Do examiners and test-takers imitate each other?

Dialogic resonance in second language testing

Abstract

Do Test-takers imitate Examiners' language use? What about the other way around? This paper is centred on the impact of verbal imitation in spoken L2 English language tests. We assessed this by measuring Test-takers' and Examiners' degrees of dialogic resonance (Du Bois, 2014; Tantucci, 2023), a key mechanism for learning and engagement. Resonance involves speakers' ability to re-use words and expressions uttered by their interlocutors during an interaction. It is often creative and can be reliably measured as a continuous variable on a large scale (Tantucci & Wang, 2021, 2022a, 2024; Tantucci & Lepadat, 2024). We retrieved 2,564 turns from the Spoken Dialogues of the *International Corpus Network of Asian Learners of English* (ICNALE). We fitted a multifactorial mixed-effects linear regression of resonance between Examiners and Test-takers and found that verbal imitation plays different roles in language testing. First, resonance values are persistently high both in Test-takers and Examiners. Second, learners' imitation is key in L2 pragmatic competence and proficiency: the more proficient learners are, the higher the resonance with their interlocutors. Most decisively, Examiners' resonance improves Test-takers' performance: the more an Examiner resonates with a Test-taker, the longer the Test-takers' utterance in the following turn. We discuss implications for second language learning and language testing, and practical applications for Examiners' training and language teaching materials development.

Keywords: Imitation, resonance, engagement, language learning, language testing

Introduction

This study is centred on the role played by verbal imitation in second language (L2) learning and language testing. It fills an important gap in the literature on L2 pragmatics, language learning and language testing: the lack of a systematic index that accounts for how Examiners and Test-takers influence each other's speech in language testing. We focus on dialogic resonance (Du Bois, 2014; Tantucci, 2023), the way speakers re-use one another's utterances, often creatively, to express something new. We are interested not only in whether Test-takers resonate with Examiners across different tasks and topics, but also in whether Examiners verbally imitate what is said by Test-takers. We analysed and annotated 2,564 turns from the Spoken Dialogues of the *International Corpus*

Network of Asian Learners of English (ICNALE). We fitted a mixed-effects linear regression of resonance between Examiners and Test-takers and controlled for illocutionary force, learners' language background, their L2 proficiency, task type, task topic, gender, and other socio-demographic variables. We found that verbal imitation plays different roles in language testing. This study produced three key findings informing research in second language learning and language testing:

1. Test-takers and Examiners resonate with one another during L2 spoken tests. Test-takers do it to engage and learn from what is said by Examiners. The latter do it to engage with what Test-takers have said but also to motivate them to continue the ongoing interaction (e.g., recast).
2. Resonance increases with L2 proficiency levels. The more proficient the learners, the higher the degree of resonance with what the Examiners say.
3. Examiners' resonance boosts Test-takers' performance. The more Examiners resonate with learners' speech, the higher the Test-takers' word count in subsequent turns.

Our paper is organised as follows: In section 2, we discuss verbal imitation and resonance. Section 3 introduces the literature on pragmatic competence and engagement in second language learning and testing. In section 4, we illustrate the data retrieval (4.1) and the annotation of all utterances (4.2). Section 5 is devoted to our analysis and the statistical results addressing our research questions. Section 6 discusses our findings in relation to second language learning and testing. Our conclusions are given in section 7.

Imitation, resonance and engagement

We constantly imitate one another in interaction (Tantucci, 2023; Chen et al., in press). We use similar gestures (Galati et al., 2019; Dresang et al., 2023), our prosodic patterns and accents align (Ladd et al., 2009), we often mirror our interlocutors' facial expressions (Dimberg et al., 2000) and respond to their gaze (Kendrick & Holler, 2017). Not only that, speakers of all cultures, ages, and language proficiency levels constantly re-use and re-combine words and expressions they just heard in conversation (Tantucci, forthcoming; Tantucci & Wang, 2021).

There are several reasons why people imitate the communicative behaviours of their interlocutors. Learning is one of them (Tomasello et al., 1993; Speidel & Nelson, 1989). In both first language acquisition (FLA) and second language acquisition (SLA), speakers imitate others to acquire new words, expressions, and establish new categories of form and meaning (Tantucci & Wang, 2022b). Speakers also imitate others' speech to conform socially (Cialdini & Goldstein, 2004): Communicating like others enables individuals to integrate into cultures and communities of practice,

enhance the appropriateness of their communicative behaviours, and adhere to social norms. Finally, and perhaps most intriguingly, speakers imitate each other's language as a form of engagement (Tantucci & Wang, 2024). This kind of alignment in conversation is called **resonance** (Du Bois, 2014; Tantucci, 2023). It is not limited to cognitive coordination (cf. Pickering & Garrod, 2021), as it – more specifically – involves speakers' dialogic ability and overt attempts to re-use one another's utterances.

Resonance is central in Dialogic syntax (Du Bois, 2014), a usage-based framework centred on the emergence of constructions as a joint effort of (at least) two speakers engaging in interaction. A key component of resonance is that it is often creative. It goes beyond merely repeating what others say (cf. Sidnell, 2010). It often occurs when the form and meaning others produce in conversation are innovatively recombined to express something new. When speakers resonate with one another, they demonstrate that what their interlocutors said is relevant for continuing an interaction. When this happens, they somewhat 'cite' others' speech in conversation and make them aware that what they said is worth re-using for the progressivity of the current exchange (Tantucci & Lepadat, 2024). When speakers consistently resonate with each other in dialogue, they show their involvement with their interlocutors' language. Conversely, when resonance levels are persistently low, that is a sign of interactional detachment, which is distinctive of autistic speech (Hobson et al., 2012; Du Bois et al., 2014; Tantucci & Wang, 2022).

Consider the following naturalistic example drawn from the demographically sampled section of the British National Corpus 2014 (BNC2014; Love et al., 2017), where three students talk about cakes:

(1) [*Three friends talking about cakes*]

A: What, she made them?

B: Yeah, **they're like polenta mix**

C: oh I see, **that's like that's not wheat, polenta's not wheat**.

BNC2014/S2ZU/S0327

In (1), speaker C spontaneously reuses a construction similar to B's. She reformulates the pronoun *they* as *that*, the verb inflexion *'re* of the lemma BE as *'s*, she repeats verbatim the comparative marker *like,* and then replaces the object referring to the type of flour *polenta mix* with *not wheat*. By doing so, C expresses overt interest in what B just said, as she not only engages with her words but also with the constructions (Goldberg, 2006; Ungerer & Hartmann, 2023) that she used. Constructions are pairings of form and meaning that are traditionally ascribed to one single speaker. However, from the angle of Dialogic Syntax, these emerge pragmatically, semantically, phonetically and

morphosyntactically from the combined efforts of both speakers (Tao & Suzuki, 2022). In the case of example (1), the dialogic construction [PRON BE *like* OBJ_{flour}][1] emerges as a byproduct of C's engagement with what B said.

Consider now the following example, also taken from the BNC2014 where two speakers in their fifties discuss a trial:

(2) [*Mundane discussion about a trial*]

A: Work commitment three times without.

B: Yeah.

A: Providing evidence.

B: Yeah.

A: You know er he can only go on for.

B: No.

A: So long (.) so he sent him to prison for eight weeks.

BNC2014/S575

As shown in (2), B's overt engagement with A's speech is much poorer. All there is here is backchanneling, as B merely acknowledges what is being said. However, none of what A says is overtly resonated in B's turns. In this case, there is no propositional effort to show that what A is saying is relevant for continuing this ongoing interaction.

As we mentioned, resonance is a form of alignment hinging on re-using and re-adapting an interlocutors' constructions, potentially including prosodic and gestural components (cf. Tantucci, forthcoming). On the one hand, it overtly stipulates that what an interlocutor said is relevant for the continuation of an interaction. On the other hand, it is a key mechanism for language learning (e.g. Köymen & Kyratzis 2014; Tantucci & Wang 2022b, 2023), as it involves morphosyntactic, pragmatic and semantic categorisation as a dialogic effort of both interlocutors. Alignment is a more general ability to coordinate cognitively and communicatively with an interlocutor, but does not necessarily involve complex imitation and (creative) adaptation of someone's communicative behaviour:

(3)

[Mundane discussion about someone living far from the interlocutors' place]

A: I'm not actually when you think about it.

B: What do you mean exactly?

