

RESEARCH ARTICLE

Complex adaptive interventions: The challenge ahead for instructed second language acquisition research

Phil Hiver¹ and Charlie Nagle²

¹Florida State University, USA and ²The University of Texas at Austin, USA

Corresponding author: Phil Hiver; Email: phiver@fsu.edu

Abstract

The effects of deliberately and selectively manipulating instructional conditions are at the heart of instructed second language acquisition (ISLA) research and, ideally, are designed to inform practice. Knowing how an intervention works, by what mechanisms and processes the treatment is beneficial—and for whom—are complex questions. In this piece, we problematize intervention-based research paradigms that do not account for context, individuals and their proactivity, or temporal variation. We highlight several key challenges that remain for ISLA research and propose a more reflexive approach to intervention that attends to these central considerations in implementing study designs.

Keywords: complex adaptive interventions; instructed second language acquisition; research–practice nexus; intervention research; implementation science; complex dynamic systems

Instructed second language acquisition (ISLA) research aims to uncover how deliberately and selectively manipulating instructional conditions influences learning. An increasing number of meta-analyses and synthetic reviews provide cumulative evidence for the consensus effects of particular instructional practices (e.g., Boers & Faez, 2023; Kang et al., 2019; Loewen & Sato, 2018; Ren et al., 2023; Saito & Plonsky, 2019; Vuogan & Li, 2023; Yanagisawa & Webb, 2021), in turn, informing and benefiting practice. Enacting knowledge through praxis this way advances our ethical imperative as researchers in an applied social science (Ortega, 2005). Knowing how to align instruction with students' learning by design leads many to pursue empirical evidence of what works best or what matters most (Education Endowment Foundation, 2023; What Works Clearinghouse, 2022). Indeed, pursuing such objective accounts through rigorous causal study designs such as randomized controlled trials (RCTs) remains the ultimate benchmark of implementation-based research in education (Eccles & Mittman, 2006; Nilsen, 2020), as in all fields of inquiry. However, recent critiques

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<https://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

emphasize that much of this (quasi)experimental research ignores the role of individuals and of social and spatial contexts and does not “catch necessary complexity, contingency, ... situatedness and conditionality” (Morrison, 2021, p. 213) underlying effective instruction. Learning is also undeniably a developmental phenomenon, and the study of developmental processes involves grappling with nonstationary parameters and characteristics of data (Molenaar et al., 2014). However, most research to date has taken a relatively narrow approach to measuring the efficacy of an instructional intervention in relation to L2 learning outcomes rather than the process of L2 development, examining performance before and shortly after the intervention without considering the complex ways in which intervention may stimulate language development well beyond the closure of the immediate instructional window.

Understanding how effects under investigation in instructed L2 learning research might be affected by changes over time, vary across different socio-geographic environments, and be conditional on different instructional contingencies is a key part of making appropriate claims about the wider applicability of a study’s findings. Typically, instruction has been treated as a single, isolated, fixed set of practices enacted upon learners with the goal of quantifying average effects. Yet, the reality is far more complex. Thus, a second, but equally important, goal is to situate instruction within a developmental framework that prioritizes individual proactivity.

In this paper, we discuss the need to widen the aperture of what is typically thought of as (quasi)experimental research under the broader rubric of intervention science (i.e., any design that deliberately implements specific interventions, strategies, or treatments to influence an outcome or generate change) and reorient around complex and dynamic principles that can inform intervention research (Al-Hoorie et al., 2023; Verspoor & de Bot, 2022). We discuss a number of principles at play in ISLA research that warrant thinking of any such intervention designs as targeting effects that are inherently relational, situated, and adaptive. We frame these as a series of challenges the field will need to engage and confront in order to counter an “illusion of certainty” (Godfroid & Andringa, 2023, p. 12) in our research findings and advance the sophistication of intervention study designs.

Challenge 1: Effects are contingent on individuals and contexts

Decades of research show that the greatest source of variance in learning across most domains relates to learners themselves and what they bring to the learning equation as individuals (Hattie, 2023). Early arguments to this effect can be found in Cronbach’s (1975) work describing the inevitable limitation of (quasi)experimental designs, given the inherent interactions a treatment has with myriad psychosocial and personal characteristics—in his words, the “aptitude”—of the participant(s). While the need to account for aptitude–treatment interactions has seen broad uptake in ISLA research, the driving purpose for Cronbach’s original proposal seems to have been lost. Cronbach (1975), in fact, argued:

Instead of making generalization the ruling consideration in our research, I suggest that we reverse our priorities. An observer collecting data in one particular situation is in a position to appraise a practice or proposition in that setting,

observing effects in context [But] a general statement about a treatment effect is misleading because the effect will come or go depending on the kind of person treated. (pp. 119–125)

Treatment effects are expected to be multiple and heterogeneous (Bryan et al., 2021). Strategies to avoid aggregating sites, persons, stimuli, and trials in analytical models (e.g., by separating out random effects, estimating between- and within-learner variance, modeling time-variant and -invariant predictors) are a useful first step (Nagle, 2023). However, profiling L2 learners and providing a spectrum of distinct interventions based on those groupings are likely to lead to multiple aggregate effects that remain incomparable and context-bound (cf. Li et al., 2022).

