

## Research Article

### COMPLEX DYNAMIC SYSTEMS THEORY IN LANGUAGE LEARNING

A SCOPING REVIEW OF 25 YEARS OF RESEARCH

**Phil Hiver**  \*

*Florida State University, USA*

**Ali H. Al-Hoorie** 

*Royal Commission for Jubail and Yanbu, Saudi Arabia*

**Reid Evans** 

*University of Massachusetts Medical School, USA*

#### Abstract

A quarter of a century has passed since complex dynamic systems theory was proposed as an alternative paradigm to rethink and reexamine some of the main questions and phenomena in applied linguistics and language learning. In this article, we report a scoping review of the heterogeneous body of research adopting this framework. We analyzed 158 reports satisfying our inclusion criteria (89 journal articles and 69 dissertations) for methodological characteristics and substantive contributions. We first highlight methodological trends in the report pool using a framework for dynamic method integration at the levels of study aim, unit of analysis, and choice of method. We then survey the main substantive contribution this body of research has made to the field. Finally, examination of study quality in these reports revealed a number of potential areas of improvement. We synthesize these insights in what we call the “nine tenets” of complex dynamic systems theory research, which we hope will help enhance the methodological rigor and the substantive contribution of future research.

#### INTRODUCTION

All theories, if they are to avoid becoming passing academic fads or bandwagons, must contribute something of substance that is new and worthwhile—something that pushes

---

We would like to thank Alyssa Vuono, Janice Wu, and Hyejin An for their assistance with coding the report pool.  
\* Correspondence concerning this article should be addressed to Florida State University, School of Teacher Education, College of Education, 1114 W. Call St., G128 Stone Building, Tallahassee, FL, 32306. E-mail: [phiver@fsu.edu](mailto:phiver@fsu.edu)

This article has been updated since its original publication. See <https://doi.org/10.1017/S0272263122000262>.

© The Author(s), 2021. 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.

the field forward. Nearly three decades have passed since complex dynamic systems theory (CDST) was first introduced into the field of language learning (Larsen-Freeman, 1994), and since then CDST perspectives and approaches have permeated many areas of applied linguistics research. The uptake of CDST in applied linguistics research has continued to accelerate, pushing further and faster even than in related fields such as education and theoretical linguistics (Koopmans, 2020; Kretzschmar, 2015). As recent work (e.g., Larsen-Freeman, 2017) synthesizing current strands of applied linguistics that have been informed by CDST shows, CDST has made important contributions to language development/acquisition (Lowie et al., 2010; Verspoor et al., 2008), language attrition (Schmid et al., 2013), language ecology (Cowley, 2011; Kramsch & Whiteside, 2008), language evolution (Ke & Holland, 2006; Mufwene et al., 2017), language policy and planning (Bastardas-Boada, 2013; Larsen-Freeman, 2018), language pedagogy (Han, 2020; Levine, 2020), bilingualism and multilingualism (Herdina & Jessner, 2002), sociolinguistics (Blommaert, 2014), educational linguistics (Hult, 2010), and communication studies (Massip-Bonet et al., 2019), among other areas of applied linguistics.

Considering this mainstream interest in CDST, it seems that it is not just appropriate but also necessary to assess this body of empirical work and evaluate the strength of its contribution to the field. Systematic and scoping reviews are uniquely positioned to afford a new vantage point on an area of research, and assessing the nature and quality of previous work has the potential to shape the future of research and practice (Alexander, 2020). Scoping reviews in particular are relevant when an area of research has not yet been extensively reviewed or when it is of a complex or heterogeneous nature (Pham et al., 2014)—arguably the case with CDST research. Scoping reviews share a number of procedural characteristics with systematic reviews, but where these two approaches to synthesis diverge is in their purposes and aims. The purpose of a systematic review is to identify the best available research on a specific question or a precise topic of research, and this often leads to answers of the appropriateness or effectiveness of some practice (Munn et al., 2018). Scoping reviews, however, look at what a field has done and how. Their aim is to examine how research is conducted in a certain field and provide an overview of the types of available evidence from that research (Arksey & O'Malley, 2005). As a result, scoping reviews generally evaluate patterns of knowledge and research methods from a greater range of study designs (Levac et al., 2010).

In the present scoping review and methodological synthesis of 25 years of CDST research, we had several objectives. In light of the growing methodological guidance available, our primary aim was to look back at the methodological characteristics of all previous empirical CDST studies in the field to note trends and tendencies in designs and analytical choices. By defining the shape of existing research designs, the field can take stock and chart a path forward. In addition to methodological characteristics, we were also interested in the substantive contributions this sizeable body of CDST research has made to the field, and what evidence it has provided for the language learning research enterprise. Given the readily apparent heterogeneity of research topics under the rubric of CDST research in language learning, we wondered what conclusions this empirical work allows us to draw and whether such a review could speak directly to broader issues and shared concerns in the field. Finally, we were interested in the rigor of this body of empirical work. Although CDST research has made many advances, we intended to explore whether this orchestrated search of the literature would reveal potential areas for

enhancing the quality of this body of research. We, thus, sought to identify future directions for CDST research that will help it continue to push the field forward with more coherent evidence and sharper insights.

