

## What Counts as Evidence of Pragmatic Transfer in University EFL? A Critical Narrative Review

Dwi Sloria Suharti <sup>\*(1)</sup>

Department of English Language Education  
[dwisloria@umt.ac.id](mailto:dwisloria@umt.ac.id)\*

Edi Ramawijaya Putra <sup>(2)</sup>

Institut Agama Buddha Negeri Sriwijaya  
[ediramawijayaputra@stabsn-sriwijaya.ac.id](mailto:ediramawijayaputra@stabsn-sriwijaya.ac.id)

### ABSTRACT

University learners of English as a foreign language (EFL) need to interpret interpersonal meanings, manage academic relationships, and adapt communication across institutional and intercultural encounters. Although existing syntheses establish that second-language pragmatics is teachable, a more specific question remains unresolved. When do measured improvements justify claims that learning extends beyond the instructional task? This critical narrative review evaluates evidence of learning transfer in university EFL pragmatics instruction. It distinguishes direct university EFL interventions from adjacent settings, non-intervention studies, and potentially overlapping research syntheses. The analysis separates gains in pragmalinguistic resources, sociopragmatic judgment, and interactional responsiveness from their application to untrained tasks or contexts. Retention after a delay is treated as a separate temporal dimension. Meta-analytic evidence supports instruction overall, but neither a universal advantage for explicit instruction nor automatic learning transfer from controlled assessments is established. University studies of e-mail instruction, oral feedback, computerized learning, and chatbot assistance show promising task-specific gains, while differences in comparison conditions and measurement constrain stronger claims. The review proposes an evidence-to-transfer framework linking identifiable learner needs to plausible instructional mechanisms, corresponding assessment indicators, and explicit boundary conditions. The principal contribution is a distinction among learning an assessed pragmatic target, retaining that learning, and applying it in new university situations. Longitudinal, cross-task, multi-institutional studies are needed to test these links.

**Keywords:** Interactional Competence; Interlanguage Pragmatics; Learning Transfer; Pragmatic Assessment; University EFL

### INTRODUCTION

Research on student–faculty e-mail requests and academic-discussion routines documents pragmatic demands relevant to university communication (Bardovi-Harlig et al., 2015; Chen, 2015; Economidou-Kogetsidis, 2011; Usó-Juan, 2022). These studies motivate the present review's focus on communication beyond grammatical correctness. Depending on institutional and disciplinary contexts, students may need to seek feedback from lecturers, negotiate deadlines, disagree with peers, or respond to criticism. Whether an utterance is successful depends partly on the linguistic resources selected, partly on how

institutional relationships are interpreted, and partly on how interlocutors respond as an encounter unfolds. Pragmatic competence therefore cannot be equated with a repertoire of polite expressions. It involves socially meaningful choice, contextual interpretation, and responsiveness to emerging conversational contingencies (Thomas, 1983; Taguchi & Ishihara, 2018).

The importance of teachable pragmatics is not in dispute. Taguchi (2015) reviewed 58 instructional studies, while Ren et al. (2023) synthesized 29 primary studies involving 1,898 learners and found a large positive pooled effect of instruction. Research has also moved beyond face-to-face classroom

activities into computer-mediated exchanges, web modules, and automated feedback (Yousefi & Nassaji, 2019; Alrefaee & Al-Shammari, 2026). These findings have encouraged proposals for integrating pragmatic awareness, explicit explanation, practice, and reflection into language curricula. However, a favorable result on a written discourse completion task, an e-mail judgment activity, or a trained speech act is not the same observation as successful negotiation in an unfamiliar conversation.

This distinction is consequential for university teaching. Students may learn to identify directness or add conventional mitigation to a written request while remaining unable to interpret a reluctant reply or reformulate the request without escalating interpersonal tension. Likewise, an academically appropriate e-mail is only one genre of communication; it does not establish transferable competence in seminar discussion, disciplinary collaboration, or intercultural problem-solving. The difference is not merely a question of whether the intervention is effective. It is a question of which construct has been measured, what process plausibly generated improvement, and how far the observed performance can be generalized (Roever, 2011; Youn, 2018).

Four major syntheses illustrate both the maturity of the field and the need for a narrower contribution. Taguchi (2015) maps instructed pragmatics; Ren et al. (2023) quantify and moderate instructional effects; Li et al. (2026) chart recent interlanguage pragmatic development in EFL contexts; and Alrefaee and Al-Shammari (2026) map digital innovations across second-language pragmatics. Building on these syntheses, this review asks a more specific analytic question: what evidence is necessary before a study of university EFL instruction can be said to demonstrate learning transfer, rather than success on the particular task used to teach or test pragmatics?

Answering that question requires attending to theoretical distinctions commonly blurred in pedagogical recommendations. Hymes (1972) situated communicative competence within socially appropriate language use. Canale and Swain (1980) subsequently offered an influential account of communicative competence, while Bachman (1990) distinguished components of communicative language ability relevant to use and assessment. Thomas (1983) separated pragmalinguistic from sociopragmatic sources of misunderstanding. These traditions help explain why knowledge of linguistic forms, understanding of social expectations, and demonstrated conversational adaptability may be related without being interchangeable. Interactional scholarship further emphasizes that linguistic action is collaboratively accomplished rather than fully specified before a speaker takes a turn (Kramsch, 1986; Hall, 1995; Young, 2011).

This conceptual clarification matters because instructional comparisons frequently aggregate or juxtapose different outcomes. Ren et al. (2023) found that explicit instruction generated numerically larger effects than implicit instruction, but the difference was not statistically significant. Their effect estimates also varied by assessment format. Classroom studies may report improvement in awareness, recognition, written production, or role-play performance, whereas longitudinal development and unprompted transfer are often less directly documented. Treating all such findings as equal evidence for a single general competence risks overstating what universities should expect from a particular pedagogy.

The review therefore uses university EFL as its primary population and differentiates three sources of warrant: setting-specific instructional evidence, adjacent or descriptive evidence that clarifies plausible mechanisms, and higher-level syntheses that inform the field without constituting independent replications

of the same primary studies. Taguchi (2011, 2026) likewise emphasizes the continuing importance of connecting pragmatic theory, developmental research, and pedagogical problems. This distinction also prevents apparent consensus from being manufactured by counting a primary article and the reviews that may already include it as separate confirmations.

In this review, learning transfer means applying pragmatic knowledge or performance developed through instruction to tasks or communicative contexts not directly practised. Near transfer concerns new scenarios that retain the trained speech act and genre; far transfer involves greater changes in genre, social relationship, or modality. These are operational distinctions relative to the instructional tasks, not fixed categories of activity. This use concerns transfer of learning rather than cross-linguistic influence. Retention concerns maintenance after a delay; a delayed score alone does not establish transfer.

This review aims to evaluate the evidence required to infer learning transfer from university EFL pragmatics instruction. It addresses three research questions: (RQ1) Which pragmatic outcomes have been demonstrated in university EFL interventions, and with what limits on causal interpretation? (RQ2) How do theories of pragmatic and interactional competence explain the gap between controlled task gains and adaptive communication? (RQ3) Which conditional instructional mechanisms and assessment designs would provide stronger evidence of near and far learning transfer? The resulting evidence-to-transfer decision matrix links learner difficulties, proposed instructional mechanisms, observable outcomes, and contextual limits on inference. Its pathways are selected according to learner needs rather than presented as a uniform sequence of activities.

