Studies in Second Language Acquisition (2026), 1–23
doi:10.1017/S0272263126101946

RESEARCH ARTICLE

A critical examination of the relationship between explicit and implicit knowledge and language teaching

Yu Tamura[1] and Junya Fukuta[2]

[1]Faculty of Foreign Language Studies, Kansai University, Osaka, Japan and [2]Faculty of Science and Engineering, Chuo University, Tokyo, Japan
Corresponding author: Yu Tamura; Email: yutamura@kansai-u.ac.jp

(Received 24 December 2025 UTC; Revised 16 July 2026 UTC; Accepted 06 August 2026 UTC)

Abstract

This paper offers a critical perspective on the implicit/explicit knowledge (IE) distinction in second language acquisition (SLA) research, with particular attention to the conceptual and methodological problems that arise when this distinction is applied to language teaching. First, the IE framework exhibits systematic measurement instability: individual tasks are extracted from test batteries without independent validation; factorial separation may reflect methodological artifacts, and researchers disagree on which tasks measure which knowledge types. These patterns suggest that the distinction functions as a formative rather than a reflective construct. Second, intervention studies face a measurement-pedagogy dilemma: measures with strong construct validity are ecologically distant from classroom performance, while ecologically valid measures cannot reliably isolate implicit knowledge. Although the IE framework does not support a realist cognitive explanation, it can still be used instrumentally. Despite these limitations, we offer minimal conditional principles for those who choose to employ it.

Introduction

The distinction between implicit and explicit (IE) knowledge has become foundational to second language acquisition (SLA) research over the past two decades. Since Ellis's (2005a) influential psychometric validation, researchers have increasingly relied on this dichotomy to explain three persistent phenomena: differences between native speakers and learners, variation among learners themselves, and inconsistencies in individual performance across contexts (Ellis, 2005a; Ellis et al., 2009). The framework addresses questions central to instructed language learning: Why do explicitly taught grammatical rules often fail to appear in spontaneous communication? How does knowledge acquired through classroom instruction differ from knowledge developed through naturalistic exposure? What role should explicit grammar teaching play in developing implicit knowledge?

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

Since Ellis (2005a), many researchers have investigated IE knowledge, particularly through psychometric methods. While some have proposed a third factor, such as automatized explicit knowledge (Suzuki, 2017), a number of studies have reported that, despite differences in the tasks used, the two-factor structure separating implicit from explicit knowledge is replicated, with acceptable reliability (e.g., Bowles, 2011; Ellis et al., 2009; Kim & Nam, 2017; Vafaee, Suzuki, & Kachinske, 2017; Zhang, 2015).

However, this paper examines two interrelated problems inherent in the IE knowledge framework: measurement instability and the measurement-pedagogy dilemma. Regarding the first problem, nearly two decades after Ellis's (2005a) attempt to operationalize these constructs psychometrically, fundamental disagreements persist about which tasks validly measure which knowledge types. Individual tests originally designed as components of composite instruments are now treated as standalone measures of implicit or explicit knowledge, despite lacking independent validation (Kusanagi, 2018). These patterns reflect three interconnected problems: the test battery fallacy, the method variance problem, and definitional inconsistency. These measurement problems suggest that, in practice, the IE distinction often functions more like a formative model, operationally defined by researcher choices about measurement procedures, than like a reflective model that captures pre-existing cognitive constructs (Alamer, Schuberth, & Henseler, 2024). We return to this distinction the Double bind section to examine its implications for the epistemological status of the IE framework.

The second problem arises when the IE framework is extended to instructional research, where researchers investigate the effectiveness of language teaching. Intervention studies encounter a specific measurement problem. Tasks designed to minimize the use of explicit knowledge, such as reaction-time measures, bear little resemblance to authentic classroom performance, whereas tasks that more closely reflect real language use leave open the question of which knowledge type underlies the observed performance. This dilemma proves particularly problematic when the IE distinction is applied to pedagogical settings, where researchers and practitioners have actively pursued such application, and where the goal is to inform teaching practice rather than conduct experimental research.

These problems (psychometric, conceptual, and practical) have consequences that extend beyond technical debates. When explaining teaching outcomes based on the IE distinction, the reliability of pedagogical implications drawn from it becomes questionable. Recent publications on IE knowledge development exhibit internal contradictions, with different studies classifying the same GJT as measuring implicit knowledge in one context and explicit knowledge in another (see the section on definitional inconsistency for detailed analysis). Such inconsistencies reflect systematic disagreement across the field about what tasks measure what constructs (e.g., Godfroid & Kim, 2021; Suzuki, 2017; Zhang, 2015). Teachers and curriculum designers thus face conflicting guidance from researchers who have extended this framework to pedagogical contexts, calling into question whether the IE distinction can sustain such application at all.

This paper is organized as follows. We first review the evolution of measurement approaches from Bialystok's (1978) early functional model through Ellis's (2005a) psychometric validation, which remains the dominant approach in validation studies of IE knowledge. In doing so, we clarify what phenomena the IE distinction was designed to explain and establish why measurement validity matters for pedagogical application. We then examine measurement instability in the IE framework through three interconnected problems: the test battery fallacy, method variance, and definitional inconsistency. Together, these problems indicate that the IE framework exhibits

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

formative characteristics. Next, we examine the measurement-pedagogy dilemma, in which researchers face trade-offs between construct validity and ecological validity when applying this distinction to intervention studies. We further argue that this measurement problem reveals a double bind in which the IE distinction falls short on both fronts: viewed as an instrumental framework, it does not support the realist explanation that scientific inquiry demands, and it lacks the practical utility that pedagogical application requires (where simple performance-based assessments suffice). We then propose three conditional principles (conceptual transparency, methodological honesty, and pedagogical justification) for researchers who nevertheless choose to employ the IE framework in instructional research. We also emphasize that although our critique questions the theoretical necessity of the distinction, the IE framework still retains instrumental value for organizing empirical work and facilitating professional communication. We conclude by synthesizing the argument and considering implications for future research on the IE framework in instructed SLA.

Background

To understand why the IE distinction became so influential in SLA, it is useful to return to the empirical puzzles that originally motivated it. The distinction did not emerge as a purely theoretical construct, but as a response to recurring patterns in learner performance that existing models struggled to explain. Three sets of observations were particularly important: (a) the performance gap between native speakers and L2 learners, (b) systematic differences among learners at different proficiency levels, and (c) marked variability in the same learner's performance across tasks and contexts. Taken together, these findings suggested that learner performance might draw on qualitatively different kinds of linguistic knowledge operating under different conditions.

Early proposals of a dual-knowledge framework in SLA emerged from attempts to explain recurrent patterns in learner performance. Bialystok (1978) defined explicit linguistic knowledge as information that learners can consciously articulate, and implicit linguistic knowledge as the intuitive, automatic information that supports spontaneous comprehension and production. Her empirical work using time-pressured and delayed grammaticality judgments (Bialystok, 1979) suggested that learners rely on fast, intuitive processes to judge well-formedness, while slower, conscious analysis becomes relevant when identifying specific errors. These early findings established the plausibility of two qualitatively different modes of linguistic knowledge without claiming that any single task directly measures one type or the other.

Building on this line of work, Han and Ellis (1998) took an initial step toward a psychometric approach by applying factor-analytic techniques to a battery of language measures. Their study was exploratory in scope, but it anticipated later attempts to validate the IE distinction through patterns of covariance among tasks.

A more systematic shift occurred with Ellis (2005a), who explicitly framed the IE distinction as a construct to be validated psychometrically through a coordinated task battery and factor analysis. His study operationalized implicit knowledge as performance on time-pressured, meaning-focused tasks and explicit knowledge as performance on untimed, form-focused metalinguistic tasks, and it helped reposition the distinction as something that could be measured through task performance rather than inferred solely from theory. This psychometric turn catalyzed a large body of subsequent research attempting to classify tasks according to the type of knowledge they were presumed to access and strongly influenced later measurement practices in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

SLA (e.g., Bowles, 2011; Godfroid & Kim, 2021; Gutiérrez, 2013; Isemonger, 2007; Kim & Nam, 2017; Suzuki, 2017; Vafaee et al., 2017; Zhang, 2015).

