Vol. 24 (2026): Proceedings of the 2nd International Conference on Linguistics, Psychology and Creative Education (ICLPCE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/iclpce.2026154161

Online and Offline Measures in L2 Pragmatics: Rethinking Interpretation, Processing, and Tolerance

Tiantong Wang*

University College London, London, United Kingdom

**Corresponding author: Tiantong Wang.*

Abstract

Research in second language (L2) pragmatics increasingly relies on both online and offline measures to assess learners' pragmatic competence. Yet these measures are often treated as broadly interchangeable, despite capturing different aspects of pragmatic performance. Offline tasks such as acceptability judgement tasks and discourse completion tasks primarily reflect learners' final interpretive outcomes, whereas online measures such as self-paced reading and eye-tracking provide access to real-time processing dynamics. This paper argues that treating these measures as equivalent indices of the same underlying ability is theoretically unwarranted. Drawing on recent experimental evidence from L2 scalar implicature research, it argues that offline target-like performance does not entail native-like online processing: L2 learners may reach pragmatically appropriate interpretations through slower, more effortful, and more explicitly inferential routes than native speakers. The paper further suggests that offline judgements may be shaped not only by interpretive success, but also by individual differences in pragmatic tolerance, that is, learners' tolerance for underinformative or pragmatically sub-optimal utterances, regardless of whether the relevant inference has actually been computed. To address these issues, the paper distinguishes three constructs that current research often conflates: interpretive success, processing efficiency, and pragmatic tolerance. A more differentiated account of what online and offline measures actually capture is therefore necessary for both accurate theoretical interpretation and more principled research design in L2 pragmatics.

Keywords

L2 pragmatics, online measures, offline measures, scalar implicature, processing efficiency

1. Introduction

A central concern in second language (L2) pragmatics research is how learners acquire the ability to interpret meaning that goes beyond what is literally encoded in an utterance. Pragmatic competence, in this sense, involves not only knowing the conventional meaning of linguistic forms, but also the capacity to derive contextually appropriate interpretations in real time, a capacity that depends on the integration of linguistic, cognitive, and social knowledge [1]. To investigate this ability, researchers have drawn on a wide range of methodological tools. Offline tasks such as acceptability judgement tasks, truth-value judgement tasks, and discourse completion tasks have long been used to elicit learners' interpretive judgements and productive responses [2]. More recently, online measures including self-paced reading, eye-tracking, and reaction time paradigms have been introduced to capture the moment-by-moment dynamics of pragmatic processing. Both

types of measures are now well established in the field, and both are routinely treated as valid indices of learners' pragmatic competence. However, the assumption that online and offline tasks index the same underlying construct deserves closer scrutiny. The two measure types differ not only in format, but in what they are theoretically sensitive to: offline tasks reflect a learner's final interpretive outcome, arrived at under conditions that permit reflection, reanalysis, and strategic reasoning; online tasks, by contrast, place demands on real-time processing and are sensitive to the timing, cost, and automaticity of pragmatic inference. If these tasks tap into different aspects of pragmatic ability, then a learner who performs in a target-like manner on an offline task may not necessarily be processing pragmatic information in a native-like way. Recent experimental evidence supports this concern. Studies examining L2 scalar implicature computation, for instance, have found that learners can match native speaker accuracy on offline interpretation tasks while still exhibiting significantly longer reading times or elevated processing costs on online measures, suggesting that the route to a correct interpretation, not just the interpretation itself, differs systematically between native speakers and L2 learners. If such divergence is real and consistent, it cannot be dismissed as methodological noise; it demands theoretical explanation. This paper argues that online and offline measures in L2 pragmatics research are not interchangeable, and that treating offline target-like performance as straightforward evidence of native-like pragmatic competence is theoretically unwarranted. To develop this argument, we focus on L2 scalar implicature as a test case, drawing on a growing body of experimental research that has employed both measure types within comparable populations and tasks. We further argue that the divergence between online and offline results points to the need to distinguish three constructs that current research tends to conflate: interpretive success, whether a learner arrives at the pragmatically appropriate interpretation; processing efficiency, whether that interpretation is derived automatically and at low cognitive cost; and pragmatic tolerance, a learner's willingness to accept pragmatically infelicitous expressions, which may influence offline judgements independently of whether a pragmatic inference has been computed. Clarifying what each type of measure actually captures, we suggest, is a necessary step toward both more accurate theoretical interpretation and more principled methodological choices in L2 pragmatics research.