Interventions are contested spaces that take place in complex and multilayered activity systems (Engeström, 2011). Indeed, recent work taking stock of RCTs in education has cautioned that aggregated findings mask the true variability that exists because the average treatment effects of educational interventions are, in fact, highly contingent on the individuals and the contexts in these accounts (Kaplan et al., 2020; Lortie-Forgues & Inglis, 2019). In essence, accounting for what works empirically is only possible when that research foregrounds and attends to the unique contextual and individual factors at play (Deaton & Cartwright, 2018; Hedges, 2018). Rather than simply adding nuance to observed effects, considerations of context and the individual are integral to interpreting and making sense of evidence-based practices. We acknowledge that examining multiple, situated levels of effects is challenging and, as such, demands a multipronged approach. Addressing heterogeneity directly at the level of the individual (e.g., through random effects structures) and at the level of the group or site (e.g., through multisite designs) is one approach that has gained traction. Yet, it is our position that this approach alone is insufficient because it continues to treat interventions as static, assuming that what works in one context for one group of learners should work (equally well) in another. In fact, it may be that an intervention works precisely because it was provided to learners in a given setting, at a specific level or stage of development, with a particular learning profile.

Parsimony in theory and research about treatment effects is valuable, and like other researchers, we recognize that the aims of applied social science include uncovering common patterns and transferable effects that extend beyond unique instances (Byrne & Callaghan, 2023). However, researchers can run into difficulties when looking for implications from statistical findings when there is an inherent mismatch between the conceptual and analytical units of analysis in research—what some statisticians refer to as a Type III error (Stark, 2022). To illustrate, the intervention sciences have problematized the practice of exclusively testing probabilistic and aggregate effects of treatments and training paradigms, all while assuming parity of groupwide or population-level participant characteristics across conditions (Cook, 2018). We should not expect to transfer findings of such effects to new settings, groups, and individuals with new sets of contextual characteristics unless we have carefully accounted for context and individuality in the effects under investigation (Hiver et al., 2022).

Yet, the practice of ignoring the “configuration of background and enabling conditions [that] provides necessary scaffolding for causal connections to occur” (Diener et al., 2022, p. 1104) is conspicuous in educational research that investigates

average, group-level treatment effects on processes and outcomes occurring at the level of the individual student—learning, engagement, and achievement. Implicitly assuming that group-based estimates, relationships, and inferential findings can generalize and apply equally to each individual, in effect flouting the ecological fallacy and ergodic principles, risks rendering that research little more than “irrelevant ... pseudo-applications” (Al-Hoorie et al., 2023, p. 280). Thus, “while rules of thumb or golden numbers ... [are] pedagogically appealing” in intervention research, they will vary in different contexts with different samples and are likely to require a high degree of contextual adaptation (Godfroid & Andringa, 2023, p. 7). Researchers are, indeed, often aware that heterogeneity exists in their data and attempt to handle it analytically by expanding the random effects structure of multilevel models or examining measurement invariance in latent variable models. Random effects can paint a more complete picture of heterogeneity by estimating between-subjects variation in relationships between predictors and outcomes, and measurement invariance is useful to establish that individuals across groups and times respond to items and questions the same way prior to substantive analyses.

However, the results of such analyses are often interpreted in ironclad terms without necessarily reflecting upon their theoretical basis. For instance, if measurement invariance is not upheld, it may simply indicate that the heterogeneity we are hoping to rule out cannot be discounted but rather should be addressed directly (Nagle, 2023). Rather than assuming that heterogeneity is noise obscuring a mean effect and focusing on the analytical options available to minimize that “noise,” our position is that research should explicitly address such heterogeneous effects and study them by design (Bryan et al., 2021). Thus, we advocate for moving such considerations from analysis to study design, considering the ways in which instructional practices can be made more flexible and adaptive from the very start (rather than checking whether the effects they produce are fixed across individuals, groups, etc., at the analytical stage).

Challenge 2: Take a developmental view and widen the window of time

Jay Lemke, the social semiotician, once observed that “every process, action, social practice, or activity occurs on some timescale ... [and often] on more than one timescale” (Lemke, 2000, p. 275). Learning, too, is undeniably a developmental phenomenon, and the study of developmental processes involves grappling with nonstationary characteristics and parameters of data (e.g., power laws of learning) (Molenaar et al., 2014). Many research designs in ISLA, however, ignore the processual nature of L2 learning and reveal a commitment to studying it from a perspective of stasis (Larsen-Freeman, 2015).

Consider, for example, a study investigating the comparative effect(s) of different training paradigms (intervention A vs intervention B) on a specific L2 feature. A typical two- or three-wave design may illustrate a larger effect for intervention B at immediate and delayed posttest and would be deemed an informative success when a probability value or an average domain-specific effect size is discovered (Plonsky & Oswald, 2014). We might consider the question more or less resolved and move on to other targets in our research. But what if we discovered that over time, the effect of intervention A is 10× stronger than the effect of intervention B given that it sparks the learner to engage

in, for example, more selective and sustained attention to that L2 feature, greater self-monitoring, and to subsequently proceduralize their existing declarative knowledge? This pattern of learning, perhaps unfolding over weeks or months, would not be captured by the first design. To our knowledge, no single experimental study has examined such effects across distinct windows of time simultaneously, and this highlights the challenge of studying learning as a process and as a product simultaneously. The fact that intervention A proves better in the long term does not invalidate intervention B but instead suggests that they are effective over different developmental windows and for different purposes. Thus, as this hypothesized example illustrates, ISLA requires a broader focus on development that reorients around several (wider) windows of time (de Bot, 2015).

