Received: 22 December 2022 | Revised: 18 July 2023 | Accepted: 18 July 2023

DOI: 10.1111/ijal.12494

ORIGINAL ARTICLE

WILEY

Speaking anxiety and task complexity effects on second language speech

Joan C. Mora | **Ingrid Mora-Plaza** | **Gonzalo Bermejo Miranda**

Department of Modern Languages and Literatures and English Studies, Universitat de Barcelona, Barcelona, Spain

Correspondence
Joan C. Mora, Universitat de Barcelona, Department of Modern Languages and Literatures and English Studies, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain.
Email: mora@ub.edu

Funding information
Spanish Ministry of Science, Innovation and Universities, Grant/Award Number: PID2019-107814GB-I00

Abstract

The association between speaking anxiety and L2 speech production, including L2 pronunciation, remains largely under-researched, especially in relation to task complexity. The present study investigates the effect of task complexity on speaking anxiety and their impact on specific dimensions of L2 speech production: speaking fluency (speed, breakdown, and repair) and accuracy (grammar, lexis and pronunciation); and global assessments of L2 speaking performance: accentedness and comprehensibility. Forty-two Spanish learners of English performed simple and complex versions of a monologic oral narrative task. The results indicated that task complexity affected learners' anxiety levels and was detrimental to their L2 speaking fluency, pronunciation accuracy, and accentedness. Moreover, higher self-perceived anxiety was associated with lower breakdown fluency and less lexico-grammatical accuracy. Last, once the contributions of L2 proficiency and working memory were controlled for, anxiety accounted for a significant 13%–15% of variance in breakdown fluency.

KEYWORDS
individual differences, oral production, second language pronunciation, speaking anxiety, task complexity

Resumen

La asociación entre la ansiedad al hablar y las dimensiones de la producción oral en L2, incluyendo la pronunciación, no se ha investigado suficientemente, especialmente en relación a la complejidad de una tarea oral. Este estudio investiga el efecto de la complejidad de la tarea en la ansiedad al hablar y su impacto sobre dimensiones específicas de la producción oral tales como la fluidez, la corrección gramatical, léxica y fonológica, y las dimensiones perceptivas del habla en L2 (acento y comprensibilidad). Cuarenta y tres estudiantes españoles de inglés realizaron una narrativa oral monológica simple y una compleja. Los resultados mostraron que la complejidad de la tarea aumentó sus niveles de ansiedad y fue perjudicial para su fluidez, su pronunciación y su acento. Además, niveles subjetivos de ansiedad más elevados se relacionaron con una menor fluidez y corrección léxico-gramatical. Por último, al analizar los resultados controlando por competencia lingüística y memoria operativa, la ansiedad explicó 13%–15% de la varianza en la fluidez oral.

PALABRAS CLAVE

diferencias individuales, producción oral, pronunciación de una segunda lengua, ansiedad al hablar, complejidad de la tarea

1 | INTRODUCTION

Foreign language anxiety (FLA) is one of the most widely investigated affective variables in both second (L2) and foreign language (FL) acquisition research, together with other related affective factors such as motivation (e.g., Dörnyei & Ushioda, 2013) and willingness to communicate (e.g., MacIntyre, 2007). FLA research has been mainly concerned with explaining the extent to which the type of situational state anxiety that occurs in language classrooms, such as anxiety originating in fearing negative evaluation when speaking the L2, is detrimental to L2 learning (Horwitz, 2017; Horwitz, Horwitz, & Cope, 1986), whereas research on the detrimental effects of speaking anxiety on dimensions of L2 speaking performance (e.g. fluency), such as when performing an oral narrative task, is relatively scarce (Pérez Castillejo, 2019), as is the effects of the task complexity (and difficulty) on L2 speaking anxiety (Brennan, 2016; Donate, 2022). The current study extends this line of research by focusing on the relationship between task complexity, speaking anxiety, and oral performance.

In addition, most FLA research examining the effect of anxiety on L2 oral performance has obtained measures of anxiety based on L2 learners' questionnaire-based self-reports of feelings during task performance (e.g., Saito & Samimy, 1996; Saito et al., 2018) or through scalar judgements (e.g. Nagle et al., 2022) obtained after task completion, often comparing the performance of low- and high-anxiety learners (Gregersen et al., 2014; Sheen, 2008) under various anxiety-generating conditions (MacIntyre & Gardner, 1994). Such subjective perceptions of anxiety gathered after task performance are likely to be influenced by feelings of success (or failure) in performing the oral task or other individual factors (self-esteem, over/under-estimation). Therefore, electrophysiological measures of anxiety obtained through sensors at the time of performing the speaking task (heart rate, electro-dermal activity) might be better suited than questionnaire-based measures to investigate the effect of anxiety on speech production under different task complexity conditions, as they are dynamic and synchronous with task performance and might be related to subjective perceptions of anxiety (Gregersen et al., 2014; MacIntyre et al., 2010). However, to the best of our knowledge, no studies to date have explored the impact that cognitive task demands may have on physiological measures of anxiety and how these might be related to L2 learners' speaking performance. The current study investigated the relationship between anxiety and oral performance using heart rate measures indexing general levels of anxiety during L2 learners' speaking performance in comparable tasks differing in cognitive complexity. In this way, we aim at contributing significantly to our current understanding of how affective individual differences factors like speaking anxiety may affect L2 learners' oral production, potentially explaining one of the sources of inter-learner variability in speaking performance and development. Does increased task complexity lead to increased anxiety levels during L2 learners' oral performance? How does task complexity affect L2 learners' fluency, accuracy, comprehensibility, and accentedness?

The present study investigated the potential effects of task complexity (performing a simple vs. a complex problem-solving task) on L2 learners' individual levels of objective physiological anxiety (HR measures) and subjective perceived anxiety (performance questionnaire) during task performance and examined the impact of task complexity and anxiety levels on fluency, accuracy, comprehensibility, and accentedness in L2 learners' speech.

2 | LITERATURE REVIEW

2.1 | Anxiety and oral production

FLA as a construct in second language acquisition is a type of state anxiety that subsumes various negative emotions learners experience in L2 classrooms. It includes feelings related to certain language-learning tasks, but it also develops over time because of the repeated experience with anxiety-generating situations linked to L2 classrooms (MacIntyre & Gardner, 1989), such as communicating in the L2, fear of negative evaluation during language performance and language testing (Horwitz, 2017; Horwitz et al., 1986). This is a situational type of state anxiety uniquely associated with L2 learning that has been shown to be independent from anxiety as an individual trait (Horwitz, 2017; MacIntyre, 2017), although L2 learners having high trait anxiety tend to experience state anxiety when using their L2 (MacIntyre & Gardner, 1989). The present study focuses on L2 speaking anxiety, the kind of anxiety learners experience when communicating orally in their L2, either in-class during communicative tasks or out-of-class in real communicative situations (Woodrow, 2006). As we examine the potential effects of task complexity on L2 speaking anxiety and how these relate to the quality of L2 learners' oral production, we could more precisely identify it as a sub-type of task anxiety or TA (Donate, 2022), L2 speaking anxiety related to features of communicative task design (difficulty, complexity).

Research has shown that anxiety is detrimental to language learning in both L2 and FL learning contexts, overall explaining about 13% of the variance in language learning according to a recent metanalysis (Teimouri et al., 2019). FLA also interferes negatively with L2 learners' performance (MacIntyre & Gardner, 1994) across various socio-linguistic conditions (Gkonou, 2014; Sevinç, 2018) and proficiency levels (Liu, 2006; Saito & Samimy, 1996), and has been shown to negatively impact reading (Saito et al., 1999), listening (Elkhafaifi, 2005), writing (Cheng, 2004), and speaking abilities (Hewitt & Stephenson, 2012; Pérez Castillejo, 2019; Young, 1990). This negative impact is partly due to anxiety interfering with attentional and encoding processes responsible for building linguistic representations with retrieval processes affecting speech production, and consequently with task performance. Attention-related psycholinguistic processes known to underlie language learning, such as the noticing of linguistic features in the linguistic input, or the proceduralization and later automatization of declarative knowledge, are all debilitated by FLA (Kormos, 2015). For example, Zuniga and Simard (2022) found a positive relationship between anxiety and the number of self-initiated self-repairs in L2 speakers performing a picture-cued oral narrative, but it was only for anxious speakers with low attentional control that language anxiety was positively related to the production of self-repairs. Therefore, anxiety can be considered an attention-draining factor that detrimentally affects L2 speech production (Kormos, 2015), including all the dimensions of L2 learners' speech we focus on in the current study: fluency, accuracy, comprehensibility, and accentedness.

Speaking fluency, defined as the smooth delivery of speech utterances in terms of their temporal properties (speech rate, articulation rate), breakdowns (pauses, hesitations), and repairs (repetitions, reformulations), has been shown to be negatively affected by speaking anxiety, as has perceived fluency (listeners' perception of speakers' speech fluidity). For example, in MacIntyre and Gardner (1994), higher-anxiety learners were judged to be less fluent (lower perceived fluidity in speech) than lower-anxiety learners in a 1-min self-description task. Pérez Castillejo (2019) found learners with higher speaking anxiety to pause more frequently within major syntactic units and to have lower phonation time ratios in a 2-min open-ended narrative monologue on their weekend activities. Lindberg et al. (2021) found learners' fluency self- and partner-ratings to be negatively associated with medium and high anxiety arousals (as measured through galvanic skin response sensors) in a 10-min interactive communicative oral task (goal-oriented discussion on international students' challenges).