2. Online and Offline Measures: Constructs and Theoretical Assumptions

The distinction between online and offline measures in psycholinguistics and language acquisition research is well established, yet its implications for L2 pragmatics have not always been made explicit. Before examining why the two measure types may yield divergent results, it is necessary to clarify what each type is designed to capture and what assumptions underlie its use.

2.1 Online Measures and Real-Time Processing Dynamics

Online measures are designed to capture language processing as it unfolds in real time [3,4]. The defining feature of such measures is temporal sensitivity: they record responses at or near the moment of processing, making them sensitive to the cognitive demands placed on the language system during comprehension. Online measures are defined by their sensitivity to processing time and cognitive load, providing evidence about the immediacy and cost of pragmatic inference. Methods such as self-paced reading, eye-tracking, reaction-time tasks, and ERP paradigms are therefore valuable not simply because they record responses online, but because they make visible how efficiently pragmatic interpretations are computed during comprehension. What these methods share is a commitment to measuring processing cost and processing immediacy, that is, how much cognitive effort a given interpretation requires and how quickly it is initiated [5]. In the context of pragmatic processing, online measures are particularly well suited to capturing whether a pragmatic inference is computed automatically and incrementally, or whether it is delayed, effortful, or dependent on post-hoc reanalysis [6]. Under this logic, online measures are taken to index not simply whether an inference is ultimately available, but how efficiently and incrementally it is computed during comprehension. A learner who consistently shows elevated reading times or delayed responses at pragmatically critical regions of an utterance, for example, provides evidence that pragmatic enrichment is not yet fully automatized, even if the learner is ultimately capable of arriving at the correct interpretation.

2.2 Offline Measures and Reflective Interpretive Judgements

Offline measures, by contrast, are designed to capture a participant's final interpretive judgement or response, typically after the full stimulus has been processed and without the time pressure characteristic of

online tasks [7]. Common offline methods in L2 pragmatics include acceptability judgement tasks (AJTs), in which participants rate the appropriateness or naturalness of an utterance in a given context; truth-value judgement tasks (TVJTs), in which participants evaluate whether an utterance is true or false relative to a described situation; discourse completion tasks (DCTs), which elicit productive pragmatic responses; and interpretation questions, which ask participants to select or describe the intended meaning of an utterance [8]. The defining feature of offline tasks is that they permit reflection: participants may reconsider their initial response, draw on explicit pragmatic knowledge, and supplement initial interpretation with deliberate reasoning [9]. This makes offline tasks particularly useful for assessing whether learners have access to pragmatic knowledge in principle, that is, whether they know what the appropriate interpretation is, even if that knowledge is not yet deployed efficiently under real-time conditions. As such, offline measures index access to pragmatic knowledge under reflective conditions rather than its real-time deployment.

2.3 The Assumed Equivalence of Measurement Constructs

Taken individually, neither measure type is inherently problematic. The difficulty arises when the two are treated as equivalent windows onto a single underlying construct labelled pragmatic competence [10]. An offline task that shows target-like performance tells us that a learner can, under reflective conditions, access the appropriate pragmatic interpretation. An online task that shows native-like processing tells us something more demanding: that pragmatic enrichment is occurring automatically, incrementally, and at low cognitive cost. These are related but distinct achievements, and conflating them obscures precisely the kind of developmental information that L2 pragmatics research aims to uncover. Crucially, these differences are not merely methodological, but point to distinct dimensions of pragmatic performance and interpretation. As we argue in the following section, the gap between these two types of evidence is not a measurement artefact to be minimised, but a theoretically meaningful signal that deserves careful interpretation [11].

3. The Divergence: Three Sources of Dissociation

If online and offline measures were simply two routes to the same destination, we would expect them to converge systematically: learners who perform in a target-like manner on offline tasks should also show native-like processing signatures on online tasks, and vice versa. The empirical record, however, does not support this expectation. A growing body of research in L2 pragmatics has documented cases in which offline and online performance dissociate, where learners match native speaker accuracy on judgement or interpretation tasks while simultaneously exhibiting elevated processing costs on real-time measures. This section examines three sources of such dissociation, each of which helps explain why online and offline measures should not be expected to converge straightforwardly.