## METHODOLOGY

### *Design and bounded evidence base*

This article is a critical narrative review, not a systematic review, scoping review, or meta-analysis. Its primary purpose is conceptual comparison and interpretation of existing evidence relevant to university EFL pedagogy. The initial citation-led corpus comprised 34 distinct works spanning foundational theory and empirical or review scholarship published through 2026. Six additional foundational sources were incorporated to address gaps in communicative competence and interactional theory: Hymes (1972), Bachman (1990), Kramsch (1986), Hall (1995), Galaczi (2008), and Young (2011). The resulting 40-work bibliography should be understood as a bounded, purposive corpus, not the denominator of an exhaustive retrieval process.

The corpus was assembled through targeted scholarly searches and bibliographic tracing rather than an exhaustive, logged database search; complete search logs, original search dates, retrieval totals, and deduplication records are unavailable. Targeted bibliographic checks clarified selected source attributions but did not constitute a systematic update.

The thematic search expressions used in the targeted checks concerned 'second language pragmatics instruction meta-analysis', 'university EFL pragmatic instruction e-mail requests', 'chatbot feedback pragmalinguistic sociopragmatic', 'interactional competence language teaching and testing', and named foundational scholars.

### *Analytical distinctions and comparative logic*

Evidence was compared using five coding questions. First, is the population demonstrably university EFL, a different educational or L2 setting, or unspecified? Second, is the source a controlled intervention, quasi-experiment, observational study,

conceptual argument, or research synthesis? Third, what is the outcome actually elicited: linguistic strategy knowledge, contextual judgment, performance across conversational turns, or application to untrained tasks or contexts, and is retention assessed separately after a delay? Fourth, what comparison permits the observed change to be attributed to instruction, a particular instructional component, or only to time and exposure? Finally, could a cited primary study also form part of a cited synthesis, limiting evidentiary independence? This approach is qualitative and critical rather than a formal risk-of-bias instrument.

Claims in the following sections are accordingly marked by their inferential reach. 'Direct' evidence refers to an intervention with identified university EFL learners and an identifiable outcome measure. 'Adjacent' evidence concerns other populations, observational designs, or work that illuminates a construct without establishing a university EFL treatment effect. 'Synthesis-level' evidence summarizes bodies of research across varying contexts. Such labels communicate

population and design relevance; they do not certify that a particular study is methodologically flawless. No new pooled effects, study-quality scores, or full overlap matrices are calculated.

Dependence among reviews and what the corpus cannot establish

Taguchi (2015), Yousefi and Nassaji (2019), Ren et al. (2023), Li et al. (2026), and Alrefae and Al-Shammari (2026) address overlapping subject matter but differ in periods, eligibility rules, technologies, and outcomes. Inclusion of both a primary article and one or more research syntheses does not create new independent evidence. In particular, the 29 studies in Ren et al. (2023), 56 in Li et al. (2026), and 118 in Alrefae and Al-Shammari (2026) must not be added together as if mutually exclusive; the extent of shared primary studies has not been checked study by study here. We use these syntheses to describe research patterns, and selected primary studies to inspect design and construct details, without treating their convergence as an independently replicated causal effect.

**Table 1**

*Functions and independence limits of major research syntheses*

| Research synthesis           | Coverage and principal value                                                            | Limit on independent inference                                                                                 |
|------------------------------|-----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Taguchi (2015)               | 58 instructional studies; overview of teachability and instructional approaches.        | May include primary investigations subsequently cited in other reviews; not a university-only dataset.         |
| Yousefi & Nassaji (2019)     | 39 studies (2006–2016); instructional and feedback moderators, including delivery mode. | Cross-setting evidence; possible shared primary studies with older instructional reviews.                      |
| Ren et al. (2023)            | 29 studies; 1,898 learners; 54 between-group effects; pooled $g = 1.656$ .              | Heterogeneous outcomes; explicit–implicit difference nonsignificant; not a university-specific estimate.       |
| Li et al. (2026)             | 56 empirical studies (2021–2025); recent EFL development and contextual moderators.     | Scoping map, not a university-intervention meta-analysis; potential overlap with contemporary primary studies. |
| Alrefae & Al-Shammari (2026) | 118 studies (2006–2024); historical and thematic map of digital L2 pragmatics.          | Technology and L2 scope broader than university EFL; potential overlap with other reviews.                     |

*Note.* Study counts originate from the individual publications. Their samples can overlap and must not be summed.

The bounded source set introduces selection risk: relevant negative, unpublished, or institution-specific work may be absent. Several included sources also provide limited design information in the accessible record. Where population or design details cannot be established, conclusions remain restricted to the documented setting and outcome. The corpus therefore cannot support estimates of the global prevalence of instructional approaches.

## RESULTS AND DISCUSSION

### *Theoretical Architecture: Knowledge, Judgment, and Interaction*

#### *Communicative competence and pragmatic knowledge*

Hymes (1972) challenged the idea that knowing a language consists only of grammatically possible utterances; communicative capacity also concerns the circumstances in which forms are feasible, appropriate, and used. Canale and Swain's (1980) model broadened the language-teaching perspective to multiple components of competence. Bachman's (1990) account of communicative language ability, including pragmatic aspects of language use, brings the question of construct definition directly into assessment: a score is meaningful only to the extent that the task elicits the ability being claimed. This lineage does not imply that 'pragmatic competence' has one fixed empirical manifestation.

Thomas's (1983) distinction remains useful for instruction and assessment. Pragmalinguistics concerns the resources available to enact a function—for instance, request formulations, mitigators, stance markers, and routines. Sociopragmatics concerns how participants understand relational distance, authority, imposition, genre, and local expectations. A grammatically well-formed request may fail because of mismatched expectations even when the learner knows how to make requests;

conversely, a learner may perceive a situation sensitively while lacking expressions for acting on that judgment. Explicitly separating these difficulties guards against designing identical lessons for qualitatively different needs.

A third dimension is metapragmatic awareness: the learner's ability to notice and explain why expressions have different interpersonal implications (Ishihara & Cohen, 2010; Bardovi-Harlig, 2013). Reflective judgments are valuable outcomes in their own right, particularly when intercultural communication offers more than one defensible way to act. However, a correct verbal explanation does not prove that a learner can select and revise language under real-time processing constraints. The validity of conclusions depends on resisting that inferential leap.

#### *From isolated speech acts to co-constructed interaction*

Teaching requests, apologies, compliments, and disagreement remains practical because these actions can be observed and compared. Yet a single turn cannot capture the whole of interactional competence. Kramsch (1986) argued for shifting attention from static proficiency toward meaning constructed with others; Hall (1995) showed how classroom interactional routines can differ from social practices outside the classroom. Young (2011) subsequently conceptualized interactional competence in terms of resources deployed jointly to accomplish discursive practices. These perspectives identify turn-taking, topic management, repair, and adaptation as constituents of performance rather than accessories to linguistic correctness.