Study quality has become central to many subdomains of SLD research (Gass et al., 2021). For instance, many syntheses have demonstrated that design tendencies related to measurement and sampling in the field leave much to be desired (Brown et al., 2018; Nicklin & Plonsky, 2020; Vitta & Al-Hoorie, 2021). Others have highlighted the need for greater transparency in checking and reporting assumptions (Hu & Plonsky, 2021), and increased rigor in data analytical strategies and reporting results (Al-Hoorie & Vitta, 2019; Larson-Hall & Plonsky, 2015; Marsden et al., 2018; Paquot & Plonsky, 2017; Plonsky, 2013, 2014). In the context of synthetic work such as this, study quality can refer to quality of the implementation of the methods or to the quality of inferences made from the methods (see also Gass et al., 2021), and as commonly observed, sound implementation of methods is orthogonal to whether those methods support a given inference. To our knowledge, this is the first methodologically oriented review of CDST research in language learning (but see Larsen-Freeman, 2017 for a detailed substantive synthesis). We are also not aware of any methodological reviews or syntheses of CDST research even in the wider social sciences or educational research literature. Thus, a critical appraisal of study quality can help to shed light on the transparency of this research, the relevance of the research targets and questions under investigation, and the appropriateness of methods of data analysis and presentation.

Of course, critical appraisal of research methods is not the pursuit of some form of elusive and idealized methodological perfection. Evaluating the methods adopted by a body of research serves a much more nuanced and meaningful purpose: to assess whether that body of work is “evidentially adequate” (Petticrew & Roberts, 2006). When considering methodological aspects and study quality, we followed recommendations to examine broader and more general methodological issues first as these can inform later reviews that assess more fine-grained aspects of study quality (Siddaway et al., 2019). In this scoping review we aimed to survey the methods employed by CDST researchers broadly, looking at generic characteristics such as research objectives, design and methodological orientation, sampling characteristics, data elicitation measures, and analytical strategies. We turn now to outlining the topic, scope, and rationale for the present review.

## **WHAT IS CDST RESEARCH?**

CDST is a meta-theory that provides an ontological position (i.e., principles of reality) for understanding language, language use, and language development in complex and dynamic terms (Hulstijn, 2020). It also captures epistemological ideas (i.e., principles of knowing) that aid scientific thinking and theorizing. In the field of language development, CDST underpins and contextualizes object theories consistent with these principles (Larsen-Freeman & Cameron, 2008), and these object theories address proximate questions about processes and outcomes of development. With regard to language, CDST proposes that language is a complex adaptive system, exhibiting both stability and dynamic change (Ellis & Larsen-Freeman, 2009). Language use is an iterative process of coadaptation in which language users adapt to the context and other interlocutors to realize the semiotic potential of language (Han, 2019). Language development is a

nonlinear, emergent process that draws on local-to-global processes of construction and global-to-local processes of constraint (de Bot, 2008). Whereas object theories (i.e., theories of language, language use, and language development/learning) are provisional, and their predictions must constantly be falsified and evaluated against observations of new evidence, the CDST meta-theory is broader in scope and relates to notions of what phenomena, questions, and aspects of inquiry should be investigated and why they merit research (Hulstijn, 2020; Overton, 2007).

In an applied field like ours, the entry point to CDST research is likely to be methodological and phenomenological, rather than at the more abstract level of theory (Larsen-Freeman, 2016b). That is, studies may set out to investigate constructs or questions pertaining to complex connections and dynamic processes of change, but are likely less concerned with disentangling the ontology and epistemology that underlies that mode of thinking (see also Ushioda, 2021). Research informed by CDST is different from other, more conventional research in two main ways: the basic assumptions that underlie it and the designs and methods that follow from those assumptions (Verspoor et al., 2011, p. 123). All research methods and paradigms<sup>1</sup> have a number of inherent assumptions, some of which are unstated or implicit in the techniques of data elicitation and analysis. CDST research takes a systems view as its point of departure (see e.g., Larsen-Freeman, 2015). CDST posits that the reality of the human and social world is one in which, first, everything counts and everything is connected (i.e., the *relational principle*) and second, everything changes (i.e., the *adaptive principle*) (Overton & Lerner, 2014). CDST research reconceptualizes the core of language, language use, and language development as systems or systemic phenomena grounded in a context-dependent and dynamic view of development. This reorientation challenges many of the field's existing assumptions and suggests new approaches to inquiry (Hiver et al., 2021).

There are multiple ways of approaching a topical area in our field. Primarily, the study of complex systems entails a focus on processes of change, and one way of doing so is through *dynamics-dominant* research using time-intensive methods (see also Van Orden et al., 2003 for a related framing). The question of how complex systems adapt to their environment to maintain their functioning over time is in fact relevant to nearly every part of applied linguistics (Larsen-Freeman & Cameron, 2008). Complex macrobehaviors, dynamic microinteractions within a system, and the emergence of new patterns of behavior are all of great interest (Ellis & Larsen-Freeman, 2009). Dynamics-dominant research includes a focus on relational dynamics, trajectories of change and development, self-organized processes, and emergent outcomes. Of course, because complex systems also have constituent parts that together make up the system, another basic approach is *interaction-dominant* research using relation-intensive methods. These designs describe systems' parts and their interactions, providing a focus on the complex underlying structure of interdependent relations (Hilpert & Marchand, 2018).

Especially important for our purposes, meta-theories such as CDST function as the necessary intellectual blueprint for conducting and evaluating research (Overton, 2015). For instance, Hiver and Al-Hoorie (2016) suggested that the core objectives of CDST research in applied linguistics should be to

(a) represent and understand specific complex systems at various scales of description; (b) identify and understand dynamic patterns of change, emergent system outcomes and behavior in the

environment; (c) trace, understand and where possible model the complex mechanisms and processes by which these patterns arise; and (d) capture, understand and apply the relevant parameters for influencing the behavior of systems.

(p. 752)

These broad objectives may serve as guiding parameters for study design as well as a way to gauge the overall contribution of a study or body of work.