Practitioners with significant experience in the classroom acknowledge that instruction often has further-reaching effects that may take some time to manifest. Yet, intervention studies rarely take into account lagged or incubation effects (Hiver & Al-Hoorie, 2020), perhaps due to the logistical challenges to collecting this longer-range data, preferring instead to employ designs that focus on immediately detectable outcomes. This research task is often considered complete by using large, representative samples of participants (Nesselroade & Molenaar, 2010). However, this also reveals a narrow understanding of external validity and generalizability confined to statistical or population generalizability (Byrne & Callaghan, 2023). Among other ways, generalizability can be achieved through the application of study findings across different time periods and timeframes (*temporal generalizability*) if results remain valid and applicable to different environments or settings (*setting generalizability*, see, e.g., Moranski & Zalbidea, 2022), and when research findings extend to different tasks or situational domains (*task generalizability*).

As recent advances in the field illustrate, taking a developmental view and widening the window of time under investigation has implications for the very research questions we investigate and, by extension, the measurement, design, and modeling we undertake in studies of L2 learning (Lowie et al., 2020; van Geert, 2023). Developmental claims cannot be made in the absence of data with a time-series element. Yet, even when treating time as a continuous dimension, the time window adopted may be more an artifact of study design than a genuine indicator of the actual timescale of change. That is, logistical constraints may dictate when data are collected, irrespective of hypothesized developmental timelines. In fact, it may be that various outcome measures of L2 learning that are in widespread use develop on different time scales, necessitating the use of multiple hierarchical temporal levels of analysis (Evans, 2020).

Understanding how findings might be affected by changes over time, differences in socio-geographic environments, and instructional situations is a key part of making appropriate claims about the wider applicability of a study's findings. Instructional treatments, too, are understood to have differential effects over time, and studies searching for an average treatment effect neglect to consider how the intervention impacts learners across the duration (e.g., beginning, middle, end) of that intervention, how such learning effects come about, where and when they unfold, and the developmental patterns this temporal variation results in. Theoretical accounts or models of language development exist that make explicit the temporal nature of learners' trajectories of learning. However, a wider reorientation in ISLA research designs has

yet to occur (Verspoor & de Bot, 2022). We would argue that, in studies of instruction and learning, uncovering how mechanisms of change and learning processes unfold is more insightful than illustrating typical outcomes and states. Thus, the specification of timing characteristics in instructional research—the scale, frequency, interval, and duration of timing effects—should be a critical component of study design.

Challenge 3: Treatment fidelity should be balanced with individual proactivity

In the context of intervention research, treatment fidelity refers to the extent to which an intervention or treatment is implemented as intended or designed (Sanetti & Luh, 2020). In practice, researchers may outline a procedural plan to achieve a desired outcome. Once implemented, this intervention protocol can then be assessed for fidelity. Fidelity is thought to be crucial to the validity of research results as it provides an indication of whether the observed effects are due to the intervention itself rather than any other extraneous or confounding factors. The extent to which program drift (i.e., unplanned deviations from the intervention) and other changes made to the content or delivery of the intervention can be avoided will potentially determine the intervention's effectiveness (Noell & Gansle, 2006). Maintaining treatment fidelity also enables researchers to replicate an intervention with more confidence, compare the effectiveness of different interventions more reliably, and identify which aspects of the intervention are critical to its success and which may need refinement (Sanetti & Collier-Meek, 2019). Not surprisingly, this metric is key to evaluating the success of educational intervention frameworks such as response to intervention, multi-tiered systems of supports, and sequential multiple assignment randomized trials (Scott et al., 2019).

While funding agencies and stakeholders in scholarly publishing make treatment fidelity a standard part of the conduct and evaluation of educational intervention research, this position is contested by the increasingly complex designs and interventions used in practice, which render effects heterogenous by default (see, e.g., Bryan et al., 2021 and Norouzian & Bui, 2023). The reality is that educational treatments and interventions tend to be multidimensional, variable-dense, and highly context-dependent, involving consideration of much more than simple delivery of discrete behaviors. Deliberately reducing the many possible options for action under the guise of fidelity of implementation discounts the consensus that educational research is a complex, dynamically evolving, human-centered endeavor unfolding in unique environments (Morrison & van der Werf, 2019). Indeed, success is often contingent on the productive adaptation and uptake of intervention by local stakeholders who must sustain programs through a dialectic of resistance, revision, and accommodation (Gutiérrez & Penuel, 2014). Consequently, research should document how students and teachers change and adapt interventions in interactions with each other in relation to their dynamic local contexts, as these represent important sources of evidence for generalizability (Harn et al., 2013). At first glance, this approach of forgoing treatment fidelity in exchange for greater adaptive tailoring appears to fly in the face of open science initiatives, including replication. However, when experimental controls are loosened, and stakeholders and learners are given more agency in shaping the intervention model, what needs to be replicable is not the exact intervention practices

(which may be tailored and site-specific) but rather the underlying set of principles that guide decision-making with respect to why, how, and when practices get adapted. It is reasonable for two studies to look very different despite following a single adaptive protocol because it allows for individual proactivity. It is still primarily the core of the systems being implemented (i.e., the adaptive protocol itself) rather than the specific adaptations and modifications that needs to be replicable.

Equally, language education research acknowledges the important role of individual differences in learning and development (Li et al., 2022). The default approach is to consider which individual factors might mediate (i.e., act as an intermediary between two other variables and explain how or why an X-Y relationship works) the effects of a given treatment or, indeed, play a moderating role (i.e., explain when, how strongly, or for whom an X-Y relationship works). Typically, fidelity to an intervention is paramount, and individuals' agency and engagement, important hallmarks of proactive language learning, are controlled for in the experimental design (Sanetti et al., 2021), relegating much of what constitutes individual variation to peripheral noise. Learning, however, is an intentional, goal-oriented, strategic, and effortful pursuit embodied and accomplished by individual language learners in social contexts of language use. As Larsen-Freeman (2013) proposed, all "instruction is motivated by the assumption that students can transfer their learning ... to another setting" (p. 107) and that learners' agency and proactive inclinations necessarily influence "what students attend to and ... how students generalize their learning experiences" (p. 115).

