). Even if SL learners’ pronunciation seems to have fossilized after many years of L2 immersion, instruction can be beneficial (Derwing et al., 2014).

Where opportunities for L2 interaction are plentiful in the SL context, the primary source of L2 input and interaction for FL learners is often the L2 classroom. Fortunately, technology offers FL instructors many tools for creating additional opportunities for meaningful L2 interaction, such as telecollaboration (Akiyama & Saito, 2016). Given positive links between motivation, self-efficacy, and learner engagement, instructors should also create opportunities for learners to reflect upon their pronunciation learning, explore different learning strategies, and develop an awareness of L2 pronunciation features (Saito, 2019; Sardegna et al., 2018). Moreover, given the importance of effort to improving pronunciation, students should be encouraged to practice their pronunciation. Here too, technology can be useful. For instance, learners can visualize their pronunciation using acoustic analysis software, and when acoustic analysis assignments avoid technical terminology, they can be successfully carried out in the L2 (Olson, 2014). Teachers can also take advantage of smartphones and other devices by asking learners to shadow L2 speakers (Foote & McDonough, 2017). Through these activities, students can develop confidence in their ability to improve their pronunciation, which could lay the groundwork for long-term success in pronunciation learning.

**Future Directions**

The state of the art in L2 pronunciation research is shifting toward a more dynamic, interactive view of IDs. As a result, researchers must begin to adopt longitudinal designs that align with this perspective, which would entail sampling IDs longitudinally to understand how they change and
co-vary. Because IDs such as L2 use and motivation are inherently time-varying, they must be sampled on more than one occasion. A wider variety of longitudinal designs would be beneficial for understanding the extent to which IDs interact to shape L2 pronunciation learning on different timescales. This would include long-range studies that sample IDs and L2 pronunciation infrequently over longer periods (i.e., years) and dense studies that sample IDs frequently over shorter periods (i.e., weeks or months). There is also a need for longitudinal research that examines how different types of IDs interact with one another. For example, more motivated learners may be more likely to seek out additional opportunities for L2 interaction, and they may be more likely to scrutinize the input they receive, both of which could catalyze their pronunciation learning. Motivation might also interact with cognitive IDs such as PSTM and aptitude. Learners with higher aptitude or PSTM might find L2 speech processing easier, making them more inclined to interact in the L2. If success in L2 pronunciation learning makes L2 learning more enjoyable, then the positive emotions that learners experience could feed back into their motivation. In short, it is easy to imagine how different types of IDs might work together to create a self-reinforcing developmental loop for L2 pronunciation learning (Moyer, 2014a). These are the intriguing questions future research must begin to explore.

There has been a tendency to adopt a quantitative approach to IDs and L2 pronunciation learning. On the one hand, quantitative research is valuable because it allows researchers to draw comparisons between ID effects in different contexts of learning and at different points in the developmental process. On the other hand, purely quantitative studies run the risk of portraying learners as bundles of variables rather than as active agents whose choices affect their L2 pronunciation development (Moyer, 2017). Qualitative or mixed-methods approaches could prove ideal for illuminating how IDs interact and how L2 learners actually engage in the pronunciation learning process.

Most L2 pronunciation studies have focused on L2 English. The de facto status of English as a global lingua franca means that one might anticipate qualitative differences in the socio-affective profiles of individuals learning L2 English compared to individuals learning other L2s. This does not mean that learners of L2 English and learners of other L2s do not share common goals, motivations, and approaches, but rather that L2 English speakers may have specific reasons for learning English that are not frequently documented in other learner populations. A similar issue arises with respect to participant samples, which are skewed toward adult learners, especially college-aged individuals. This issue is not trivial given that many IDs co-vary with age (Moyer, 2014b). Thus, future work should aim to broaden sampling practices both in terms of participant characteristics and language pairs.

In closing, it is worth repeating that current models of L2 sound learning were not developed with IDs in mind. While L2 pronunciation researchers have been able to draw broad hypotheses from these models, the time is right to begin formulating new models that directly address IDs and L2 pronunciation learning. In order to do that, future work must embrace longitudinal designs to study how IDs interact in different learning contexts and learner populations.

References


Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and