In parallel, a separate tradition developed around implicit learning and the role of awareness in acquiring linguistic knowledge (e.g., Leung & Williams, 2012; Rebuschat, 2013; Rebuschat & Williams, 2012; Williams, 2005). This body of work, which we refer to as the CU (Consciousness–Unconsciousness) framework, conceptualizes implicit knowledge as genuinely inaccessible to awareness even under unlimited time and explicit prompting, and explicit knowledge as the conscious, reportable knowledge that learners can verbalize. This perspective connects to broader cognitive psychology, particularly work in the implicit learning tradition such as Dienes (2007), which emphasizes the direct measurement of awareness. These approaches assess participants' conscious access to knowledge through explicit awareness rather than inferring it from patterns of task performance. These two traditions, psychometric IE (Implicit–Explicit) and awareness-based CU (Conscious–Unconscious), use the same terminology but differ conceptually and methodologically. IE approaches distinguish knowledge types based on performance characteristics, whereas CU approaches ground the distinction in mental states. Because the two frameworks target different constructs, findings are often difficult to compare, and the field lacks consensus on how IE knowledge should be defined or measured. This conceptual tension sets the stage for the measurement instability examined in the two sections that follow.

Measurement instability in the IE framework

As reviewed in the Background section, Ellis (2005a) transformed the IE distinction from a theoretical proposal into a psychometrically operationalized framework by developing a battery of tasks designed to elicit different types of knowledge under varying performance conditions. His battery included three tests targeting implicit knowledge (timed GJT, elicited imitation, and oral narrative tasks) and two tests targeting explicit knowledge (e.g., untimed GJT and metalinguistic knowledge tasks). Ellis employed principal component analysis (PCA), which he referred to as "principal component factor analysis," to demonstrate that these tasks clustered into two factors.[1] He then interpreted it as evidence for distinct IE knowledge systems. This test battery approach subsequently became the standard reference for researchers investigating the IE distinction, with Ellis et al. (2009) providing refined versions and validation studies extending the methodology across different learner populations and linguistic structures.

However, this psychometric tradition created a critical measurement problem that has gone largely unrecognized in subsequent research. While Ellis designed his battery as an integrated instrument in which multiple tasks collectively provide evidence for

[1]Ellis (2005a) repeatedly referred to his analytic procedure as "principal component factor analysis," a term that conflates PCA with common factor analysis (FA). PCA extracts components based on total variance and is a data reduction technique without assumptions about causal relationships between constructs and indicators. In contrast, FA extracts factors assumed to be latent variables that cause observed covariances among indicators. Ellis used PCA but interpreted the results using factor-analytic terminology, treating extracted components as if they were latent factors reflecting underlying knowledge systems. The ambiguous terminology in Ellis (2005a) reflects broader ambiguity about whether the IE distinction represents correlation-based performance clustering or theoretical claims about cognitive architecture—an ambiguity that has propagated through subsequent research and contributed to the measurement problems discussed in this section (see Alamer et al., 2024, for the detailed discussion of the formative and reflective approaches).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

knowledge types through their covariation patterns, individual tasks are now routinely extracted from this battery and treated as standalone measures. Researchers cite Ellis (2005a) or Zhang (2015) to justify using a single task, such as timed GJT or elicited imitation, as a direct measure of implicit knowledge and untimed GJT or metalinguistic knowledge test as a direct measure of explicit knowledge, even though these tasks were originally validated only as components of a composite assessment. This practice, which we call the *test battery fallacy*, undermines the psychometric logic on which the IE framework was established and raises fundamental questions about construct validity. The following subsections examine how this fallacy operates: the first clarifies the battery and documents how task extraction manifests across the literature; the second considers whether the factorial evidence supporting the battery reflects genuine construct differences or methodological artifacts; and the third examines persistent disagreement about task classifications even among methodologically careful researchers.

The test battery fallacy

Ellis's (2005a) approach exemplifies what we call correlation-based construct inference: constructs are not measured directly but inferred from patterns of covariation across multiple performance measures. Under this logic, implicit and explicit knowledge gain empirical support through systematic clustering of tasks: when time-pressured, meaning-focused tasks correlate with each other but not with untimed, form-focused tasks, researchers infer that these task families access different knowledge systems. The battery thus functions as an integrated instrument in which no single task directly measures a construct; rather, shared variance among tasks constitutes the inferential basis for construct existence. Multiple heterogeneous indicators are not merely desirable for measurement precision; they are logically necessary because the construct itself is defined by their pattern of covariation.

This correlation-based inference reflects a fundamental problem in measurement validity. As Fukuta, Nishimura, and Tamura (2023) argue, drawing on Borsboom, Mellenbergh, and van Heerden's (2004) ontological realist perspective, valid measurement requires demonstrating that test responses are causally produced by the attribute being measured—not merely that tests correlate with each other. Ellis's (2005a) battery produced clear factorial separation: oral imitation, oral narrative, and timed GJT loaded on Factor 1, while untimed GJT and metalinguistic test loaded on Factor 2. Ellis interpreted these factors as implicit and explicit knowledge, respectively. However, factor-analytic evidence of covariation among tasks does not establish what psychological entities, if any, cause this covariation. When Task A and Task B correlate, this could reflect: (a) both tasks are caused by the same latent construct, (b) both tasks are caused by different constructs that happen to correlate, (c) both tasks share methodological characteristics, or (d) both reflect general proficiency. Correlation-based validation cannot distinguish among these explanations because it demonstrates association without establishing causation. The IE framework thus faces what Fukuta et al. call an ontological problem: it infers psychological constructs from performance patterns without demonstrating that these constructs causally produce the observed performance.

This ontological problem has critical implications for task extraction. If correlation-based validation cannot establish causal relationships between hypothetical knowledge types and task performance—if we can only observe that certain tasks cluster together without understanding why they cluster or what causes the clustering—then extracting a single task from the battery eliminates even the minimal correlational evidence on

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

which construct claims depend. When researchers cite Ellis (2005a) to justify using timed GJT alone as an *implicit knowledge measure*, they face two compounded problems. First, the original battery provided only correlational evidence that timed GJT clusters with other time-pressured tasks, not causal evidence that implicit knowledge produces this clustering. Second, by removing the task from the battery context, researchers eliminate even this correlational pattern, leaving no empirical basis—neither causal nor correlational—for the construct claim. The practice thus violates both the minimal requirements of correlation-based validation (maintaining the covariation pattern observed in the battery) and the more rigorous requirements of ontological validity (demonstrating causal relationships between constructs and measurements).

Moreover, as Fukuta (2025) demonstrates, the IE framework's actual logic is performance-first: it infers knowledge types from task performance patterns rather than measuring pre-theorized mental states. This inferential direction has important implications. When researchers extract different task combinations from the original battery—one study using timed GJT alone, another using elicited imitation, a third using oral narrative—they are not merely using different measures of the same construct but potentially working with different construct definitions. Each task provides unique performance characteristics (timing, modality, metalinguistic demand) that contribute to what implicit knowledge or explicit knowledge means in operational terms. This explains the persistent definitional inconsistency documented below: researchers cannot agree on task classifications because the correlation-based approach leaves the construct definition dependent on which performance patterns are included in the inferential base.

A survey of the literature reveals that the test battery fallacy is not an isolated error but a systemic issue manifesting in various forms, ranging from arbitrary task extraction to methodological compromise and statistical misinterpretation. The following examples illustrate how researchers across different studies and contexts have violated both the minimal requirements of correlation-based validation and the more rigorous requirements of ontological validity.

First, the most common manifestation is the arbitrary selection of single tasks. Numerous intervention studies treat a single task as a unitary measure simply by citing previous factor-analytic studies. Khezrlou (2021) exemplifies this practice. In her study investigating the effect of task repetition, she chose an elicited imitation task (EIT) to measure implicit knowledge and an untimed GJT to measure explicit knowledge. She justified these choices by citing Ellis (2005a) and Zhang (2015). On this basis, she concluded that the explicit instruction combined with task repetition facilitated the development of implicit knowledge. However, this interpretation rests on the unvalidated assumption that the EIT alone measures implicit knowledge. The improvement of the EIT score could reflect automatized explicit knowledge, as cautioned by Suzuki (2017) and Suzuki and DeKeyser (2015). This exemplifies the problematic assumption that if a task was claimed to measure implicit or explicit knowledge in previous studies, it can be extracted and used as a standalone measure, violating the latent-variable logic.

Michaud and Ammar (2023) present a related case in which the limitation was explicitly acknowledged rather than left unaddressed. They utilized a paper-based GJT for explicit knowledge and an EIT for implicit knowledge, and acknowledged the critique by Suzuki and DeKeyser (2015) that EIT may tap into automatized explicit knowledge. Relying on DeKeyser's (2017) remark that the two are *functionally equivalent*, however, they proceeded to label EIT scores as implicit knowledge and concluded that their intervention led to an improvement of both explicit and implicit knowledge.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

Acknowledging the limitation in this way does not resolve it: it leaves open whether the EIT tapped implicit or automatized explicit knowledge, even as the scores were interpreted as the former.