[1]

A: Erm I think it's not it's not rea- a given that all the sort of the most brilliant best people are gonna end up at.

BNC2014/S2ZU/S0327

In (3), there is clear cognitive and communicative alignment: Speaker B's clarification request *What do you mean exactly* leads to A's elaboration on the topic, so that speakers coordinate turn-taking smoothly, and work together to establish mutual understanding. Here speakers align but do not resonate. Although the speakers are aligned in meaning and are communicatively cooperating, B does not re-use A's linguistic constructions, nor does A build on B's language in a way that shows dialogic engagement with respective linguistic choices. The interaction is collaborative, but not formally imitative or creatively resonant.

When measured on a large scale, resonance is a reliable indicator of verbal engagement (Tantucci & Lepadat, 2024), as it counts as an index of sustained involvement (or lack thereof) with an interlocutor's speech. In this study, we will operationalise resonance in naturalistic interaction between Test-takers and Examiners during a spoken English language test. We will focus on the degree of resonance from each party, whether this correlates with speakers' language proficiency and, most crucially, whether resonance favours Test-takers' performance in oral tests.

Pragmatic and interactional competence in second language acquisition and testing

Resonance is decisive for pragmatic competence (Tantucci, 2023), which, in turn, is key in (second) language learning. Pragmatic competence is what allows language users to participate in communication, drawing from linguistic, social, and cultural knowledge to express meaning in context (Archer & Grundy, 2011). Research on interlanguage pragmatics (e.g., Liu, 2006, 2010; Kasper & Blum-Kulka, 1993) has explored the acquisition of second language (L2) users' pragmatic competence and the effect of the communicative context on L2 learners 'pragmatic choices and social identity (Taguchi, 2019). L2 pragmatics has primarily focused on the link between pragmatics and grammar, with a focus on speech acts (e.g., complaining, (dis)agreeing, and opinion-giving) (e.g., Bardovi-Harlig, 2003; Kasper & Rose, 2002) and L2 speakers' engagement (e.g., Mercer, 2019). Few studies have investigated other features of L2 pragmatics, for example, implicature (e.g., Bouton, 1999), routine formulae (e.g., Bardovi-Harlig, 2009), and politeness (e.g., Liu, 2006). Little research in L2 pragmatics has focused on the effect of Teachers/Examiners' speech on learners' language production (e.g., van Lier & Matsuo, 2000). Relatively little attention has also been paid to how teacher or examiner talk shapes learners' language production (e.g., van Lier & Matsuo, 2000), or to

the ways in which meaning in focused interactions is co-constructed by participants through the situated interpretation of both linguistic and nonverbal behavior (Jenkins & Parra, 2003: 91). Notably, Kasper and Ross (2006) focus on the interactional competence that needs to be enacted in contexts of language testing (e.g., Young, 2008). They bring attention to power asymmetries and found that Test-takers interpret examiner alignment or disalignment as cues to recalibrate their behavior, which introduces additional interactional demands beyond language use itself. This project responds directly to McNamara's (1996) call for a research agenda that takes interaction seriously in language testing and investigates variables that are theoretically central to performance (pp. 85–86). Similarly, Brown (2003) examines how interviewer variation during IELTS Speaking Tests shapes the co-construction of speaking proficiency, emphasizing that interactional competence is not an individual trait but jointly produced through interaction. Brown identifies interviewer styles—ranging from accommodating to controlling—that significantly influence test-takers' opportunities to demonstrate skills such as turn-taking, repair, and elaboration. Accommodating interviewers, who build rapport and adjust their questioning, foster more dialogic and extended candidate responses, while controlling interviewers constrain interaction. Following Jacoby and McNamara (1999), we adopt the view that interactional competence is not an individual possession but something that emerges in the interaction itself—"jointly constructed" through practices such as turn-taking, repair, and the negotiation of topics and roles. As Galaczi & Taylor (2018) emphasise, interactional competence is a sociocognitive mechanism, encompassing turn-taking, topic management, repair, and interactive listening, as well as kinesthetic abilities to use visual behaviors like gaze and gestures effectively.

In second language acquisition (SLA), experimental studies explored L2 pragmatic competence in the classroom, e.g. comparing implicit and explicit instruction (e.g., Eslami & Eslami-Rasekh, 2008; Fordyce, 2014), collaborative and individual learning (e.g., Taguchi & Kim, 2016), and learners' metalinguistic awareness of pragmatic competence, e.g., (audio-visual) acceptability judgement tasks (e.g., Bardovi-Harlig, 2013) and verbal reports (e.g., Cohen & Olshtain, 1993).

Spoken/written discourse completion tasks (DCT) and multiple-choice tasks have been used to examine learners' acquisition of speech acts (e.g., Roever, 2011). However, these tasks do not provide learners with opportunities for authentic language use or extended interaction (Golato, 2003; Bardovi-Harlig, 2013; Cohen, 2020). By contrast, Role-Plays and interviews elicit spontaneous dialogic language production (e.g., Gablasova et al., 2017).

Our study aims to shed new light on a neglected area in the literature of interactional and pragmatic competence: the role of grammatical constructions as a dialogic byproduct of engagement. Specifically, we explore how dialogic resonance and the joint realisation of dialogic constructions as pairings of form and meaning can be analysed as key quantifiable dimensions of L2 pragmatic and interactional competence.

The 'role' of Role-Plays

Role-Plays are based on the description of a communicative context, including details about speakers' characteristics and communicative goals (Félix-Brasdefer, 2004, 2007). They allow the investigation and assessment of L2 pragmatic competence across various communicative contexts and events. Role-Plays are considered "the closest elicitation tasks to conversation" (Bardovi-Harlig, 2013:74) and elicit speech characterised by linguistic features similar to spontaneous interaction: even if speakers can plan the first turn, they can't usually plan the following ones (Bardovi-Harlig, 2013). Most research on Role-Plays and interviews relied on specific queries from a corpus, that is, on expressions decided 'a priori'. Typical measures include frequency of speech acts (e.g., Hudson, 2001) across proficiency levels and task types (e.g., Aijmer, 2004; Gablasova et al., 2017; McEnery & Kifle, 2002), the analysis of over/under/misuse of formulaic expressions (e.g., Takahashi, 2010), and register appropriateness (Biber, 2006; Gilquin & Paquot, 2008).

Engagement in L2 pragmatics

Engagement is a key component of (L2) pragmatic competence. In this field, it is defined as "the amount (quantity) and type (quality) of learners' active participation and involvement" in conversation and in different learning tasks or activities (Hiver et al., 2024:202). Active engagement varies depending on intra- and inter-cultural conventions (Aubrey, 2017) and is pivotal in education and language learning (Mackey, 1999; Philp & Duchesne, 2016; Phung, 2017; Mercer, 2019). It influences academic participation, performance, and success. By contrast, disengagement may correlate with learners' poor mental health, academic dropouts, and high-risk behaviours (Christenson et al., 2012).

In second language acquisition (SLA), diverse engagement indexes are addressed as predictors of language learning, e.g., attention, language awareness, and motivation (e.g., Schmidt, 2001; Svalberg, 2009; Rebuschat, 2015). Behavioural, cognitive, and social aspects of engagement are closely linked to resonance (Hiver et al., 2024; Philp & Duchesne, 2016). One way to account for spontaneous engagement has been by counting learners' words and turns, with higher scores indicating more involvement. Conversational features – e.g., turn-taking and backchanneling – and sociopragmatic ones – e.g., opinion-giving and reciprocity (Culpeper & Tantucci, 2021; Culpeper et al. 2025) – have been used as measures of social engagement (Hiver et al., 2024).

Most studies on engagement in language learning rely on qualitative learners' self-reports, e.g. questionnaires or focus groups (Hiver et al., 2024). Few studies have explored engagement in

dialogue, using quantitative measures on L2 proficiency (e.g., Dörnyei & Kormos, 2000; Garcia-Ponce & Tavakoli, 2022) and task complexity (e.g., Michel, 2011). Some studies also looked at alignment in conversation (e.g., Costa et al., 2008; Tavakoli & Foster, 2008), emphasising that aligning speech to the interlocutor's (e.g. copying their constructions) "greatly simplifies production and comprehension" (Pickering & Garrod, 2004: 169).