Research on assessment supports this distinction. Galaczi (2008), using paired speaking-test interactions, identified differing collaborative, parallel, and asymmetric interaction patterns, showing that observable competence depends in part on how interaction is jointly organized. Youn (2018) similarly

treats interactional pragmatics assessment as a task-design and validation problem. Although these are not causal studies of university EFL instructional effectiveness, they justify measuring the response to another speaker instead of only the opening move.

Taguchi and Barón (2024) provide a more directly relevant demonstration of the adaptation construct: 65 L2 English users in Spain performed video-based requests or apologies with differing simulated listener reactions, and strategy adjustments were associated with proficiency and experience. Because this was cross-sectional and used prerecorded responses, it does not show that a proposed teaching sequence creates flexibility in university students. Its value lies in specifying an observable behavioral target that traditional isolated-item tests may miss: an appropriate change of strategy after social feedback.

### ***Intercultural variation and learner agency***

The criterion of appropriateness is itself situated. Thomas (1983) drew attention to pragmatic failure across expectations, but more recent work questions using idealized native-speaker norms as the single benchmark for multilingual encounters. Taguchi and Ishihara (2018) emphasize negotiation in English as a lingua franca, while McConachy (2019) foregrounds intercultural interpretation and critical awareness. A lecturer's expectations can be institutionally consequential without becoming an interculturally universal rule.

Consequently, pragmatic learning should not be equated with ever-increasing indirectness or unqualified compliance with one local norm. A student can reasonably choose directness to convey urgency or ethical dissent if the choice is understood, justified, and interactionally managed. The pedagogical goal is not norm-free relativism either: learners need information about foreseeable consequences of specific forms in specific communities. The

unresolved issue is how to observe both informed choice and communicative success without masking power imbalances or penalizing legitimate variation. This issue shapes the framework and measurement proposals below.

### **What Is Specific to University EFL Communication?**

#### ***Academic e-mail as a well-specified but narrow domain***

Student-faculty correspondence provides a relatively clear university task, and several sources address it directly. Economidou-Kogetsidis (2011) documented potentially problematic patterns of directness, address, openings, and closings in university e-mail requests within a particular institutional context. That study is descriptive: it identifies situated communication issues but does not estimate a teaching effect. Chen (2015) reported improvements in aspects of Chinese EFL students' e-mail request literacy following instruction, while Usó-Juan (2022) examined naturally occurring e-mail requests before and after strategy instruction among Spanish EFL learners. These findings support the educational relevance of contextualized e-mail writing, but the strength of causal claims differs by comparison design.

More recently, López-Serrano (2025) found improved appropriateness judgments and explanation of problems among 70 first-year university EFL learners after a self-access web module. López-Serrano et al. (2025) reported gains in writing performance and confidence among 68 first-year Spanish university students. These articles address related pedagogical environments and should not be assumed to represent wholly independent evidence or distinct intervention mechanisms without checking participant and module overlap. Their measures also differ: noticing problematic e-mails, producing a better e-mail, and confidently evaluating an e-mail are related but non-identical outcomes. None

alone establishes spontaneous oral adaptation in a seminar or research meeting.

### ***Discussion, requests, and oral feedback***

University oral communication may require disagreement, clarification, feedback negotiation, and sequential management. Bardovi-Harlig et al. (2015) examined 37 ESL learners in an intensive English program and a university academic-discussion course. Their findings support corpus-informed instruction for discussion routines, but the ESL setting makes the study adjacent rather than direct university EFL evidence. Nguyen et al. (2012), by contrast, studied 69 third-year English-major pre-service teachers in Vietnam; this study is classified here as direct university EFL evidence. Its constructive-criticism outcomes were assessed through discourse completion, role-play, and oral peer-feedback tasks; delayed testing examined retention. Alcón Soler (2005) provides further mechanism-relevant evidence on requests. These studies address specific pragmatic targets rather than establishing cross-genre learning transfer.

Another direct university EFL intervention is Pham et al. (2024), which involved 122 Vietnamese university learners receiving explicit instruction on English requests while groups differed in oral corrective feedback. Gains in recognition and production were reported, including at delayed assessment, but differences among feedback-type groups were not significant. Because the groups shared explicit instruction and the pragmatic target was requests, the study does not demonstrate one feedback strategy's universal advantage or transfer across discourse genres. Its importance is the careful separation of a common instructional effect from a differential component effect.

### ***Population boundaries and curricular inference***

A key problem in reviews of 'university EFL' is the silent promotion of adult L2, secondary-school, and non-intervention results into

university recommendations. Siddiqi and Whyte (2025), for example, investigate classroom input and interactional opportunities in French secondary education, not university EFL. Taguchi (2022) investigates effects of task medium on pragmatic performance among native and nonnative speakers, not a long-term instructional treatment. Taguchi and Barón (2024) address language-use experience and adaptive performance, not a university intervention. Such studies can inform task design or construct definition, but their findings cannot be read as university-specific treatment estimates.

Disciplinary and institutional variation adds further limits. Requests to a lecturer, disagreement in a medical simulation, critique in a literature seminar, and collaboration on an engineering project do not share identical participation structures. A general course can introduce transferable interpretive resources, but which transfer claims are warranted depends on whether assessment spans genres, partners, and communicative contingencies. University-specific guidance must therefore identify the communicative task and the expected outcome rather than invoking 'higher education' as a homogeneous context.

### ***Critical Comparison of Instructional Effectiveness***

#### ***Overall instructional effects: strong average, heterogeneous construct***

The most robust quantitative conclusion available in the cited evidence is that instruction often improves assessed pragmatic outcomes. Ren et al. (2023) reported Hedges'  $g = 1.656$  (95% confidence interval [1.387, 1.924]) across 29 primary studies and 54 between-group effects. Yet this number is an average across different learners, instructional targets, durations, and instruments. It should not be presented as the expected effect of an integrated university curriculum, nor as a directly transferable prediction for an untested model.

The measurement moderator is especially instructive. Ren et al. (2023) found the largest effects for written discourse completion tasks, followed by oral production and multiple-choice measures. A written task can make newly taught forms readily available and reduce the uncertainty of negotiating an actual interlocutor response. This does not invalidate the observed learning. Instead, it indicates that the measured gain may concern controlled production under particular elicitation conditions. The difference between 'a learner has acquired a trained response' and 'a learner can repair an unfamiliar interaction' is an empirical question, not a rhetorical preference.

The meta-analysis also reported larger effects with longer treatments, in foreign-language rather than second-language settings, and across particular educational levels; the high-school category produced the largest average educational-level effect. It would therefore be incorrect to cite the overall pooled result as special evidence that university learners necessarily benefit most. Taguchi (2015) and Yousefi and Nassaji (2019) likewise establish a broader instructional field but not an independent replication of every primary study in Ren et al. (2023). Their different synthesis windows and eligibility rules make direct comparison informative but not additive.