Second, researchers often face a pragmatic dilemma where they recognize measurement limitations but proceed regardless. Andringa, de Glopper, and Hacquebord (2011) explicitly admitted that their untimed GJT could not be considered a "pure test of explicit knowledge" (p. 897) because it allows recourse to implicit knowledge. Yet, citing the factor-analytic evidence from Han and Ellis (1998) and Ellis (2005a), they proceeded to use it alongside a free written response task to operationalize the distinction. Andringa et al. (2011) reported that the effects of instruction on the target structures manifested differently across the GJT and the free written response task. They interpreted this discrepancy as evidence that two tasks measure distinct constructs to some extent. Still, however, given their own admission that the GJT is not a pure measure of explicit knowledge, their findings do not allow a clear determination as to whether the results reflect a genuine explicit-implicit knowledge difference or whether implicit knowledge partly contributed to GJT performance. Similarly, Pawlak and Biedroń (2021) illustrate both good practices and potential measurement concerns. They used four different tasks to operationalize the productive and receptive dimensions of IE knowledge while explicitly acknowledging serious methodological limitations, particularly the poor internal consistency of their receptive measures ($\alpha = .55$ for explicit, $\alpha = .44$ for implicit). They also openly discussed the potential lack of unidimensionality in these receptive measures. At the same time, their correlational analyses suggested that phonological short-term memory was associated with implicit productive knowledge (measured by a focused communication task), while working memory capacity was associated with explicit productive knowledge (measured by a verb form completion task). However, their regression analyses showed that these relationships were weak and substantially mediated by learners' general grammatical proficiency. Thus, although the study is notably transparent about its limitations and employs multiple measures, it still leaves unresolved the interpretive problem of drawing construct-level conclusions about IE knowledge from measures whose psychometric adequacy is partly in question.

Third, even when a battery is used, statistical and interpretive errors persist. Akakura (2012) serves as a prime example. Although she administered four tasks (EIT, oral production, GJT, metalinguistic knowledge) and conducted PCA, her validation logic remains open to question. PCA produces weighted composites and follows a formative, data-reduction logic rather than testing a reflective latent-variable model. As such, it does not directly evaluate the theoretical assumption that implicit and explicit knowledge are latent traits.[2] Furthermore, the correlations between the measures taken to represent IE knowledge raise concerns about discriminant validity. Even so, she interpreted the results at the task level, claiming that explicit instruction improved *implicit knowledge* simply because scores on the elicited imitation task increased.

[2]As emphasized throughout this paper (see also Fukuta, 2025; Tamura & Fukuta, in press), the IE distinction is more coherently treated as a formative model rather than a reflective one. Crucially, while Akakura (2012) adopts a reflective, latent-trait interpretation, she validates her measures using PCA, which is a formative method. Our critique, therefore, targets the internal inconsistency between her theoretical assumption (reflective) and statistical method (formative), not the use of PCA per se. Fundamentally, the IE distinction lacks the sound reflective grounding she assumes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

Finally, even methodologically sophisticated studies are not immune to unresolved task-to-construct mapping problems. Godfroid (2016) represents a much better case in terms of methodology and transparency. Unlike Khezrlou (2021) or Pawlak and Biedroń (2021), Godfroid used three different measures across different time points as well as concurrent RT measures during learning, employing awareness interviews to identify truly unaware learners. This methodological triangulation clearly demonstrates attention to the risk of using a single measure. Nevertheless, the central task-to-construct mapping remained assumed rather than independently validated within the study. The word monitoring task and controlled oral production task were still treated as single measures of implicit and explicit knowledge, respectively. This does not invalidate the study's empirical findings, but it does limit the extent to which the results on those tasks can be interpreted as direct evidence for distinct underlying knowledge systems. Adequate validation of such task-to-construct mappings would ideally require multiple indicators per construct rather than reliance on a single task, together with evidence that these indicators behave as a reflective model, for example, stable factor structures and discriminant validity in confirmatory factor analysis. Such psychometric evidence is necessary but not sufficient, however. As noted above, covariation among indicators does not by itself establish that a latent construct causally produces them (Borsboom et al., 2004; Fukuta et al., 2023). What would close this gap is evidence of that causal link, such as showing that an independently identified or manipulated construct yields the predicted pattern across multiple indicators, rather than inferring the construct from task covariation alone.

Taken together, these cases illustrate recurring instances of validity threats. The test battery fallacy is not merely an occasional oversight; rather, it represents a fundamental methodological problem in SLA research. Whether through arbitrary extraction, acknowledge-but-proceed labeling, or pragmatic compromise, the persistent reliance on individual tasks, even within sophisticated designs, undermines the very psychometric logic that the IE framework rests upon. This kind of research practice suggests that simply combining tests into a battery does not guarantee construct validity if the interpretation of those tests reverts to single-task reliance.

Method variance and the measurement illusion

Even if we set aside the ontological problems identified in the preceding subsection and accept correlation-based validation as sufficient, the IE framework faces an additional threat to construct validity. The test battery fallacy is closely related to the method variance issue raised decades ago by Campbell and Fiske (1959). In their classic analysis of convergent and discriminant validity, they warned that if measurement variance (i.e., method effects) exceeds convergent validity, the resulting discriminant validity may be illusory. This argument is particularly relevant to recent psychometric studies using reaction-time (RT) measures (e.g., Godfroid & Kim, 2021; Suzuki, 2017; Vafaee et al., 2017). The observed factorial separation might reflect methodological differences, such as reaction-time versus accuracy-based tasks rather than genuine construct differences between implicit and explicit knowledge.

In fact, even proponents of the IE framework have expressed skepticism regarding the construct validity of these measures. Ellis and Roever (2021) pointed out that psycholinguistic tasks such as word monitoring, self-paced reading, and the visual-world paradigm showed weak inter-correlations, despite being theoretically grouped as measures of implicit knowledge. They also reported that these tasks typically exhibit greater variability and lower stability compared to judgments or metalinguistic tests.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

Such patterns suggest that factorial separation may stem from the measurement characteristics of RT tasks rather than their access to a qualitatively distinct type of linguistic knowledge.

Godfroid and Kim (2021) also provided counterevidence against treating RT-based tasks as measures of implicit knowledge. Their best-fitting model was a two-factor model (implicit and explicit knowledge), and, surprisingly, both accuracy-based measures such as timed GJTs and RT-based measures such as SPR and WMT loaded onto the implicit knowledge factor, which was predicted by a measure of implicit-statistical learning aptitude. What is more, they found that WMT and SPR correlated weakly (rs = .178, p = .098) and that both loaded poorly onto the implicit knowledge factor. They concluded that performance on the RT-based tasks was less reliable than performance on the accuracy-based tasks, and that RT-based tasks therefore "are not a homogeneous whole" (p. 629). From the standpoint of validity, they recommended interpreting WMT and SPR separately rather than grouping them together as measures of a single construct.

Even the WMT, which Godfroid and Kim (2021) singled out as the more promising of the two reaction-time measures, does not clearly behave as an implicit measure when examined directly. Suzuki et al. (2023b), whose first author is a proponent of RT-based measures of implicit knowledge, investigated the validity of the WMT using both behavioral and neural (fMRI) evidence. They found that brain regions related to procedural memory were activated when L2 speakers processed ungrammatical sentences, whereas regions related to declarative memory were not associated with the grammatical sensitivity index (GSI). They interpreted this neural evidence as supporting the hypothesis that the WMT is a potential measure of implicit knowledge. Nevertheless, GSI was predicted not by their implicit learning aptitude measure (a serial reaction time task) but by their explicit, language-analytic aptitude measure (LLAMA_F, r = .44), and no aptitude measure predicted brain activation. This is the reverse of the pattern Godfroid and Kim (2021) reported. Based on these findings, Suzuki et al. (2023b) acknowledged that "[t]he word-monitoring task (itself) cannot be simply considered as an implicit or explicit knowledge test, as its completion is likely to recruit different types of knowledge depending on learners' proficiency and experience" (p. 131).