There are other criteria to use when designing and evaluating CDST research in applied linguistics (Larsen-Freeman & Cameron, 2008; Verspoor et al., 2011). A useful point of entry are the *operational considerations* such as deciding what to case as a complex system, the boundaries of this unit of analysis, and the level of resolution and timescale(s) at which to analyze that system. *Contextual considerations* delineate the spatiotemporal frame of reference for the system and environmental features that are empirically salient to the system and its development. *Macrosystem considerations* account for dynamic outcomes or states in which a system has stabilized and help to pursue a temporal understanding of adaptive change and trajectories of development. *Microstructure considerations* define the makeup of a complex system, describing the functional whole, its constituents, and their relationships and interactions. Together these considerations provide a window into interpreting system behavior and inducing change in a complex system (Hiver & Al-Hoorie, 2016).

Very recent methodological advances have emerged in the field that strive to do justice to the complex, nonlinear learner development data. The main goal of these CDST-inspired studies is to develop multifactorial, nonlinear, and probabilistic models that are a better fit for such complex and dynamic language learner data than those currently available. For instance, a number of recent studies informed by CDST (e.g., Kliesch & Pfenninger, 2021; Murakami, 2016, 2020; Pfenninger, 2020; Verspoor et al., 2021) use generalized additive (mixed) modeling (GAMM) to (a) tease apart spatially distributed between- and within-learner variation, (b) disentangle mechanisms that have differing inherent time-courses (e.g., what aspects have the strongest impact on ongoing L2 writing development and over what timescale?), and (c) examine a system's interconnected structure as well as its dynamic behavior (e.g., what interactions occur between various cognitive and noncognitive ID variables across time). This approach examines variability as an informative data point in its own right (Verspoor & de Bot, 2021) and includes variability in its algorithms; it is, thus, ideal for analyzing nonlinear change over time in iterated learning experiments.

## AN INTEGRATIVE FRAMEWORK FOR CDST RESEARCH

Many applied linguists have recognized that the issues they are tackling are fundamentally complex, broad, and systemic (Han, 2020; see Larsen-Freeman, 2017, for a conceptual review). With CDST methods, the debate around the merits of qualitative versus quantitative research has been superseded by concern for the merits of individual versus group level (i.e., high- or low-*N*) designs and analyses, and the timescale or number of occasions (i.e., high- or low-*T*) appropriate for these designs. CDST encourages design decisions at several distinct levels—aim, unit of analysis, and method (Figure 1)—that reorient research toward processes of learning and development rather than exclusively focusing

on the product of learning (see also Larsen-Freeman, 2020). An integrative framework that combines these elements of research assumptions and design choices can be used to evaluate the contribution of CDST research.

Starting with the aim, an integrative design might be exploratory or may attempt to test certain understandings or expectations, including observationally and (quasi-) experimentally. Although the complex social world does not lend itself to universals that can be applied across all settings and populations, it is nevertheless possible to form probabilistic predictions by comparison to other similar systems, under similar conditions and contexts, with similar outcomes (Hiver & Al-Hoorie, 2020b). Consequently, when using CDST research tools in applied linguistics, there is no reason to shy away from making predictions and then subjecting these predictions to empirical test. As the double-headed arrow shows (Figure 1), integrative CDST designs should take both of these aims into account. Adopting a dual exploratory–falsificatory approach can radically reorient researchers and their aims, making them actively seek negative, disconfirming results rather than exclusively celebrating positive ones and experiencing disappointment when encountering inevitable negative findings.

A second level where a study can be designed in an integrative way is the unit of analysis, which has to do with whether the level of granularity in a CDST study is at the individual or the group level. Here some have contrasted an idiographic, person-centered, individual-level approach with a nomothetic, variable-centered, group-level approach (see e.g., Lowie & Verspoor, 2019). The former is focused on finding what is unique in each individual, while the latter looks for generalizations that apply across many individuals. This unit also applies to timescales and processes of change in which the nomothetic approach emphasizes general profiles of interindividual variability—often using cross-sectional data—and the overall mean trajectory of all cases, whereas the idiographic approach emphasizes intraindividual variability—often longitudinally—and the unique developmental trajectories of each individual (Verspoor et al., 2011).

Group-based research<sup>2</sup> remains popular in applied linguistics research, though individual-based designs may allow researchers to more readily operationalize the assumptions of CDST (Lowie, 2017) because this type of research holds a close lens to development and change without averaging away individual idiosyncrasies (Molenaar & Campbell, 2009). The utility of both individual and group based designs also squares with Molenaar's (2015) thinking on the appropriate level of granularity in such research—not requiring an exclusive focus on the individual case, but instead centering the objective to build more adequate models that take into account individual factors without giving up the search for general patterns and tendencies: “analyses of intra-individual variation does not preclude valid generalization across subjects.... In this way nomothetic knowledge about idiographic processes can be obtained” (Molenaar, 2015, p. 37). Individual-based research designs allow meticulous analyses of single cases while group-based results uncover broader tendencies that can show how these results vary in the population. If a group is the system chosen as the unit of analysis for research, or if it is any higher-level system than an individual, then it may be that group-level data are more relevant for that particular study. An integrative design at the unit of analysis would attempt to draw from both the individual-level and group-level of analysis which are complementary from a CDST perspective.