3.1 Interpretative Outcomes versus Processing Routes

The most fundamental source of dissociation lies in the difference between what a learner ultimately arrives at and how they get there. Two individuals may produce identical final interpretations of an utterance while having followed entirely different processing routes to reach that outcome. In first language processing, pragmatic enrichment is often assumed to proceed relatively rapidly and with lower processing cost than in non-native comprehension. In L2 processing, the same enrichment may be achieved, but through a slower, more deliberate, and more explicitly inferential route. The learner may, for instance, initially register the literal meaning of an utterance, recognise that it is pragmatically underinformative, and consciously reason toward the enriched interpretation, arriving at the correct answer, but via a qualitatively different path. This distinction between outcome and route is not merely theoretical. For instance, studies on scalar implicature processing have shown that participants may ultimately arrive at pragmatically appropriate interpretations, while online measures reveal delayed or effortful computation [6]. Slabakova [11] likewise shows that advanced L2 learners can display substantial success in deriving pragmatic inferences, yet such success does not entail that these inferences are computed through the same incremental and low-cost processing route observed in native speakers. Taken together, these findings suggest that offline accuracy, on its own, is an insufficient basis for concluding that pragmatic processing has reached a native-like level of efficiency. This dissociation shows that interpretive success does not uniquely determine the nature of the underlying processing route.

3.2 Time Pressure and the Automaticity of Pragmatic Inference

A second, related source of dissociation concerns the role of time pressure in distinguishing automatic from controlled processing. Online tasks, by their nature, impose processing demands in real time: participants cannot pause, reconsider, or revise their initial response in the way that offline tasks permit. This means that online measures are selectively sensitive to the automaticity of a given cognitive operation, whether it is triggered rapidly and without deliberate effort, or whether it requires time and attentional resources to complete. In the domain of scalar implicature, this distinction is particularly consequential. Scalar implicatures, inferences such as deriving some but not all from the use of some in a context where all would have been more informative, are known to be computationally demanding even for native speakers, requiring the consideration of alternative expressions and the suppression of the literal reading [12]. For L2 learners, this demand may be further compounded by the broader pressures of processing in a non-native language. Previous research has shown that while L2 learners can perform accurately on scalar implicature tasks under untimed conditions, online measures often reveal reduced automaticity relative to native speakers [5,6]. The practical implication is straightforward: a learner who has learned to compute scalar implicatures through deliberate reasoning may perform well on offline tasks that allow sufficient time for this reasoning to run to completion, while still showing non-native processing profiles when time pressure prevents the full inferential sequence from being executed. Offline tasks, by allowing reflection and reanalysis, may therefore overestimate the degree to which pragmatic inference has been automatized.

3.3 Pragmatic Tolerance as a Source of Variation in Offline Judgements

The third source of dissociation is conceptually distinct from the first two and concerns a potential confound that is especially salient in offline judgement tasks. When a participant accepts or rejects an utterance on an acceptability judgement task or endorses a particular interpretation on a truth-value judgement task, the response they produce is not a pure measure of whether a pragmatic inference has been computed. It is also shaped by that participant's pragmatic tolerance, their willingness to accept utterances that are pragmatically infelicitous, underinformative, or suboptimal in a given context. Katsos and Bishop [13] explicitly distinguish pragmatic tolerance from inferential computation, showing that acceptability judgements cannot be assumed to directly reflect whether a scalar implicature has been derived. A learner who gives a target-like judgement on a scalar implicature task, for instance, rating an utterance as infelicitous when a stronger scalar term would have been more appropriate, may be doing so because they have successfully computed the implicature and recognised the violation. Alternatively, they may have a low tolerance for pragmatic infelicity in general, leading them to reject a wide range of pragmatically suboptimal utterances regardless of whether the specific scalar inference has been derived. Conversely, a learner with high pragmatic tolerance may accept underinformative utterances not because they have failed to compute the implicature, but because they are disinclined to penalise speakers for pragmatic suboptimality. In either case, the offline judgement does not straightforwardly reflect the presence or absence of pragmatic inference computation. Individual differences of this kind may be particularly relevant in L2 populations, whose metalinguistic awareness, orientations to pragmatic norms, and familiarity with target-language conventions may differ both from native speakers and from one another [14]. This means that group-level comparisons between L2 learners and native speakers on offline tasks may obscure important within-group variation, and that apparent convergence between learner and native speaker judgements may sometimes reflect tolerance alignment rather than equivalent pragmatic processing.