Due to the detrimental effect of anxiety on the attention-related processes governing speech production (e.g., lexical activation and grammatical encoding), higher levels of anxiety while performing an oral task in the L2 may be expected to result in decreased lexical and grammatical accuracy. However, while low anxiety is associated with the production of grammatically more complex speech (Robinson, 2007), the use of a wider variety of vocabulary (Kormos & Dörnyei, 2004), and learners' ability to self-repair production errors (Sheen, 2008) and high anxiety has been linked to word-retrieval difficulties (Donate, 2022), the detrimental effects of speaking anxiety on specific measures of grammatical and lexical accuracy have not been demonstrated. For example, Révész (2011) found anxiety to be unrelated to measures of grammatical and lexical accuracy (errors per AS-unit, error-free AS units, self-repairs) and Phillips (1992) found anxiety to be unrelated to a measure of percent of error-free communication units, whereas negative correlations between anxiety and global assessments of spoken performance are more widely attested (Hewitt & Stephenson, 2012).

Although negative relationships between language anxiety and pronunciation accuracy and development are well attested in the literature (e.g., Baran-Łucarz, 2011, 2022; Szyszka, 2017), and there is empirical evidence for a *pronunciation anxiety* construct ("the feeling of apprehension experienced by non-native speakers in oral-communicative situations, due to negative/low pronunciation self-perception and to beliefs and fears related to pronunciation", Baran-Łucarz, 2014, p. 453), research on the effect of anxiety on accentedness (listeners' perception of degree of nativelikeness in L2 speech) is scarce to date. However, language anxiety is negatively associated with self-perception of pronunciation competence (Szyszka, 2017), and more anxious learners have been found to speak with less of a target accent (MacIntyre & Gardner, 1994).

Research on the effect of anxiety on comprehensibility (the subjective perception of ease or difficulty in understanding speech) is also scarce to date. Saito et al. (2018) found anxiety (assessed through an emotion questionnaire) to negatively interfere with speech comprehensibility gains in Japanese high-school students learning English as a foreign language over a 3-month period. Similarly, Nagle et al. (2022) used an interactive oral task and found the interlocutors' anxiety levels (assessed through a 100-point continuous scale) to be negatively related to their speech comprehensibility. However, to the best of our knowledge, no studies have investigated the impact of anxiety on accentedness and specific L2 pronunciation measures.

2.2 | Physiological measures of anxiety

Speech-related language anxiety has been shown to lead to changes in psychophysiological stress reactivity measurable through cardiovascular (heart rate, HR; heart rate variability, HRV) or electrodermal measures (skin conductance level and response, SCL and SCR) (see Kreibig, 2010). These measures have been used widely in studies investigating anxiety in public speaking (e.g., Beatty & Behnke, 1991; Bodie et al., 2010; Croft et al., 2004; Felnhofer et al., 2014; Witt et al., 2006) as well as, though less frequently, in foreign language anxiety studies (e.g., Gregersen et al., 2014; MacIntyre et al., 2010). Interestingly, research has shown that speakers' self-reports of experienced public speaking anxiety during a speech do not necessarily correspond with physiological arousal measures of HR and electrodermal activity (EDA; Gallego et al., 2022; Schwerdtfeger, 2004), while Gregersen et al. (2014) observed a strong relationship between L2 learners' perceptions of anxiety and physiological HR measures during a speaking task. They also found L2 Spanish learners with high trait anxiety to experience higher arousal during a classroom presentation than their low anxiety peers. Whereas HR increases as a function of increased anxiety levels, HRV (the standard deviation of inter-beat intervals) has been positively associated with speakers' ability to regulate emotions to better cope with speaking anxiety (MacIntyre et al., 2010). To the best of our knowledge, no studies to date have used physiological measures to explore task complexity effects on speaking anxiety.

2.3 | Anxiety and task complexity in speaking tasks

Task difficulty, moderated by learners' L2 proficiency level, is known to trigger anxiety effects on speech processing and output production (MacIntyre & Gardner, 1994), but research on task complexity effects on speaking anxiety, potentially also moderated by learners' L2 proficiency level, is relatively scarce and has so far produced mixed results. For example, Kim and Tracy-Ventura (2011) found low-anxiety learners to outperform high-anxiety learners across tasks varying in cognitive complexity, but they did not find task complexity to significantly affect learners' anxiety levels. Révész (2011) did not find her anxiety measure (based on three self-report questionnaire items) to significantly correlate with measures of complexity and accuracy in speech production (in neither the simple nor the complex version of a speaking task), whereas Robinson (2007) found that a higher output anxiety learner group produced less complex syntax than a lower output anxiety learner group. However, Brennan (2016) found the complex version of the fire chief task (Gilabert, 2007; Gilabert et al., 2009; Révész et al., 2016) to trigger higher self-reported anxiety (SRA) levels than its simple counterpart. Donate (2022) examined the effects of manipulating task complexity on the speaking performance of low-level learners of Spanish through two picture-based oral narrative tasks (simple vs. complex, differing in the number of elements), and qualitatively determined (through an analysis of participants' comments) strong associations between task anxiety, cognitive processes (e.g., word retrieval problems), and oral performance on two L2 levels of task complexity. The current study implements similar task complexity manipulations, but unlike Donate (2022), potential links between task complexity and speaking anxiety were also examined quantitatively through physiological measures (HR) and related such measures to the quality of learners' oral production in terms of speaking fluency, accuracy, accentedness, and comprehensibility measures.

Task-based language teaching (TBLT) research on task complexity has shown that more cognitively complex tasks draw learners' attention to linguistic form and result in oral productions of increased linguistic complexity and accuracy, thus promoting L2 grammatical, lexical, and pragmatic development (Kim & Taguchi, 2015; Robinson, 2005, 2011; Robinson & Gilabert, 2007). Whether the extra attention to form driven by increased task complexity will extend to pronunciation is still unclear. Recent studies suggest that this may be the case in interactive map tasks (Mora & Levkina, 2018; Solon et al., 2017) or in communicative decision-making tasks (Mora-Plaza et al., 2018) where the use of phonetic targets is essential for task resolution. For example, Gordon (2021) showed that a complex TBLT intervention focusing on phonetic form led to significant improvement in L2 learners' speech comprehensibility. In contrast, other research has shown that in tasks where no specific pronunciation features are targeted, pronunciation accuracy assessed through accentedness ratings either decreases or shows no improvement as task demands increase (Crowther et al., 2018). Therefore, it is uncertain at present what impact anxiety may have on learners' ability to benefit from the attention to form induced by increased task demands.

2.4 | Individual differences in proficiency and working memory

Individual differences in proficiency and working memory constitute sources of variability in determining the effects of task complexity on oral production performance, playing an especially important role in cognitively complex tasks (Robinson, 2011). Research has identified L2 proficiency as one of the causes of anxiety, as lower levels of language ability are likely to lead to higher levels of anxiety (Baran-Łucarz, 2022; Donate, 2022; Jiang & Dewaele, 2020; Teimouri et al., 2019). In the context of the current study where task complexity is manipulated, learners with higher proficiency may be better able than lower proficiency learners to cope with the challenging linguistic demands of a more complex speaking task. Higher proficiency may also provide an advantage to learners in helping them control or diminish anxiety levels while speaking, which may originate in L2 learners' negative self-perceptions of speaking task performance (Gkonou, 2014). In this study, we used an elicited imitation task (EIT), which has been demonstrated to have the potential to distinguish effectively and reliably different proficiency levels (McManus & Liu, 2020; Yan et al., 2016 for a meta-analysis), independently of phonological short-term memory (Kim, Tracy-Ventura, & Jung, 2016).

High working memory capacity may facilitate L2 oral production, especially in tasks requiring high cognitive demands (Kormos & Sáfár, 2008). However, Kormos and Trebits (2011) found that working memory only had an effect for the task they operationalized as simple (a cartoon description task), whereas no effect was found for the one operationalized as complex (a picture narration task). On the contrary, Kim et al. (2015) found that the task where working memory capacity made a difference was the one they had operationalized as complex, whereas the effect of task complexity in terms of recast noticing and question development was nonsignificant. Despite such lack of consistency in determining the role of working memory in complex speaking tasks, working memory is an important individual differences factor related to speaking performance (see Kormos, 2015 for discussion). In addition, working memory may be related to speaking anxiety because the former is implicated in regulating attention during the speaking process, and the latter is detrimental to L2 learners' ability to focus their attention on performance during the speech production process. According to Eysenck et al.'s (2007) attentional control theory, the adverse effects of anxiety on performance are larger on tasks that impose substantial demands on the processing and storage capacity of working memory. As a consequence, tasks that make high demands on working memory will be negatively affected by high levels of anxiety and individual differences in both L2 proficiency and working memory may interfere with measuring task complexity effects on speaking anxiety and on the quality of the speaking performance. In the current study, we aim to control for individual differences in L2 proficiency and working memory when assessing the effects of anxiety on L2 speech production.

The aim of the current study is twofold: (1) to explore the role of cognitive task complexity on L2 learners' subjective (self-rated anxiety, SRA) and objective physiological (HR) measures of anxiety while performing speaking tasks, and (2) to assess the effects of anxiety and task complexity on learners' L2 speech production (fluency, accuracy, accentedness, and comprehensibility) while controlling for individual differences in L2 proficiency and working memory. To that end, we had L2 learners perform a simple and a complex version of a monologic decision-making oral narrative task: the *fire-chief task*, (Gilabert et al., 2009), which previous research has shown to trigger different levels of perceived anxiety (Brennan, 2016), and we obtained measures of learners' L2 proficiency and working memory to control for their potential effects when assessing speaking anxiety effects on L2 speech production. This study was guided by the following research questions (RQ):

RQ1. Does task complexity manipulation (differential cognitive demands in a simple and complex version of a decision-making speaking task) affect physiological (heart rate) and self-reported measures of L2 speaking anxiety?

RQ2. Does task complexity manipulation affect dimensions of L2 speech production (fluency, accuracy, accentedness, and comprehensibility)?

RQ3. Are L2 learners' physiological and SRA levels related to dimensions of L2 speech production (fluency, accuracy, accentedness, and comprehensibility)?