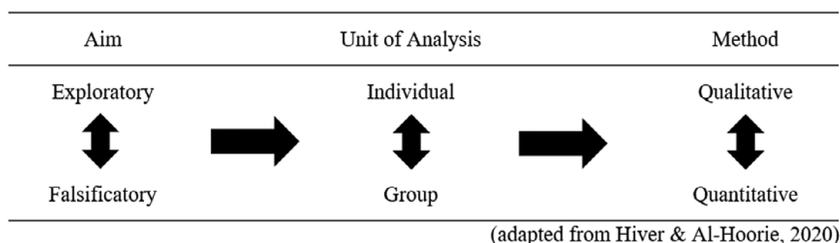


FIGURE 1. Integrative framework for CDST research designs proposed by Hiver and Al-Hoorie (2020b).

The final choice is the method. Integrative CDST designs can draw from both qualitative and quantitative methods to advance knowledge in a particular area of applied linguistics. CDST research encourages mixing quantitative or qualitative methods to investigate broad questions of interest (Hiver et al., 2021; Larsen-Freeman & Cameron, 2008). Whether quantitative, qualitative, or some integrated combination, CDST methods deal primarily with longitudinal data if they operate using the adaptive principle, but may also apply to cross-sectional data if concerned with the relational principle. Longitudinal data and designs are usually more CDST compatible because these focus on the outcomes or patterns that are reached at different points in time as well as the mechanisms that explain how an outcome is reached. Additionally, it is nearly impossible to study change and development (the adaptive principle) without also accounting for context and interconnectedness (the relational principle).

## THE PRESENT STUDY

As 25 years has passed since CDST was introduced to the field, it is time to look back at this body of research and systematically review it. As mentioned in the preceding text, we approached this scoping review project with two parallel objectives—one descriptive and one substantive. These correspond with our research questions. Given this body of research spanning the 25 years from 1994 to 2019, we asked the following research questions:

- RQ1. What are the methodological characteristics of CDST studies in the field (including participants, contexts, timescales, and analytic strategy)?
- RQ2. What are the substantive contributions of these CDST studies to the field?
- RQ3. What areas for improving CDST study quality are apparent?

## METHOD

### INITIAL SEARCH

We conducted a search for studies spanning the 25-year period of interest (1994–2019). We chose this period because 1994 marks the date of the very first contribution on the topic of complexity theory/dynamic systems theory in the field—a conference paper delivered by Larsen-Freeman (1994) at the Second Language Research Forum. Our scope covered peer-reviewed articles, book chapters, conference papers and proceedings, and

doctoral dissertations. We conducted our search in databases relevant to our field (i.e., ERIC, MLA, ProQuest, and PsycINFO) using the search terms shown in Figure 2. As we describe, we also looked beyond the results of the database searches at this stage to ensure that important and pertinent research reports were not overlooked. Figure 3 shows this entire process.

As Alexander (2020) proposes, when constructing a report pool, a robust search procedure must justify the specific delimitations instituted with consideration of the potential consequences of those decisions. With this in mind, we first specified where

("complex\* theor\*" OR "complex dynamic system\*" OR "dynamic system\* theor\*" OR "CDST") AND ("applied linguistic\*" OR "foreign language" OR "second language" OR "L2") AND at.exact("Book Chapter" OR "Dissertation/Thesis" OR "Conference Proceeding" OR "Conference Paper" OR "Article") AND la.exact("English") AND PEER(yes)

FIGURE 2. Database search terms.

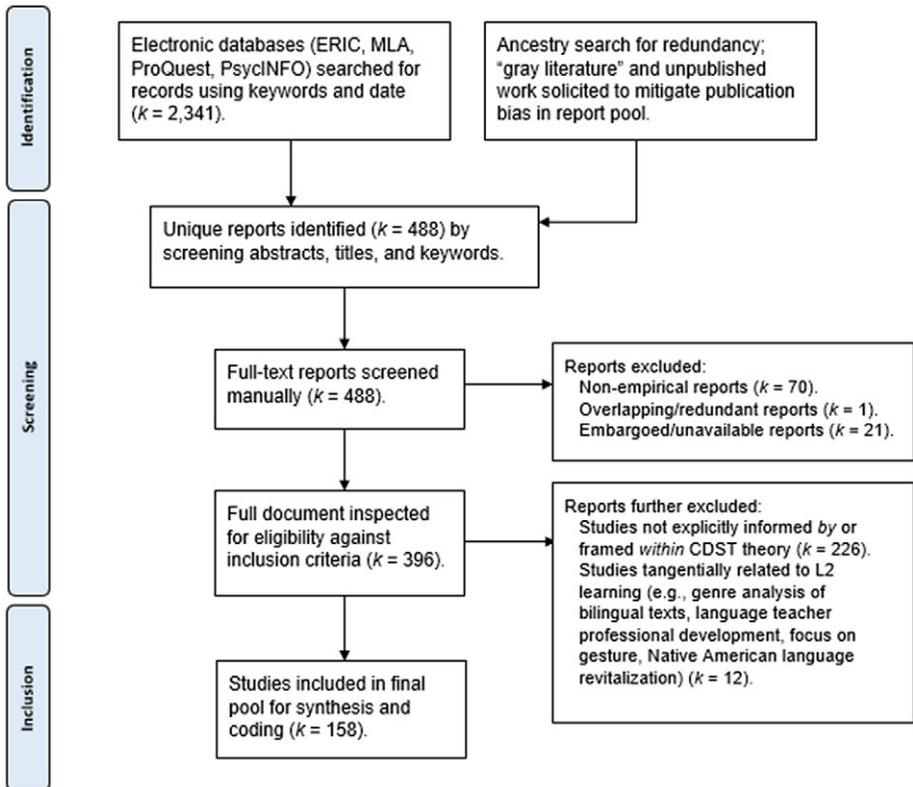


FIGURE 3. PRISMA flow chart illustrating identification of studies through database search. Source: Page et al. (2021).

search terms should appear (i.e., in one or both the abstract or main text) to avoid the false negatives likely to arise from either more generic or polysemous use of the term *complexity* (e.g., used to denote a measure of language production) exclusively in titles and keywords. This restriction also enhances the replicability of our approach. This search returned a total of 2,341 hits from the combined database. We then supplemented this pool by a Google Scholar search and an ancestry search as redundancy checks.