Beyond their contested construct validity, these measures also raise questions of practical utility. Ellis and Roever (2021) argued that while psycholinguistic tasks may be suitable for experimental purposes, they lack face validity as language tests. They questioned the value of these instruments if the goal is to design *standalone tests* of IE knowledge. This terminology deserves closer examination. If researchers expect individual tests to function independently with construct validity, that is, to directly measure implicit or explicit knowledge without requiring triangulation from multiple instruments, this contradicts the foundational premise of the test battery approach, which assumes that no single test directly measures implicit or explicit knowledge. The battery was designed precisely because individual indicators are indirect and require triangulation. Criticizing tests for lacking standalone validity suggests a confusion about whether tests are meant to be direct measures (committing the test battery fallacy) or convergent indicators (following the battery logic).

This fundamental disagreement remains unresolved two decades after Ellis (2005a). While some researchers argue that psycholinguistic measures finally isolate true implicit knowledge (Suzuki, 2017; Vafaee et al., 2017), others urge caution in the use of the RT measures due to the weak coherence among them (Godfroid & Kim, 2021). Moreover, still others suspect that factorial separation reflects method variance rather than construct variance (Ellis & Roever, 2021). This measurement disagreement leads to the definitional inconsistency, which will be further discussed in the next section.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

More importantly, it also indicates that factor-analytic evidence alone cannot determine whether implicit and explicit knowledge represent distinct cognitive constructs or merely reflect families of tasks with different measurement properties.

Definitional inconsistency and its theoretical roots

The two preceding subsections demonstrated that researchers misapply psychometric logic by treating individual tasks as construct measures and that factorial evidence for the battery is itself unsettled: the apparent separation may be an artifact of method variance, and the reaction-time measures cohere only weakly among themselves. This section examines a related problem: even when researchers attempt to follow proper measurement principles, they cannot agree on which tasks validly measure implicit versus explicit knowledge. The same timed-grammaticality judgment task classified as an implicit measure in one study appears as an explicit measure in another (Godfroid & Kim, 2021; Suzuki, 2017). Individual tests validated in one population yield different factor structures in another context (Schurz, 2024). However, competing definitions of a single construct are themselves not uncommon, and their existence alone does not render the construct invalid. In the IE framework, by contrast, the definitions rely on the choice of tasks, and definitional differences are thereby reified into different task selections and classifications, manifesting as the measurement instability documented above rather than as merely terminological disagreement. This definitional instability is thus a symptom of the formative nature of IE. Nevertheless, it should be stressed that the formative nature itself does not undermine the usefulness of the IE framework.

The problem operates at two levels. At the theoretical level, researchers invoke different conceptual definitions without clarifying which aspect of the IE distinction their measures target: some emphasize awareness (i.e., conscious vs. unconscious; e.g., Rebuschat, 2013; Rebuschat & Williams, 2012), others processing characteristics (i.e., speed, automaticity; e.g., Ellis, 2005a; Suzuki, 2017), still others memory systems (i.e., declarative vs. procedural; e.g., Paradis, 2009). At the empirical level, these conceptual differences manifest as systematic disagreement about task classifications. We first examine the theoretical confusion at the heart of this instability (the unwarranted conflation of processing speed with consciousness), then document the empirical consequences: cross-study incomparability and persistent classification disagreements that resist resolution even through sophisticated validation efforts.

The IE framework in SLA research assumes implicit knowledge exhibits two defining features: it operates unconsciously and activates rapidly, producing the fluent, accurate performance characteristic of native speakers. This theoretical alignment (unconscious with fast, conscious with slow) grounds the widespread use of time pressure to operationalize the IE distinction (Ellis, 2005a). However, this assumed correspondence between speed and consciousness lacks empirical warrant (e.g., Maie & DeKeyser, 2020; Tamura, Harada, Kato, Hara, & Kusanagi, 2016). In cognitive psychology, this is called an alignment problem (Melnikoff & Bargh, 2018): features theoretically clustered together do not always co-occur empirically.

Using grammaticality judgment tasks combined with self-report awareness measures, Tamura et al. (2016) demonstrated that L2 learners exhibited unconscious knowledge of target grammatical structure (English tough-movement) while producing slow reaction times. This finding directly contradicts the framework's assumption that unconscious knowledge necessarily operates quickly. Furthermore, Maie and DeKeyser (2020) compared objective measures (reaction-time tasks) with subjective

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

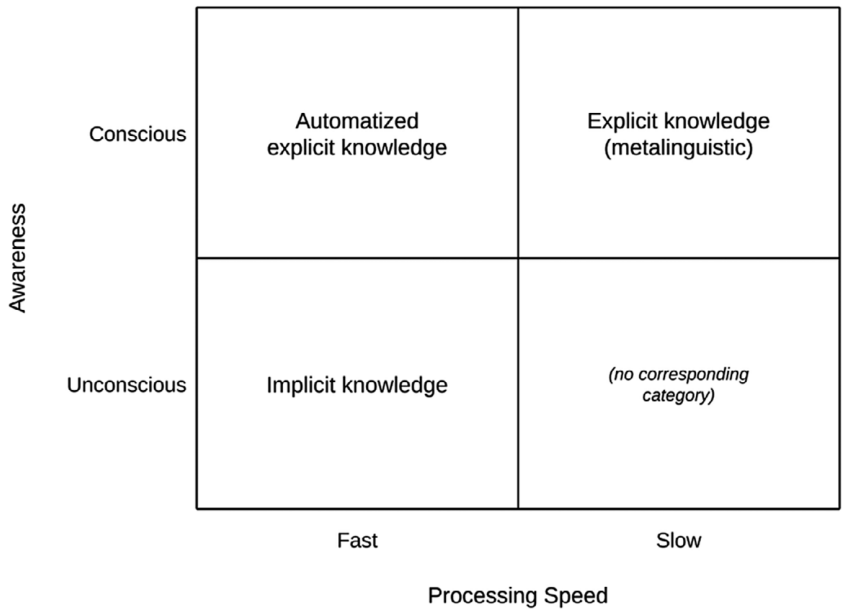


Figure 1. The alignment problem in the IE framework (speed × awareness).

measures of awareness and found conflicting patterns. Knowledge classified as implicit by one criterion appeared explicit by another, further illustrating that the field lacks a unified standard for validating these constructs.

This misalignment creates two empirical patterns the IE framework cannot accommodate (see Figure 1). First, as Suzuki and DeKeyser (2015) and Suzuki (2017) documented, explicit knowledge can become automatized through practice, producing fast, accurate performance in highly fluent L2 speakers. This represents conscious knowledge that operates with implicit-like speed. Second, as Tamura et al. (2016) demonstrated, learners can possess unconscious grammatical knowledge that activates slowly. Beyond these two patterns, and most critically, Maie and DeKeyser (2020) revealed that even when researchers attempt to classify knowledge, objective and subjective measures often yield contradictory results, suggesting that the very coordinates of this theoretical map remain unstable.

Figure 1 illustrates this complexity: the four quadrants show that fast/slow processing cross-cuts conscious/unconscious awareness, creating theoretical possibilities that the IE distinction collapses into false equivalences. If time-pressured tasks can reflect either unconscious knowledge (lower-left quadrant) or automatized conscious knowledge (upper-left quadrant), and if untimed tasks can reflect either conscious knowledge (upper-right quadrant) or slowly accessed unconscious knowledge (lower-right quadrant), then task performance alone cannot determine knowledge type. This conceptual confusion explains why the field cannot agree on task classifications, as the following examples demonstrate.

The conceptual ambiguity identified above produces systematic disagreement about what tasks measure. The most widely used case, GJTs, shows that this instability is not limited to disagreement across task labels but extends to the internal interpretation of a single task. Ellis (2005a) classified timed GJT as measuring implicit knowledge based on

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

factor-analytic results. Gutiérrez (2013) argued that within GJT, responses to grammatical sentences reflect implicit knowledge while responses to ungrammatical sentences reflect explicit knowledge—a claim that subdivides the same task into different construct measures depending on item type. Rebuschat and Williams (2012) offer a conceptually distinct source of instability. By incorporating trial-by-trial subjective measures such as source attribution (intuition, guessing, memory, and rule-based responses) and confidence ratings (binary or graded), their methodology presupposes that learners' awareness states can shift from one trial to the next within the same GJT session. In other words, the task itself is assumed to involve a mixture of conscious and unconscious processes, leaving no basis for treating GJT performance as indexing a single underlying construct. These are not minor methodological variations but fundamental disagreements about what the task measures.

Adding another layer of complexity, Mehraein, Marefat, and Karami (2022) found that, on GJTs, responses to easy structures loaded on a factor they labeled automatized knowledge, whereas responses to difficult structures loaded on an explicit knowledge factor, showing again that the same task is classified differently depending on the structure tested. If task classification varies by grammatical structure tested, the construct being measured is either structure-specific or the task does not measure a unified construct at all.