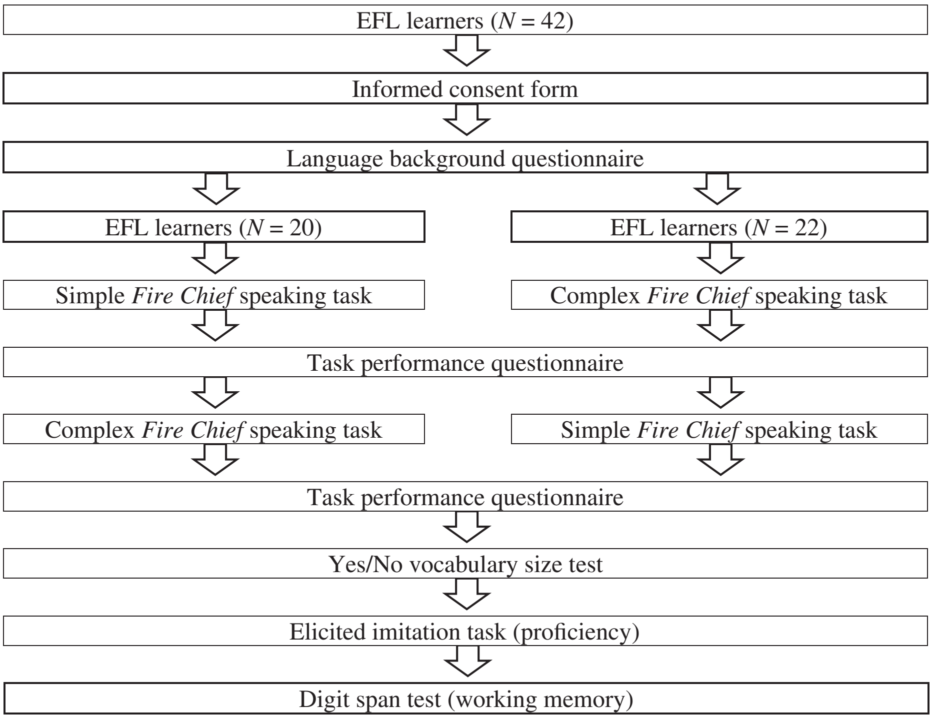


FIGURE 1 Research design.

RQ4. How do physiological and SRA levels contribute to explaining inter-learner differences in L2 speech production (fluency, accuracy, accentedness, and comprehensibility) while controlling for individual differences in L2 proficiency and working memory?

3 | METHODS

Data collection took place individually in a quiet lab in a single 45-min session. Participants first signed an informed consent form and filled in a language background questionnaire. Then, the EFL learners performed a simple and a complex version of a decision-making speaking task (*Fire Chief*) in a counterbalanced order while wearing HR sensors. For each task, participants read the instructions and looked at the picture for 30 s before speaking. Immediately after task completion, they assessed their perceived level of task difficulty, effortfulness, and anxiety on 9-point Likert-scale questionnaires (Révész et al., 2016). Finally, they completed vocabulary size, L2 oral proficiency, and working memory tests on individual laptops. Their oral productions were recorded on a Marantz PMD-661 solid-state digital recorder with an external Shure SM58 voice microphone at a 44.1-KHz sampling rate (16-bit resolution; 1 channel). These speech recordings were manually transcribed and annotated in Praat (Boersma & Weenink, 2016) for silent and filled pauses above 250 ms (de Jong & Bosker, 2013), and their location (either mid-clause or end-clause) and measures of speed, breakdown, and repair fluency (Skehan, 2003) as well as accuracy (i.e., lexical, grammatical, and pronunciation errors) were obtained. In addition, we obtained measures of receptive vocabulary size, L2 oral proficiency, and working memory. Thirty-second excerpts from each task by each participant were then rated for comprehensibility and accentedness by a panel of seven native English listeners (Figure 1 illustrates the research design and the experimental procedures).

TABLE 1 Participants' demographics: means (M), standard deviations (SD), and 95% confidence intervals (CI).

	M	SD	Range	95% CI
Age at testing	22.44	7.56	18–57	20.12–24.77
Age of onset of L2 learning	6.13	3.72	3–24	4.91–7.35
Instruction (h/week)	3.18	1.99	0.67–12	2.52–3.84
Input (h/week)	9.79	11.49	1.17–44	6.01–13.57
Output (h/week)	7.69	8.41	0–32	4.93–10.46
Vocabulary size (0–10 K)[a]	6289	1269	1950–9250	5872–6706
Self-estimated proficiency (0–9)[b]	6.77	1.62	1.8–9	6.24–7.31
Oral proficiency (0–120)[c]	94.76	18.57	47–119	88.65–100.86

[a]Obtained through X/Y Lex vocabulary size tests (Meara & Milton, 2003; Meara & Miralpeix, 2006).
[b]Averaged self-estimated ability to speak spontaneously, understand, read, write, and pronounce English (1 = very poor - 9 = native-like).
[c]Obtained through an Elicited Imitation Task (Ortega et al., 2002).

3.1 | Participants

Forty-two (38 females, 4 males) undergraduate EFL learners volunteered in the study for course credit. They were either balanced Catalan-Spanish bilingual speakers ($N = 33$) or Spanish-dominant speakers ($N = 9$) living in Spain (see Table 1 for demographics). They had started learning English as a foreign language at school since around age 6 and differed widely in terms of exposure and use of English outside class. Their oral proficiency in English indicated on average a B2 level in the Common European Framework of Reference for Languages as assessed through an EIT (Ortega et al., 2002). No participants reported hearing impairments or speech pathologies.

3.2 | Tasks

3.2.1 | L2 speaking task

Learners' performed a simple and a complex version of the *fire-chief task* in English (Gilabert et al., 2009). This task was deemed appropriate both to generate a sufficiently long sample of unprepared extemporaneous speech from which to obtain measures of speaking performance, and to induce inter-learner variability in speaking anxiety. The simple and complex versions of the task differed in terms of their cognitive demands along resource-directing variables (i.e., ± reasoning demands) and were designed to be equivalent at the level of resource-dispersing variables (i.e., planning time: 1 min, and prior knowledge of the target items; Robinson, 2005, 2007, 2011). Thus, the complex version of the task was expected to generate increased focus on form relative to the simple version of the task (Robinson, 2005, 2011; Robinson & Gilabert, 2007). Regarding anxiety, Brennan (2016) had found the complex version of this task to generate significantly higher state anxiety ratings than the simple version.

In this task, learners were instructed to handle a challenging fire emergency situation (Gilabert et al., 2009). They were given a drawing of a building on fire where a number of people needed help from rescue teams (see Appendix A) and were prompted to describe and justify the actions they would undertake. Cognitive complexity was operationalized in terms of the number of resources available to extinguish the fire and the number of elements to be factored in when making decisions. The *fire-chief task* has been widely employed in previous L2 oral (Gilabert et al., 2009; 2011) and writing studies (Vasylets et al., 2017), and it has been empirically validated in terms of the cognitive complexity required to complete each version of the task (Gilabert, 2007).

3.2.2 | L2 oral proficiency

The EIT from Ortega et al. (2002), which included 30 English sentences ranging in length (7–19 syllables) and grammatical and lexical complexity, was presented auditorily over headphones. Participants were instructed to repeat each sentence as accurately as they could after a 250 ms beep signal, which occurred 2000 ms after the sentence ended, and had 6.8 s to repeat the sentence after the beep. Following Ortega et al.'s (2002) rubric, the learners' recorded productions were then scored (0–4 points) depending on how much of the original sentence was repeated and how accurately. The second author rated the learners' sentence productions following the rubric and rating examples provided in Ortega et al. (2002). Thirty percent of the data was rated by an additional applied linguistics' researcher, who was previously familiarized with the speech samples and trained to provide scores to the learner productions in the EIT. *Cohen's kappa* coefficient indicated good substantial agreement ($\kappa = 0.74$) between the two raters.

3.3 | Working memory

Learners carried out a forward- and a backward-digit span task (Montero Perez, 2020) in Inquisit 5 Lab (Draine, 1999). The forward-digit span task assesses storage capacity in the short-term memory phonological loop, whereas the backward-digit span requires both storage and manipulation of information. For each trial, participants were visually presented with a random sequence of digits, which appeared one by one on the computer screen with a 100 ms pause between items. Learners were then asked to recall the sequence of digits in the forward or reverse order in which they had appeared by typing the answer into a textbox. The number of digits in a sequence started at two in the first trial and increased until the participants failed to correctly recall two consecutive sequences of the same length. Participants' scores on each of the digit span tests equaled the number of sets for which they could repeat the sequence correctly. A global working memory measure was computed from the sum of both forward and backward digit span mean scores.

3.4 | Oral production and anxiety measures

Here, we describe the objective and subjective measures of speaking performance, speaking anxiety, and individual differences in L2 English (see Table 2 for a summary of the measures and how they were calculated).

3.5 | Speaking fluency and accuracy

L2 speaking fluency is a multidimensional construct consisting of utterance fluency (the temporal characteristics of speech production, including hesitations and repairs), cognitive fluency (the efficient use of the psycholinguistics mechanisms underlying speech production), and perceived fluency (listeners' perception of how fluent a speaker is) (Segalowitz, 2010). In the current study, we are concerned with the measurable aspects of utterance fluency that may be indicative of the lack of automaticity in the formulation (lexical, grammatical, morphophonological, and phonetic encoding) articulation and monitoring processes characteristic of L2 speech production (Kormos, 2006; Segalowitz, 2010). All such processes may be affected by task factors, such as task complexity, or speaking anxiety, leading to less fluent speech characterized by slower speed, increased pause frequency and duration, and more frequent hesitations and repetitions (Robinson, 2011). Following Skehan's (2003, 2009) and Tavakoli and Skehan's (2005) taxonomy, we operationalized L2 utterance fluency into three components: *speed*, *breakdown*, and *repair*. Following De Jong et al. (2015b), we chose a set of measures to uniquely represent each dimension and avoided using measures like overall speech rate (number of syllables per minute of speaking time, including pausing time) that would encompass two components into a single measure. Thus, as in De Jong et al. (2015b) and Kahng (2020), we operationalized *speed fluency*

TABLE 2 Objective and subjective measures of speaking performance, speaking anxiety and individual differences in L2 English.