***Explicit versus implicit instruction: a conditional hypothesis, not a winner***

Alcón Soler (2005) and Nguyen et al. (2012) offer evidence consistent with benefits from explicit form-focused or metapragmatic instruction for selected targets. Takimoto (2008), however, found that both deductive and inductive approaches can generate pragmatic learning when they direct attention to relevant features. Crucially, Ren et al. (2023) found an explicit-instruction advantage that was numerical rather than statistically significant in the pooled comparison. The defensible claim is thus not 'explicit teaching is always best', but that instructional effectiveness depends on which learner

difficulties are addressed and how the comparison is constructed.

Explicit explanation might be particularly useful when a learner cannot identify contextual cues or lacks metalanguage for analyzing an interaction. Repeated exposure and inductive comparison might be useful when learners need to discriminate patterns in varied input. Interactional practice might be required when the challenge is retrieving and adapting already known resources under conversational pressure. These are mechanism proposals that follow from construct distinctions, not treatment rankings established by head-to-head university EFL trials. Future studies must match dosage, practice opportunities, input quality, and assessment tasks before attributing differences to 'explicitness' itself.

***Feedback, tasks, and mediating conditions***

Task-based approaches make communicative goals and contextual consequences salient. Taguchi and Kim (2018) provide a foundation for task-based teaching and assessment, while Ishihara (2010) stresses the interface between research and teacher development. A role-play may simply rehearse a script or may require contingency-sensitive responses, depending on how the interlocutor behaves. This distinction makes task design more informative than labels such as 'authentic', 'communicative', or 'interactive'.

Feedback research also cautions against isolating a visible classroom feature from the treatment package. Pham et al. (2024) varied feedback forms within shared explicit instruction and did not find significant differences between feedback conditions. In an AI-mediated setting, Ajabshir (2025) compared teacher and chatbot feedback with 87 university EFL learners; both treatment groups improved, chatbot feedback yielded stronger pragmalinguistic gains, and sociopragmatic differences were not significant. Such studies can support particular

affordances, but they do not warrant a universal ordering of feedback types. The outcome instrument, targeted speech act,

intervention length, and common instructional components remain alternative explanations.

**Table 2**

*Illustrative evidence map: population, outcome, and inference ceiling*

| Study / evidence tier                   | Population and design                                                                                             | Outcome the finding can support                                                                           | Primary inference limit                                                                      |
|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Nguyen et al. (2012) · direct           | 69 third-year English-major pre-service teachers in Vietnam; explicit, implicit, and control conditions.          | Constructive-criticism production across DCT, role-play, and oral peer-feedback tasks; delayed retention. | Single pragmatic target; delayed repeated tasks examine retention, not cross-genre transfer. |
| Chen (2015) · direct                    | 28 Chinese EFL learners; e-mail request instruction.                                                              | Selected dimensions of student–faculty e-mail literacy.                                                   | Small genre-specific study; no warrant for general oral transfer.                            |
| Usó-Juan (2022) · direct                | Spanish EFL e-mail strategy instruction; 110 authentic e-mails analyzed.                                          | Changes in contextualized written request strategies.                                                     | E-mail-specific evidence; causal interpretation depends on comparison structure.             |
| Zhang (2021) · direct                   | 59 Chinese university EFL learners; CMC with/without data-driven instruction.                                     | Improved compliment-response performance with guidance.                                                   | Speech-act-specific; does not test new academic genres.                                      |
| Zhong & Zhang (2023) · direct           | 180 medical-university EFL learners; mediated vs conventional explicit vs control.                                | Tested recognition and production of selected speech acts.                                                | Delivery components and test format may contribute to group differences.                     |
| Pham et al. (2024) · direct             | 122 Vietnamese university EFL learners; shared explicit instruction, four feedback conditions.                    | Request recognition/production gains and delayed maintenance.                                             | Feedback-type contrasts nonsignificant; not feedback-specific superiority.                   |
| Ajabshir (2025) · direct                | 87 university EFL learners; teacher vs chatbot feedback vs control.                                               | Pragmalinguistic gains; no significant differential sociopragmatic gain.                                  | Five-week request/DCT focus; no human–interaction transfer outcome.                          |
| López-Serrano (2025) · direct           | 70 first-year Spanish university EFL learners; web self-access module.                                            | E-mail appropriateness judgments and awareness.                                                           | Judgment is not production or spontaneous interaction.                                       |
| López-Serrano et al. (2025) · direct    | 68 first-year Spanish university EFL learners; five-week web module.                                              | E-mail production and confidence.                                                                         | No untreated control; potential relatedness to other module studies.                         |
| Bardovi-Harlig et al. (2015) · adjacent | 37 ESL learners in an intensive English program and a university academic-discussion course; routine instruction. | Possible value of corpus-informed discussion input.                                                       | ESL rather than EFL setting; discussion-task gains do not establish cross-genre transfer.    |
| Taguchi & Barón (2024) · adjacent       | 65 L2 users in Spain; cross-sectional simulated reactions.                                                        | Language adaptation associated with experience/proficiency.                                               | No instructional causal effect or university-only population.                                |
| Taguchi (2022) · adjacent               | 62 native/nonnative English speakers; task-medium comparison.                                                     | Assessment setting can alter pragmatic performance.                                                       | Medium effect is not a long-term treatment effect.                                           |

*Note.* “Direct” indicates a university EFL intervention; “adjacent” indicates broader populations, observational designs, or methodological evidence. Classifications are not formal risk-of-bias scores.

The studies support targeted learning under instructed conditions more securely than they support the generalization of that learning across genres and interlocutors. Stronger conclusions require observing an outcome beyond the trained task, documenting whether the comparison condition controlled for time and practice, and distinguishing maintenance of a trained response from transfer to a substantively new communicative demand.

### ***Digital Instruction: Affordance Is Not Mechanism***

Digital delivery matters in university EFL education because learners may have limited exposure to varied English-speaking academic interactions outside the classroom. Computer-mediated communication can provide repeatable input, asynchronous reflection, interaction records, and access to feedback that face-to-face teaching cannot always supply at comparable frequency. Yet an online environment is not a single pedagogical treatment. It may provide authentic peer interaction, rule-based drills, self-access explanations, teacher dialogue, automated scoring, or a combination of these mechanisms. Therefore, treating 'technology' as one explanatory variable obscures what learners actually do.

Yousefi and Nassaji (2019), in a meta-analysis of 39 studies, considered the role of instructional mode and related moderators in second-language pragmatics. Their findings are useful for considering computer-mediated affordances, but that review involves varied learner populations and a synthesis window that may overlap with other instructional surveys. It cannot by itself establish the superiority of computer-delivered pedagogy for contemporary universities. Zhang (2021), studying 59 Chinese university EFL learners, found that combining computer-mediated communication with data-driven instruction supported compliment-response performance

more than communication alone. This is more theoretically informative than a simple 'online versus offline' comparison because it suggests that input analysis and guided attention may constitute an active ingredient.

Zhong and Zhang (2023) investigated 180 learners recruited from a Chinese medical university and compared computer-mediated explicit instruction, conventional explicit instruction, and a control condition. Their findings favored the computer-mediated condition for the assessed speech acts. However, outcomes were measured as speech-act recognition and production; institutional specificity, engagement, interface design, and measurement fit remain relevant to generalization. A medical-university sample does not establish that the same approach will transfer to negotiation with a supervisor or intercultural collaboration in another discipline.