Change can be prompted by instruction, but learners should also be expected to proactively engage with an intervention, adapt to it, and mobilize their learning beyond the classroom (Chow et al., 2022). This adaptive transformation of learning based on the affordances of the instructional treatment also reframes what "successful" effects look like, as there may be instances where effects are nonlinear, indirect, or require an incubation period for certain individuals. Individual engagement during a training paradigm is challenging to make sense of statistically, even though engagement measures are often indicators of the psychological and behavioral impacts that educational researchers want interventions to have (e.g., greater risk-taking, increased tolerance of ambiguity, more self-regulation in learning; Hiver & Wu, 2023). Our position is that understanding the reciprocal interface between fidelity of treatment on the one hand and the agentic engagement of individuals with that intervention on the other is a challenge not yet taken up by the field (Larsen-Freeman, 2017). For instance, flexibility around contextual modifications based on the needs of practitioners and students is likely to result in more authentic buy-in from stakeholders, leading to greater innovations in practice and more desired outcomes (Harn et al., 2013). Thus, attending to the ways that learners actively engage throughout an intervention can clarify the mechanisms and processes through which the treatment is beneficial, thereby allowing the field to reconcile the inherent proactivity of individual learners with the more normative assumptions of instructional interventions (Larsen-Freeman, 2014).

Challenge 4: Knowing about the "who" is primary in knowing "what works"

Synthetic work shows that close to 70% of published language learning research is conducted exclusively with university student samples in the global north

(Andringa & Godfroid, 2020) and that the volume of research related to English as the target L2 approaches a staggering 80%. As Andringa and Godfroid (2020) cautioned, “Our sampling choices not only create problems for generalizability but also pose ethical dilemmas. Our science may not provide answers to questions for a vast majority of language learners” (p. 139). As recent initiatives demonstrate (e.g., SLA for All), we know less than we think we do about our phenomena of interest, and sweeping claims from this narrow evidence base about what works and for whom border on scholarly hubris (Godfroid & Andringa, 2023). In light of this unsustainable state of affairs, we should practice greater epistemic humility (Hedges, 2018). One means of doing so is to include in research reports a detailed discussion of sample characteristics and the specific ways in which findings may (not) generalize to other samples and research sites (see, e.g., Rachels & Rockinson-Szapkiw, 2018 and Simons et al., 2017).

Establishing knowledge about particular populations is also crucial given considerations we have highlighted earlier—namely, that effects are contingent on context and individuals in those contexts and that individual proactivity adds layers of complexity to treatment implementation. The result is that all interventions must be transparent about this inherent complexity. Complex interventions are those that deliberately attempt to account for how multiple interacting components—including at the individual level—lead to various target outcomes over time. Among other things (Skivington et al., 2021; Steenbeek & van Geert, 2015; van Geert & Steenbeek, 2014), the complexity of an intervention resides in:

- the different individuals targeted by the intervention and the groups or hierarchical levels in which they reside and operate;
- the number of interacting components within the so-called experimental and control conditions;
- the number and difficulty of behaviors and activity types required by individuals receiving or delivering the intervention;
- the number and variability of desirable and undesirable outcomes from those target individuals; and
- the degree of flexibility or tailoring of the intervention permitted for those target individuals.

Though we acknowledge the value in group-based designs and do not advocate that L2 learning research must center exclusively on the empirical study of the individual, our position is that important notions about who does the learning, how they accomplish that learning, and why they benefit from interventions in different ways and to different degrees are inextricably interrelated. We stand to gain much by centering the individual in cross-setting interventions that leverage and foreground diversity in our research.

The rationale for more diverse sampling procedures that reduce our field’s reliance on WEIRD samples (Henrich et al., 2010) from single sites is relatively straightforward. In addition to more statistical power and greater external validity, broader and increasingly diverse participant pools allow us to know more about more people. Researchers may be tempted to sample narrowly since more homogeneous sampling often enables the detection of larger effect sizes. However, multisite sampling is a strategy that aims to

mitigate the “range restriction” that results from narrow sampling and undercuts any attempts to establish generalizable and transferable links between interventions and learning outcomes (Moranski & Ziegler, 2021). That is, our knowledge base is less likely to be confined to explanations of how interventions work with populations of literate, highly educated, affluent language learners in well-resourced classroom settings of the global north (Godfroid & Andringa, 2023). Still, diverse sampling should be based on reciprocal, mutually beneficial engagements and equitable knowledge exchange with underrepresented stakeholders rather than extractive “parachute research” (Gewin, 2023), which perpetuates historic disadvantages through one-off or short-term projects planned, executed, and reported without seeking substantive local expertise or input.

Challenge 5: The research–practice interface remains paramount

Scholars engage in intervention research for the primary goal of promoting student learning (Hattie, 2023). The imperative for relevance to practice—an important criterion for scientific rigor (Gutiérrez & Penuel, 2014)—grounds research of L2 learning and instruction. However, of the many hundreds of empirical articles published on L2 instruction each year, only a small percentage of findings eventually impact classroom practice (Marsden & Kasprovicz, 2017; Sato & Loewen, 2019). Beyond technical questions of which studies deserve practitioners’ attention, practitioners and policymakers may have strong preferences for internal sources of evidence and may be justifiably skeptical of the relevance of evidence developed outside their context (Jackson, 2022). Still, the research–practice interface remains fraught, and advocating for more robust designs and greater rigor is likely to trigger concerns about what applications intervention research provides or what takeaways practitioners will be able to hold on to (Farley-Ripple et al., 2018).