Measures		Calculation
Fluency ([1] = speed, [2] = breakdown, [3] = repair)		
MSD[1]	Mean syllable duration	Phonation time[a]/no. syllables (syll)
PD[2]	Pause duration (ms)	Pausing time/no. filled and silent pauses
SPD[2]	Silent pause duration (ms)	Pausing time/no. of silent pauses
PF[2]	Pause frequency	Silent and filled pauses x 100 syll
MCPF[2]	Mid-clause pause frequency	Mid-clause silent and filled pauses x 100 syll
MCSPF[2]	Mid-clause silent pause frequency	Mid-clause silent pauses x 100 syll
REP[3]	Repetition and corrections	Disfluencies x 100 syll
Accuracy		
Err	Error rate	Errors x 100 syll
ErrGL	Lexico-grammatical error rate	Grammatical and lexical errors x 100 syll
ErrP	Pronunciation error rate	Pronunciation errors x 100 syll
Speech ratings		
COM	Comprehensibility	Ease of understanding (1 = very difficult, 9 = very easy)
ACC	Accentedness	Degree of foreign accent (1 = not accented, 9 = strongly accented)
Anxiety		
HRm	Heart rate mean	Mean beats per minute
HRsd	Heart rate standard deviation	Standard deviation of mean beats x minute.
SRA	Self-reported anxiety	Rating of anxiety (1 = no anxiety, 9 = great anxiety)
Individual differences		
EI	Elicited imitation (proficiency)	Sum of 0–4 score x 30 sentences (0-120)
WM	Working memory	Forward and backward digit span test score

[a]Phonation time is the total speaking time minus silent time (De Jong et al., 2015b).

as mean syllable duration (MSD: speaking time in ms excluding pausing time divided by the number of syllables produced), *breakdown fluency* as pause duration (PD: mean duration of silent and filled pauses; and SPD: mean duration of silent pauses) and pause frequency (PF: number of silent and filled pauses per 100 syllables; and MCPF: number of mid-clause silent and filled pauses per 100 syllables), and *repair fluency* (REP) as the number of repetitions and corrections per 100 syllables. In addition, we included a measure of mid-clause silent pause frequency (MCSPF), as this measure (together with REP and MSD) has been found to correlate strongly with perceived fluency (Suzuki & Kormos, 2019) and to be strongly related to L2-specific knowledge and cognitive skills and only weakly influenced by L1 speaking style (Kahng, 2020).

Speaking accuracy measures were obtained by computing error rates (number of errors × 100 syllables) based on the accuracy-annotated orthographic transcriptions in Praat. The three research assistants who orthographically transcribed the oral productions were instructed to annotate lexical, grammatical, and pronunciation errors (phonemic substitutions and mispronunciations). Three error rates were computed: total errors × 100 syllables (Err), lexico-grammatical errors × 100 syllables (ErrGL), and pronunciation errors × 100 syllables (ErrP).

3.6 | Comprehensibility and accentedness

Comprehensibility and accentedness of L2 speech were measured globally through native speakers' perceptual judgements (Saito & Plonsky, 2019). Seven native speakers of English (*M* age = 28.57, *SD* = 4.86, *Range* = 22–37), graduate students in applied linguistics and experienced EFL teachers familiar with Spanish-accented speech, were recruited as expert raters. They all assessed 86 speech samples (43 participants by two L2 narratives, one simple and one complex) on 9-point Likert scales for comprehensibility (1 = very difficult to understand, 9 = very easy to understand) and accentedness (1 = not accented, 9 = strongly accented) presented randomly through *Praat*. Inter-rater reliability (Cronbach's alpha intra-class correlation coefficients) of native listeners' ratings was high for comprehensibility ($\alpha = 0.88$) and accentedness ($\alpha = 0.90$), so average comprehensibility and accentedness ratings were computed across listeners for each learner.

3.7 | Anxiety

Individual physiological indices of anxiety and heart rate (HR) were obtained through NeuLog sensors (https://neulog.com/) by averaging the physiological response of each participant (beats per minute) across the duration of their speaking performance (HRm), which was 85 seconds on average, and by calculating an individual HR standard deviation measure (HRsd) representing the magnitude and frequency of anxiety peaks. Their self-perceived level of anxiety (SRA) was measured on a 9-point Likert scale (1 = no anxiety at all; 9 = great anxiety). See Table 2 for a summary of speaking performance, anxiety, and individual differences measures.

3.8 | Analytical procedures

All statistical analyses were conducted in SPSS 26. Relationships between anxiety measures and between anxiety measures and oral production measures (RQ3) were assessed through non-parametric correlations (*Spearman's* rank-order correlation coefficients), as these data were found not to be normally distributed. Task complexity effects on speaking anxiety (RQ1) and oral production (RQ2) were assessed through linear mixed-effects models, which included a random intercept for subject to control for inter-subject variability. Finally, hierarchical multiple regressions were used to assess the independent contribution of L2 proficiency (EIT scores), working memory (WM), and speaking anxiety to explaining inter-learner variability in L2 oral production (RQ4).

4 | RESULTS

The overall results (Table 3) indicate that, as expected, learners were more fluent and comprehensible in the simple than the complex task, whereas they were more accurate lexically and grammatically in the complex than the simple task, in accordance with the cognition hypothesis (Robinson, 2005). Pronunciation accuracy, however, was lower in the simple than the complex task, both in terms of pronunciation errors and perceived accentedness. The frequency and magnitude of anxiety peaks (HRsd) was larger in the simple than the complex task. Unexpectedly, the learners' ratings of task performance in the simple task for anxiety (*M* = 4.77, *SD* = 2.09), difficulty (*M* = 4.86, *SD* = 1.97), and mental effort (*M* = 4.95, *SD* = 1.98) were slightly higher than in the complex task (*M* = 4.58, *SD* = 2.06; *M* = 4.70, *SD* = 1.94; *M* = 4.81, *SD* = 2.03; respectively). A closer look at these data confirmed that this was due to the order in which participants had done the task. Learners who did the simple task first (S > C) found the complex task to generate greater anxiety and to be more difficult and effortful than the simple task (see Table 3).

TABLE 3 Means (M), standard errors (SE), and 95% confidence intervals (95% CI) of speech measures in the simple and the complex task.

	Simple			Complex		
	M	SE	95% CI	*M*	SE	95% CI
Self-ratings (SR)						
Anxiety (S > C)	4.70	0.46	3.8–5.7	5.30	0.46	4.4–6.2
Anxiety (C > S)	4.83	0.43	4.0–5.7	3.96	0.43	3.1–4.8
Difficulty (S > C)	4.40	0.42	3.6–5.3	5.45	0.42	4.6–6.3
Difficulty (C > S)	5.26	0.39	4.5–6.1	4.04	0.39	3.3–4.8
Mental Effort (S > C)	4.65	0.44	3.8–5.5	5.55	0.44	4.7–6.4
Mental Effort (C > S)	5.22	0.41	4.4–6.0	4.17	0.41	3.4–5.0
Anxiety						
HRm	88.24	1.83	84.6–91.9	86.06	1.83	82.4–89.7
HRsd	7.51	0.90	5.7–9.3	9.16	0.90	7.4–11.0
Fluency						
MSD	277.96	4.34	269–287	282.57	4.34	274–291
PD	902.04	65.07	772–1033	985.88	65.07	855–1116
SPD	732.56	52.81	627–838	785.92	52.81	680–892
PF	19.14	0.84	17.5–20.8	20.18	0.84	18.5–21.9
MCPF	12.68	0.78	11.1–14.2	13.34	0.78	11.7–14.9
MCSPF	9.59	0.68	8.2–11.0	10.59	0.68	9.2–12.0
REP	2.52	0.28	2.0–3.1	3.25	0.28	2.7–3.8
Accuracy						
Err	5.605	0.347	4.9–6.3	5.024	0.347	4.3–5.7
ErrGL	4.60	0.32	4.0–5.2	3.75	0.32	3.1–4.4
ErrP	1.00	0.18	0.6–1.4	1.28	0.18	0.9–1.6
Pronunciation						
Comprehensibility	6.68	0.23	6.2–7.1	6.46	0.23	6.0–6.9
Accentedness	4.42	0.26	3.9–4.9	4.76	0.26	4.2–5.3

Abbreviations: HRm, mean heart rate; HRsd, standard deviation of heart rate; MSD, mean syllable duration; PD, pause duration; SPD, silent pause duration; PF, pause frequency; MCPF, mid-clause pause frequency; MCSPF, mid-clause silent pause frequency; REP, repetitions and restarts; Err, error rate; ErrGL, grammatical and lexical error rate; ErrP, pronunciation error rate.

As higher proficiency learners might be better able to cope with task complexity and get less anxious than lower proficiency learners, we controlled for individual differences in L2 proficiency by including the EIT scores as a covariate in the linear mixed-effects models we report on below. In this way, we examined the effects of task complexity on L2 learners' speaking anxiety and oral performance while controlling for individual differences in L2 proficiency.

4.1 | Task complexity and speaking anxiety

Prior to answering RQ1, we investigated to what extent our speaking anxiety measures were related to one another. The mean heart rate score in beats per minute (HRm) and the HR standard deviation score (HRsd) were unrelated to one another ($r = -0.028$, $p = 0.800$), suggesting that they measured different dimensions of speaking anxiety: overall anxiety level during speaking performance (HRm) versus frequency and magnitude of anxiety peaks during speaking performance (HRsd). As it is possible for learners to obtain an average low anxiety level (HRm) while experiencing occasional anxiety peaks (e.g., caused by lexical retrieval difficulties) leading to high HRsd scores, HRm and HRsd can but are not necessarily expected to correlate with one another. L2 learners' self-perceived levels of speaking anxiety (SRA) were only weakly related to HRm ($r = 0.238$, $p = 0.027$) and unrelated to HRsd ($r = 0.054$, $p = 0.623$). This is likely due to HRm (but not HRsd) reflecting more closely L2 learners' perception of how anxious they felt during task performance and to such perceptions being affected by their perception of success in completing the speaking task well.