The web-based e-mail studies offer a similarly useful caution. López-Serrano (2025) assessed awareness and appropriateness judgments, whereas López-Serrano et al. (2025) assessed production and confidence in a related self-access instructional line. They show that a well-scaffolded digital module can be embedded within first-year university teaching and be associated with measurable improvements. Yet the later production study had no untreated comparison group, and neither design measured naturally occurring negotiation across unrelated discourse genres. It is also important to examine possible shared instructional materials or learner pools before regarding the two studies as independent demonstrations of a single technology effect.

Immersive environments raise a different inferential issue. Taguchi (2022) compared pragmatic behavior in computer-based and virtual-reality role-play conditions. Because task medium influenced aspects of

performance, the study primarily informs measurement and ecological representation rather than proving that virtual reality instruction produces durable competence. The simulated environment can alter what is elicited, which means differences between a virtual task and a written measure may reflect task demands as much as instruction.

Generative artificial intelligence adds adaptive-looking conversations and immediate textual feedback, but fluency of output is not a guarantee of valid sociopragmatic interpretation. Ajabshir (2025) reports promising chatbot-supported pragmalinguistic development, with no statistically significant chatbot advantage for sociopragmatic gains. Alrefaee and Al-Shammari's (2026) synthesis maps the rapid diversification of digital pragmatics research, but broad adoption trends should not be confused with established developmental mechanisms. Chatbot feedback may overgeneralize one politeness convention, suppress context-specific identities, or misinterpret interpersonal intent. As Ishihara (2010) and Taguchi and Ishihara (2018) imply for teacher mediation and intercultural choice, feedback needs contextual justification and an opportunity for learner disagreement, not merely automated correction. Privacy, retention of student communications, and errors in generated examples require institutional oversight.

A critical distinction therefore emerges between delivery affordances (access, immediacy, repeatability), learning processes (attention, comparison, strategic trial, contingent negotiation), and targeted outcomes (form choice, judgment, responsiveness, transfer). Studies that manipulate delivery while leaving the processes unspecified offer weaker explanations than studies that measure which process changes behavior. The framework below consequently treats technology as a possible means of activating a mechanism, rather than as a learning mechanism in itself.

## **Assessment Validity and the Missing Transfer Test**

### ***What common measures can and cannot claim***

Assessment is central to the review's argument because pedagogical effectiveness is inseparable from the definition of achievement. Bachman (1990) established the importance of specifying communicative ability and validity arguments, while Roever (2011) warned against restricting the construct of second-language pragmatics to isolated speech acts. A written discourse completion task (DCT) can show that a student can formulate a request in response to a standardized prompt. A contextual judgment test can show that the student recognizes problems in a given exchange. Neither alone establishes that the learner can manage competing interpersonal goals when an interlocutor changes stance mid-conversation.

Ren et al. (2023) make the threat to overgeneralization visible: effects were largest for written DCT outcomes. An intervention may directly teach the same linguistic modifications or decision principles needed to complete that measure. A large score difference can therefore be compatible with modest improvement in rapid, multi-turn communication. This is not an argument for abandoning standardized tests; they provide controlled comparisons and diagnostic information. The problem lies in asserting a broader construct than the method samples.

A classroom role-play also has limits. It elicits prosody, timing, negotiation, and repair that written prompts usually cannot show, but it remains an elicited situation influenced by roles, partner familiarity, and the authenticity of the stakes. Youn (2018) shows why pragmatic interaction assessments require carefully justified task design. Galaczi (2008) demonstrates that co-participant behavior shapes the interaction one can observe and score. Assessors should therefore record interlocutor contingencies and consider paired-

performance dependence rather than attributing every turn to one individual's stable trait.

### ***Pragmatic outcomes, learning transfer, and retention***

For analytic clarity, this review distinguishes three pragmatic performance domains and asks whether improvement in each transfers to untrained tasks. Pragmalinguistic repertoire concerns the diversity, availability, and functional accuracy of expressions used for a communicative goal; relevant indicators include varied request heads, mitigation, stance, and framing moves. Sociopragmatic judgment concerns reasoning about social distance, rights, obligation, power, stakes, genre, and cultural variation; indicators include context-sensitive explanations and recognition of alternative defensible strategies. Interactional responsiveness concerns modifying action after a partner's move; indicators include relevant repair, strategic reformulation, acknowledgment of resistance, and participation in multi-turn negotiation. Learning transfer requires evidence of applying these resources to untrained scenarios or contexts. Retention concerns maintenance after a delay and can be assessed on trained or untrained tasks.

These domains need not develop in a fixed order. Learners may demonstrate competent repair with familiar peers before they can explain the strategy metapragmatically; another learner may write nuanced explanations but hesitate in real-time speech. Scores should thus be reported as a profile rather than compressed into a single 'pragmatic competence' index without validation. Sociopragmatic measures should permit justified variation instead of always equating closeness to one lecturer's preference with universal correctness (McConachy, 2019; Taguchi & Ishihara, 2018).

A stronger university intervention design can compare trained targets with untrained tasks at both immediate and delayed assessment. After

instruction on e-mailing a lecturer about an extension, a new e-mail request about feedback within a comparable student–lecturer relationship can serve as a near-transfer probe. A live negotiation with a hesitant supervisor can probe farther transfer by changing modality and interactional demands. The task differences must be specified relative to instruction. Delayed assessment of the trained target tests retention, whereas delayed untrained tasks examine retention of transferred learning. Independent raters can code both locally appropriate forms and turn-contingent responses; rating rubrics should make clear whether they reward convention, intelligibility, achievement of the goal, or effective negotiation. Test tasks must not require cultural assimilation as a hidden scoring criterion.

### ***Measurement as an explanation for conflicting findings***

Apparently conflicting instructional results may reflect different constructs rather than contradictions about learning. A DCT-based study can favor an explicit form-focused intervention because the outcome makes the targeted form salient, while another study may favor task interaction for responsive role-play performance. Similarly, chatbot assistance may improve predictable linguistic formulations without showing corresponding improvement in complex situational judgment (Ajabshir, 2025). Rather than averaging away these differences, the interpretive task is to map treatment ingredients to what the test actually requires. The causal question becomes whether a change persists when superficial task cues, partners, or demands change.

### ***Integrative Argument: Why Learning Gains and Adaptive Competence Diverge***

#### ***Three distinctions that organize the evidence***

The literature is often summarized as agreement that pragmatics is teachable, contextualized instruction is desirable, and

feedback helps. These statements are broadly plausible but analytically thin. A more informative synthesis distinguishes availability from selection, selection from enactment, and enactment from transfer. Availability concerns whether the learner possesses usable expressions. Selection concerns whether the learner can judge their implications in a particular context. Enactment concerns whether selected actions can be managed sequentially with a partner. Learning transfer concerns application under changed circumstances; retention concerns maintenance over time. None of these entailments should be assumed merely because the preceding level is observed.