Central to recent discussions about the research–practice divide is the notion of a shared dialogue between stakeholders (McKinley, 2019). This argument proposes that one of the primary issues preventing research from crossing the divide is a lack of communication and information. Thus, researchers need to communicate more clearly, frequently, and responsibly with relevant classroom stakeholders to help increase the uptake of intervention-based findings (Sato & Loewen, 2022). Translational initiatives intended to address this imbalance, such as TESOLgraphics (<https://tesolgraphics.wixsite.com/tesolgraphics>), are gaining traction in the field. When research is both applicable to the classroom and visibly useful for practitioners, it is more likely to inform practice. However, what this “communication and dialogue” proposal overlooks is that the research–practice divide is not only an issue of communication and adequacy of available information but of values (Bryk et al., 2011; Reincke et al., 2020). As related work cautions, “to believe that education research is value-free [...] or that theories succeed purely on the merit of their evidence base is to misunderstand how educational research becomes pedagogical practice” (Schuetze, 2022, p. 93).

Values alignment is an important but neglected aspect of whether educational research crosses into mainstream practice (Ball, 2012; Schneider, 2014). However clearly and convincingly it is conveyed, when evidence is incongruent with a community’s belief systems, it will be filtered out before it can lead to meaningful change

in practice (Lewis & Wai, 2021). Conversely, when research findings align with the existing values of practitioners, that research feels more intuitive because it fits existing classroom practices and beliefs about pedagogy (Luong et al., 2019). There is broad evidence that stakeholders selectively attend to evidence; when instructors' overall educational philosophy matches the instructional approach of the intervention, they implement the intervention more effectively than those for whom this match does not exist (Harn et al., 2013). As such, values-alignment impacts not only the efficacy of an intervention more narrowly but also its eventual reach and broader uptake in practice. To ignore the importance of values alignment is to perpetuate the status quo of researchers and practitioners talking past each other (Jackson, 2022).

Centering the research–practice interface can be accomplished, to some extent, by involving practitioners as competent partners in multiple stages of intervention research, acknowledging their insider knowledge and their concerns (i.e., problems of practice) as legitimate starting points for empirical intervention (Sato & Loewen, 2022). One research template intended to do this is *design-based intervention research*, which involves forming collaborative partnerships, co-designing with communities, negotiating joint focus and goals, building capacity at scale, and engaging in continuous improvement. This deliberative and participatory process is rare in intervention research. Clearly, methodological expertise alone is insufficient to democratize the development of evidence (Jackson, 2022). The realization that teachers ultimately adopt pedagogical strategies that reinforce their existing views of language education is key.

Necessary buy-in that will support the research–practice interface can also be meaningfully encouraged by transparently acknowledging the range of limitations, boundary conditions, and contingencies that apply to a set of empirical findings (Al-Hoorie et al., 2023; Simons et al., 2017). This form of epistemic humility is lacking from the design and reporting of many studies in our field that nevertheless demonstrate an eagerness to convey broad and ambitious implications for practice. Another important avenue to recalibrate the research–practice interface is to actively counter the misguided notion that teachers are mere technicians (Kubanyiova, 2020) and accept that in complex classroom settings, they exercise professional judgment and improvisational capability as they “wrestle with the fact that there are no ‘right’ [instructional] answers, only appropriate [instructional] choices” (Johnson, 2019, p. 172). Entering the research partnership with practitioners from a place of acceptance that they constitute an important “for whom” and “for what” of our work is key to delivering on our ethical imperative as researchers (Ortega, 2005, 2013).

A protocol for intervening differently

Constructing a program of research that tackles the challenges we outlined above requires new approaches to study design that reconsider traditional notions of effectiveness. At the same time, such a research program is also likely to yield sharper insights that push instructed L2 learning research forward. Here, we propose one template for doing so through complex adaptive interventions.

Adaptive interventions describe treatments that use an implementation protocol with explicit guidelines on when, for whom, how, and in what sequence a given treatment can be implemented most effectively. Adaptive interventions detail whether, how,

when, and which measures to use in the process of tailoring an educational intervention based on information about an individual learner or subgroups of learners at specific time points (see also Roberts et al., 2021). Similarly, adaptive approaches to classroom L2 instruction are not new; what is new is much of the conceptual and methodological apparatus around systematically constructing and studying interventions in this way. Such interventions avoid the one-size-fits-most approach common in group-based intervention research. Instead, their objective is to provide what is needed to those who need it when it is needed.

Building on this framework, complex adaptive interventions combine the idea of adaptability and responsiveness to the dynamic and evolving nature of the system in which the intervention is implemented. In a complex adaptive system, the interactions between components are multiple and dynamic, and they may change over time in response to various factors. The need to adapt an intervention may arise because what works for one individual may not work for another (between-person heterogeneity), and what works now may not work in the future for the same individual (within-person heterogeneity). A complex adaptive intervention is designed to respond to this complexity by being flexible, self-organizing, and capable of adapting to changing circumstances.