To mitigate selection and publication biases, we also set out to intentionally incorporate so-called gray literature (Rothstein & Hopewell, 2009) in our report pool. This includes nontraditional research documents that are found outside of typical publishing venues such as organizational reports, working papers, and conference proceedings. Finally, we put out a call to solicit unpublished work, edited volumes not cataloged by the search engines, or preprints we might have missed in our search. We then examined this total report pool against the inclusion criteria described in the next section.

### **INCLUSION CRITERIA**

To be eligible for inclusion in this scoping review, the report had to satisfy the following criteria:

1. It must involve an empirical design (whether quantitative, qualitative, or mixed method). Methodological and conceptual articles<sup>3</sup> were excluded.
2. It must explicitly identify itself as operating within, or informed by, CDST or its terminological antecedents.
3. It must be related to language learning. Reports on either nonlanguage education or theoretical linguistics were excluded.
4. It must be in English.
5. It must be available before August 2019.

Here we must add several caveats about our inclusion criteria. First, work in the field over the past two decades has emerged from two related theoretical frameworks—complexity theory (CT) and dynamic systems theory (DST) (see e.g., Larsen-Freeman, 2007). As many readers and scholars in this domain will suspect, it is unlikely that those working with “complexity theory” in SLD were doing different work than those working with “dynamic systems theory.” Consensus simply had not yet been reached on terminology. CDST is a more recent amalgam<sup>4</sup> that reflects the self-organization of nomenclature. While it has become the field’s theoretical umbrella term of choice, it is an emergent entity with both new and existing properties of CT and of DST (e.g., Larsen-Freeman, 2017). Because we locate much of our own work within this paradigm, we were hyperaware of this terminological diversity and were explicit about looking for these terminological antecedents in the report pool.

Second, unlike some synthetic work in our field guided by specific questions (e.g., “how effective is form focused instruction?”; see Kang et al., 2018), here it is a theoretical framework that drives our inclusion criteria. As a result, an element of self-selection is inherent when filtering out all studies that did not self-identify as being CDST research. The procedural challenge in creating such a report pool is, of course, that the decision of which studies to include or exclude markedly influences the outcome of the review. For instance, we are aware of several empirical studies routinely cited by CDST scholars as exemplars of this approach that are nevertheless not framed by the original authors as

CDST research, or that never mention being informed by CDST (see e.g., Eskildsen, 2009). However, a scoping review with search terms and inclusion criteria like ours could not subjectively include such studies on a case-by-case basis as report pool construction would become arbitrary and lack reproducibility. We cast a wide net with this inclusion criterion and sampled self-labeled CDST studies without prefiltering how robust this self-labeling was or whether studies focused exclusively on CDST. Because of this, the report pool included a heterogeneous array of topics and themes. We, therefore, acknowledge this limitation, and are cautious in interpreting this report pool as a flawless representation of CDST research on second language learning.

**CODING**

Applying our inclusion criteria first to the title-abstract-keyword of all unfiltered reports, we obtained a total of 488 reports (see Figure 4 for a yearly breakdown). No proceedings, conference papers, edited book chapters, or unpublished work met all our inclusion criteria, primarily due to lack of explicit detail as to how they were informed by CDST. Journal articles in this pool were primarily, though not exclusively, from SSCI and SCOPUS indexed journals. While these journals have been observed to present high-quality research, which the field trusts as both robust and consistent (Andringa & Godfroid, 2020), restricting reports to such journals may present a representativeness limitation. For this reason, we did not undertake any further filtering of journal articles. Presumably because CDST is perceived as a comparatively novel theoretical orientation that has high potential for application in empirical work, there were many dissertations in our pool. For the sake of a comprehensive sample and parsimonious analysis, here we combined dissertations with journal articles—though we acknowledge that dissertations often do not tend to follow conventional journal preferences, are broader in scope, and often include innovative ideas, but also undergo a somewhat different review process than peer-reviewed journal articles. This pool of studies was then manually inspected against the inclusion criteria by all three authors and discussed until 100% agreement was

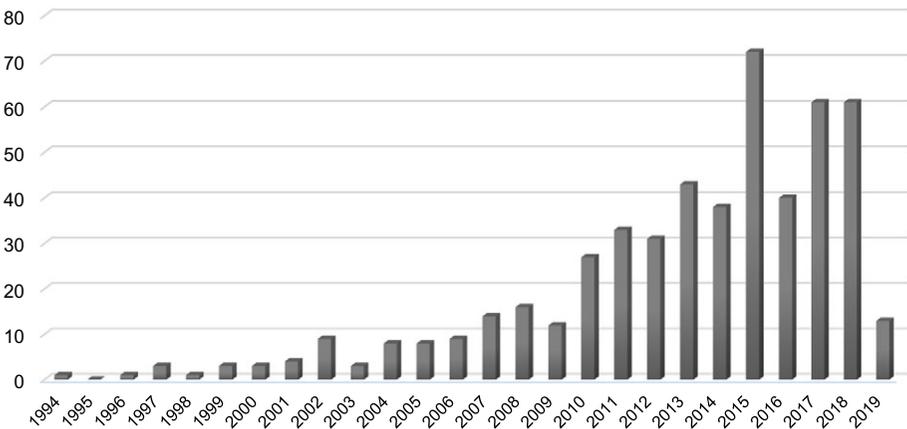


FIGURE 4. Studies filtered from initial search ( $k = 488$ ) by year of publication.

reached. As a result, 158 reports were retained in the final pool (89 journal articles and 69 dissertations).