The first distinction is reflected in Li et al. (2026), whose scoping review reports more consistent attention to pragmalinguistic development than to sociopragmatic and intercultural dimensions. The second appears in interactional scholarship: Kramsch (1986), Hall (1995), and Young (2011) explain why communication is jointly produced and contingent on a partner. The third follows from construct-validity analyses and the medium-sensitive findings in Taguchi (2022). These strands converge conceptually, but they do not constitute three independent university EFL intervention replications. Their value is explanatory alignment, not a numerical accumulation of support.

### ***Contrasts that delimit causal and comparative claims***

The most revealing contrasts are not between technologies as product categories but between designs and outcomes. Alcón Soler (2005) and Nguyen et al. (2012) are compatible with explicit explanation being beneficial for selected pragmatic targets, while Ren et al. (2023) do not establish its universal superiority. Pham et al. (2024) show gains under shared instruction but no reliable differences between compared oral feedback types; Ajabshir (2025) suggests the relative value of chatbot feedback may differ between

pragmalinguistic and sociopragmatic outcomes. These do not resolve into a single ranking because learner populations, treatment combinations, pragmatic targets, and tests differ.

The university e-mail findings offer a second contrast. López-Serrano (2025) identifies gains in evaluative awareness, while López-Serrano et al. (2025) report gains in production and confidence. Even when both results are valuable, they are not interchangeable measures, and possible relatedness of intervention materials constrains claims of independent convergence. A third contrast concerns digitally mediated language-use settings: Zhang (2021) suggests that instructional guidance adds value to online communication, whereas Taguchi (2022) primarily shows that virtual environments affect elicited performance. The former concerns an instructional comparison; the latter concerns measurement conditions. Placing them in the same 'technology works' category erases the distinction.

### ***The inference chain as the review's conceptual contribution***

The central inference chain can be expressed as instructional opportunity → proposed learning mechanism → proximal outcome → application to untrained tasks or contexts. Retention is evaluated separately by repeating assessment after a delay. The first link requires specifying what the learner encounters and does, not simply the mode of delivery. The second requires a plausible account of noticing, form-function mapping, strategic retrieval, or feedback-based adaptation. The third requires an instrument genuinely aligned with that process. The final link requires explicit task contrasts and a fair comparison to support claims of near or far learning transfer.

The resulting critical claim is not that existing interventions fail, nor that every curriculum requires maximal realism. It is that an instructional result should be described at the

strongest level justified by its design and no higher. A favorable e-mail judgment test legitimately supports awareness for that e-mail domain. A multi-turn role-play supports task-elicited responsiveness. An untrained interaction can provide evidence of learning transfer; a delayed version can additionally assess its retention. Preserving these distinctions makes the review more useful to curriculum designers and researchers than merely combining recommendations for exposure, explanation, practice, and assessment.

### **An Evidence-to-Transfer Framework for University EFL Pragmatics**

#### ***Rationale and status of the proposal***

The evidence-to-transfer framework is a conditional matrix connecting a diagnosed learner difficulty, a proposed instructional mechanism, an appropriate evidence source, and the boundary of the inference. Its contribution is the explicit alignment of instructional claims with construct, comparison, and learning-transfer requirements within university EFL communication. The framework is an analytic and research-design proposition, not an empirically validated instructional model.

The framework treats instructional components as potentially iterative and selectively combined. In an e-mail-writing class, learners may first need more linguistic options and later encounter unexpected institutional constraints. In a discussion seminar, learners may already know appropriate disagreement formulations but lack experience responding to opposition. A single fixed sequence would treat both classes identically; a conditional model prompts instructors to identify what is missing before selecting a task.

#### ***Four mechanism–outcome pathways***

A context discrimination pathway addresses learners who can produce expressions but misread social stakes. Its proposed mechanism

is comparison of variants across changed relationships, levels of imposition, genres, and interlocutor responses. Predicted proximal changes include more differentiated appropriateness explanations and improved recognition of consequences. Such progress would not by itself establish fluent conversation. Teacher mediation remains necessary to avoid converting local conventions into universal cultural rules (Thomas, 1983; McConachy, 2019).

A form–function repertoire pathway targets learners whose awareness exceeds their expressive resources. Corpus samples, contrastive exemplars, deliberate retrieval, and explanation may expand linguistic options for requests, feedback, and disagreement. Predicted proximal changes include appropriately varied heads, modifiers, sequencing, and framing in fresh but comparable prompts. This pathway is consistent with research on corpus-informed routines and instructed forms (Bardovi-Harlig et al., 2015; Alcón Soler, 2005).

An interactional contingency pathway targets learners who perform effectively when a situation is fully specified but struggle after a partner resists, misunderstands, or reinterprets a move. Tasks should vary interlocutor reactions, require multi-turn continuation, and support analysis of turn-by-turn decisions. Predicted outcomes include appropriate repair, strategy switching, explicit clarification, and negotiated resolution, evaluated against more than one interlocutor script. This pathway is conceptually informed by interactional competence and adaptation research (Young, 2011; Taguchi & Barón, 2024).

An intercultural choice and transfer pathway targets narrow reliance on one prescribed standard. Learners compare plausible norms, justify choices, and conduct tasks with

unfamiliar partners or social identities. The proposed mechanism is flexible goal-oriented reasoning rather than memorization of a preferred native-speaker response. Observable

outcomes include defensible alternative strategies, evidence-sensitive revision after misunderstanding, and successful cross-context communication. However, the judgment of 'success' must be explicitly

negotiated and independently rated to avoid rewarding conformity in the name of flexibility (Taguchi & Ishihara, 2018; McConachy, 2019).

**Table 3**

*Evidence-to-transfer decision matrix*

| <b>Diagnosed need &amp; hypothesized mechanism</b>                                             | <b>Proximal indicator</b>                                                           | <b>Transfer probe / timing</b>                                                                    | <b>Scope and testable boundary</b>                                                         |
|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Context misinterpretation → contrast contextual variants, guided comparison and justification. | Distinguishes authority, imposition and genre effects with reasoned alternatives.   | New university scenario varying roles and communicative stakes; immediate and delayed assessment. | Judgment gains do not imply oral fluency; norms require local/intercultural justification. |
| Limited expressive options → form–function mapping, exemplars and retrieval.                   | Broader functional repertoire; suitable request, stance and mitigation choices.     | Fresh genre-matched production task without taught wording; immediate and delayed assessment.     | May not help learners who already possess the forms but cannot respond in real time.       |
| Weak turn-contingent response → variable partner reactions, repair and feedback.               | Relevant reformulation, clarification and repair after changed interlocutor stance. | Unrehearsed multi-turn interaction with an unfamiliar partner; immediate and delayed assessment.  | Interactional scores depend on interlocutor contribution and task realism.                 |
| Rigid normative rule → intercultural comparison, agency and negotiation.                       | Explains alternative defensible choices and consequences.                           | New intercultural scenario; immediate and delayed reflection and performance.                     | Do not equate success with conformity to a single native-speaker norm.                     |

*Note.* The rows are conditional pathways, not chronological stages. Probes distinguish changes in task or context from the timing of assessment.