We then tested task complexity effects on speaking anxiety (RQ1) by submitting the HRm, HRsd, and SRA scores separately to a set linear mixed-effects models with *Task* (simple, complex) as a fixed factor, a random intercept for *Subject*, and proficiency as a covariate (see Appendix B for parameter estimates). For HRm, the slightly lower HR learners obtained in the complex (86 beats per minute) than the simple task (88 beats per minute), reached significance in the Type III tests of fixed effects ($F(1, 41) = 5.79$, $p = 0.021$). This unexpected result might be due to task order effects. In fact, participants who did the simple task first (but not those who did the complex task first) obtained a slightly higher mean HR in the simple ($M = 90.34$) than the complex task ($M = 86.19$). However, the HRsd scores were substantially higher in the complex ($M = 9.16$) than the simple task ($M = 7.51$), yielding a significant *Task* effect ($F(1, 41) = 7.41$, $p = 0.009$), which indicated that, as expected, changes in HR during task performance were of a significantly larger magnitude in the complex than the simple task ($\beta = -1.64$, $SE = 0.61$, $t = -2.72$). This effect was independent of task order effects as HRsd was higher in the complex than the simple task in both task orders (S > C: C = 8.11, S = 7.14; C > S: C = 9.91, S = 7.72). For SRA (and perceived difficulty and effortfulness), differences between task conditions did not reach significance ($F(1, 41) = 0.850$, $p = 0.632$; see Appendix B for parameter estimates). Taken together, these results suggest that task complexity did not affect learners' overall speaking anxiety, whether measured physiologically (HRm) or through self-reports (SRA). However, task complexity significantly affected HRsd, suggesting that the frequency and magnitude of changes in anxiety levels were higher during the performance of the complex than the simple task.

4.2 | Task complexity and oral production

Research question two (RQ2) enquired whether task complexity affected the dimensions of fluency, accuracy, comprehensibility, and accentedness in L2 speech production. Oral production measures were submitted to linear mixed-effects models with *Task* (simple, complex) as a fixed factor and a random intercept for *Subject* (see Appendix B for parameter estimates). As expected, task complexity was found to be detrimental to speaking fluency (see Table 2): all measures consistently denoted lower fluency (longer mean syllable duration [MSD] and pause durations [PD, SPD], and higher pause frequency [PF, MCPF, MCSPF] and disfluency ratios [REP]) in the complex than the simple task. Despite the consistency of this trend, task effects did not reach significance for PD ($F(1, 41) = 3.26$, $p = 0.078$) and MCSPF ($F(1, 41) = 3.66$, $p = 0.063$). However, the number of repetitions and corrections (REP) was significantly higher in the complex than the simple task ($F(1, 41) = 9.55$, $p = 0.003$).

Grammatical and lexical accuracy was, overall, higher in the complex than in the simple task (see Table 2), suggesting that the more complex task generated more attention to form than the simple task. Interestingly, error rates for grammar and lexis (ErrGL) differed from pronunciation error rates (ErrP) in that learners made significantly more grammatical and lexical errors in the simple than in the complex task ($F(1, 41) = 5.67$, $p = 0.022$), but more pronunciation errors in the complex task, though not to a significant extent ($F(1, 41) = 2.19$, $p = 0.146$). It is also worth noting that the ErrGL and ErrP measures were inversely related to one another ($r = -0.123$, $p = 0.257$).

We next assessed the effects of *Task* on the global speech dimensions of comprehensibility and accentedness. Comprehensibility was lower in the complex than the simple task, but the difference did not reach significance (*F*

(1, 41) = 2.00, $p = 0.165$). For accentedness, however, a significant main effect of *Task* (F (1, 41) = 8.09, $p = 0.007$) confirmed that learners obtained stronger accent ratings in the complex than the simple task. In fact, ErrP was positively and moderately related to the strength of accent native listeners perceived in the learners' speech ($r = 0.385$, $p < 0.001$).

4.3 | Speaking anxiety and oral production

Our third research question (RQ3) asked to what extent individual levels of anxiety during task performance were related to fluency, accuracy, comprehensibility, and accentedness. Correlational analyses between the two sets of scores (see Table 4) revealed weak to moderate associations between some of the measures of speaking anxiety and some measures of speaking performance. For example, moderate associations emerged between learners' SRA and breakdown fluency (PF, MCPF, MCSPF) and more strongly so in the complex than the simple task, whereas associations with repair fluency (REP) and lexico-grammatical accuracy were weaker, and for the latter, it only occurred in the complex task. This indicates that learners who had produced more repetitions, pauses, and grammatical and lexical errors in their speech had also perceived higher anxiety levels during task performance. However, associations involving physiological measures of anxiety were weak and only occurred between HRm and mid-clause pause frequency (MCPF) in the complex task, suggesting that learners who had produced more mid-clause pauses experienced higher levels of anxiety in the complex task, as measured in beats per minute. No other associations between anxiety and speech production measures reached significance[1].

4.4 | L2 proficiency, working memory, and speaking anxiety

As illustrated in Table 4, L2 proficiency was weakly to moderately associated with our dependent measures. More proficient learners reported having experienced lower anxiety levels after task performance; they produced shorter and less frequent pauses, and less repetitions and errors. As expected, they were also perceived to be more comprehensible and less strongly accented in both the simple and the complex task. WM was found to be weakly related to L2 proficiency, suggesting that learners with larger WM had achieved higher proficiency levels in English, and it was negatively associated with pause duration measures (in the simple task), suggesting that learners with stronger WM capacity needed less time to formulate clauses when speaking.

The relative contribution of L2 proficiency (EIT scores), working memory (WM), and speaking anxiety as individual learner factors (independent of task complexity effects on cognitive demands) in explaining variance in L2 oral production (RQ4) was examined through a series of hierarchical multiple regressions. These were run on the L2 speech dimensions that had been found to be related to L2 proficiency, WM, and L2 speaking anxiety (HRm and SRA) in the simple and the complex task (see Table 5), which were measures of breakdown fluency (PD, SPD, PF, MCPF, MCSPF).

In these regression models, PD, SPD, PF, MCPF, and MCSPF scores were used as dependent measures and the predictors of L2 proficiency and WM were entered separately at steps 1 and 2, respectively, whereas the speaking anxiety measures (SRA and HRm) where entered in step 3. In this way, we could evaluate the contribution WM makes to explaining variance in the breakdown fluency scores additionally to that explained by L2 proficiency, and we could also assess the amount of variance that could be uniquely attributed to speaking anxiety, with individual differences in L2 proficiency and working memory being controlled for.

The outcome of the hierarchical regression models (Table 5) indicates that L2 proficiency explains a significant amount of variance (between 10.1% and 18.6%) in all of the breakdown fluency measures, whereas WM only makes an additional significant contribution ($p = 0.026$) of 5.3% in the case of MCSPF. This suggests that L2 proficiency, which is also more strongly linked to pause frequency than to pause duration measures, is a much stronger predictor of L2 breakdown fluency than WM. Interestingly, once the effect of L2 proficiency and WM are controlled for, speaking

TABLE 4 *Spearman's* rank-order correlation coefficients between speaking anxiety and oral production measures in the simple (S) and complex (C) task (asterisks indicate significant correlations).

			EIT		WM		HRm		HRsd		SRA	
			r	p	r	p	r	p	r	p	r	p
	EIT	S	–	–	0.337*	0.002	−0.164	0.137	0.022	0.840	−0.225	0.039
	WM	C	–	–	–	–	0.104	0.340	−0.102	0.352	−0.100	0.358
Fluency	MSD	S	−0.116	0.463	0.080	0.608	0.048	0.761	0.066	0.673	0.293	0.056
		C	−0.258	0.099	0.070	0.654	0.225	0.147	0.005	0.972	0.283	0.066
	PD	S	−0.439*	0.004	−0.538*	<0.001	−0.102	0.514	0.043	0.786	0.240	0.122
		C	−0.083	0.601	−0.199	0.200	−0.033	0.836	−0.075	0.631	−0.041	0.794
	SPD	S	−0.162	0.304	−0.359*	0.018	−0.281	0.068	0.043	0.783	−0.016	0.917
		C	−0.090	0.570	−0.255	0.099	−0.122	0.434	0.037	0.812	−0.008	0.960
	PF	S	−0.240	0.126	−0.252	0.102	0.123	0.432	0.185	0.234	0.341*	0.025
		C	−0.462*	0.002	−0.185	0.234	0.259	0.093	−0.012	0.940	0.440*	0.003
	MCPF	S	−0.376*	0.014	−0.176	0.259	0.265	0.086	0.153	0.329	0.271	0.079
		C	−0.469*	0.002	−0.233	0.133	0.312*	0.042	−0.016	0.920	0.548*	<0.001
	MCSPF	S	−0.207	0.189	−0.257	0.096	0.134	0.392	0.278	0.071	0.025	0.875
		C	−0.294	0.059	−0.276	0.073	0.154	0.324	0.111	0.479	0.480*	0.001
	REP	S	−0.324*	0.037	0.092	0.559	0.036	0.817	−0.064	0.682	0.343*	0.025
		C	−0.460*	0.002	−0.094	0.548	0.028	0.861	−0.137	0.381	0.107	0.496
Accuracy	Err	S	−0.348*	0.024	−0.135	0.387	0.190	0.223	−0.232	0.134	−0.011	0.943
		C	−0.15	0.342	−0.065	0.678	0.012	0.939	0.079	0.613	0.347*	0.023
	ErrGL	S	−0.165	0.296	−0.145	0.352	0.261	0.091	−0.110	0.481	0.138	0.379
		C	−0.032	0.841	0.062	0.695	−0.009	0.954	0.166	0.288	0.355*	0.020
	ErrP	S	−0.368*	0.017	−0.112	0.476	−0.128	0.415	−0.336*	0.028	−0.261	0.091
		C	−0.067	0.673	−0.088	0.575	0.004	0.979	0.023	0.883	−0.033	0.832
Pronunciation	Comp	S	0.578*	<0.001	0.069	0.660	−0.270	0.080	−0.053	0.738	−0.254	0.100
		C	0.536*	<0.001	0.213	0.171	0.000	0.999	0.143	0.360	−0.194	0.213
	Acc	S	−0.585*	<0.001	−0.024	0.879	0.094	0.549	−0.064	0.682	0.080	0.608
		C	−0.474*	0.002	−0.137	0.380	−0.116	0.460	0.104	0.508	0.077	0.622