These 158 reports were then coded individually using a descriptive categorization scheme (see Supplementary Material) that included detailed markers such as study design and length as well as more substantive descriptors such as empirical contribution and study limitations. Each researcher coded a third of the final pool. To validate these judgments a second researcher along with a team of two trained coders independently coded 30% of all reports. The observed interrater agreement (83.6%) across coding categories was above the conventional 80% threshold (McHugh, 2012) and the observed kappa ( $\kappa = .67, p < .001$ ) approached conventional agreement standards. While kappa is a conservative estimate of interrater agreement, especially as possible categories increase (Brutus et al., 2010), we consider the reliability of our coding to be acceptable, but we acknowledge that future researchers may improve upon it.

## RESULTS

### METHODOLOGICAL CHARACTERISTICS

Starting with the characteristics of participants found in CDST research, varied sample sizes and participant age groups were included. Figure 5 shows that 14 studies included an  $N$  of 1, and that in this pool there were fewer studies as sample size increased. When combined with several other design characteristics, this highlights the increasing importance of individual-based and idiographic research. Though a handful of studies included larger samples, perhaps due to CDST's interest in the individual learner sample sizes tended to be modest. Roughly 40% of all studies featured a sample size of  $N \leq 10$ , and only 13 studies in the entire report pool included a sample of  $N > 100$ . The largest sample size in the article pool was  $N = 924$  ( $Mdn = 13.5$ ,  $IQR = 31$ ), while the largest sample size in the dissertation pool was  $N = 1,723$  ( $Mdn = 16$ ,  $IQR = 28.5$ ). Within this pool, studies with younger participants were clearly the minority (Table 1), with 112 studies (70.8%)

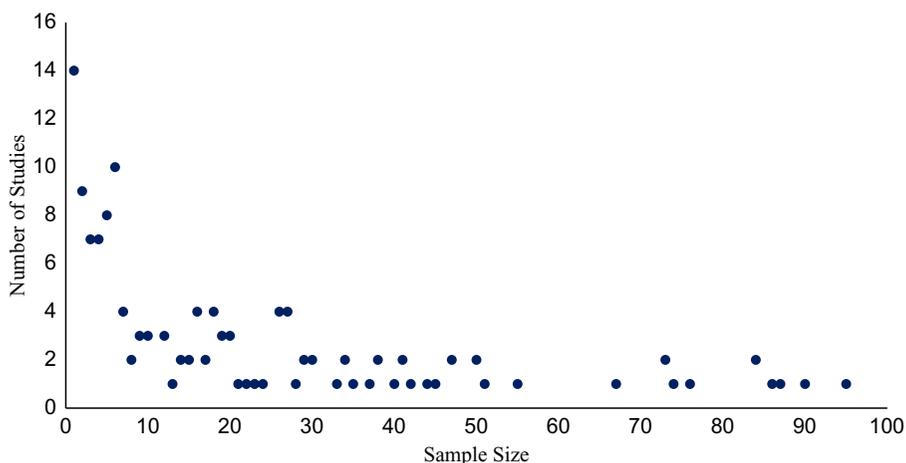


FIGURE 5. Sample size of studies in the report pool.

TABLE 1. Participant characteristics in CDST research

	<i>k</i>	%
Age		
young children (< 7)	4	2.5%
primary school (7–12)	10	6.3%
secondary school	15	9.4%
university (18+)	48	30.3%
other adult (25+)	64	40.5%
multiple, mixed age groups	8	5%
unspecified	9	5.6%

Note: *k* = 158.

sampling either university students or adults aged 18 or older. The rarest were studies with participants aged seven years and younger (4 studies) followed by those with respondents aged 7–12 (10 studies). Eight studies featured multiple, mixed age groups, while the age of participants was unspecified in nine studies. While this may reflect some of the field's sampling tendencies in general these characteristics have remained unexplored in CDST research to date.

Because CDST is a relational-contextual perspective in which spatiotemporal context plays an integral role in making sense of empirical findings, we expected adequate depth of contextual detail to feature in the studies we reviewed. Table 2 shows that a wide range of research contexts were represented in the study pool, with foreign and second language learning contexts accounting for 132 studies (83.5%) of the total. Other research contexts were only minimally present, including bilingual language contexts, heritage language contexts, and a mix of several of these within the same study.

Various instructional settings were also part of this pool. In addition to the 79 studies (50%) that took place in conventional instructed language settings, our pool showed that only a handful CDST studies have been conducted in online learning, in immersion environments, in study abroad contexts, or in language for specific purposes classrooms. Only three studies investigated untutored, naturalistic language learning. Considering the importance of context in CDST research, the number of studies that left unspecified either the research context (14 studies; 8.8%) or the instructional setting (41 studies; 26%) was large—a point we turn to in our discussion.

Participants also represented various L1 backgrounds and target L2s (Table 2). We categorized a total of 24 different L1s here based on their geographical origin for the sake of parsimony (i.e., some studies featured multiple languages). Far fewer target languages were featured. Among these, what stands out is the dominance of L2 English as a target language, accounting for nearly 70% in the pool. Though we only included reports written in English, this imbalance is perhaps to be expected given the global importance of L2 English. It also stands in contrast to the relatively low frequency of other languages that are, arguably, equally widespread and important target languages. Spanish was the second most represented L2 in our pool (10.1%), while some world languages were featured in just a single study. Finally, eight studies did not specify the target language in question.