***Illustrative application, indicators, and falsifiable expectations***

Consider a first-year university EFL course in which students regularly e-mail faculty and participate in project meetings. Diagnostic tasks first distinguish limited linguistic repertoire from contextual misinterpretation or poor conversational follow-through. Students who cannot phrase a request plausibly may need corpus-based comparison and short, targeted production practice. Students who can write a suitable opening but do not respond to a lecturer's refusal may need contingently scripted role-plays and guided analysis of sequential choices. Students who consistently mark a direct strategy as 'wrong' regardless of context may need intercultural cases that permit reasoned alternatives. The diagnostic

profile guides the selection of activities for each learner need.

For measurement, learners can complete a baseline e-mail production task and video-supported role-play, followed by trained, near-transfer, and farther-transfer tasks at immediate and delayed assessment. Near-transfer tasks retain the trained speech act and genre while changing scenario details; farther-transfer tasks change genre, social relationship, or modality more substantially. Linguistic range, contextual rationale, response contingency, and interactional outcome should be separately rated. Evidence for the form–function pathway would consist of improved expressive options; evidence for the contingency pathway would require relevant revisions after an unanticipated partner move. Performance on untrained tasks would support

transfer, while maintenance after a delay would support retention.

A falsifiable prediction is that an intervention explicitly manipulating interlocutor contingencies, with instructional time and pragmatic targets matched, should yield better performance on untrained contingent interactions than otherwise comparable single-turn rehearsal, even if the two conditions produce similar written DCT scores. Another is that contextual comparison should improve justification accuracy more directly than uninstructed exposure when learners' initial difficulty is sociopragmatic discrimination. Null results would require revising the corresponding mechanism assumptions.

### ***Boundary conditions, equity, and implementation requirements***

The framework is most directly relevant to university EFL programs with identifiable academic genres, a feasible source of contextualized input, and a way to assess both immediate and changed-task performance. Low proficiency may limit the expressiveness of any adaptive task; high proficiency may create ceiling effects on routine judgments. Disciplines differ in stakes and genres, and repeated work with familiar lecturers may not generalize to intercultural contacts. Raters' local norms can also make apparent development an artifact of conformity. Studies should therefore document language proficiency, linguistic background, institutional practices, task familiarity, partner behavior, and whether the outcome matches the claimed dimension.

Teacher preparation is a practical prerequisite, not a trivial afterthought. Ishihara (2010) and Ishihara and Cohen (2010) emphasize research-informed pragmatic teaching; Vellenga (2004) questioned the sufficiency of pragmatic information in language textbooks. Siddiqa and Whyte's (2025) secondary-school analysis likewise cautions that exposure and materials may provide uneven opportunities, although it should not be cited as direct

evidence for university course provision. Departments should support teacher analysis of authentic discourse, consent-aware collection of examples, culturally fair scoring, and access to appropriately moderated digital activities.

### ***Research Agenda Derived from Unresolved Comparisons***

First, studies should test component mechanisms, not only whole instructional packages. A university EFL experiment could compare explicit explanation with guided inductive comparison while holding targets, practice time, input, and feedback constant. A second factor could manipulate whether interlocutor reactions vary. Proximal awareness and repertoire would be assessed alongside contingency-sensitive role-plays. Such work could determine whether apparent explicit-instruction effects result from increased noticing, greater opportunities to retrieve forms, or alignment with the outcome task. Intact classroom allocation should be handled with analyses sensitive to clustering.

Second, measurement and transfer require deliberate design. Studies could compare trained, near-transfer, and farther-transfer tasks through written prompts, task-based oral performance, and naturally occurring or plausibly simulated academic communications, with immediate and delayed assessment. Reliability and validity evidence should address rater agreement, interlocutor effects, disciplinary norms, and the stability of scores across contexts (Bachman, 1990; Youn, 2018). The key question is whether learners retain the ability to negotiate meaning when social relationships, communicative stakes, genre, and available preparation differ.

Third, more research should be population specific. Multi-institutional university EFL samples should be distinguished from secondary learners, English as a second language populations, and professional

multilingual users. Within universities, studies should compare disciplines and local cultures rather than treat all students as interchangeable. Such designs would show where a mechanism is transportable and where curricular adaptation is necessary. Observational work on authentic seminar interaction should complement intervention designs rather than be treated as equivalent causal evidence.

Fourth, digital feedback should be investigated with proper mechanism controls. Research might compare teacher and chatbot feedback for the same target and exposure, independently evaluate feedback appropriateness across culturally diverse scenarios, and assess whether improvements extend to human interaction after the technology is removed. Ajabshir (2025) provides a useful starting point because its pragmalinguistic and sociopragmatic outcomes diverged. Future protocols should also document erroneous feedback, student reliance, confidentiality safeguards, and opportunities to contest machine-generated advice.

Finally, cumulative reviews should manage overlap explicitly. A future registered update could retrieve studies systematically from Scopus, Web of Science, ERIC, and subject databases; publish complete search strings and dates; and map every primary study to each relevant synthesis. A citation-overlap matrix would allow analysts to distinguish genuine replication from repeated review coverage. This step is important because an expanding number of narrative and systematic reviews does not necessarily mean that comparable evidence on university EFL transfer is expanding at the same rate.

## CONCLUSION

For RQ1, university EFL interventions support gains in assessed pragmatic awareness, written production, and selected speech-act performance, with retention documented in some studies. The strength of causal attribution depends on comparison conditions, shared instructional components, and assessment format. These results do not by themselves establish learning transfer to untrained academic genres or interactions. Likewise, the nonsignificant pooled contrast between explicit and implicit instruction does not establish a universal ranking of methods.

For RQ2, pragmatic and interactional theories explain the gap by distinguishing available linguistic resources, contextual judgment, and turn-contingent enactment. A learner may know an expression or judge its appropriateness without being able to adapt it after an interlocutor's response. Because interaction is jointly produced, performance on a controlled task cannot automatically stand for flexible communication across partners and settings.

For RQ3, stronger evidence would link diagnosed learner needs to specified mechanisms, including contextual comparison, form–function mapping, contingent interaction, and intercultural choice, and assess the corresponding outcomes on trained and untrained tasks. Explicit contrasts between near and farther transfer, assessed immediately and after a delay, would distinguish generalization from retention. Matched instructional comparisons, independent ratings, and documentation of interlocutor behavior would strengthen causal and measurement claims.

The evidence-to-transfer matrix organizes these requirements for university EFL pragmatics research and pedagogy. Comparative, cross-task, longitudinal studies across institutions are needed to test its proposed links and establish when task-

specific gains support communicative flexibility in academic life.