In addition to reflecting the principles of complex adaptive systems (e.g., distributed control, nonlinearity, etc.), there are several key elements of complex adaptive interventions that are designed to enhance their effectiveness and sustainability, and we describe these briefly below. Importantly, we see complex adaptive interventions as a complementary mode of conducting ISLA research, not a framework that replaces existing research designs. Additionally, as mentioned above, it is reasonable for two studies that follow a single adaptive protocol and operate on consolidated principles that guide decision-making about why, how, and when practices get adapted to still look very different. The decision tree for implementing the intervention—which could be seen as analogous to the DNA of the study—is the same (see Figure 1), but the specific adaptations and modifications offered to different learners over time and in varied settings will differ depending on all the things listed below:

- **Sensitivity to Individuals:** Complex adaptive interventions personalize and tailor the specific strategies, supports, and routines available (e.g., the intensity, frequency, type, or timing of interventions) to learners' unique characteristics. Adjusting the content and pace of instruction to students' needs, ongoing progress, and their responses to the intervention recognizes that what works for some learners may need to be adapted for others.
- **Sensitivity to Contexts:** Complex adaptive interventions are designed to be sensitive to the particular contexts in which they are implemented. Adjusting an intervention to account for specific contextual demands and the environmental factors that can impact the effectiveness of the intervention recognizes that what works in one setting may need to be adjusted for another.
- **Dynamic Adjustment:** Complex adaptive interventions are designed to change over time as needed, responding to changes in the environment, participant characteristics, and evolving circumstances and responses of the individual. This responsiveness can be continuous or occur at specific decision points, and it

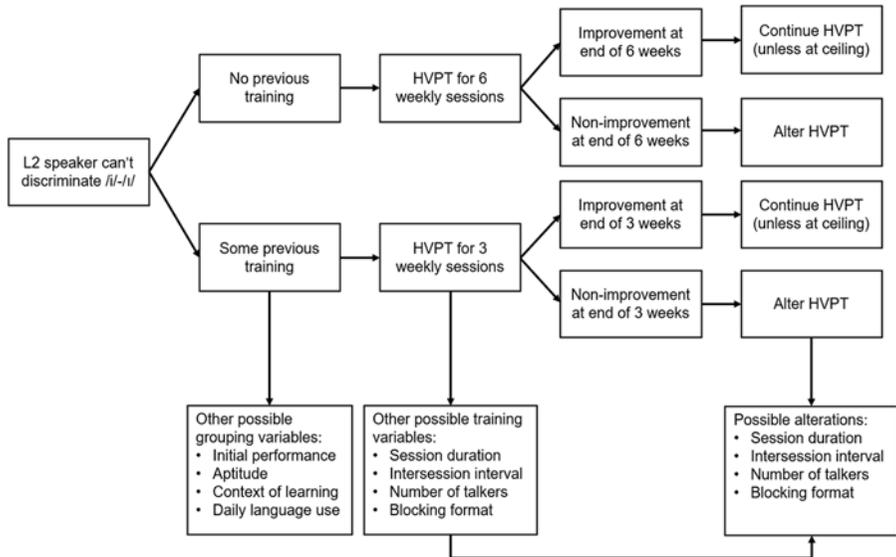


Figure 1. Example decision tree for a complex adaptive intervention for High-Variability Pronunciation Training (HVPT).

allows the intervention to adapt its strategies and even its target outcomes as the situation evolves.

- **Feedback Loops:** Complex adaptive interventions collect, analyze, and integrate training data and feedback from individuals' responses, learning, or ongoing progress to guide decision-making and iterate the treatment. These feedback loops enable providers to monitor and evaluate the intervention based on the information received.
- **Optimization:** Complex adaptive interventions are designed to improve and maximize student learning through continuous optimization. Multiple intervention components interact while the intervention maintains its functionality and continues to adapt and evolve beyond its initial implementation. Based on these diverse patterns and behaviors, intervention strategies are refined over time to optimize individuals' learning.

Complex adaptive interventions do not involve randomization to experimental conditions. Instead, the methods and study designs available extend beyond group-assignment comparisons and include design-based intervention research, formative experiments, single-case designs, and experimental ethnography, among others. We envision complex adaptive interventions as a template explicitly designed to investigate "the specific mechanisms by which outcomes for certain individuals [can be] accomplished within specific structural and ecological circumstances" (Gutiérrez & Penuel, 2014, p. 19). Complex and adaptive in this sense mean that intervention is applied conditionally to examine which combinations of components and which sequences of treatments are most effective (Hedges, 2018). This will usually entail

digging into the training data to track developmental trajectories that learners take and staying closely connected to each step of the process while implementing an intervention (see Figure 1). Complex adaptive interventions provide a protocol for intervention research to attend to the challenges we have highlighted above, and they allow researchers to respond adaptively to various levels of the system targeted for change and implement an instructional treatment in response to specific relational components in a specific context.

Conclusion

“What works” and “what matters” depend on many factors, and, as we outlined above, there are many challenges to the deterministic view of effects that dominate intervention research in ISLA. In the broader behavioral sciences, this realization is increasingly framed as a “heterogeneity revolution” (Bryan et al., 2021, p. 986). Confronting these challenges can refocus our analytical lens and “mov[e] us away from imagining interventions as fixed packages of strategies with readily measurable outcomes and toward more open-ended socially embedded experiments that involve ongoing mutual engagement” (Gutiérrez & Penuel, 2014, p. 20). Research that tackles these challenges head-on, such as complex adaptive interventions, will enable us to achieve greater sophistication and rigor and a sharper focus than has been possible so far. Importantly, we do not advocate for abandoning group-based RCTs, nor do we dismiss the impact such studies have had on building the current knowledge base of the field. Instead, we propose that the field invest in a necessary and complementary research model centered on exploring and documenting individual, time- and context-sensitive effects using flexible and responsive intervention methods.