Quantitative research on L2 engagement indicates that the more proficient the speakers, the higher their engagement with the interlocutors, with increasing turn-takings, discourse markers, word repetitions, and text length (e.g., Crossley & McNamara, 2013; Iwashita et al., 2008; Kang & Wang, 2014; Garcia-Ponce & Tavakoli, 2022). Finally, lexical diversity (i.e., the range of different words produced) also plays a role in engagement. Lower lexical diversity was found in dialogic speech compared to monologic speech in L1 and advanced L2 learners (Michel, 2011; Bottini, 2022): proficient speakers produce more repetitions in dialogues to create interpersonal involvement (Tannen, 1990).

Engagement and authenticity in language testing

In language testing, authentic engagement involves Test-takers' cognitive, linguistic, and social skills typical of real language use (Spence-Brown, 2001; Burton, 2020). This is important for test validation and construct validity (Alderson et al., 1995; Brown, 2000) as test-score interpretations can be generalised to communicative competence in the real world.

Learners', teachers', and examiners' perceptions of authentic engagement have been analysed with interviews and questionnaires (e.g., Behizadeh & Engelhard, 2014; Emerick, 2019). Authenticity in Test-takers' language performance has also been compared with real-world language use using corpus methods (e.g., Staples et al., 2018), conversational and discourse analysis (e.g., Woodward-Kron & Elder, 2016; Gates, 2018). However, little research tackled authentic engagement in spoken language tests, for instance, in assessing scriptedness, spontaneity, and relevance (Burton, 2020; Roever, 2011).

Overall, most existing studies on L2 engagement have used self-reports and qualitative methods, while only few used quantitative measures. Among these, none operationalised learners' and teachers'/examiners' engagement measuring how they affect each other's speech in interaction, especially in a language testing context. Philp & Duchesne (2016) highlight the importance of extending the contexts where engagement is studied in SLA since most existing studies are limited to the classroom context. Similarly, Hiver et al. (2024) call for further research to investigate the effect of examiner talk, task type, and other contextual variables on Test-takers' engagement.

In the context of our study, we refer to engagement as learners' verbal and pragmatic involvement in interaction (e.g. Hiver et al. 2024), hinging on the quantity and quality of a speaker's participation in a dialogic exchange, encompassing both constructional categorisation and interpersonal involvement. We will then treat authenticity, as the resemblance of test interactions to real-world language use (e.g., Spence-Brown, 2001; Burton, 2020). This includes the naturalness of the language, the spontaneity of the interaction, and the degree to which it reflects actual communicative contexts. Authentic engagement thus arises in interactions that are both spontaneous and socially consequential. It underpins situated, mutual involvement between interlocutors that goes beyond the mere goal of a task completion and involves, in the specific case of this paper, overt attempts to (creatively) resonate with an interlocutor's speech. This is not a condition explicitly given to improve the task performance and emerges naturally as a byproduct of speakers' intentions to engage with one another's speech. More specifically, we will focus on dialogic engagement in a low-stakes English spoken language test. We examine whether and how Test-takers and Examiners verbally imitate each other by measuring their degree of (creative) resonance. These are our research questions:

1. Do Examiners and Test-takers verbally imitate each other?
2. Does resonance increase across L2 proficiency levels?
3. Does Examiners' resonance enhance Test-takers' performance?

Methodology

In this section, we illustrate the retrieval (4.1) and the annotation (4.2) of our data.

Data retrieval

For this study, we selected the *International Corpus Network of Asian Learners of English* (ICNALE; Ishikawa, 2023). ICNALE includes 3.5 million words produced by users of English as a first language (L1) and Asian learners of English as a second language (L2) whose proficiency ranges from pre-intermediate to upper-intermediate (corresponding to CEFR levels A2 to B2). The L2 participants are college students from ten regions in Asia – four regions where English is used as a second language (ESL) (Hong Kong, Pakistan, the Philippines, and Singapore/Malaysia) and six where English is used as a foreign language (EFL) (China, Indonesia, Japan, Korea, Thailand, and Taiwan). The L1 participants include "college students, teachers, and businesspersons in more than five countries in the world"[2] (Ishikawa, 2023:18). The corpus contains written essays, spoken monologues and

dialogues' transcripts and video recordings. Data collection was controlled for topic and production time. L2 participants' metadata includes age, gender, education and academic background, language background (L1), L2 learning history, L2 proficiency, and English learning motivation. L2 proficiency metadata comes from high-stakes exams the Test-takers took before data collection.

We used data from the Spoken Dialogues component. Each dialogue occurs between a Test-taker and an Examiner during a face-to-face oral proficiency interview (Ishikawa, 2023:20). It lasts approximately 30-40 minutes and consists of:

1. A conversation about learning English.
2. A picture description, Role-Play and Q&A on the topic of part-time jobs.
3. A picture description, Role-Play and Q&A on the topic of smoking in public places.
4. A final reflection on the interview performance.

We only considered tasks 2 and 3. For each of them, we extracted the language produced during Role-Play and following Q&A and disregarded picture description tasks, so that we could focus exclusively on dialogic spontaneous interaction (Bardovi-Harlig, 2013). During each Role-Play, the Examiner gives the Test-taker a card describing an imaginary scenario. Figures 1 and 2 show the cards about the part-time job and the smoking Role-Play, respectively.

Figure 1.

Instructions about the Role-Play on the topic of part-time jobs (Ishikawa, 2023:53)

Figure 2.

Instructions for the Role-Play on the topic of smoking (Ishikawa, 2023:53)

The register of Role-Plays is formal, and the Examiner is instructed to refute the Test-taker's claims. This assesses the Test-taker's ability to persuade an interlocutor in a position of power who holds a different opinion. The Q&A session is informal and occurs after the Role-Play. The Examiner asks questions about the Test-taker's personal experience with the Role-Play topic. The Examiners' contributions are mostly **unscripted**, a key factor of this exam, but follow some prompts with sample questions in the Q&A task (Ishikawa, 2023:53).

ICNALE Examiners are teachers and speakers of English as a second language (L2). This is important for two reasons: to ensure the Test-takers felt at ease during the interviews and to represent the use of English as a lingua franca. The ICNALE team selected teachers who passed a mock interview to act as Examiners during the oral proficiency interviews.

We extracted a random sample of transcripts from the Role-Play and Q&A tasks. This was balanced for Test-takers across proficiency levels and language backgrounds. Our final dataset consists of 2,564 turns and 41,676 words (mean number of words per turn = 16.25). Table 1 provides descriptive statistics about the dataset.

Table 1.

Number of speakers, turns, and tokens per proficiency level which were analysed in this study[3]

The sample included twenty dialogic interactions: five per proficiency level. The number of tokens varies across proficiency levels, with the lowest value (*M*=8.54, Total=2,212) in the A2 group.

Annotation

Each observation in our dataset corresponded to a different speaking turn. We manually coded the transcript for speaker's ID, gender, language background (LB), speaker role (Examiners (E) vs Test-takers (T)), task type (TTy), task topic (TTo), resonance, English proficiency level (PL), illocutionary force (IF), intonation (a PRAAT[4] spectrogram showing whether the final intonation of each turn was rising, falling or level), and word count (W). A sample line of the input of all these dimensions is given in Table 2:

Table 2.

Sample input of annotation[5]

In the following subsections, we describe the annotation rationale for resonance and illocutionary force with reference to examples from our dataset.

Annotating resonance

Resonance is the dependent variable of our study. It accounts for the verbal efforts (if any) made by Examiners (E) and Test-takers (T) to engage with one another's utterances. Resonance occurs when speaker B re-uses some of the speaker A's words or expressions and can be reliably measured as a continuous variable (e.g. Tantucci & Wang, 2022a, 2022b, 2023, 2024; Tantucci & Lepadat, 2024). To do this, we relied on Tantucci's (2023) Dialogic Categorisation Model (DCM), which includes two conditions:

i. Resonance can be identified when there is at least one word – including interjections or pragmatic markers – that is repeated verbatim from interlocutor A to B[6].
ii. The measurement of resonance is based on the number of internal constituents of the dialogic construction(s) that emerges from both A and B's utterances.

We can look at example (4) from our dataset and see how to operationalise this (resonance cases will be marked in bold in all extracts).