4. A Three-Construct Framework for L2 Pragmatic Competence

The three sources of dissociation outlined in the previous section converge on a common theoretical problem: the construct of pragmatic competence, as it is typically operationalised in L2 research, is not sufficiently differentiated to accommodate the range of phenomena that online and offline measures jointly reveal. When a learner produces a target-like judgement on an offline task, we cannot straightforwardly conclude that they possess native-like pragmatic competence, because such a judgement conflates at least three distinct achievements: arriving at the correct interpretation, doing so efficiently and automatically, and responding in a way that reflects actual inference computation rather than a general disposition toward pragmatic strictness or leniency. To make theoretical progress, these achievements need to be treated as separable constructs, each with its own developmental trajectory, its own methodological requirements, and

its own relationship to the broader construct of pragmatic competence. The researcher proposes, accordingly, a three-construct framework that distinguishes interpretive success, processing efficiency, and pragmatic tolerance. These three constructs are not intended as exhaustive components of pragmatic competence; rather, they represent the minimum theoretical distinctions required to interpret online and offline data in a principled way.

4.1 Interpretive Success

Interpretive success refers to whether a learner arrives at the pragmatically appropriate interpretation of an utterance, that is, whether the output of the comprehension process is target-like in content. This is the construct that offline tasks are most directly designed to measure, and it corresponds most closely to what is traditionally understood as pragmatic knowledge: knowing what an utterance means in context, knowing what speech act is being performed, and knowing what implicature is generated by a particular scalar expression. Interpretive success can be assessed relatively straightforwardly through accuracy-based offline measures such as truth-value judgement tasks and interpretation questions, and it represents the most accessible entry point into a learner's pragmatic competence. However, interpretive success is also the most limited of the three constructs. A learner may achieve target-like interpretive success through a variety of routes, including deliberate metalinguistic reasoning, reliance on explicit pragmatic rules, or even educated guessing based on contextual cues, none of which necessarily implies that pragmatic inference is proceeding in a native-like manner. Interpretive success, in other words, is a necessary but not sufficient condition for native-like pragmatic competence.

4.2 Processing Efficiency

Processing efficiency refers to the cognitive cost and temporal profile of pragmatic inference, whether a given pragmatic interpretation is derived rapidly, automatically, and without placing substantial demands on attentional resources [5,15]. This is the construct to which online measures are most directly sensitive [4,16]. A learner who shows native-like reading times at pragmatically critical utterance regions, or who exhibits no response-time delay relative to native speakers under time pressure, provides evidence that pragmatic inference is being carried out with a degree of efficiency that approaches native-language processing [6,17]. Processing efficiency is theoretically important because it reflects the degree to which pragmatic knowledge has been proceduralized, integrated into the real-time language processing system rather than remaining accessible only through deliberate reflection [18-20]. Experimental evidence suggests that scalar implicatures are not computed automatically but instead require cognitive resources and time, indicating that the shift from effortful to more efficient processing constitutes a meaningful dimension of development that offline tasks are not well suited to capture [6,11,17,21]. From a developmental perspective, then, processing efficiency represents a more demanding and potentially more revealing index of pragmatic acquisition than interpretive success alone [11,22].

4.3 Pragmatic Tolerance

Pragmatic tolerance refers to a participant's general disposition toward pragmatically infelicitous or suboptimal utterances, their willingness to accept such utterances as acceptable, appropriate, or true, independently of whether the relevant pragmatic inference has been computed. As discussed in Section 3.3, pragmatic tolerance functions as a confound in offline judgement tasks because it influences response patterns in ways that are orthogonal to inference computation. A participant with low pragmatic tolerance will tend to reject pragmatically infelicitous utterances even when they have not explicitly computed the implicature that renders them infelicitous; a participant with high pragmatic tolerance may accept such utterances even when they have computed the implicature and recognised the violation. Treating pragmatic tolerance as a distinct construct has two important consequences. First, it alerts researchers to the possibility that convergence between learner and native speaker performance on offline tasks may sometimes be spurious, reflecting similar tolerance profiles rather than equivalent pragmatic processing. Second, it opens the possibility of measuring pragmatic tolerance directly, for instance through tasks that systematically vary the degree of pragmatic infelicity and examine whether participant responses scale accordingly. Identifying and controlling for pragmatic tolerance is particularly important in studies that use offline judgement data as the primary evidence for claims about implicature computation or pragmatic development.