References

- Al-Hoorie, A. H., Hiver, P., Larsen-Freeman, D., & Lowie, W. (2023). From replication to substantiation: A complexity theory perspective. *Language Teaching*, 56(2), 276–291.
- Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics*, 40, 134–142.
- Ball, A. F. (2012). To know is not enough: Knowledge, power, and the zone of generativity. *Educational Researcher*, 41, 283–293.
- Boers, F., & Faez, F. (2023). Meta-analysis to estimate the relative effectiveness of TBLT programs: Are we there yet? *Language Teaching Research*, 1–19. <https://doi.org/10.1177/13621688231167573>
- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 5(8), 980–989.
- Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 127–162). Springer.
- Byrne, D., & Callaghan, G. (2023). *Complexity theory and the social sciences: The state of the art* (2nd ed.). Routledge.
- Chow, J. Y., Davids, K., Button, C., & Renshaw, I. (2022). *Nonlinear pedagogy in skill acquisition: An introduction*. Routledge.
- Cook, T. D. (2018). Twenty-six assumptions that have to be met if single random assignment experiments are to warrant “gold standard” status: A commentary on Deaton and Cartwright. *Social Science & Medicine*, 210, 37–40.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30(2), 116–127.

- Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2–21.
- de Bot, K. (2015). Rates of change: Timescales in second language development. In Z. Dörnyei, P. D. MacIntyre & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 29–37). Multilingual Matters.
- Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, 17(4), 1101–1119.
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science*, 1(1), 1–3.
- Education Endowment Foundation. (2023). The education endowment foundation: Building the role of evidence in the education system. In M. Sanders & J. Breckon (Eds.), *The what works centres: Lessons and insights from an evidence movement* (pp. 54–69). Policy Press.
- Engeström, Y. (2011). From design experiments to formative interventions. *Theory & Psychology*, 21(5), 598–628.
- Evans, R. (2020). On the fractal nature of complex syntax and the timescale problem. *Studies in Second Language Learning and Teaching*, 10(4), 697–721.
- Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. *Educational Researcher*, 47(4), 235–245.
- Gewin, V. (2023). Pack up the parachute: Why global north–south collaborations need to change. *Nature*, 619, 885–887.
- Godfroid, A., & Andringa, S. (2023). Uncovering sampling biases, advancing inclusivity, and rethinking theoretical accounts in second language acquisition: Introduction to the special issue SLA for All? *Language Learning*, 73, 981–1002.
- Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, 43(1), 19–23.
- Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, 79(2), 181–193.
- Hattie, J. (2023). *Visible learning: The sequel*. Routledge.
- Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, 11(1), 1–21.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83.
- Hiver, P., & Al-Hoorie, A. H. (2020). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.
- Hiver, P., Al-Hoorie, A. H., & Evans, R. (2022). Complex dynamic systems theory in language learning: A scoping review of 25 years of research. *Studies in Second Language Acquisition*, 44(4), 913–941.
- Hiver, P., & Wu, J. (2023). Engagement in task-based language teaching. In C. Lambert, S. Aubrey & G. Bui (Eds.), *The role of the learner in task-based language teaching: Theory and research* (pp. 74–90). Routledge.
- Jackson, C. (2022). Democratizing the development of evidence. *Educational Researcher*, 51(3), 209–215.
- Johnson, K. E. (2019). The relevance of a transdisciplinary framework for SLA in language teacher education. *The Modern Language Journal*, 103, 167–174.
- Kang, E. Y., Sok, S., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*, 23(4), 428–453.
- Kaplan, A., Cromley, J., Perez, T., Dai, T., Mara, K., & Balsai, M. (2020). The role of context in educational RCT findings: A call to redefine “evidence-based practice. *Educational Researcher*, 49(4), 285–288.
- Kubanyiova, M. (2020). Language teacher education in the age of ambiguity: Educating responsive meaning makers in the world. *Language Teaching Research*, 24, 49–59.
- Larsen-Freeman, D. (2013). Transfer of learning transformed. *Language Learning*, 63, 107–129.
- Larsen-Freeman, D. (2014). Another step to be taken – Rethinking the end point of the interlanguage continuum. In Z. H. Han & E. Tarone (Eds.), *Interlanguage* (pp. 203–220). John Benjamins.
- Larsen-Freeman, D. (2015). Saying what we mean: Making a case for “language acquisition” to become “language development”. *Language Teaching*, 48(4), 491–505.
- Larsen-Freeman, D. (2017). Complexity theory: The lessons continue. In L. Ortega & Z. H. Han (Eds.), *Complexity theory and language development* (pp. 11–50). John Benjamins.

- Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity*, 7(4), 273–290.
- Lewis, N. A., & Wai, J. (2021). Communicating what we know, and what isn't so: Science communication in psychology. *Perspectives on Psychological Science*, 16(6), 1242–1254.
- Li, S., Hiver, P., & Papi, M. (2022). Individual differences in second language acquisition: Theory, research, and practice. In L. S., P. Hiver & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 3–34). Routledge.
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285–329.
- Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166.
- Lowie, W., Michel, M., Rousse-Malpat, A., Keijzer, M., & Steinkrauss, R. (Eds.). (2020). *Usage-based dynamics in second language development*. Multilingual Matters.
- Luong, K. T., Garrett, R. K., & Slater, M. D. (2019). Promoting persuasion with ideologically tailored science messages: A novel approach to research on emphasis framing. *Science Communication*, 41(4), 488–515.
- Marsden, E., & Kasprowicz, R. (2017). Foreign language educators' exposure to research: Reported experiences, exposure via citations, and a proposal for action. *The Modern Language Journal*, 101(4), 613–642.
- McKinley, J. (2019). Evolving the TESOL teaching–research nexus. *TESOL Quarterly*, 53(3), 875–884.
- Molenaar, P. C., Lerner, R. M., & Newell, K. M. (Eds.) (2014). *Handbook of developmental systems theory and methodology*. Guilford.
- Moranski, K., & Zalbidea, J. (2022). Context and generalizability in multisite L2 classroom research: The impact of deductive versus guided inductive instruction. *Language Learning*, 72(1), 41–82.
- Moranski, K., & Ziegler, N. (2021). A case for multisite second language acquisition research: Challenges, risks, and rewards. *Language Learning*, 71(1), 204–242.
- Morrison, K. (2021). *Taming randomized controlled trials in education: Exploring key claims, issues and debates*. Routledge.
- Morrison, K., & van der Werf, G. (2019). Making “what works” work: The elusive research butterfly. *Educational Research and Evaluation*, 25(5–6), 225–228.
- Nagle, C. L. (2023). A design framework for longitudinal individual difference research: Conceptual, methodological, and analytical considerations. *Research Methods in Applied Linguistics*, 2(1), 100033.
- Nesselroade, J. R., & Molenaar, P. C. (2010). Emphasizing intraindividual variability in the study of development over the life span: Concepts and issues. In R. M. Lerner, M. E. Lamb & A. M. Freund (Eds.), *The handbook of life-span development* (pp. 30–54). John Wiley.
- Nilsen, P. (2020). Making sense of implementation theories, models, and frameworks. In B. Albers, A. Shlonsky & R. Mildon (Eds.), *Implementation science 3.0* (pp. 53–79). Springer.
- Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention*, 32(1), 32–39.
- Norouzian, R., & Bui, G. (2023). Meta-analysis of second language research with complex research designs. *Studies in Second Language Acquisition*, 1–25. <https://doi.org/10.1017/S0272263123000311>
- Ortega, L. (2005). For what and for whom is our research? The ethical as transformative lens in instructed SLA. *The Modern Language Journal*, 89(3), 427–443.
- Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, 63, 1–24.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912.
- Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1–2), 72–89.
- Reincke, C. M., Bredenoord, A. L., & van Mil, M. H. (2020). From deficit to dialogue in science communication: The dialogue communication model requires additional roles from scientists. *EMBO Reports*, 21(9), e51278.
- Ren, W., Li, S., & Lü, X. (2023). A meta-analysis of the effectiveness of second language pragmatics instruction. *Applied Linguistics*, 44(6), 1010–1029.
- Roberts, G., Clemens, N., Doabler, C. T., Vaughn, S., Almirall, D., & Nahum-Shani, I. (2021). Multi-tiered systems of support, adaptive interventions, and SMART designs. *Exceptional Children*, 88(1), 8–25.

- Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69(3), 652–708.
- Sanetti, L. M. H., & Collier-Meek, M. C. (2019). *Supporting successful interventions in schools: Tools to plan, evaluate, and sustain effective implementation*. Guilford.
- Sanetti, L. M. H., Cook, B., & Cook, L. (2021). Treatment fidelity: What it is and why it matters. *Learning Disabilities: Research and Practice*, 36(1), 5–11.
- Sanetti, L. M. H., & Luh, H. J. (2020). Treatment fidelity in school-based intervention. In A. Reschly, A. J. Pohl & S. L. Christenson (Eds.), *Student engagement: Effective academic, behavioral, cognitive, and affective interventions at school* (pp. 77–87). Springer.
- Sato, M., & Loewen, S. (2019). Do teachers care about research? The research–pedagogy dialogue. *ELT Journal*, 73(1), 1–10.
- Sato, M., & Loewen, S. (2022). The research–practice dialogue in second language learning and teaching: Past, present, and future. *The Modern Language Journal*, 106(3), 509–527.
- Schneider, J. (2014). *From the ivory tower to the schoolhouse: How scholarship becomes common knowledge in education*. Harvard Educational Publishing Group.
- Schuetze, B. A. (2022). The research–practice divide is not only an issue of communication, but of values: The case of growth mindset. *Texas Education Review*, 10(1), 92–104.
- Scott, T. M., Gage, N. A., Hirn, R. G., Lingo, A. S., & Burt, J. (2019). An examination of the association between MTSS implementation fidelity measures and student outcomes. *Preventing School Failure*, 63, 308–316.
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128.
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, 374, n2061.
- Stark, P. B. (2022). Pay no attention to the model behind the curtain. *Pure and Applied Geophysics*, 179(11), 4121–4145.
- Steenbeek, H., & van Geert, P. (2015). A complexity approach toward mind–brain–education (MBE): Challenges and opportunities in educational intervention and research. *Mind, Brain, and Education*, 9, 81–86.
- van Geert, P. L. (2023). Some thoughts on dynamic systems modeling or L2 learning. *Frontiers in Physics*, 11, 1186136.
- van Geert, P., & Steenbeek, H. (2014). The good, the bad and the ugly? The dynamic interplay between educational practice, policy and research. *Complicity: An International Journal of Complexity and Education*, 11, 22–39.
- Verspoor, M., & de Bot, K. (2022). Measures of variability in transitional phases in second language development. *International Review of Applied Linguistics in Language Teaching*, 60(1), 85–101.
- Vuogan, A., & Li, S. (2023). Examining the effectiveness of peer feedback in second language writing: A meta-analysis. *TESOL Quarterly*, 57(4), 1115–1138
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved March 15, 2024, from <https://ies.ed.gov/ncee/wwc/Handbooks>.
- Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, 71(2), 487–536.

Cite this article: Hiver, P and Nagle, C. (2024). Complex adaptive interventions: The challenge ahead for instructed second language acquisition research. *Annual Review of Applied Linguistics*, 1–16. <https://doi.org/10.1017/S0267190524000060>