(4) [Q&A: *Part-time job*]
E: Okay, right, I understand. So please introduce **the job you were doing before**.
T: Uh, **my job I was doing before** was a cleaning job.
ICNALE CHIN001 (B1.2)

The pair in (4) is from a Q&A task where a Chinese Test-taker (T) answers Examiner's (E) questions on part-time jobs and resonates with his construction. Here, T pro-actively engages with E's turn as he verbatim re-uses several of his words, namely *job, doing* and *before*. This satisfies the condition (i) for the identification of resonance. From this, we can see whether a more schematic construction can be derived from T's engagement with E's speech, as per condition (ii). In the same turn, T re-combines the determiner *the* as *my* and the personal pronoun *you* as *I.* This creates the affordances for the joint categorisation of a more schematic construction [DET[7] *job* $\text{PRON}_{\text{personal}}$ $\text{BE}_{\text{past-tense}}$ *doing before*]. The internal constituents of this are 6: this is the resonance value to annotate for T's turn. The same procedure is then repeated for all the dataset turns so that a large-scale resonance value across speakers in different contexts can be obtained.

Dialogic constructions that emerge from resonance are best represented as diagraphs (Du Bois & Giora, 2014: 354), i.e., syntactic structures emerging across turns, as given in Table 3. When the original ad hoc construction is modified, that is marked as underlined text (in case of <u>replacement</u>) and in brackets (in case of (<u>addition</u>)):

Table 3.

Diagraph: [DET *job* $PRON_{personal}$ $BE_{past\text{-}tense}$ *doing before*]

The top row of Table 3 is the dialogic construction jointly realised by both Examiner and Test-taker. Here, most of T's engagement with E's speech is via verbatim repetition of what he heard. There is, in other words, little **recombinant creativity** (Tantucci, 2023), as only 3 constituents (*the* > *my*; *you* > *I*; *were* > *was*) out of 6 are re-adapted to engage with E's speech innovatively. Recombinant creativity favours language learning. It creates the conditions for establishing abstract categories of form and meaning as components of the constructions being used, in this case, DET, $PRON_{personal}$, and $BE_{past\text{-}tense}$. Recombinant creativity is also at play at the illocutionary level: T innovatively turns E's directive speech act (Searle, 1976) into a new Representative one (cf. Schegloff, 2007 on conditional relevance of adjacency pairs).

We can now annotate a second example, based on the Role-Play task, between an Examiner (E) and an Indonesian Test-taker (T).

(5) [Role-play: *Part-time job*]

E: Yeah, I heard from many of your friends that now **you are working part-time**.

T: **I am working** in the – **as a waitress** at the restaurants.

ICNALE/IDN001 (B1.1)

Table 4.

Diagraph: [$PRON_{personal}$ BE *working* $ADV_{restaurant}$]

Similar to (4), we now can identify resonance due to verbatim repetition of at least one lexical item from E to T, namely *working*. This may create the conditions for more schematic analogies across turns and thus satisfy condition (ii). That is indeed the case, as the personal pronoun ($PRON_{personal}$) *you* is recombined as *I*, the copula BE, inflected as *are* in E's turn, is reformulated as *am* by T, and the adverbial referring to a restaurant job ($ADV_{restaurant}$) *part-time* is turned into *as a waitress*. The degree of resonance here is 4: the number of internal constituents of the dialogic construction [$PRON_{personal}$ BE *working* $ADV_{restaurant}$] emerging from T's engagement with E's speech. Recombinant creativity here is 3, the number of constituents replaced from E to T. This is because

resonance accounts for engagement with an interlocutor's linguistic material and comprises all the 4 slots of the emerging dialogic construction [$PRON_{personal}$ BE *working* $ADV_{restaurant}$]. Recombinant creativity is more specific. It has to do with pro-active reshaping of an interlocutor's construction and involves the internal constituents that are re-adapted by T when re-using E's construct. This creates affordances for schematic categorisation and, ultimately, language learning. The speech act is also recombined, as E's rogative speech act (one eliciting a response, cf. Leech, 1983, 2014) becomes a representative in T's turn.

This annotation process required three stages of inter-rater reliability. Three independent annotators disambiguated resonance values throughout our 2,564 observations. The rate of accuracy across annotators was measured in Krippendorff's Alpha and corresponded, respectively, to $\alpha = 0.74$, $\alpha = 0.89$, and finally $\alpha = 0.93$. At each stage, a 25% sample of the data was independently annotated. Cases of disagreement were resolved among the annotators before moving to the annotation of a new randomised sample.

Annotating illocutionary force

Illocutionary force (IF) is an important variable of our study. This is not because we anticipated that IF would answer our RQs, but because the roles of a Test-taker (T) and an Examiner (E) are very different. E will be more likely to ask questions or make requests than T. The latter, in turn, is more likely to assert things. This may influence resonance, e.g., it is more likely to resonate with someone's opinion (A: *today* ***is far too hot to work*** B: *I am also sweating,* ***it's impossible to work***) than in response to a question or a request (A: *Are you ready?* B: *Yes*). We controlled for IF as a random effect and annotated it as the head act of every turn of our dataset. Our taxonomy draws on Searle (1976) and Leech (2014), comprising 5 categories: Commissives, Directives, Expressives, Representatives and Rogatives[9]. With Commissives, the speaker commits to future actions, as in promises, offers, and so on. As face threats (Brown & Levinson, 1987) were unsurprisingly scarce in our dataset, Commissives were mainly used politely. They could be identified by whether thanking for the offer/help was a plausible response.

Directives are performed to get the hearer to do something, such as orders, requests, and similar elicited behaviours. They are expressed grammatically with imperatives/hortatives (cf. Traugott & Dasher, 2002). In contexts of pragmalinguistic politeness (Leech, 2014), they are (or can be) idiomatically marked with the pragmatic marker *please*, as in (6). Here T tries to persuade E to return the money spent in a restaurant full of smokers:

(6) [Role-play: *Smoking*]

E: **So**, let's go, **please begin**.

T: **So, please return my money,** it is uncomfortable to eat food.

ICNALE/IDN007 (A2)

Excerpt (6) includes a matching of directives. Even in this case, resonance is present with the [*So, please* VP] construction emerging from T's engagement with E's speech.

Rogatives involve seeking information from the hearer (cf. Leech, 1983). For this we looked at transcriptions' question marks *?*, Subject-Auxiliary Inversions (SAI), and turn-final rising intonations. Matching of rogatives across turns was more common, e.g., in contexts where T asks for some clarification about E's question, as in example (7):

(7) [Role-play: *Part-time job*]

E: Okay, so are you satisfied with your job?

T: Yeah I was – you mean?

E: Satisfied? So, do you think it's a good job for you?

ICNALE/CHN002 (B1.1)

Representatives describe the state of affairs of a situation. They include assertions (e.g., *it's raining*), evaluatives (e.g., *I think it may be raining*) and information based on some evidence (e.g., *apparently it's raining*) (cf. Tantucci, 2016a, 2016b). Finally, expressives communicate feelings and affects. They were identified as conventionalised constructions in contexts of thanking, praising, greeting and less compositional strategies to express feelings.

(8) [Q&A]

T: Yeah, **bye**.

E: **Bye**. Well, good job.

ICNALE/IDN003 (B1.2)

Illocutionary force depends on formal but also functional diagnostics and thus requires careful inter-rater reliability. Like what we did for resonance, three independent annotators disambiguated resonance values throughout the 2,564 observations of our dataset. The rating accuracy was measured in Krippendorff's Alpha and corresponded, respectively, to $\alpha = 0.68$, $\alpha = 0.74$, and finally $\alpha = 0.81$. At each stage, a 25% sample of the data was independently annotated. Cases of disagreement were resolved among the annotators before moving to the annotation of a new randomised sample.

Analysis

In this section, we address our three research questions (RQ) and provide the statistical and qualitative analysis of our data.

Examiners and Test-takers imitate each other

We can start with descriptive statistics addressing RQ1: *Do examiners and test-takers verbally imitate each other?*

Figure 3.

Half-violin plots of resonance produced by Test-takers and Examiners

The half-violin plots in Figure 3 represent values of resonance (y-axis) spontaneously produced by Test-takers (blue) and Examiners (orange) in the two tasks of our study (x-axis): Role-Play (left) and Q&A (right). This visualisation includes the Kernel distribution density of resonance (the larger the area, the more the instances around that specific value on the y-axis) and the boxplots, with the means given as a black dot.