Elicited imitation demonstrates similar instability. While some studies classified EIT as implicit based on online processing demands (Ellis, 2005a; Erlam, 2006; Spada, Shiu, & Tomita, 2015; Zhang, 2015), other studies challenged this, finding that EI scores correlated with metalinguistic knowledge but not with implicit learning aptitude, and concluding that EI taps automatized explicit knowledge rather than implicit knowledge, since learners can consciously apply rules during repetition (Suzuki & DeKeyser, 2015). Furthermore, Suzuki et al. (2023a) found a comparable tension between behavioral and neural evidence. On the one hand, from a behavioral viewpoint, learners relied on automatized explicit knowledge during the EI task. On the other hand, fMRI data showed that both explicit and implicit knowledge were used during the task and interacted dynamically. Taken together, these findings suggest that it is difficult to conclude what the EI task measures.

The instability also extends to the target population and the broader research context. Schurz (2024) tested early-teenage EFL learners in Austria, a traditional, explicit EFL environment, and Sweden, a more implicit, exposure-rich one, asking whether the factor structure derived from multiple grammar tasks differs between these two learning environments. They did: a two-factor model distinguishing automatized/implicit knowledge from explicit knowledge fit best for the Austrian learners, whereas for the Swedish learners a single-factor model fit slightly better than the two-factor model, though not significantly so.

Recently, researchers have attempted to model automatized explicit knowledge as a distinct construct from implicit and non-automatized explicit knowledge, using confirmatory factor analysis (Mehraein et al., 2022, 2026; Schurz, 2024; Suzuki, 2017). Nevertheless, even these studies have not agreed upon the best-fitting factor model. Suzuki (2017) proposed that some of the measures originally classified as implicit knowledge may in fact be measures of automatized explicit knowledge. In this line of research, Mehraein et al. (2022, 2026) tested three-factor models separating implicit, automatized explicit, and (non-automatized) explicit knowledge. In Mehraein et al. (2026), the three-factor model achieved adequate fit, but automatized explicit and metalinguistic knowledge correlated almost perfectly (r = .99); in Mehraein et al. (2022), the three-factor model did not achieve adequate fit, and the two factors that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

emerged (automatized vs. explicit) correlated at $r = .93$. Either way, automatized explicit knowledge could not be sustained as a distinct construct. It should be noted that although these studies introduce automatized explicit knowledge as a theoretical construct, several of them conclude that a two-factor solution fits best, separating only implicit from explicit knowledge and grouping automatized explicit knowledge together with explicit knowledge rather than treating it as a separate factor (Mehraein et al., 2026). Schurz (2024) likewise set out to model these knowledge types, but the battery included no psycholinguistic task such as word monitoring or self-paced reading. As a result, implicit and automatized explicit knowledge were not separated, and the three-way distinction was never actually tested. In sum, although recent confirmatory work has tried to model these knowledge types, the discriminant validity among them remains unstable.

In this section, we have shown how unstable the IE framework is across multiple studies. At the theoretical level, the inconsistencies stem in part from different conceptualizations: some researchers focus on awareness, others on processing characteristics. This conceptual ambiguity produces instability in task classification, even in validation studies that rely on confirmatory factor analysis, where the same task is classified as a different construct depending on the linguistic structure or learning environment. Even though some recent studies have attempted to model automatized explicit knowledge as a distinct construct, they have not yet established discriminant validity (e.g., Mehraein et al., 2022, 2026). In conclusion, the review in this section suggests that explicit and implicit knowledge, as theoretical constructs, might not perform as stable latent variables.

The measurement-pedagogy dilemma

In the preceding section, we discussed the problem of extracting individual tasks from a test battery, which we termed the test battery fallacy. In this section, we turn to another structurally rooted problem in the IE framework. When researchers draw pedagogical implications from the IE framework, it is widely accepted that the development of implicit knowledge should be the primary concern of language instruction, while still acknowledging the role of explicit knowledge in the development of implicit knowledge (e.g., Ellis, 2005b). Once this goal was established, intervention studies naturally sought to examine whether instruction had succeeded in developing implicit knowledge, which are precisely the studies reviewed in the section on the test battery fallacy (e.g., Akakura, 2012; Andringa et al., 2011; Khezrlou, 2021; Michaud & Ammar, 2023).

The test battery developed by Ellis (2005a) was designed to establish construct validity for research purposes, but it provided the only available instruments for such investigations. As we pointed out earlier in this paper, however, the measurement basis required to support this goal is not well grounded, even in research contexts. This problem becomes even more serious in pedagogical settings, mainly because measures of implicit knowledge are difficult to implement in the classroom. Together, these issues constitute what we call the measurement-pedagogy dilemma.

This dilemma appears difficult to resolve within the IE framework: measures with strong construct validity (e.g., psycholinguistic tasks) lack ecological validity, while ecologically valid measures (e.g., oral production tasks) cannot reliably isolate implicit knowledge. Recent psychometric studies (Suzuki, 2017; Suzuki & DeKeyser, 2015) propose a three-way distinction: implicit knowledge, automatized explicit knowledge, and non-automatized explicit knowledge. This refinement, while enhancing construct precision for experimental research, reveals the fundamental measurement challenge facing pedagogical applications of the IE framework.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

If researchers cannot reliably distinguish automatized explicit knowledge from implicit knowledge using time-pressured tasks—because both produce fast, accurate performance—then the measurement instruments available to teachers (who lack access to laboratory equipment) cannot possibly make this distinction either. The problem becomes circular: psychometric validation studies designed to establish which tasks measure implicit knowledge increasingly resort to specialized psycholinguistic instruments (word monitoring, self-paced reading, visual-world paradigm) that demonstrate poor inter-correlations even when theoretically measuring the same construct (Ellis & Roever, 2021). These instruments may achieve greater construct purity by minimizing opportunities for conscious reflection, but they sacrifice the ecological validity needed for classroom application.

It is natural for teachers and researchers interested in instructional effectiveness to draw on the IE framework when evaluating the effects of their instruction or when seeking to understand the nature of their students' grammatical knowledge. However, measuring IE knowledge with the rigor that the IE framework demands is not feasible in educational settings. The assessment tools teachers can feasibly use, such as oral interviews, communicative tasks, and written production, allow sufficient processing time for automatized explicit knowledge to operate, making them unsuitable for isolating implicit knowledge according to the framework's own criteria.

This measurement difficulty extends beyond technical limitations to practical indeterminacy. Consider two learners who produce accurate relative clauses during a classroom speaking activity: one has acquired the structure implicitly through extensive reading, while the other has automatized explicitly learned rules through deliberate practice. Both produce fluent, accurate performance under real-time communicative pressure. What assessment could a teacher administer to distinguish their underlying knowledge types? A timed grammaticality judgment might reveal processing speed, but fast performance could reflect either implicit knowledge or well-automatized explicit knowledge. An oral production task captures authentic performance but cannot determine cognitive source. A metalinguistic interview might identify conscious rule knowledge, but learners with implicit knowledge may still be able to articulate rules they have induced through pattern recognition.

The three-way distinction thus highlights a paradox: the more refined the theoretical taxonomy becomes, the more demanding the measurement requirements, and the less feasible pedagogical application. Suzuki and DeKeyser's (2015) proposal aimed to resolve measurement ambiguity in the IE framework, but it reveals instead that the framework imposes analytical requirements that are increasingly demanding even for controlled research settings, let alone for evaluating instructional effectiveness.

The double bind

The measurement problem identified in the two preceding sections are not independent difficulties. Together, they point to a deeper conceptual problem that we call the double bind: the IE distinction meets neither the practical demands of pedagogical application nor the explanatory demands of scientific inquiry. Identifying this double bind, however, does not lead us to discard the IE distinction. Used for appropriate purposes and under the explicit conditions we set out in the Conditional principles section, the framework retains genuine value. We consider each side in turn.

From a pedagogical perspective, the distinction between implicit and automatized explicit knowledge is irrelevant. The two learners described in the preceding section who produce accurate relative clauses in spontaneous communication are functionally indistinguishable from an instructional standpoint. Whether performance derives from

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

implicit knowledge or from rules automatized through practice has no bearing on instructional decision-making. As DeKeyser (2020) acknowledges, the IE distinction "in this narrow sense should be of little concern to second language learners and teachers" (p. 97) precisely because it does not map onto pedagogically consequential differences in performance. Instructional assessment should employ ecologically valid performance measures such as oral narratives, role-plays, and communicative tasks, without attempting to decompose performance into underlying knowledge components.