Note: Err, error rate; ErrGL, grammatical and lexical error rate; ErrP, pronunciation error rate; Comp, comprehensibility; Acc, accentedness; EIT, elicited imitation task; WM, working memory; MSD, mean syllable duration; PD, pause duration; SPD, silent pause duration; PF, pause frequency; MCPF, mid-clause pause frequency; MCSPF, mid-clause silent pause frequency; REP, repetitions and restarts.

anxiety explains an additional significant amount of variance in pause frequency measures (PF = 13.1%; MCPF = 15.8%; MCSPF = 6%), but not in pause duration measures, suggesting that speaking anxiety is reflected in the speech of L2 learners mainly in the amount of pauses they make, rather than in how long they are. In the break-down fluency measures where speaking anxiety uniquely explained a significant amount of variance (PF and MCPF), SRA appeared to contribute more strongly and significantly (PF: 8.2%, $\beta = 0.299$, $t = 2.97$, $p = 0.004$; MCPF: 8.1%, $\beta = 0.296$, $t = 3.14$, $p = 0.002$) than HRm did (PF: 2.3%, $\beta = 0.159$, $t = 1.57$, $p = 0.119$; MCPF: 4.4%, $\beta = 0.218$, $t = 2.31$, $p = 0.024$), suggesting that SRA was a more sensitive measure than HRm in capturing individual variability in pause frequency.

TABLE 5 Hierarchical multiple regression models on breakdown fluency measures with L2 proficiency (EIT), working memory (WM), and L2 speaking anxiety (SRA, HRm) as predictor variables (asterisks indicate significance).

Variable	Predictors	R	R^2	ΔR^2	ΔF	p
PD	EIT	0.364	0.133	0.133	12.527	0.001*
	WM	0.409	0.167	0.035	3.373	0.070
	SRA, HRm	0.413	0.170	0.003	0.142	0.868
SPD	EIT	0.318	0.101	0.101	9.194	0.003*
	WM	0.356	0.127	0.026	2.388	0.126
	SRA, HRm	0.385	0.148	0.022	0.999	0.373
PF	EIT	0.338	0.114	0.114	10.594	0.002*
	WM	0.362	0.131	0.017	1.580	0.212
	SRA, HRm	0.512	0.262	0.131	7.018	0.002*
MCPF	EIT	0.431	0.186	0.186	18.688	<0.001*
	WM	0.440	0.194	0.008	0.798	0.374
	SRA, HRm	0.593	0.352	0.158	9.658	<0.001*
MCSPF	EIT	0.327	0.107	0.107	9.794	0.002
	WM	0.400	0.160	0.053	5.121	0.026
	SRA, HRm	0.469	0.220	0.060	3.031	0.054

Abbreviations: EIT, elicited imitation task; WM, working memory; SRA, self-rated anxiety; HRm, mean heart rate; PD, pause duration; SPD, silent pause duration; PF, pause frequency; MCPF, mid-clause pause frequency; MCSPF, mid-clause silent pause frequency.

5 | DISCUSSION

The current study investigated whether task complexity affects speaking anxiety levels in L2 learners as they performed a decision-making oral narrative task and explored the effects of task complexity and speaking anxiety on the L2 speech production dimensions of speaking fluency, accuracy, accentedness, and comprehensibility. In addition, we assessed the independent contribution of individual differences in L2 proficiency, working memory and speaking anxiety to the quality of L2 learners' oral production.

5.1 | Task complexity effects on L2 speaking anxiety

When examining task complexity effects on speaking anxiety (RQ1), we found the complex task to yield significantly higher levels of anxiety variability (HRsd) during learners' oral performance than the simple task. The higher range of anxiety levels within learners' speaking performance in the complex task indicated a higher frequency of arousal peaks departing from the learners' overall HR mean. We interpret this effect as the emotional consequence of the more frequent formulation problems and the higher error rates the complex task generated compared to the simple task, as it imposed higher reasoning demands on the learners' conceptualization process. This finding extends, for a physiological measure of anxiety (HRsd), the results Brennan (2016) had obtained through a perceived anxiety scale using the same oral narrative task and a comparable sample population, and it is also consistent with Robinson's (2007) finding that higher output anxiety participants could not benefit from increased task complexity to produce more complex speech. However, unlike HRsd, HRm was not significantly affected by task complexity manipulations. This could be due to HRsd being more sensitive in capturing anxiety-generating formulation issues in speech production as well as the HRm measure being more sensitive to task order effects (anxiety levels were higher on the version of the task (either simple or complex) done first, whereas this was not the case for the HRsd measure. Contrary to our expectations, task complexity was not found to affect learners' perception of anxiety during task performance (SLA). This is likely due to our post-task performance questionnaire only including a single question (in the form of a 9-point Likert type scale) about anxiety, rather than a set of open-ended questions covering both learners' positive and negative emotions related to their performance as well as a report of which linguistic factors were found to induce anxiety during their speaking performance (as in Donate, 2022). However, our results are consistent with those of Donate (2022), the only study we know of that directly manipulated task complexity and assessed its effects on anxiety during speaking performance, in that the cognitive demands of the task did not affect task anxiety, which she attributed to individual, contextual, task-related factors potentially interfering with task complexity effects.

5.2 | Task complexity effects on L2 speech production

Task complexity was found to affect L2 speaking fluency and accuracy (RQ2) in various ways. First, a tendency was found for L2 speech to be slower and to contain more and longer pauses and repetitions in the complex than the simple task, a detrimental effect of task complexity on speaking fluency that only reached significance for number of repetitions and repairs (REP). We were expecting a larger effect of task complexity on speaking fluency, in accordance with previous research within the framework of Robinson's (2005, 2011) Cognition Hypothesis. However, the differences in task complexity between the simple and the complex version of the task we used may not have posed a real challenge for the relatively advanced L2 learners in our study.

In accordance with the notion that higher complexity leads to greater attention to form (Robinson, 2005, 2011), learners did produce significantly less lexico-grammatical errors in the complex than the simple task. Importantly, this positive effect of task complexity on speaking accuracy was not found for pronunciation, which followed the opposite pattern, with pronunciation errors being more frequent (though not to a significant extent) in the complex than the simple task. Thus, it seems that increasing task complexity in a pronunciation-unfocused speaking task might benefit accuracy at the lexical and grammatical level at the expense of pronunciation. In line with Crowther at al. (2018), L2 learners' speech from the complex task was significantly more strongly accented than the speech from the simple task, suggesting that increased task complexity was detrimental to pronunciation accuracy by directing learners' attentional resources to lexical and grammatical form. This finding contrasts with the findings from previous research using pronunciation-focused tasks where pronunciation features were essential to task completion (Mora & Levkina, 2018, Mora-Plaza et al., 2018; Solon et al., 2017) and increased task complexity led to more accurate segmental pronunciation resulting from increased attention to phonetic form. Thus, it is difficult for learners to allocate attention to pronunciation when the nature of the speaking task requires them to divide their attentional resources amongst various linguistic dimensions, including lexis and grammar (Derwing et al., 1998). Therefore, in the absence of a specific focus on phonetic form induced by task design manipulations, increased task complexity may negatively affect pronunciation accuracy, especially as the enhanced focus on form generated by the higher linguistic demands of a more complex task may result in more accurate oral production at the lexical and grammatical levels (e.g., Housen et al., 2012).

5.3 | Relationship between speaking anxiety and L2 speech production

As regards the relationship between speaking anxiety and L2 speech (RQ3), our findings show that learners' perception of anxiety (SRA) during their speaking performance significantly aligned with breakdown fluency and lexico-grammatical accuracy and more strongly so in the complex than the simple task. However, physiological measures of anxiety (HRm and HRsd) were generally disassociated with speaking fluency, accuracy, comprehensibility, or accentedness. We found HRm to be significantly related to mid-clause pause frequency in the complex task, suggesting that speaking anxiety might have interfered with the efficient operation of the formulator during lexical and grammatical encoding (Kormos, 2015). This would support recent findings by Pérez Castillejo (2019), who found foreign language anxiety (as measured through the FLCAS questionnaire; Horwitz et al., 1986) to predict, among other utterance fluency measures, pause frequency, and duration. However, this deserves further research, given the weak strength of this relationship ($r = 0.312$). Therefore, we could only partly replicate through HR measures the effects of task complexity Brennan (2016) had found with the same tasks using a questionnaire-based anxiety scale. Although our results are in accordance with previous findings (e.g., Kim & Tracy-Ventura, 2011), the difference in cognitive complexity operationalized through resource-directing variables between the simple and the complex version of the tasks (Robinson & Gilabert, 2007) might not have led to sufficiently challenging conceptual demands to trigger distinct levels of anxiety detectable through HR measures.

HR and SRA measures were also unrelated to comprehensibility and accentedness, suggesting that the extent to which L2 learners' levels of anxiety might have been reflected in their speech (e.g., through breakdown fluency) went undetected by native listeners when assessing L2 speech globally. Thus, our data do not show the detrimental effects of anxiety on comprehensibility other studies have reported using questionnaire scores (Saito et al., 2018) or judgements of self-perceived anxiety based on a continuous scale (Nagle et al., 2022), which might be due to the relatively limited variability in the comprehensibility and accentedness scores we obtained.