The answer to RQ1 is **yes**: Test-takers (T) and Examiners (E) frequently imitate one another during spoken language tests. Even more surprising is that levels of resonance of T and E are very close. One may expect T to imitate E more due to power imbalance (cf. Culpeper & Tantucci, 2021), to value what is said by E, and to increase their chances of passing the test. That is indeed the case, but only very marginally. Tables 5 and 6 include resonance means across speakers.

Table 5.

Resonance means across task type (when resonance was present)

Table 6.

Resonance means across task type (in the whole dataset)

Table 5 includes the means in the Role-Play and Q&A tasks of both T and E when resonance is present. Table 6 is based on the same conditions but now includes all turns produced in our data (not just resonating ones). Two very important observations are in order here:

i. Contexts of Role-Play allow remarkably higher levels of resonance for both Test-takers and Examiners.
ii. Examiners resonate less than Test-takers, but just barely.

Concerning (i), the context of Role-Play differs greatly from Q&A. In Role-plays speakers perform a conversation much less bound to a sequential question-answer structure typical of an interview (c.f. 'scriptedness' in Goffman, 1981; Levinson, 1979). Both speakers 'freestyle' in a real-world dialogue where T's communicative goal is to persuade their supervisor to allow them to do a part-time job (cf. Tantucci, 2021 on perlocutionary effects). Dialogic environment and sequence organisation (Schegloff, 2007) are less fixed and predictable. They allow for much more creative intervention in one another's speech, serving agreement and disagreement (cf. Tantucci & Wang, 2021, 2022a). This indicates that 'scripted' interactions (i.e., where speakers' turns at talk are prescribed a priori, e.g., Goddard, 2004) inhibit engagement. The latter increases when interlocutors can 'freely' take turns.

The other aspect worth expanding on is (ii), as it is far from obvious why E imitate T's speech to such a high degree of resonance. Consider example (9) below:

(9)

E: So, uh, there are different types of speaking such as, uh, you know, face-to-face, one-to-one conversation and then a **group discussion**. Which do you like better?
T: Uh, of course, the um, group uh, the **group talking**.
E: Hmm, so why do you prefer a **group discussion**, **group talking**?

ICNALE/CHN002/(B1.1)

In (9), T resonates with the compound *group discussion,* as he recombines it as *group talking*. The more schematic [*group* NP_{speech}] construction thus results from T's engagement with E's speech. At this point, E also resonates with T and re-uses the original compound *group discussion* to show that this is a preferable lexical choice. He finally self-expands on the same lexeme by resonating with the student's less idiomatic expression *group talking*.

The Examiner here shows a pedagogical preference for the idiomatic option (he recasts, cf. Ammar & Spada, 2006), so that T understands that the best collocate of the [*group* NP_{speech}]

construction is *discussion*. At the same time, E also imitates what T said to engage with his/her speech, boost his/her positive face (cf. Brown & Levinson, 1987; Tantucci et al., 2022), and motivate him/her to continue the interaction. This illustrates how, for different reasons, resonance is as important a resource for E as it is for T. In section 5.3, we will see that E's resonance has a remarkable effect on T's exam performance.

Resonance and proficiency

We can now address RQ2: *Does resonance increase across L2 proficiency levels?*

To do so, we fitted a mixed-effects linear regression using the *lme4* package (Bates et al., 2015). Our dependent continuous variable was Resonance, our predictors were Proficiency (A2, B1.1, B1.2, B2, L1 test-taker, L2 examiner), Task type (Role-Play, Q&A), Task Topic (Part-time job, Smoking), Role (Examiner, Test-taker). As random effects, we included Illocutionary force (IF) and Test-taker ID nested into language background (LB) (Chinese, English, Indonesian, Japanese, Taiwanese, Thai). Including "Examiner" and "L1 Test-taker" as levels of the Proficiency predictor allowed us to model interactional variation across all speaker types in the corpus and capture differences in resonance that arise not only from linguistic proficiency, but also from distinct roles and interactional goals. The results of the best-fitting linear regression model are shown in Table 7. The model explains 20% of the variance of resonance values (R^2 = .198)[10]:

Table 7.

Results of the best-fitting mixed-effects regression model predicting resonance values[11]

The random effects at the top of Table 7 comprise the standard deviation, i.e. the variability from the predicted values. The fixed effects are reported below, with the estimate column showing the coefficients that predict resonance relative to each predictor. Significant values are marked at the bottom left with one or more asterisks (*), while tendencies that approximate significance are marked with a dot (.).

The mean value of resonance at the intercept is 3.486. This then increases across proficiency levels: **the more proficient the Test-taker, the higher the resonance value**. This is a key result that directly answers our RQ2. In fact, there is a positive coefficient for B1.1 Test-takers, approximating significance (β = 0.810, p = 0.081), then significantly increasing for B1.2 (β = 1.019, p = 0.034) and even more for B2 (β = 1.165, $p < 0.001$). L1 Test-takers' and L2 Examiners' values are also positive, albeit not significantly. This is presumably due to scarcer motivation in passing the test in L1 Test-

takers (our control group) and the very role of Examiners, using different resonance strategies from the ones used by Test-takers (cf. Section 5.1).

Our fixed effects also include task type. Here, resonance values are significantly lower in the Q&A compared to the Role-Play (β = -1.024, p < .001). This is due to interactional scriptedness: the more constrained the interlocutors' behaviours, the lower their resonance values. A Role-Play context allows speakers to take turns and engage with each other more freely than in Q&A. The fixed effect of task topic is not statistically significant. However, when it comes to speakers' role, Examiners significantly resonate 'slightly' albeit significantly less than Test-takers (β = -0.49, p = 0.037).

Back to our main finding: resonance increases across proficiency levels. This is also evident in Figure 4, showing the error bars of resonance values (y-axis) predicted by proficiency (x-axis) and task types (colours):

Figure 4.

Predicted values of resonance per proficiency level and task type

Figure 4 shows a steady increase in resonance values across all Test-takers, from A2 to B2. There is then a mild decrease with L1 Test-takers (presumably less concerned about proving their speaking skills) and Examiners (due to their different roles).

Something interesting can be said about the random effects. Figure 5 displays the intercepts of resonance predicted by language background.

Figure 5.

Predicted resonance values per linguistic background

Figure 5 shows that the intercepts for Taiwanese and Chinese speakers of L2 English are the highest for resonance, which confirms that resonance tends to be cross-culturally higher among L1 speakers of Chinese (including Mandarin and Minnan, cf. Tantucci & Wang, 2021, 2022a).

Resonance and test performance

In this final section of our analysis, we address RQ3: *Does examiners' resonance enhance test-takers' performance?*

To investigate this, we isolated all instances where a test-taker's (T) turn immediately followed an examiner (E) utterance containing resonance. We then fitted a Bayesian regression model[11], using T's word count as the dependent variable and E's resonance as a continuous predictor. A Bayesian approach was chosen to better handle the reduced sample size (n = 283) and to quantify uncertainty more effectively in this exploratory analysis. We used the *brms* package (Bürkner, 2017), fitting a Gaussian family regression including four chains with 10,000 iterations for Markov Chain Monte Carlo (MCMC) sampling, with the first 1,000 iterations used as a warm-up. For the Bayesian model we used weakly informative priors to regularise estimates and capture uncertainty (Normal(0, 5) prior for the intercept and a Normal(0, 2) prior for the slope). The model converged with all R-hat values at 1.00. Results indicated a positive linear relationship between examiner resonance and subsequent learner word count. While this does not confirm causality, it clearly suggests that examiners' dialogic alignment cues greater learner engagement in the next turn. We visualized this relationship in Figure 6, with a regression line and 95% credible intervals. For clarity, the plot is based on a posterior predictive distribution centered at the estimated intercept ($\beta_0 \approx 24.94$) and slope ($\beta_1 \approx 3.57$), with standard errors used to calculate uncertainty bands. The results of our Bayesian model are presented in Table 8:

Table 8.

Bayesian regression of E's resonance predicting T's next utterance length

The intercept indicates that when the Examiners' resonance is zero, the expected Test-takers' word count in the following turn is approximately 24.94 (first column). The coefficient for resonance is 3.57, which means that for every unit increase in resonance, the Test-takers' word count in the next turn is expected to increase by 3.57 (i.e., something between 3 and 4 words). The lower and upper credible intervals (l-95% CI: 2.79, u-95% CI: 4.33) comprise the range of this increase, suggesting high certainty in this estimate.

These results can be captured visually in Figure 6, showing the predicted values of Test-takers' word count depending on Examiners' resonance:

Figure 6.