These difficulties, however, are not simply a consequence of extending the IE framework beyond its intended scope. The measurement problems identified above point to a separate question about the framework's adequacy for scientific inquiry. Addressing this question requires distinguishing between two fundamentally different approaches to measurement: reflective models and formative models (Alamer et al., 2024). Reflective models assume that latent constructs causally produce observable indicators: test performance reflects pre-existing knowledge states (Borsboom et al., 2004; Fukuta et al., 2023). Under this logic, sufficient validation should yield consensus about which tasks measure which constructs, because researchers are measuring the same underlying entities. Formative models, by contrast, define constructs operationally through researcher choices about measurement procedures: the construct is constituted by the selected indicators rather than existing independently of them. As Fukuta (2025) argues, the IE framework exhibits formative characteristics: researchers cannot agree on task classifications because the constructs themselves are defined by whichever performance patterns researchers include in their operational definitions. Different task selections do not merely measure the same constructs differently; they instantiate different construct definitions.

The IE framework is often treated as operating under a reflective measurement logic, in which shared variance among multiple tasks provides evidence for underlying constructs. However, the epistemological status of those constructs, whether they represent distinct cognitive systems or merely useful categories for organizing performance data, remains underspecified. One way to understand this underspecification is through the distinction between substantial and instrumental interpretations of dual-process theories, as identified by Augusto (2024). Substantial interpretations make realist claims about cognitive architecture, positing that theoretical distinctions correspond to actual, separate mental systems or representations with distinct neural substrates and processing characteristics. Instrumental interpretations, conversely, treat theoretical distinctions as useful organizing frameworks for empirical research without necessarily claiming that they reflect distinct cognitive mechanisms. An instrumental interpretation remains agnostic about underlying mental architecture, using the distinction as a heuristic for organizing behavioral observations. Under this instrumental interpretation, IE knowledge can function as disposition concepts rather than theoretical concepts in the sense presupposed by a substantial interpretation (Tamura and Fukuta, in press).

Building on Augusto's framework, Tamura and Fukuta (in press) demonstrate how this epistemological distinction applies to competing measurement traditions in SLA research. They argue that the CU framework exemplifies a substantial interpretation: it grounds the dichotomy in mental states—unconscious knowledge is genuinely inaccessible to awareness even under unlimited time and explicit prompting, while conscious knowledge consists of reportable metalinguistic information (Rebuschat, 2013; Williams, 2005). The IE framework, conversely, defines the distinction operationally through task performance: implicit knowledge is what gets used in time-pressured, meaning-focused tasks while explicit knowledge is what gets accessed in untimed,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

form-focused tasks (Ellis, 2005a). This operational approach exemplifies an instrumental interpretation—it provides a functional categorization based on observable behavior without necessarily committing to claims about separate mental representations or distinct cognitive systems.

Fukuta (2025) further shows that the difference between IE and CU extends beyond their formative or reflective character to the direction of inference between performance and mental states. The IE framework adopts a performance-first approach, inferring mental states from task performance patterns and treating consciousness as merely one factor among many that might distinguish those patterns. The CU framework reverses this priority, taking consciousness as the foundational construct and examining how it causally explains performance differences.

One might object that this reversal simply relocates the measurement problem. If the distinction is grounded in mental states rather than task performance, how are those mental states themselves to be accessed? The objection, however, treats CU as a measurement alternative to IE. The shift from IE to CU is not a change of instruments addressing the same question, but a change in what the measurement is designed to detect. What characterizes CU research is not the assumption that rapid, time-sensitive, or systematic responses can be taken as evidence of implicit knowledge, but a strategy that targets awareness itself: how conscious access can be excluded, isolated, or experimentally manipulated. To this end, CU research employs paradigms such as dual-task procedures with blinding, subliminal or masked stimulus presentation, and the visual-world paradigm, along with other techniques designed to restrict or detect conscious access to the stimulus. Performance is then interpreted not in terms of how fast or accurate it is, but in terms of whether and how it depends on conscious access. In this way, the CU framework is addressed to scientific inquiry into cognitive mechanisms, the purpose of which is to explore the mental state of knowledge, not to provide guidance for classroom practice.

Fukuta's (2026) Hierarchical-Layer (H-L) Model of Conscious-Unconscious Knowledge offers one theoretical framework for pursuing CU-based inquiry in SLA by treating awareness as the foundational construct rather than inferring it from performance patterns. This emphasis on awareness is not new to the model but continues a line of research on learning without awareness, including Williams (2005), Rebuschat and Williams (2012), and Rebuschat (2013). That is, the H-L Model is a recent formalization of a research tradition that has taken a particular interest in the role of awareness in language learning. Drawing also on work in cognitive psychology, the model posits that linguistic knowledge comprises two components: structural knowledge, which captures regularities and supplies the basis for judgments, and judgment knowledge, the evaluative sense that a given item conforms to those regularities. Each component may be either conscious or unconscious, and this combination gives rise to distinct mental states that may underlie similar observable performance.

The H-L Model also helps address a specific question about how the CU framework treats constructs like automatized explicit knowledge and low-confidence explicit knowledge.[3] Low-confidence explicit knowledge can be understood as a combination of conscious structural knowledge and judgment knowledge that is either unconscious

[3]This question was raised by an anonymous reviewer. It presupposes the IE framework's taxonomic logic, in which knowledge types can be represented along crossed dimensions of speed (fast/slow) and awareness (conscious/unconscious). The necessity of differentiating automatized explicit knowledge or low-confidence explicit knowledge from implicit knowledge arises from this taxonomy, not from the CU framework itself. The CU framework adopts a different analytic priority: rather than classifying performance along

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

or, when conscious, decoupled from accuracy. Automatized explicit knowledge, by contrast, does not map cleanly onto a single mental state. When explicit rules are retained consciously and their application simply speeds up, the resulting performance reflects efficiency gains within a rule-based metalinguistic state. When judgment becomes fast and intuitive, however, this does not necessarily indicate that the original explicit rule has become unconscious; the performance may instead rely on independently developed unconscious structural knowledge supporting conscious judgment. From the perspective of the H-L Model, the IE construct of automatized explicit knowledge thus conflates mental states that cannot be distinguished from observable behavior alone.

The conceptual coarseness revealed by the H-L Model is not resolved when researchers try to refine constructs within the IE framework itself. Suzuki and DeKeyser's (2015) three-way taxonomy, for instance, treats automatized explicit knowledge as distinct from both implicit knowledge and non-automatized explicit knowledge, attempting to add construct precision on this basis. Yet this refinement remains within the instrumental logic identified above. It provides organizational utility for classifying task performance, but scientific explanation of cognitive architecture requires realist commitments about the mental systems that produce such performance. Nor does the three-way taxonomy provide pedagogically useful distinctions. When instructional studies invoke the IE distinction to explain teaching outcomes, both forms of inadequacy directly affect how findings are interpreted. A study reporting that explicit instruction enhanced explicit knowledge but not implicit knowledge communicates that learners improved on untimed metalinguistic tasks but not on time-pressured production measures. But what does this pattern mean for instruction? The labels, *implicit knowledge* and *explicit knowledge*, or *automatized explicit knowledge* if a measure targeting it were included, add no explanatory depth beyond what the task descriptions already convey, and the task descriptions themselves (performance on untimed metalinguistic judgment versus time-pressured production) provide teachers with little information about learners' actual communicative abilities.

The problems identified above are not separate pedagogical and scientific difficulties but manifestations of a single structural issue: researchers attempt to use an instrumental framework (IE framework) for purposes requiring either substantial claims (scientific explanation) or practical evidence (pedagogical guidance). As Fukuta (2025) observes, the epistemological difference between the IE and CU frameworks is often overlooked, which makes it "challenging to determine which framework is most appropriate for the specific research objectives" (p. 76). What is needed, then, is a research program that takes the epistemological difference between IE and CU seriously and matches methodological choices to the specific purposes of inquiry.

For scientific inquiry pursuing explanatory adequacy about cognitive architecture, frameworks grounded in mental states rather than task performance would likely be better suited. Fukuta's (2026) H-L Model, introduced above, offers one research program in this direction. Readers interested in a detailed discussion of the H-L Model are referred to Fukuta (2026), as such a discussion is beyond the scope of the present paper. Faced with the limitations of the IE framework, one might consider supplementing it with CU-based methods rather than abandoning it altogether. However, such hybrid approaches combining IE and CU methods fall short of epistemological

---

the speed × awareness grid (see Figure 1), it asks whether learners have conscious access to the knowledge underlying their performance, and how that access shapes what they do.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

coherence. As argued above, the two frameworks adopt fundamentally different inferential directions, performance-first inference versus awareness-centered analysis. Partial integration of them does not resolve the problems this paper has identified.