Turning to study design characteristics, we looked at the general approach to study design as well as the timescale of data collection in the reviewed studies (Table 3).

TABLE 2. Contextual characteristics in CDST research

	<i>k</i>	%
Research Context		
foreign language	90	56.9%
second language	42	26.5%
bilingual	2	1.2%
heritage	1	0.6%
mixed	9	5.6%
unspecified	14	8.8%
Instructional Setting		
generic instructed setting	79	50%
online/VLE	15	9.5%
immersion	10	6.3%
language for specific purposes	5	3.1%
uninstructed setting	3	1.9%
study abroad	2	1.2%
mixed	3	1.9%
unspecified	41	26%
Participant L1		
European languages	95	60.1%
Asian languages	45	28.4%
Middle Eastern languages	10	6.3%
specified as “mixed”	23	14.5%
unspecified	7	4.4%
Target L2		
English	110	69.6%
Spanish	16	10.1%
French	9	5.6%
German	8	5.0%
Chinese (inclusive)	7	4.4%
Japanese	6	3.8%
Arabic	2	1.2%
Swedish	2	1.2%
Finnish	1	0.6%
Italian	1	0.6%
Persian	1	0.6%
Polish	1	0.6%
Portuguese	1	0.6%
Vietnamese	1	0.6%
unspecified	8	5.0%

*Note:*  $k = 158$ . Numbers for participant L1s may not sum to the total number of studies due to some studies including multiple samples. Numbers for target L2 may not sum to the total number of studies due some studies including multiple target languages. VLE = virtual learning environment.

Whereas over a third (59 studies) were cross-sectional, more than 53% of studies (84 studies) were longitudinal in design. In relation to the field more generally, this is a substantially higher proportion (Al-Hoorie & Vitta, 2019). The overall approach to data collection or data sampling was ambiguous in the remaining 15 studies. Examples of these include analyses of users’ asynchronous chat messages, video observations of classroom interaction patterns, computer-assisted corpus analysis, and analysis of classroom

pedagogical artifacts. With regard to study length, data elicitation took place most often over a span of months (54 studies), followed by studies with a timespan of weeks (33 studies), years (32 studies), hours (9 studies), and days (5 studies). Comparative data from other reviews in the field indicates that this proportion of studies with a time window of months and years is markedly higher in CDST studies (Vitta & Al-Hoorie, 2020). Study length in our report pool ranged from 90 minutes to 4 years. Note that these numbers do not refer to the frequency of data elicitation but to the duration of the study. More often than not, details regarding the frequency of data collection were not specified in these studies, which made it difficult to determine, for instance, if studies with a timespan measured in weeks elicited data from participants daily over this period, twice (at the start and end of this period), or only once per participant over the course of the study.

With reference to dynamic method integration (Table 4), CDST research entails design decisions at several distinct levels: study aim, unit of analysis, and choice of method. It is perhaps notable that more than 80% of studies (130 studies) were exploratory and only 28 studies had a falsificatory aim, that is to test hypotheses empirically that are related either to CDST principles (e.g., that intraindividual variation is informative about development) or topically circumscribed predictions (e.g., that there are regularities in trajectories of L2 development). No single study we reviewed combined both exploratory and falsificatory aims, a finding that seems counter to the hybrid nature of a great deal of research in the field. However, by necessity we coded these notions (confirmatory vs. exploratory) from the research objectives formulated by studies in the report pool and from characteristics of their research designs, not by examining claims made by authors that their data “confirmed” or “supported” certain conclusions after the fact.

The choice of unit of analysis was also straightforward for many studies in this pool. The unit of analysis in 73 studies was the group, and in 70 studies it was the individual. Six studies specified the unit of analysis as texts (i.e., learner language), and the unit of analysis was unspecified in four studies. There were five studies in this pool that included both individual analyses and group analyses as explicit comparisons across levels. These we classified as having more than one unit of analysis. While this is a very small subset of studies, they illustrate the extent to which relying exclusively on group-level data and

TABLE 3. Timescales in CDST research

	<i>k</i>	%
Approach		
longitudinal	84	53.1%
cross-sectional	59	37.3%
unspecified	15	9.5%
Study Length		
hours	9	5.6%
days	5	3.1%
weeks	33	20.8%
months	54	34.1%
years	32	20.2%
unspecified	26	16.4%

Note: *k* = 158.

TABLE 4. Elements of the dynamic method integration framework in CDST research

	<i>k</i>	%
Study Aim		
exploratory	130	82.3%
falsificatory	28	17.7%
both	0	0%
Unit of Analysis		
group	73	46.2%
individual	70	44.3%
texts	6	3.8%
more than one unit	5	3.1%
unspecified	4	2.5%
Methodology		
qualitative	74	46.8%
quantitative	46	29.1%
mixed	36	22.8%
unspecified	2	1.2%

Note: *k* = 158.

insights may impoverish the field's understanding of various phenomena (see also Lowie & Verspoor, 2019).

Table 4 further shows that choice of method was split across qualitative (74 studies), quantitative (46 studies), and mixed methods (36 studies). Here we adopted an inclusive definition of *methodology* related to the purpose, focus, design, procedures (e.g., means of sampling, data collection, and analysis) of studies in the report pool. Two studies in the total pool did not describe their methodological choices clearly. The large number of purely qualitative studies may reflect the general tendency for newcomers (e.g., graduate students or scholars newly interested in CDST) attempting to apply methods for investigating interconnectedness and dynamic development to default to methods that “capture rich dense datasets” (Ushioda, 2021, p. 252). This is borne out in our data, with roughly 80% of dissertations in our pool drawing heavily on qualitative designs. While our review in no way suggests that exclusively qualitative methods are poorly suited to studying complexity and dynamicity, we did find particular limitations in the present pool of studies, two of which relate to collecting data and adopting analyses that do not lend themselves to either investigating connections in context or to dynamic change and development. We leave discussion of these issues until later.