Predicted Test-takers' utterance length depending on Examiners' resonance

Figure 6 clearly answers our RQ3: **Examiners' resonance improves Test-takers' performance**. Increasing Examiner resonance values on the x-axis leads to a dramatic rise in the Test-takers' word count on the y-axis. Put simply, the more the Examiners resonate, the more the language produced by the Test-takers immediately after. This is a remarkable result, as it demonstrates the pedagogical and transformative impact of Examiners/Teachers' resonance on Test-takers/Learners' verbal production and success at dialogic tests. Significantly, text length correlates with proficiency (e.g., Crossley & McNamara, 2013).

Discussion

Our study advances theoretical, methodological, and pedagogical practice in language learning, language testing and L2 pragmatics with three findings:

1. Verbal imitation plays a decisive role among Test-takers and Examiners as they constantly resonate with one another's speech.
2. Resonance increases with L2 proficiency. The more L2 learners resonate with an Examiner, the higher their communicative competence.
3. Resonance boosts Test-takers' performance. The more an Examiner resonates, the longer the utterances produced by the Test-taker as a result.

These findings are inherently transformative. They inform the discoursive practice of L2 Learners/Test-takers and Teachers/Examiners, showing that resonating with an interlocutor's speech is a booster for language learning, intercultural engagement, and exam performance. Resonance is a pivotal component of (L2) interaction and spoken language tests. It is bidirectional in unscripted and semi-scripted language tests, and a key element of authentic engagement. Test-takers resonate to engage with Examiners' speech. This serves to acquire new categories of form and meaning jointly constructed across turns: L2 speakers' resonance serves learning in the here-and-now of an interaction. Test-takers also resonate to impress Examiners, prove communicative competence, and receive higher marks. Among them, proficient L2 speakers with more developed pragmatic competence re-use interlocutors' linguistic resources effectively to create social engagement. Examiners also resonate with Test-takers. They do so for pedagogical reasons, create a positive communicative atmosphere in the test, motivate Test-takers and 'validate' what they say. They often

resonate to re-cast and provide immediate feedback on Test-takers' language production (cf. Section 5.1).

Importantly, L1 Test-takers' resonance is not as high as for upper-intermediate (B2) L2 speakers. ICNALE, L2 Test-takers are university students who can receive formative feedback and speaking practice with trained examiners. By contrast, L1 Test-takers include college students, businesspeople, and teachers who might not share the same motivation.

Another important finding is the relationship between resonance and 'scriptedness'. Activity types that comprise a scripted component – determining who needs to say/ask what and when – inhibit resonance. This was the case with Q&A sessions showing significantly lower resonance levels than Role-Play tasks. Scripted (or semi-scripted) talk involves a lower degree of proactive re-use and imitation of an interlocutor's speech and is thus less 'engaging'. This has decisive implications for test design, language teaching, and intercultural communication (e.g., Kecskés, 2014) in institutional settings.

Our most significant discovery is the impact of Examiners' resonance on Test-takers' performance: the more Examiners resonate, the longer the Test-takers' turns that follow. This indicates that Examiners' communicative behaviour has a direct effect on Test-takers' success at dialogic tasks (learners' text length is a byproduct of proficiency, e.g., Crossley & McNamara, 2013). The impact of this is twofold. It casts new light on the need to control for Examiners' resonance in dialogic tasks to secure test reliability. It also shows how crucial is Teachers' resonance for Learners' motivation and L2 verbal performance, as a key (albeit not the only) proxy of pragmatic competence.

Implications

Our findings have decisive implications in (second) language testing and language learning. They inform language test design, validation, marking criteria definition and scale construction. Designing spoken language tests with diverse dialogic tasks will enable Test-takers to display more pragmatic competence and engagement in different contexts, performing different speaker roles.

The findings from RQ3 are important for examiner training and standardisation. Examiners' awareness of the effects of resonance might be raised using sample extracts from datasets based on spoken language tests such as the ICNALE corpus. Examiners might also need regular training, guidance, and standardisation sessions on how to identify and evaluate dialogic resonance and engagement. Similarly, our results have practical applications in language teaching and language education policy, including standard settings, materials development, and teacher training (Gablasova & Bottini, 2022; Hiver et al., 2024).

Transformative and methodological implications also exist for research on (L2) pragmatic competence and engagement: resonance increases with pragmatic competence but varies inter-culturally and in neurodivergent populations (Tantucci & Wang, 2023). This should inform EDI teaching practices of conversational engagement in the classroom. Quantitative measures of resonance are a recent development in pragmatics research (Tantucci & Lepadat, 2024) with the potential to inform future research on language learning and verbal imitation.

Limitations

Some limitations of this study must be acknowledged. We investigated engagement in L2 and L1 English in a low-stakes language testing context. Future research could focus on types of language tests, such as high-stakes and/or computer-based exams. Our dataset did not include evaluations of Test-takers' performance, although we used Test-takers' L2 proficiency from high-stakes exams. The link between resonance and other marking criteria could be investigated, e.g. combining resonance with relevance and on/off-topic responses. Broader demographics (e.g. age, L2 speakers' learning background) and language instruction variables (e.g. different teaching strategies and classroom instruction) could also impact resonance and engagement.

Conclusion

This study provided a new theoretical and methodological approach to L2 pragmatics and L2 test design. It pointed to the fundamental role played by verbal imitation in dialogic tasks. It analysed Test-takers' and Examiners' dialogic resonance, a form of alignment involving the (often creative) re-use of an interlocutor's words and constructions. We provided a replicable annotation scheme to measure resonance on a large scale as an index of engagement and pragmatic competence. We found that not only Test-takers resonate with Examiners, the latter also persistently resonate and re-use Test-takers' speech. We discovered that resonance significantly increases across L2 proficiency levels: the more a learner resonates, the higher his/her speaking skills. Our analysis revealed that scriptedness inhibits engagement: the more a dialogic activity includes a scripted component, the more infrequent the speakers' attempts to resonate and engage with each other. Finally, we showed that Examiners' resonance enhances Test-takers' performance: when Examiners resonate, Test-takers immediately produce longer turns. These findings open a new research avenue into the interplay between resonance and language proficiency, suggesting potential applications in L2 educational settings, language testing, and key aspects of pragmatic competence.

References

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3*(1), 173–90. http://dx.doi.org/10.35360/njes.29

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*(4), 543–574. http://dx.doi.org/10.1017/S0272263106060268

Archer, D., & Grundy, P. (2011). Introduction. In D. Archer, & P. Grundy (Eds.), *The pragmatics reader* (pp. 1–10). Routledge.

Aubrey, S. (2017). Inter-cultural contact and flow in a task-based Japanese EFL classroom. *Language teaching research, 21*(6), 717–734. http://dx.doi.org/10.1177/1362168816683563

Bardovi-Harlig, K. (2003). Understanding the role of grammar in the acquisition of L2 pragmatics. In A. Martínez Flor, A. Fernández Guerra, & E. Usó Juan (Eds.), *Pragmatic competence and foreign language teaching* (pp. 13–25). Publicacions de la Universitat Jaume.

Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistics resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning, 59*(4), 755–795. http://dx.doi.org/10.1111/j.1467-9922.2009.00525.x

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning, 6*3(1), 68–86. http://dx.doi.org/10.1111/j.1467-9922.2012.00738.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Behizadeh, N., & Engelhard, G. (2014). Development and validation of a scale to measure perceived authenticity in writing. *Assessing Writing, 21*, 18–36. https://doi.org/10.1016/j.asw.2014.02.001

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97–116. http://dx.doi.org/10.1016/j.jeap.2006.05.001

Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.2.16, retrieved from http://www.praat.org/

Bottini, R. (2022). *Lexical complexity in L2 English speech: Evidence from the Trinity Lancaster Corpus* [PhD thesis]. Lancaster University.

Bouton, L. (1999). Developing non-native speaker skills in interpreting conversational implicatures in English: Explicit teaching can ease the process. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 47–70). Cambridge University Press.