## REFERENCES

- Ajabshir, Z. F. (2025). A comparative study of teacher feedback and chatbot feedback on second language learners' pragmalinguistic and sociopragmatic competences. *International Journal of Human-Computer Interaction*, 41(16), 9881–9892.  
<https://doi.org/10.1080/10447318.2024.2429756>
- Alcón Soler, E. (2005). Does instruction work for learning pragmatics in the EFL context? *System*, 33(3), 417–435.  
<https://doi.org/10.1016/j.system.2005.06.005>
- Alrefae, Y. M. H. A., & Al-Shammari, A. H. (2026). Digital innovations in second language pragmatics: A systematic review (2006–2024). *Social Sciences & Humanities Open*, 13, Article 103004.  
<https://doi.org/10.1016/j.ssaho.2026.103004>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63(S1), 68–86. <https://doi.org/10.1111/j.1467-9922.2012.00738.x>
- Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2015). The effect of instruction on pragmatic routines in academic discussion. *Language Teaching Research*, 19(3), 324–350.  
<https://doi.org/10.1177/1362168814541739>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.  
<https://doi.org/10.1093/applin/1.1.1>
- Chen, Y.-S. (2015). Developing Chinese EFL learners' email literacy through requests to faculty. *Journal of Pragmatics*, 75, 131–149.  
<https://doi.org/10.1016/j.pragma.2014.05.009>
- Economidou-Kogetsidis, M. (2011). “Please answer me as soon as possible”: Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics*, 43(13), 3193–3215.  
<https://doi.org/10.1016/j.pragma.2011.06.006>
- Galaczi, E. D. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly*, 5(2), 89–119.  
<https://doi.org/10.1080/15434300801934702>
- Hall, J. K. (1995). “Aw, man, where you goin'?”: Classroom interaction and the development of L2 interactional competence. *Issues in Applied Linguistics*, 6(2), 37–62.  
<https://doi.org/10.5070/L462005217>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.
- Ishihara, N. (2010). Instructional pragmatics: Bridging teaching, research, and teacher education. *Language and Linguistics Compass*, 4(10), 938–953.  
<https://doi.org/10.1111/j.1749-818X.2010.00242.x>
- Ishihara, N., & Cohen, A. D. (2010). *Teaching and learning pragmatics: Where language and culture meet*. Pearson Longman.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366–372. <https://doi.org/10.1111/j.1540-4781.1986.tb05291.x>

- Li, H., Kasim, Z. M., & Ang, L. H. (2026). A scoping review of EFL learners' interlanguage pragmatic development. *PLOS ONE*, 21(3), Article e0344811. <https://doi.org/10.1371/journal.pone.0344811>
- López-Serrano, S. (2025). The impact of self-access web-based instruction on EFL learners' pragmatic awareness of email requests to faculty. *Journal of Pragmatics*, 243, 24–41. <https://doi.org/10.1016/j.pragma.2025.05.006>
- López-Serrano, S., Martínez-Flor, A., & Sánchez-Hernández, A. (2025). The effects of self-access web-based pragmatic instruction and L2 proficiency on EFL students' email request production and confidence. *Languages*, 10(11), Article 279. <https://doi.org/10.3390/languages10110279>
- McConachy, T. (2019). L2 pragmatics as 'intercultural pragmatics': Probing sociopragmatic aspects of pragmatic awareness. *Journal of Pragmatics*, 151, 167–176. <https://doi.org/10.1016/j.pragma.2019.02.014>
- Nguyen, T. T. M., Pham, T. H., & Pham, M. T. (2012). The relative effects of explicit and implicit form-focused instruction on the development of L2 pragmatic competence. *Journal of Pragmatics*, 44(4), 416–434. <https://doi.org/10.1016/j.pragma.2012.01.003>
- Pham, H. T., Petraki, E., & Nguyen, M. T. T. (2024). Effects of oral corrective feedback combined with explicit instruction on EFL learners' recognition and production of English requests. *Applied Pragmatics*, 6(2), 196–227. <https://doi.org/10.1075/ap.21007.pha>
- Ren, W., Li, S., & Lü, X. (2023). A meta-analysis of the effectiveness of second language pragmatics instruction. *Applied Linguistics*, 44(6), 1010–1029. <https://doi.org/10.1093/applin/amac055>
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. <https://doi.org/10.1177/0265532210394633>
- Siddiqi, A., & Whyte, S. (2025). Classroom learning of English L2 requests: Input and interactional opportunities in French secondary schools. *Language Teaching Research*, 29(2), 535–559. <https://doi.org/10.1177/13621688211064932>
- Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. <https://doi.org/10.1017/S0267190511000018>
- Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching*, 48(1), 1–50. <https://doi.org/10.1017/S0261444814000263>
- Taguchi, N. (2022). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, 56(1), 308–335. <https://doi.org/10.1002/tesq.3070>
- Taguchi, N. (2026). Pragmatics in TESOL. *TESOL Quarterly*, 60(2), 635–652. <https://doi.org/10.1002/tesq.70111>
- Taguchi, N., & Barón, J. (2024). Pragmatic competence as linguistic adaptation: Adapting linguistic resources to different listener reactions. *System*, 127, Article 103509. <https://doi.org/10.1016/j.system.2024.103509>
- Taguchi, N., & Ishihara, N. (2018). The pragmatics of English as a lingua franca:

- Research and pedagogy in the era of globalization. *Annual Review of Applied Linguistics*, 38, 80–101.  
<https://doi.org/10.1017/S0267190518000028>
- Taguchi, N., & Kim, Y. (Eds.). (2018). *Task-based approaches to teaching and assessing pragmatics*. John Benjamins.  
<https://doi.org/10.1075/tblt.10>
- Takimoto, M. (2008). The effects of deductive and inductive instruction on the development of language learners' pragmatic competence. *The Modern Language Journal*, 92(3), 369–386.  
<https://doi.org/10.1111/j.1540-4781.2008.00752.x>
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2), 91–112.  
<https://doi.org/10.1093/applin/4.2.91>
- Usó-Juan, E. (2022). Exploring the role of strategy instruction on learners' ability to write authentic email requests to faculty. *Language Teaching Research*, 26(2), 213–237.  
<https://doi.org/10.1177/13621688211066997>
- Vellenga, H. (2004). Learning pragmatics from ESL & EFL textbooks: How likely? *TESL-EJ*, 8(2). <https://tesl-ej.org/wordpress/issues/volume8/ej30/ej30a3/>
- Youn, S. J. (2018). Task design and validity evidence for assessment of L2 pragmatics in interaction. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics* (pp. 217–246). John Benjamins.  
<https://doi.org/10.1075/tblt.10.09you>
- Young, R. F. (2011). Interactional competence in language learning, teaching, and testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 426–443). Routledge.
- Yousefi, M., & Nassaji, H. (2019). A meta-analysis of the effects of instruction and corrective feedback on L2 pragmatics and the role of moderator variables: Face-to-face vs. computer-mediated instruction. *ITL – International Journal of Applied Linguistics*, 170(2), 277–308.  
<https://doi.org/10.1075/itl.19012.you>
- Zhang, Y. (2021). Combining computer-mediated communication with data-driven instruction: EFL learners' pragmatic development of compliment responses. *System*, 103, Article 102624.  
<https://doi.org/10.1016/j.system.2021.102624>
- Zhong, L., & Zhang, Y. (2023). Computerized mediated explicit instruction and EFL learners' knowledge of speech acts: Investigating the role of EFL learners' language proficiency and perceptions. *Heliyon*, 9(9), Article e19738.  
<https://doi.org/10.1016/j.heliyon.2023.e19738>