5.4 | The contribution of L2 proficiency, working memory, and anxiety to L2 speech production

Finally, higher proficiency learners were found to speak more fluently and accurately and to be more comprehensible and less foreign accented than lower proficiency learners. Those with stronger WM skills produced shorter and less frequent pauses. Such relationships appeared to be stronger in the complex than the simple task, but not to significant levels. L2 proficiency was found a much stronger predictor of L2 breakdown fluency than WM. Still, when controlling for L2 proficiency and WM, speaking anxiety was found to explain a significant 13%–15% of additional amount of variance in pause frequency, indicating that independently of individual differences in L2 proficiency and working memory, higher levels of anxiety were mainly reflected in pause frequency scores (RQ4). This suggests that it is the number of pauses learners make when speaking (rather than their duration) that they associate with the levels of anxiety they perceive when speaking.

6 | CONCLUSIONS

The aim of this study was to explore the effects of foreign language anxiety and increasing task complexity on L2 speech production, and the mediating role of proficiency and working memory. First, findings suggest that increasing reasoning demands also increases the perceived levels of anxiety during the performance of a monologic speaking task. Second, this research also offers further evidence of the detrimental effects of task complexity on L2 speaking fluency (Baralt et al., 2014) and pronunciation (Crowther et al., 2018). Whereas grammar and lexis appear to benefit from enhanced attention to form resulting from increasing cognitive complexity and the concomitant prioritization of meaning, pronunciation does not, suggesting a trade-off between lexico-grammatical pronunciation accuracy that deserves further investigation. Third, learners' perception of their own anxiety level (but not physiological measures) significantly correlated with breakdown fluency and accuracy. Last, higher proficiency learners were more fluent, accurate, comprehensible, and perceived as less foreign-accented than lower proficiency learners, but WM explained little variability in L2 speech production. Interestingly, self-perceived anxiety explained a great amount of inter-individual variance in pause frequency additionally to that explained by L2 proficiency and working memory.

Further research should investigate the extent to which the occurrence of anxiety peaks and the magnitude of these peaks align with speaking disfluencies or pronunciation and lexico-grammar errors. It would be informative to temporally align the HR plot and the recording of the speaking task, as this might allow for the establishment of meaningful links between specific speech features (mistakes and disfluencies) and heart-rate peaks indicative of anxiety. In addition, the use of the idiodynamic method (Gregersen et al., 2014; MacIntyre, 2012) may prove useful in obtaining further information on learners' self-perception of their speaking anxiety. More longitudinal studies are needed to determine the extent to which task complexity interacts with L2 learners' anxiety levels. Pedagogical implications of the current study include the need to consider the relationship between task complexity and speaking anxiety in the design of pedagogical tasks and the need to develop tasks for training speaking and reducing L2-learners' speaking anxiety. At present, it is uncertain whether increased cognitive task complexity may generate anxiety to a level that would offset its benefits for pronunciation. Individual differences in learners' capacity to cope with attention-draining emotions such as anxiety may be key in determining the linguistic benefits that increasing task cognitive complexity provides. We hope the current study encourages further research on FLA and L2 oral performance, whose findings would doubtlessly be useful for teachers and task designers alike.

AUTHOR CONTRIBUTIONS

The corresponding author Joan C. Mora and co-author Ingrid Mora-Plaza are responsible for the conceptualization and design of the study and participant recruitment. Joan C. Mora and Ingrid Mora-Plaza implemented the statistical analysis of the data. All authors participated in the data collection and acoustic analysis of the oral data, and in the writing up of the manuscript.

ACKNOWLEDGMENTS

We would like to thank the participants in the study for their availability and Dr. Roger Gilabert for kindly sharing the oral tasks with us. This research was supported by grant PID2019-107814GB-I00 from the Spanish Ministry of Science, Innovation and Universities.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

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

ORCID

Joan C. Mora https://orcid.org/0000-0001-6071-2008
Ingrid Mora-Plaza https://orcid.org/0000-0003-1625-7153

TRANSPARENT PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ijal.12494

ENDNOTE

[1] Following De Jong et al. (2015a), Segalowitz (2016), and Kahng (2020), we ran the correlations described above using L2-specific residualized scores computed by partialing out L1 measures (obtained from an oral map task) from L2 measures for anxiety and speaking fluency through linear regressions. The association between self-perceived anxiety and MSD became slightly stronger ($r = 0.303$, $p = 0.005$), and it became slightly weaker for PF ($r = 0.300$, $p = 0.005$). No other significant associations emerged out of these analyses, probably due to the measures selected already being highly L2-specific and unlikely to be affected to a large extent by individual speaking styles.

REFERENCES

Baralt, M., Gilabert, R., & Robinson, P. (Eds.). (2014). *Task sequencing and instructed second language learning*. Bloomsbury. https://doi.org/10.5040/9781472593665

Baran-Łucarz, M. (2011). The relationship between language anxiety and the actual and perceived levels of foreign language pronunciation. *Studies in Second Language Learning and Teaching*, *1*(4), 491–514. https://doi.org/10.14746/SSLLT.2011.1.4.3

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. *Canadian Modern Language Review*, *70*(4), 445–473. https://doi.org/10.3138/cmlr.2666

Baran-Łucarz, M. (2022). Language Anxiety. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 83–96). Routledge. https://doi.org/10.4324/9781003022497-8

Beatty, M. J., & Behnke, R. R. (1991). Effects of public speaking trait anxiety and intensity of speaking task on heart rate during performance. *Human Communication Research*, *18*(2), 147–176. https://doi.org/10.1111/j.1468-2958.1991.tb00542.x

Bodie, G. D. (2010). A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety. *Communication Education*, *59*(1), 70–105. https://doi.org/10.1080/03634520903443849

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program]. Version 6.1.09. https://www.praat.org

Brennan, K. A. (2016). *Second language anxiety and task complexity* [Doctoral dissertation, Universitat de Barcelona].

Cheng, Y. S. (2004). A measure of second language writing anxiety. Scale development and preliminary validation. *Journal of Second Language Writing*, *13*(4), 313–335. https://doi.org/10.1016/j.jslw.2004.07.001

Croft, R. J., Gonsalvez, C. J., Gander, J., Lechem, L., & Barry, R. (2004). Differential relations between heart rate and skin conductance, and public speaking anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *35*(3), 259–271. https://doi.org/10.1016/j.jbtep.2004.04.012

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, *40*(2), 443–457. https://doi.org/10.1017/S027226311700016X

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS) (pp. 17–20). The 6th Workshop of Disfluency in Spontaneous Speech conducted at the Royal Institute of Technology, Stockholm, Sweden.

De Jong, N., Groenhout, R., Schoonen, R., & Hulstijn, J. (2015a). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243. https://doi.org/10.1017/s0142716413000210

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2015b). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *5*, 893–916. https://doi.org/10.1017/S0142716412000069

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

Donate, Á. (2022). Task anxiety, cognition and performance on oral tasks in L2 Spanish. *Journal of Spanish Language Teaching*, *9*(1), 1–18. https://doi.org/10.1080/23247797.2022.2090661

Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching motivation*. Routledge. https://doi.org/10.4324/9781315833750

Draine, S. (1999). *Inquisit*. Version 5.0.14.0 [Computer Program]. Millisecond Software. https://www.millisecond.com/

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, *89*(2), 206–220. https://doi.org/10.4324/9781315833750

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2014). Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task. *Cyberpsychology, Behavior, and Social Networking*, *17*(5), 310–316. https://doi.org/10.1089/cyber.2013.0472

Gallego, A., McHugh, L., Penttonen, M., & Lappalainen, R. (2022). Measuring public speaking anxiety: Self-report, behavioral, and physiological. *Behavior Modification*, *46*(4), 782–798. https://doi.org/10.1177/0145445521994308

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 215–240. https://doi.org/10.1515/iral.2007.010

Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 105–140). John Benjamins. https://doi.org/10.1075/tblt.2.10ch5

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, *47*(3-4), 367–395. https://doi.org/10.1515/iral.2009.016

Gkonou, C. (2014). The sociolinguistic parameters of L2 speaking anxiety In M. Pawlak, J. Bielak, & J., & A. Mystkowska-Wiertelak (Eds.), *Classroom-oriented research. Second language learning and teaching* (pp. 15–32). Springer. https://doi.org/10.1007/978-3-319-00188-3_2

Gordon, J. (2021). Pronunciation and task-based instruction: Effects of a classroom intervention. *RELC Journal*, *52*(1), 94–109. https://doi.org/10.1177/0033688220986919

Gregersen, T., MacIntyre, P. D., & Meza, M. D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *The Modern Language Journal*, *98*(2), 574–588. https://doi.org/10.1111/modl.12084

Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *The Modern Language Journal*, *96*(2), 170–189. https://doi.org/10.1111/j.1540-4781.2011.01174.x

Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31–47). Multilingual Matters. https://doi.org/10.21832/9781783097722-004

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins. https://doi.org/10.1075/lllt.32

Jiang, Y., & Dewaele, J. M. (2020). The predictive power of sociobiographical and language variables on foreign language anxiety of Chinese university students. *System*, *89*, 102207. https://doi.org/10.1016/j.system.2020.102207

Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, *41*, 457–480. https://doi.org/10.1017/S0142716420000065

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, *99*, 656–677. https://doi.org/10.1111/modl.12273

Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 287–306). John Benjamins. https://doi.org/10.1075/tblt.2.18ch11

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition*, *37*(3), 549–581. https://doi.org/10.1017/s0272263114000618

Kim, Y., Tracy-Ventura, N., & Jung, Y. (2016). A measure of proficiency or short-term memory? Validation of an elicited imitation test for SLA research. *The Modern Language Journal*, *100*(3), 655–673. https://doi.org/10.1111/modl.12346

Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum.

Kormos, J. (2015). Individual differences in second language speech production. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 369–388). Cambridge University Press. https://doi.org/10.1017/CBO9781107447257.017

Kormos, J., & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift fur Interkulturellen Fremdsprachenunterricht*, *9*(2), 1–19.