Brown, J. D. (2000). What is construct validity? *JALT Testing and Evaluation SIG Newsletter*, 4(2), 7–10.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language testing*, *20*(1), 1-25.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage* (No. 4). Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511813085

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. http://dx.doi.org/10.18637/jss.v080.i01

Burton, J. (2020). Raters' measurement of test-task authentic engagement in L2 oral-performance assessment: An exploration of scale development. *System*, *90*, 102233. https://doi.org/10.1016/j.system.2020.102233

Chen, Q.R., Tantucci, V., & Chen, R.Y. (In press). Imperfective aspect in Chinese conversation: Do speakers imitate one another's constructions? *Cognitive Linguistics.*

Christenson, S., Reschly, A., & Wylie, C. (2012). *Handbook of research on student engagement.* Springer. http://dx.doi.org/10.1007/978-1-4614-2018-7

Cialdini, R., & Goldstein, N. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(1), 591–621. http://dx.doi.org/10.1146/annurev.psych.55.090902.142015

Cohen, A. (2020). Considerations in assessing appropriateness in spoken language. *Language Teaching, 53*(2), 183–202. http://dx.doi.org/10.1017/S0261444819000156

Cohen, A., & Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Quarterly, 27*(1), 33–56. http://dx.doi.org/10.2307/3586950

Costa, A., Pickering, M., & Vorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes, 23*(4), 528–556. http://dx.doi.org/10.1080/01690960801920545

Crossley, S., & McNamara, D. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology, 17*(2), 171–192.

Culpeper, J., & Tantucci, V. (2021). The principle of (im) politeness reciprocity. *Journal of Pragmatics*, *175*, 146-164.

Culpeper, J., Tantucci, V., & Field, E. (2025). Impoliteness reciprocity online. *Journal of Pragmatics*, *242*, 216-236.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science*, *11*(1), 86–89. http://dx.doi.org/10.1111/1467-9280.00221

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research, 4*, 275–300. http://dx.doi.org/10.1177/136216880000400305

Dresang, H., Wong, A., & Buxbaum, L. (2023). Shared and distinct routes in speech and gesture imitation: Evidence from stroke. *Cortex, 162*, 81–95. http://dx.doi.org/10.1016/j.cortex.2023.01.010

Du Bois, J. (2014). Towards a dialogic syntax. *Cognitive linguistics*, *25*(3), 359–410. http://dx.doi.org/10.1515/cog-2014-0024

Du Bois, J., & Giora, R. (2014). From cognitive-functional linguistics to dialogic syntax. *Cognitive Linguistics*, *25*(3), 351–357. http://dx.doi.org/10.1515/cog-2014-0023

Du Bois, J., Hobson, R., & Hobson, J. (2014). Dialogic resonance and intersubjective engagement in autism. *Cognitive Linguistics*, *25*(3), 411–441. http://dx.doi.org/10.1515/cog-2014-0025

Emerick, M. (2019). Explicit teaching and authenticity in L2 listening instruction: University language teachers' beliefs. *System, 80*(1), 107–119. https://doi.org/10.1016/j.system.2018.11.004

Eslami, Z., & Eslami-Rasekh, A. (2008). Enhancing the pragmatic competence of non-native English-speaking teacher candidates (NNESTCs) in an EFL context. In E. Alcón Soler, & A. Martínez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 178–197). Multilingual Matters. http://dx.doi.org/10.21832/9781847690869

Félix-Brasdefer, J. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language learning, 54*(4), 587–653. http://dx.doi.org/10.1111/j.1467-9922.2004.00281.x

Félix-Brasdefer, J. (2007). Natural speech vs. elicited data: A comparison of natural and role play requests in Mexican Spanish. *Spanish in Context, 4*(2), 159–185. http://dx.doi.org/10.1075/sic.4.2.03fel

Fordyce, K. (2014). The differential effects of explicit and implicit instruction on EFL learners' use of epistemic stance. *Applied Linguistics, 35*(1)6–28. http://dx.doi.org/10.1093/applin/ams076

Gablasova, D., Brezina, V., Mcenery, T., & Boyd, E. (2017). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics, 38*(5), 613–637. http://dx.doi.org/10.1093/applin/amv055

Gablasova, D., Bottini, R. (2022). Spoken Learner Corpora for Language Teaching. In R. R. Jablonkai, & E. Csomay (Eds.), *Routledge Handbook of Corpora in English Language Teaching and Learning* (pp. 296–310). Routledge.

Galati, A., Symeonidou, A., & Avraamides, M. (2019). Do aligned bodies align minds? The partners' body alignment as a constraint on spatial perspective use. *Discourse Processes*, *57*(2), 99–121. http://dx.doi.org/10.1080/0163853X.2019.1672123

Garcia-Ponce, E., & Tavakoli, P. (2022). Effects of task type and language proficiency on dialogic performance and task engagement. *System*, *105*, 102734. https://doi.org/10.1016/j.system.2022.102734

Gates, G. (2018). *An analysis of rehearsed speech characteristics on the Oral Proficiency Interview-computer (OPIc).* Unpublished MA dissertation, retrieved from https://scholarsarchive.byu.edu/etd/6731.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41–61. http://dx.doi.org/10.1075/etc.1.1.05gil

Goddard, C. (2004). "Cultural scripts": A new medium for ethnopragmatic instruction. In M. Achard, & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 145–165). Mouton de Gruyter. http://dx.doi.org/10.1515/9783110199857

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199268511.001.0001

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics, 24*(1), 90–121. http://dx.doi.org/10.1093/applin/24.1.90

Hiver, P., Al-Hoorie, A., Vitta, J., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language teaching research, 28*(1), 201–230. http://dx.doi.org/10.1177/13621688211001289

Hobson, R., Hobson, J., Garcia-Perez, R., & Du Bois, J. (2012). Dialogic linkage and resonance in autism. *Journal of Autism and Developmental Disorders*, 42, 2718–2728. http://dx.doi.org/10.1007/s10803-012-1528-6

Hudson, T. (2001). Self-assessment methods in cross-cultural pragmatics. In T. Hudson, & J. Brown (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (pp. 57–74). University of Hawai'i at Mānoa.

Ishikawa, S. (2023). *The ICNALE guide: An introduction to a learner corpus study on Asian learners' L2 English.* Routledge. http://dx.doi.org/10.4324/9781003252528

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49. http://dx.doi.org/10.1093/applin/amm017

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for specific purposes*, *18*(3), 213-241.

Jenkins, S., & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal*, *87*(1), 90-107.

Kang, K., & Wang, L. (2014). Impact of different task types on candidates' speaking performances and interactive features that distinguish between CEFR levels. *Cambridge Research Notes, 57*, 40–49. https://www.cambridgeenglish.org/images/177881-research-notes-57-document.pdf

Kasper, G., & Blum-Kulka, S. (1993). *Interlanguage pragmatics*. Oxford University Press. http://dx.doi.org/10.1093/oso/9780195066029.001.0001

Kasper, G., & Rose, K. (2002). *Pragmatic development in a second language*. Blackwell Publishing.

Kecskés, I. (2014). *Intercultural pragmatics*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199892655.001.0001

Kendrick, K., & Holler, J. (2017). Gaze direction signals response preference in conversation. *Research on Language and Social Interaction, 50*(1), 12–32. http://dx.doi.org/10.1080/08351813.2017.1262120

Köymen, B., & Kyratzis, A. (2014). Dialogic syntax and complement constructions in toddlers' peer interactions. *Cognitive Linguistics*, *25*(3), 497-521.

Ladd, D., Schepman, A., White, L., Quarmby, L., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics, 37*(2), 145–161. http://dx.doi.org/10.1016/j.wocn.2008.11.001

Leech, G. (1983). *Principles of pragmatics*. Longman. http://dx.doi.org/10.4324/9781315835976

Leech, G. (2014). *The pragmatics of politeness*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195341386.001.0001

Levinson, S. (1979). Activity types and language. *Linguistics, 17*(5–6), 365–399. http://dx.doi.org/10.1515/ling.1979.17.5-6.365

Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners.* Peter Lang.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics, 22*(3), 319–344. http://dx.doi.org/10.1075/ijcl.22.3.02lov

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, *21*(4), 557–587. https://doi.org/10.1017/S0272263199004027

McEnery, T., & Kifle, N. (2002). Epistemic modality in argumentative essays of second language writers. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 182–95). Longman. http://dx.doi.org/10.4324/9781315838069

McNamara, T. F. (1996). Measuring second language performance. Longman.