4.4 Implications

Distinguishing these three constructs has direct implications for how L2 pragmatics studies should be designed and interpreted. Most immediately, it suggests that studies relying exclusively on offline measures are likely to conflate interpretive success with processing efficiency, and may additionally be confounded by individual differences in pragmatic tolerance. Conversely, studies relying exclusively on online measures may capture processing costs without establishing whether learners are, in fact, computing the target interpretation at all. The most theoretically informative research designs will therefore be those that incorporate both measure types within the same study and same population, allowing the three constructs to be examined in relation to one another rather than in isolation. Such designs are not merely methodologically preferable; they are, we argue, theoretically necessary if L2 pragmatics research is to make well-grounded claims about the nature and development of pragmatic competence.

5. Conclusion

This paper has argued that online and offline measures in L2 pragmatics are not interchangeable, and that treating them as equivalent operationalisations of a single construct obscures theoretically significant variation in how pragmatic competence develops. Drawing on evidence from scalar implicature research, we have identified three distinct sources of dissociation between the two measure types, differences in processing route, differences in automaticity under time pressure, and the independent influence of pragmatic tolerance on offline judgement responses, and proposed a three-construct framework that distinguishes interpretive success, processing efficiency, and pragmatic tolerance as analytically separable dimensions relevant to the study of pragmatic competence. The practical implications of this framework are clear. Researchers who rely exclusively on offline measures risk overestimating the degree to which pragmatic knowledge has been proceduralised and may draw conclusions about pragmatic acquisition that hold only under reflective, untimed conditions. Researchers who rely exclusively on online measures, meanwhile, risk misattributing processing costs without establishing whether target-like interpretation has in fact been achieved. Neither measure type, in isolation, is adequate to the theoretical demands of the field. The framework proposed here ultimately calls for a more deliberate alignment between theoretical constructs and methodological choices, a recognition that the question being asked should determine the measure being used, and that convergent evidence from multiple methods is not a luxury but a requirement for principled claims about L2 pragmatic competence. The divergence between online and offline measures, properly understood, is not a problem to be resolved but a finding to be explained. Future research should move beyond treating measurement convergence as a criterion of validity and instead examine divergence as a theoretically informative phenomenon.

References

[1] Kasper, G., & Rose, K. R. (2002). Pragmatic development in a second language. Oxford, UK: Blackwell Publishing.

[2] Kasper, G., & Dahl, M. (1991). Research methods in interlanguage pragmatics. Studies in Second Language Acquisition, 13(2), 215–247. https://doi.org/10.1017/S0272263100009955

[3] Bialystok, E. (1993). Metalinguistic awareness: The development of children's representations of language. In C. Pratt & A. F. Garton (Eds.), Systems of representation in children: Development and use (pp. 211–233). London, UK: Wiley.

[4] Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 87(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

[5] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

[6] Segalowitz, N. (2010). Cognitive bases of second language fluency. New York, NY: Routledge.

[7] Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. Journal of Memory and Language, 51(3), 437–457. https://doi.org/10.1016/j.jml.2004.05.006

[8] Roever, C. (2010). Testing of second language pragmatics: Past and future. Language Testing, 27(4), 463–481. https://doi.org/10.1177/0265532210361476

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

[10] Bialystok, E. (1993). Metalinguistic awareness: The development of children's representations of language. In C. Pratt & A. F. Garton (Eds.), Systems of representation in children: Development and use (pp. 211–233). London, UK: Wiley.

[11] Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford, UK: Oxford University Press.

[12] Slabakova, R. (2010). Scalar implicatures in second language acquisition. Lingua, 120(10), 2444–2462.

[13] Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. Cognitive Psychology, 87, 54–78. https://doi.org/10.1016/j.cogpsych.2016.05.001

[14] Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness. Cognition, 120(1), 67–81. https://doi.org/10.1016/j.cognition.2011.02.015

[15] Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. Language Testing, 23(2), 229–256. https://doi.org/10.1191/0265532206lt329oa

[16] Kormos, J. (2006). Speech production and second language acquisition. Mahwah, NJ: Lawrence Erlbaum.

[17] Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), Research methods in second language psycholinguistics (pp. 20–49). New York, NY: Routledge.

[18] Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. Cognitive Psychology, 58(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

[19] DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), Theories in second language acquisition: An introduction (pp. 97–113). Mahwah, NJ: Lawrence Erlbaum.

[20] Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

[21] Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

[22] Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. Language Learning and Development, 8(4), 365–394. https://doi.org/10.1080/15475441.2011.626386

[23] Taguchi, N. (2011). Teaching pragmatics: Trends and issues. Annual Review of Applied Linguistics, 31, 289–310. https://doi.org/10.1017/S0267190511000018

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.