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 267–289). John Benjamins.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Linck, J., Osthus, P., Koeth, J., & Bunting, M. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lindberg, R., McDonough, K., & Trofimovich, P. (2021). Investigating verbal and nonverbal indicators of physiological response during second language interaction. *Applied Psycholinguistics*, *42*(6), 1403–1425. https://doi.org/10.1017/S014271642100028X

Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System*, *34*(3), 301–316. https://doi.org/10.1016/j.system.2006.04.004

MacIntyre, P. D. (2007). Willingness to communicate in the second language. Understanding the decision to speak as a volitional process. *The Modern Language Journal*, *91*(4), 564–576. https://doi.org/10.1111/j.1540-4781.2007.00623.x

MacIntyre, P. D. (2012). The idiodynamic method: A closer look at the dynamics of communication traits. *Communication Research Reports*, *29*(4), 361–367. https://doi.org/10.1080/08824096.2012.723274

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). Multilingual Matters. https://doi.org/10.21832/9781783097722-003

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning*, *39*(2), 251–275. https://doi.org/10.1111/j.1467-1770.1989.tb00423.x

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

MacIntyre, V. A., MacIntyre, P. D., & Carre, G. (2010). Heart rate variability as a predictor of speaking anxiety. *Communication Research Reports*, *27*(4), 286–297. https://doi.org/10.1080/08824096.2010.496323

McManus, K., & Liu, Y. (2020). Using elicited imitation to measure global oral proficiency in SLA research: A close replication study. *Language Teaching*, *55*(1), 116–135. https://doi.org/10.1017/s026144482000021x

Meara, P. M., & Milton, J. (2003). *X-lex: The Swansea levels test*. Express Publishing.

Meara, P. M., & Miralpeix, I. (2006). Y_Lex: The Swansea advanced vocabulary levels test. v2. 05. *Swansea: Lognostics*.

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Mora, J. C., & Levkina, M. (2018). Training vowel perception through map tasks: The role of linguistic and cognitive complexity. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380–9566, University of Utah, September 2017 (151–162). Iowa State University, Ames, IA.

Mora-Plaza, I., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380–9566, University of Utah, September 2017 (151–162). Iowa State University, Ames, IA.

Nagle, C., Trofimovich, P., O'Brien, M., & Kennedy, S. (2022). Beyond linguistic features: Exploring behavioural and affective correlates of comprehensible second language speech. *Studies in Second Language Acquisition*, *44*(1), 255–270. https://doi.org/10.1017/S0272263121000073

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). *An investigation of elicited imitation tasks in crosslinguistic SLA research*. Article presented at the Second Language Research Forum, Toronto, Canada.

Pérez Castillejo, S. (2019). The role of foreign language anxiety on L2 utterance fluency during a final exam. *Language Testing*, *36*(3), 327–345. https://doi.org/10.1177/0265532218777783

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, *76*(1), 14–26. https://doi.org/10.1111/j.1540-4781.1992.tb02573.x

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*(1), 162–181. https://doi.org/10.1111/j.1540-4781.2011.01241.x

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*, *38*(4), 703–737. https://doi.org/10.1017/S0272263115000339

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Multilingual Matters.

Robinson, P. (Ed.) (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins. https://doi.org/10.1075/tblt.2

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 161–176. https://doi.org/10.1515/iral.2007.007

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

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, Y., & Samimy, K. K. (1996). Foreign language anxiety and language performance: A study of learner anxiety in beginning, intermediate, and advanced-level college students of Japanese. *Foreign Language Annals*, *29*(2), 239–249. https://doi.org/10.1111/j.1944-9720.1996.tb02330.x

Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal*, *83*(2), 202–218. https://doi.org/10.1111/0026-7902.00016

Segalowitz, N. (2010). *Cognitive bases of second language fluency* (1st ed.). Routledge. https://doi.org/10.4324/9780203851357

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 79–95. https://doi.org/10.1515/iral-2016-9991

Sevinç, Y. (2018). Language anxiety in the immigrant context: Sweaty palms? *International Journal of Bilingualism*, *22*(6), 717–739. https://doi.org/10.1177/1367006917690914

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

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14. https://doi.org/10.1017/S026144480200188X

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, *39*(2), 347–380. https://doi.org/10.1017/S0272263116000425

Schwerdtfeger, A. (2004). Predicting autonomic reactivity to public speaking: Don't get fixed on self-report data!. *International Journal of Psychophysiology*, *52*(3), 217–224. https://doi.org/10.1016/j.ijpsycho.2003.10.008

Szyszka, M. (2017). *Pronunciation learning strategies and language anxiety: In search of an interplay*. Springer. https://doi.org/10.1007/978-3-319-50642-5

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second Language* (pp. 239–277). John Benjamins. https://doi.org/10.1075/lllt.11.15tav

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

Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning*, *67*(2), 394–430. https://doi.org/10.1111/lang.1222

Witt, P. L., Brown, K. C., Roberts, J. B., Weisel, J., Sawyer, C. R., & Behnke, R. R. (2006). Somatic anxiety *patterns* before, during, and after giving a public speech. *Southern Communication Journal*, *71*, 87–100. https://doi.org/10.1080/10417940500503555

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal*, *37*(3), 308–328. https://doi.org/10.1177/0033688206071315

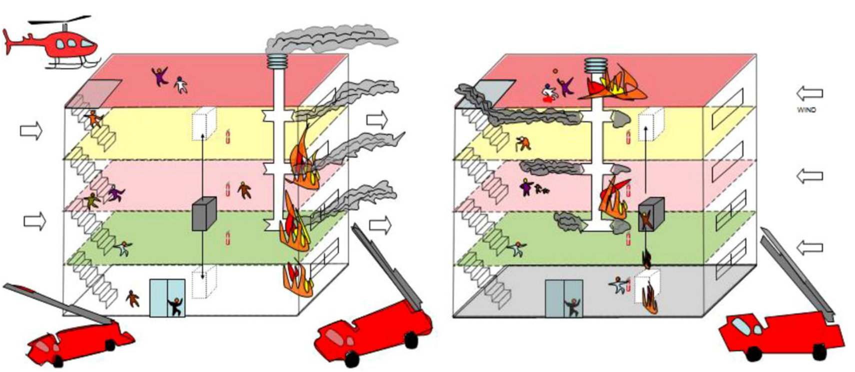
Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*(4), 497–528. https://doi.org/10.1177/0265532215594643

Young, D. J. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals*, *23*(6), 539–553. https://doi.org/10.1111/j.1944-9720.1990.tb00424.x

Zuniga, M., & Simard, D. (2022). Exploring the intricate relationship between foreign language anxiety, attention and self-repairs during L2 speech production. *System*, *105*, 102732. https://doi.org/10.1016/j.system.2022.102732

How to cite this article: Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2024). Speaking anxiety and task complexity effects on second language speech. *International Journal of Applied Linguistics*, *34*, 292–315. https://doi.org/10.1111/ijal.12494

APPENDIX A: THE SIMPLE (LEFT) AND COMPLEX (RIGHT) VERSIONS OF THE FIRE CHIEF TASK (Gilabert et al., 2009)



APPENDIX B: PARAMETER ESTIMATES OF LINEAR MIXED-EFFECTS MODELS (SHADED CELLS INDICATE A SIGNIFICANT TASK EFFECT)

							95% CI	
		β	SE	*df*	*t*	*p*	Lower	Upper
Anxiety								
HRm	Intercept	86.06	1.83	46.30	47.03	<0.001	82.38	89.75
	Task	2.18	0.90	41.00	2.41	0.021	0.35	4.00
HRsd	Intercept	9.16	0.90	51.32	10.18	<0.001	7.35	10.97
	Task	−1.65	0.61	41.00	−2.72	0.009	−2.87	−0.43
SRA	Intercept	4.60	0.32	52.96	14.24	<0.001	3.95	5.24
	Task	0.21	0.23	41.00	0.92	0.362	−0.26	0.68
Fluency								
MSD	Intercept	282.57	4.34	56.33	65.06	<0.001	273.87	291.27
	Task	−4.61	3.50	41.00	−1.32	0.195	−11.68	2.46
PD	Intercept	985.88	65.07	52.72	15.15	<0.001	855.36	1116.41
	Task	−83.84	46.45	41.00	−1.81	0.078	−177.65	9.98
SPD	Intercept	785.92	52.81	56.47	14.88	<0.001	680.14	891.70
	Task	−53.36	42.75	41.00	−1.25	0.219	−139.71	32.98
PF	Intercept	20.18	0.84	57.03	23.98	<0.001	18.50	21.87
	Task	−1.05	0.69	41.00	−1.51	0.138	−2.44	0.35
MCPF	Intercept	13.34	0.78	50.89	17.13	<0.001	11.77	14.90
	Task	−0.66	0.51	41.00	−1.28	0.209	−1.69	0.38
MCSPF	Intercept	10.58	0.68	54.65	15.53	<0.001	9.22	11.95
	Task	−1.00	0.52	41.00	−1.91	0.063	−2.05	0.06
REP	Intercept	3.25	0.28	57.25	11.57	<0.001	2.69	3.81
	Task	−0.73	0.23	41.00	−3.16	0.003	−1.20	−0.26
Accuracy								
Err	Intercept	5.02	0.35	74.06	14.48	<0.001	4.33	5.72
	Task	0.58	0.40	41.00	1.44	0.156	−0.23	1.39
ErrGL	Intercept	3.75	0.32	71.66	11.64	<0.001	3.11	4.39
	Task	0.85	0.36	41.00	2.38	0.022	0.13	1.58
ErrP	Intercept	1.27	0.18	67.62	7.17	<0.001	0.92	1.63
	Task	−0.27	0.18	41.00	−1.48	0.146	−0.65	0.10
Pronunciation								
Comprehensibility	Intercept	6.46	0.23	51.97	28.63	<0.001	6.01	6.91
	Task	0.22	0.16	41.00	1.42	0.165	−0.09	0.54
Accentedness	Intercept	4.76	0.26	45.52	18.41	<0.001	4.24	5.28
	Task	−0.34	0.12	41.00	−2.85	0.007	−0.58	−0.10