For pedagogical application, evidence that learners can communicate effectively matters more than decomposition of performance into cognitive components. The IE distinction, as an instrumental framework, retains value for organizing empirical work and facilitating professional communication, but this utility should not be confused with scientific or pedagogical necessity. Researchers who choose to employ the IE framework must therefore attend to specific conditional principles we develop in the next section.

Conditional principles

Even though the IE distinction has the structural problems discussed above, we do not argue that the IE distinction should be abandoned now and left behind. Rather, we propose three conditional principles for researchers who choose to employ the IE distinction, believing that it can still meaningfully contribute to SLA research or broader language teaching research, despite the limitations. It should be noted that our aim is to establish minimum standards for responsible use of the IE framework and that these are conditional principles, which only apply *if* researchers choose the IE framework.

Principle 1: Conceptual transparency

The first and foremost principle is that researchers need to be clear about their epistemological stance—the interpretation they adopt of the IE distinction. As discussed in the Double bind section, this distinction can be understood in two fundamentally different ways: as an instrumental framework for organizing observations or as a substantial claim about cognitive architecture (Augusto, 2024; Tamura & Fukuta, in press).

For researchers adopting an instrumental perspective, it is important to recognize both the value and limitations of this approach. As Augusto (2024) argues, instrumental interpretations of theoretical constructs can serve important coordinating functions even when substantial claims remain uncertain or unwarranted. Applied to the IE framework, this means the distinction retains value in providing shared terminology for discussing performance differences and organizing empirical findings without requiring commitment to strong claims about cognitive architecture (Tamura & Fukuta, in press).

This instrumental function should not be dismissed as valueless. Reports about which knowledge type benefited from instruction still convey useful information about task performance patterns, even when the underlying constructs remain underspecified. When teachers discuss activities that promote implicit learning, they refer to meaning-focused, input-rich tasks regardless of whether these actually develop knowledge that is genuinely unconscious versus merely automatized. This communicative utility, such as organizing research findings and facilitating professional discourse, represents the IE framework's primary contribution.

However, researchers who take this instrumental perspective must acknowledge that their investigations are limited in terms of scientific explanation. The framework provides a vocabulary for describing performance patterns but does not resolve questions about underlying cognitive mechanisms. For instance, it cannot determine whether fluent L2 production derives from genuinely unconscious knowledge or from

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

automatized explicit rules—a distinction that may be irrelevant for pedagogical purposes but crucial for theories of, for example, ultimate attainment.

Conversely, researchers who adopt a substantial perspective—treating the IE distinction as reflecting real differences in cognitive architecture—face a different set of obligations. They need to overcome the problems identified in this paper and justify what explanatory power their research adds beyond existing frameworks. This requires addressing the measurement instability documented above. Without such justification, substantial claims about implicit versus explicit knowledge lack empirical foundation.

Principle 2: Methodological honesty

The second principle is methodological honesty, which many researchers already practice to some extent: acknowledging the measurement limitations of their specific study. Researchers should be cautious when claiming that particular tasks measure implicit or explicit knowledge, even when previous empirical studies have made such claims. As reviewed earlier, there is an ongoing debate about which tasks measure which type of knowledge, and no consensus has been reached. Under these circumstances, researchers should remain restrained in assuming that a single task taps one type of knowledge rather than the other.

The test battery fallacy must also be considered. Even if researchers believe that timed and untimed GJTs tap different types of knowledge, it is premature to assert that any single measure cleanly indexes a specific construct. Psychometric approaches imply that a battery of tasks, not an isolated task, provides evidence for distinct knowledge types. Just as psychologists do not extract a single questionnaire item, analyze it in isolation, and claim that it represents a psychological construct, SLA researchers who adopt the IE framework should avoid extracting one task from a test battery and assuming it measures either implicit or explicit knowledge. What matters is the variance–covariance structure across multiple tasks, which researchers then interpret and label.

Principle 3: Pedagogical justification

As discussed above, some intervention studies adopt the IE distinction to examine the effects of instructional treatments on learners' grammatical knowledge. For such studies, it is essential to articulate why the IE distinction is relevant to the instructional question at hand. Simply assuming that learners' grammatical knowledge falls into implicit and explicit categories, and that available measures accurately tap these constructs, does not guarantee that the IE framework is useful or that the research question derived from it is meaningful. Researchers must ask what specific pedagogical decisions depend on distinguishing these knowledge types. As noted in the Double bind section, the more practical the instructional implications become, the less useful the IE framework tends to be, because performance-level characteristics do not reliably reflect the source of learners' knowledge. Slow performance may originate from slow, unconscious knowledge, while fast performance may stem from automatized or speeded-up explicit knowledge.

Because the knowledge type can be inferred only from observable performance, with no independent criterion for identifying it, using a performance-derived label to explain the very performance from which it was inferred risks circular reasoning. Consider a teacher who observes that one of her students performs well on a certain grammatical structure in an error-correction task but struggles to produce it correctly

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

in an impromptu speaking task. Assuming the implicit/explicit dichotomy, suppose the teacher infers that the student has explicit but not yet implicit knowledge. When asked why, the teacher might answer that it is because error-correction performance is good while that of the speaking task is not. This reasoning is a case of tautology: the label is derived from the very performance difference that it is then used to explain. When implicit and explicit knowledge function as disposition concepts, they describe performance patterns rather than cause them, so invoking them as explanatory mechanisms reproduces this circularity (Tamura & Fukuta, in press).

Moreover, from a pedagogical standpoint, distinguishing implicit and explicit knowledge in classroom settings is not straightforward. As reviewed above, recent validation studies such as Suzuki (2017) and Vafaee et al. (2017) converge on the idea that psycholinguistic tasks involving reaction-time measures are promising candidates for assessing implicit knowledge. However, it is neither realistic nor feasible to use self-paced reading, word-monitoring, or eye-tracking tasks as outcome measures in everyday teaching practice. Students may initially enjoy these tasks because they are novel, but their relevance to typical classroom activities—reading passages, completing grammar exercises, participating in communicative tasks, or practicing pronunciation—is uncertain.

Furthermore, implicit knowledge in the sense commonly assumed in the IE framework (fast and unconscious) does not typically develop within the short time frames of most intervention studies. Therefore, even if researchers aim to investigate the development of this type of implicit knowledge, standard intervention designs may not be capable of capturing it. In such cases, although investigating implicit knowledge may still be a worthwhile research question, the methodological constraints of classroom-based studies need to be acknowledged explicitly.

Conclusion

This paper has offered a critical examination of the IE knowledge distinction, arguing that its uncritical application to language teaching research obscures rather than illuminates the relationship between instruction and acquisition. As we have demonstrated, the framework faces two interrelated problems that compromise its scientific and pedagogical utility. First, systematic measurement instability manifests through the test battery fallacy (treating composite indicators as standalone measures), method variance confounds (factorial separation potentially reflecting measurement artifacts), and definitional inconsistency (persistent disagreement about task classifications). Given that recent confirmatory approaches have not yet established discriminant validity among these constructs, they might not perform as stable latent variables (see the section on definitional inconsistency). These patterns are likely due to the IE distinction's formative nature, in that it functions as an operational classification defined by researcher choices rather than a reflective measure of pre-existing knowledge states. Second, the measurement-pedagogy dilemma amounts to a trade-off between psychometric precision and ecological validity: measures that better isolate implicit knowledge become increasingly distant from authentic classroom performance, while ecologically valid measures leave indeterminate which type of knowledge they tap into. These two problems converge on a double bind. On the one hand, when the IE framework is seen as instrumental, as discussed above, it does not support the realist explanation required for scientific inquiry into cognitive architecture. On the other hand, the IE framework lacks the practical utility required for pedagogical assessment, where simple performance-based measures suffice.

These challenges suggest that the IE distinction often promises more explanatory power than it can empirically deliver. However, acknowledging these limitations does

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

not require abandoning the framework entirely. Instead, we have proposed three conditional principles to guide researchers who choose to employ this distinction: conceptual transparency, methodological honesty, and pedagogical justification. By clarifying their epistemological stance (instrumental vs. substantial), refraining from treating individual tasks as pure measures of latent constructs, and explicitly justifying the pedagogical relevance of the distinction, researchers can mitigate the risks of misinterpretation.