Mercer, S. (2019). Language learner engagement: Setting the scene. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 1–19). Springer. http://dx.doi.org/10.1007/978-3-030-02899-2

Michel, M. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second language task complexity. Researching the cognition hypothesis of language learning and performance* (pp. 141–173). John Benjamins.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72. https://doi.org/10.1017/S0267190515000094

Phung, L. (2017). Task preference, affective response, and engagement in L2 use in a US university context. *Language Teaching Research*, *21*(6), 751–766. https://doi.org/10.1177/1362168816683561

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–226. http://dx.doi.org/10.1017/S0140525X04000056

Pickering, M., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction.* Cambridge University Press. http://dx.doi.org/10.1017/9781108610728

Rebuschat, P. (2015). *Implicit and explicit learning of languages.* John Benjamins. http://dx.doi.org/10.1075/sibil.48

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing, 28*(4), 463–481. http://dx.doi.org/10.1177/0265532210394633

Schegloff, E. (2007). *Sequence organization in interaction: A primer in conversation analysis.* Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511791208

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524780

Searle, J. (1976). A classification of illocutionary acts. *Language in Society, 5*(1), 1–23. http://dx.doi.org/10.1017/S0047404500006837

Sidnell, J. (2010). *Conversation analysis: An introduction*. Wiley-Blackwell.

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing, 18*(4), 463–481. http://dx.doi.org/10.1191/026553201682430139

Speidel, G., & Nelson, K. (1989). *The many faces of imitation in language learning* (Vol. 24). Springer Science & Business Media. http://dx.doi.org/10.1007/978-1-4612-1011-5

Staples, S., Biber, D., & Reppen, R. (2018). Using corpus-based register analysis to explore the authenticity of high-stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. *The Modern Language Journal, 102*(2), 310–332. http://dx.doi.org/10.1111/modl.12465

Svalberg, A. (2009). Engagement with language: Interrogating a construct. *Language Awareness, 18*, 242–258. http://dx.doi.org/10.1080/09658410903197264

Taguchi, N. (2019). *The Routledge handbook of second language acquisition and pragmatics* (pp. 1–14). Routledge. http://dx.doi.org/10.4324/9781351164085

Taguchi, N., & Kim, Y. (2016). Collaborative dialogue in learning pragmatics: Pragmatic-related episodes as an opportunity for learning request-making. *Applied Linguistics, 37*(3), 416–437. http://dx.doi.org/10.1093/applin/amu039

Takahashi, S. (2010). Assessing learnability in second language pragmatics. In A. Trosborg (Ed.), *Pragmatics across languages and cultures* (pp. 391–421). De Gruyter Mouton. http://dx.doi.org/10.1515/9783110214444

Tannen, D. (1990). Gender differences in topical coherence: Creating involvement in best friends' talk. *Discourse processes, 13*(1), 73–90. http://dx.doi.org/10.1080/01638539009544747

Tantucci, V. (2016a). Textual factualization: The phenomenology of assertive reformulation and presupposition during a speech event. *Journal of Pragmatics*, *101*, 155-171.

Tantucci, V. (2016b). Toward a typology of constative speech acts: Actions beyond evidentiality, epistemic modality, and factuality. *Intercultural Pragmatics*, *13*(2), 181-209.

Tantucci, V. (2021). *Language and social minds: The semantics and pragmatics of intersubjectivity*. Cambridge University Press.

Tantucci, V. (2023). Resonance and recombinant creativity: Why they are important for research in Cognitive Linguistics and Pragmatics. *Intercultural Pragmatics*, *20*(4), 347–376. https://doi.org/10.1515/ip-2023-4001.

Tantucci, V. (In press). *Language and (Creative) Imitation: Dialogic resonance in Pragmatics and Grammar*. Cambridge University Press.

Tantucci, V., & Wang, A. (2021). Resonance and engagement through (dis-) agreement: Evidence of persistent constructional priming from Mandarin naturalistic interaction. *Journal of Pragmatics*, *175*, 94-111.

Tantucci, V., & Wang, A. (2022a). Resonance as an applied predictor of cross-cultural interaction: Constructional priming in Mandarin and American English interaction. *Applied Linguistics*, *43*(1), 115-146.

Tantucci, V., & Wang, A. (2022b). Dynamic resonance and explicit dialogic engagement in Mandarin first language acquisition. *Discourse Processes*, *59*(7), 553-574.

Tantucci, V., & Wang, A. (2023). Dialogic priming and dynamic resonance in autism: creativity competing with engagement in Chinese children with ASD. *Journal of autism and developmental disorders*, *53*(6), 2458-2474.

Tantucci, V., & Wang, A. (2024). British Conversation is Changing: Resonance and Engagement in the BNC1994 and the BNC2014. *Applied Linguistics*, amae040.

Tantucci, V., & Lepadat, C. (2024). Verbal engagement in doctor–patient interaction: Resonance in Western and Traditional Chinese Medicine. *Journal of Pragmatics*, *230*, 126-141.

Tantucci, V., Wang, A., & Culpeper, J. (2022). Reciprocity and epistemicity: On the (proto)social and cross-cultural 'value' of information transmission. *Journal of Pragmatics*, *194*, 54-70.

Tao, H., & Suzuki, R. (2022). The pragmatics of creative language use in East Asian languages: an introduction. *East Asian Pragmatics*, *7*(3), 297-305.

Tavakoli, P., & Foster, P. (2008). Task design and L2 performance. *Language Learning, 58*(2), 429–473.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child development*, *64*(6), 1688–1705. http://dx.doi.org/10.2307/1131463

Traugott, E., & Dasher, R. (2002). *Regularity in semantic change.* Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511486500

Ungerer, T., & Hartmann, S. (2023). *Constructionist approaches: Past, present, future*. Cambridge University Press. http://dx.doi.org/10.1017/9781009308717

van Lier, L., & Matsuo, N. (2000). Varieties of conversational experience: Looking for learning opportunities. *Applied Language Learning, 11*, 265–287.

Woodward-Kron, R., & Elder, C. (2016). A comparative discourse study of simulated clinical roleplays in two assessment contexts: Validating a specific purpose language test. *Language Testing, 33*(2), 251–270. https://doi.org/10.1177/0265532215607399

Young, R. (2008). Language and interaction. London, UK/New York, NY: Routledge.

Figures

Figure 1. Instructions about the Role-Play on the topic of part-time jobs (Ishikawa, 2023:53)

Figure 2. Instructions about for the Role-Play on the topic of smoking (Ishikawa, 2023:53)

Figure 3. Half-violin plots of resonance produced by Test-takers and Examiners

Figure 4. Predicted values of resonance per proficiency level and task type

Figure 5. Predicted resonance values per linguistic background

Figure 6. Predicted Test-takers' utterance length depending on Examiners' resonance

Notes

[1] The Object is semantically associated with some kind of flour: OBJ_{flour}.

[2] Australia, Canada, Ireland, New Zealand, the Philippines, the UK, and the US.

[3] The sample included 5 English, 4 Chinese, 4 Indonesian, 4 Japanese, 4 Thai, and 4 Taiwanese L1 speakers.

[4] PRAAT (Boersma & Weenink, 2023) is a software for speech processing. Spectrograms are visual representations of the frequencies (y axis) in a sound signal through time (x axis).

[5] Intonation was annotated to explore potential prosodic cues to resonance and engagement, especially in question–response sequences. However, it was excluded from the final models, as it did not improve model fit. We retained it for descriptive completeness and potential use in future multimodal analyses.

[6] This excludes items that are not ostensively recognisable as lexical repetitions, e.g. token with extremely high token frequency such as articles *the* or *a* which may occur alone (not as part of a larger construction) independently of dialogic engagement.

[7] Determiner.

[8] Personal Pronoun.

[9] The other Searlian category of declarations was absent from our data.

[10] Collinearity (VIF): All < 1.2 with no collinearity concerns. Residuals were normally distributed with slight skew. For Homoscedasticity, residual vs. fitted plot showed no major patterns.

[11] The dataset includes turns from speakers at six proficiency levels: A2 (n = 101), B1.1 (n = 136), B1.2 (n = 106), B2 (n = 117), L1 test-takers (n = 52), and Examiners (n = 42). In total, 317 turns were produced by test-takers and 237 by examiners. By task type, there were 293 Role play turns and 242 Q&A turns. The task topics were equally distributed: PTJ (n = 267) and SMK (n = 268).

[12] Intercept: Normal distribution with 0 mean and 5 standard deviation (normal(0, 5)). Slope: Normal distribution with 0 mean and 2 standard deviation (normal(0, 2)).