Ultimately, the value of the IE distinction lies not in its ability to map the exact neural or cognitive architecture of the learner's mind, but in its potential to serve as a heuristic for organizing our understanding of L2 performance. By moving away from the pursuit of *true* measurement toward a more modest, instrumental use of the framework, the field can better align its theoretical models with the practical realities of the language classroom.

Acknowledgments. Junya Fukuta is now affiliated with the Faculty of Letters, Gakushuin University, Japan. We would like to thank the editor and the three anonymous reviewers for their constructive comments and suggestions, which greatly improved this manuscript. Any remaining errors are our own. During the preparation of this manuscript, the authors used ChatGPT and Claude between September 2025 and July 2026 to translate drafts written by the authors in Japanese into English, and to refine the wording and style of English text written by the authors. The tools were not used to generate the substance of the argument, to analyze data, or to produce references. The authors reviewed and edited all output, verified all claims and citations against the original sources, and take full responsibility for the content of this article.

Funding statement. This work was supported by JSPS KAKENHI Grant Number JP25K00481.

Competing interests. The authors declared no potential conflicts of interest with respect to the research, authorship, and/ or publication of this article.

References

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research*, *16*(1), 9–37. https://doi.org/10.1177/1362168811423339

Alamer, A., Schuberth, F., & Henseler, J. (2024). When and how to use confirmatory composite analysis (CCA) in second language research. *Studies in Second Language Acquisition*, *46*(2), 597–616. https://doi.org/10.1017/S0272263124000020

Andringa, S., de Glopper, K., & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning*, *61*(3), 868–903. https://doi.org/10.1111/j.1467-9922.2010.00623.x

Augusto, R. (2024). Two kinds of process or two kinds of processing? Disambiguating dual-process theories. *Review of Philosophy and Psychology*, *15*(1), 277–298. https://doi.org/10.1007/s13164-023-00673-1

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning*, *28*(1), 69–83. https://doi.org/10.1111/j.1467-1770.1978.tb00305.x

Bialystok, E. (1979). Explicit and implicit judgments of L2 grammaticality. *Language Learning*, *29*(1), 81–103. https://doi.org/10.1111/j.1467-1770.1979.tb01053.x

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*(2), 247–271. https://doi.org/10.1017/S0272263110000756

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). Routledge.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

DeKeyser, R. M. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 83–104). Routledge, Taylor & Francis Group.

Dienes, Z. (2007). Subjective measures of unconscious knowledge. *Progress in Brain Research*, *168*, 49–64. https://doi.org/10.1016/S0079-6123(07)68005-4

Ellis, R. (2005a). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2005b). Principles of instructed language learning. *System*, *33*(2), 209–224. https://doi.org/10.1016/j.system.2004.12.006

Ellis, R., Loewen, S., Elder, C., Reinders, H., Erlam, R., & Philp, J. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters. https://doi.org/10.21832/9781847691767

Ellis, R., & Roever, C. (2021). The measurement of implicit and explicit knowledge. *The Language Learning Journal*, *49*(2), 160–175. https://doi.org/10.1080/09571736.2018.1504229

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*(3), 464–491. https://doi.org/10.1093/applin/aml001

Fukuta, J. (2025). Consciousness in second language research: How epistemological stances shape experimental design and interpretation. *Second Language*, *24*, 70–83. https://doi.org/10.11431/secondlanguage.24.0_70

Fukuta, J. (2026). Beyond a unidimensional view: The multilayered perspective on the implicit–explicit knowledge interface. *Second Language Research.* Advance Online Publication. https://doi.org/10.1177/02676583251411768

Fukuta, J., Nishimura, Y., & Tamura, Y. (2023). Pitfalls of production data analysis for investigating L2 cognitive mechanism: An ontological realism perspective. *Journal of Second Language Studies*, *6*, 95–118. https://doi.org/10.1075/jsls.21013.fuk

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, *38*(2), 177–215. https://doi.org/10.1017/S0272263115000388

Godfroid, A., & Kim, K. M. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606–634. https://doi.org/10.1017/S0272263121000085

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*(3), 423–449. https://doi.org/10.1017/S0272263113000041

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*(1), 1–23. https://doi.org/10.1177/136216889800200102

Isemonger, I. M. (2007). Operational definitions of explicit and implicit knowledge: Response to R. Ellis (2005) and some recommendations for future research in this area. *Studies in Second Language Acquisition*, *29*(1), 101–118. https://doi.org/10.1017/S0272263107070040

Khezrlou, S. (2021). Explicit instruction through task repetition: Effects on explicit and implicit knowledge development. *Language Awareness*, *30*(1), 62–83. https://doi.org/10.1080/09658416.2020.1866590

Kim, J., & Nam, H. (2017). Measures of implicit knowledge revisited: Processing modes, time pressure, and modality. *Studies in Second Language Acquisition*, *39*(3), 431–457. https://doi.org/10.1017/S0272263115000510

Kusanagi, K. (2018). Gaikokugo no bunpō chishiki ni okeru ichigensei no kenshō: Bunpōsei handan no seitōritsu, hannō jikan, shukanteki hensū o taishō ni [Testing the unitary nature of grammatical knowledge in foreign languages: Accuracy, response time, and subjective variables in grammaticality judgment] [Doctoral dissertation, Nagoya University]. Nagoya University Repository. http://hdl.handle.net/2237/00027747

Leung, J. H. C., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form–meaning connections. *Language Learning*, *62*(2), 634–662. https://doi.org/10.1111/j.1467-9922.2011.00637.x

Maie, R., & DeKeyser, R. M. (2020). Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition*, *42*(2), 359–382. https://doi.org/10.1017/S0272263119000615

Mehraein, S., Marefat, H., & Karami, H. (2022). L2 English knowledge: Implicit, explicit, or automatized? The role of structure difficulty. *The Journal of AsiaTEFL*, *19*(1), 17–34. https://doi.org/10.18823/asiatefl.2022.19.1.2.17

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946

Mehraein, S., Marefat, H., & Karami, H. (2026). Exploring implicit, automatized, and explicit knowledge: Testing competing CFA models. *Research Methods in Applied Linguistics*, *5*(1), 100303. https://doi.org/10.1016/j.rmal.2026.100303

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Michaud, G., & Ammar, A. (2023). Explicit instruction within a task: Before, during, or after? *Studies in Second Language Acquisition*, *45*(2), 442–460. https://doi.org/10.1017/S0272263122000316

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins.

Pawlak, M., & Biedroń, A. (2021). Working memory as a factor mediating explicit and implicit knowledge of English grammar. *Annual Review of Applied Linguistics*, *41*, 118–125. https://doi.org/10.1017/S0267190521000052

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626. https://doi.org/10.1111/lang.12010

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*(4), 829–856. https://doi.org/10.1017/S0142716411000580

Schurz, A. (2024). Measuring teenage learners' automatized, explicit, and/or implicit knowledge: A question of context? *Language Learning*, *74*(2), 506–541. https://doi.org/10.1111/lang.12624

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*(3), 723–751. https://doi.org/10.1111/lang.12129

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229–1261. https://doi.org/10.1017/S014271641700011X

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*(4), 860–895. https://doi.org/10.1111/lang.12138

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023a). fMRI reveals the dynamic interface between explicit and implicit knowledge recruited during elicited imitation task. *Research Methods in Applied Linguistics*, *2*(2), 100051. https://doi.org/10.1016/j.rmal.2023.100051

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023b). An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral and neural processing. *Studies in Second Language Acquisition*, *45*(1), 109–136. https://doi.org/10.1017/S0272263122000043

Tamura, Y., & Fukuta, J. (in press). Psychological approaches to "dual-system model" in SLA: A critical overview. In J. Fukuta, J. Matthews, & S. Wakabayashi (Eds.), *Theoretical issues in second language research: Challenges and new directions* (pp. 124–148). John Benjamins.

Tamura, Y., Harada, Y., Kato, D., Hara, K., & Kusanagi, K. (2016). Unconscious but slowly activated grammatical knowledge of Japanese EFL learners: A case of tough movement. *ARELE*, *27*, 169–184. https://doi.org/10.20581/arele.27.0_169

Vafaee, P., Suzuki, Y., & Kachinske, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, *39*(1), 59–95. https://doi.org/10.1017/S0272263115000455

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*(2), 269–304. https://doi.org/10.1017/S0272263105050138

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*(3), 457–486. https://doi.org/10.1017/S0272263114000370

Cite this article: Tamura, Y., & Fukuta, J. (2026). A critical examination of the relationship between explicit and implicit knowledge and language teaching. *Studies in Second Language Acquisition,* 1–23. https://doi.org/10.1017/S0272263126101946

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101946