Closely related to the design decisions we reviewed in the preceding text are the choices of data elicitation methods and data analytical strategies. In contrast with methodological work suggesting that CDST research should both innovate with existing methods and expand on these (e.g., Lowie, 2017; MacIntyre et al., 2017), we found that a range of conventional and widely used techniques for data collection were present in reviewed studies (Table 5). The technique most frequently adopted was interviews and focus groups (68 studies; 43%). Other data elicitation methods included analysis of written samples of learner language, oral language/interaction samples, and observations. Surveys, tests, and pedagogic tasks were also commonly employed by CDST researchers. Other data sources used more sparsely included think-aloud protocols, stimulated recall, and field notes.

TABLE 5. Analytical strategies in CDST research

	<i>k</i>	%
Data Collection		
interview/focus group	68	43%
written language sample	44	27.8%
survey/questionnaire	41	26%
lesson observation	35	22.1%
tests	25	15.8%
oral language/interaction sample	23	14.5%
pedagogic task	17	10.8%
field notes/memos	12	7.6%
stimulated recall	4	2.5%
think-aloud protocols	2	1.2%
other	13	8.2%
Analysis Technique		
qualitative coding/analysis	64	40.5%
conventional inferential statistics	27	17.1%
dynamic statistical analyses	24	15.2%
descriptive statistics	8	5%
advanced multivariate statistics	4	2.5%
unspecified	74	46.8%

Note:  $k = 158$ . Numbers for data collection strategy and analysis technique may not sum to the number of studies due to many studies including multiple types of data and multiple analyses.

Thirteen studies featured other types of data elicitation tools such as samples of student academic work, drawing tasks, or momentary sampling measures (e.g., the idiodynamic approach—a research template that collects data on time-dependent variation within a single individual or unit). Notably, the majority of studies in the report pool, in both the article and dissertation subsets, included multiple complementary data sources. Studies that did so included at least two but often up to four data sources in combination, and were distributed across nearly all years. This may reflect a general tendency to approach data collection in CDST research with a “more is more” mentality: because everything counts, everything is connected, and everything changes, study design may have followed the premise that more data is more appropriate to examine such phenomena fully.

Turning to analysis techniques, qualitative coding and analysis methods appeared to be those employed most often in the reviewed studies (64 studies; 40.5%), perhaps a logical extension of the large number of studies that adopted qualitative data collection techniques. This was nearly triple the frequency of the next largest category of analysis techniques. Qualitative data analysis techniques here included content and discourse analysis, ethnographic analysis, inductive thematic coding or grounded theory analysis, and metaphor analysis. Twenty-four other studies (15.2%) adopted dynamic statistical analysis such as using the coefficient of variation (2 studies), min-max graphs and moving correlations (5), recurrence quantification analysis and Monte Carlo simulation (3), growth curve modeling (1), time-series analysis (5), generalized additive mixed-effects models (1), state space plots and grids (1), fractal analysis (1), or trend analysis (3) and timeplots (2). When examining other data analytical strategies, we found that eight studies relied on descriptive statistics alone (not including studies reporting effect sizes) and a

further 27 studies adopted conventional inferential statistical analyses. These included analyses such as *t*-tests, canonical correlations, analyses of variance (ANOVA), and linear regression analysis. A handful of other advanced multivariate statistical analyses were used (four studies), including factor analysis and principal components analysis, cluster analysis, and latent variable modeling (i.e., SEM). We also found a large number of instances (74 studies; 46.8%) in which the data analysis technique was either unclear or unspecified—examples of this include unintuitive descriptions such as “we analyzed our data in Excel” or “the data were coded manually.” The finding that a large proportion of studies<sup>5</sup> did not fully establish methodological integrity for the reader is one we return to in the following text when reflecting critically on our other research questions.

### **SUBSTANTIVE CONTRIBUTIONS**

In addition to methodological characteristics of these studies, we were also interested in determining what substantive contributions this pool of studies has made to the field. Because we cast a wide net and sampled self-labeled CDST studies without prefiltering how robust this self-labeling was or whether studies focused exclusively on CDST, the report pool included a heterogeneous array of topics and themes (e.g., learners’ perceptions toward classroom tasks, how digital games mediate language use, language attrition in first generation immigrants, and the development of authorial voice and rhetorical knowledge in L2 writing, etc.). Across all these we looked at contributions in two broad areas: first, empirical contributions and, second, practical contributions (i.e., related to both research and pedagogy) to the field.

Table 6 shows that empirical contributions were demonstrated in a variety of areas. Two of the most noticeable contributions were that studies reported evidence supporting the claim that the phenomena or constructs under study were indeed complex and dynamic: Thirty-one studies (19.6%) corroborated the existence of dynamic regularities in development, and another 29 studies (18.3%) provided evidence of system interconnectedness and interaction between elements being studied. Other notable contributions included evidence of the influence of context in development, of the nonlinearity of development or the presence of nonlinear predictors, of emergent outcomes and patterns, and of system adaptation or self-organization in response to inputs or to contextual affordances. Among other contributions were studies that provided evidence of inter- and intraindividual variability, as well as studies illustrating the methodological value of applying CDST tools to advance understanding in the field and the compatibility of CDST with previous research drawing on other diverse paradigms. A small number of also established evidence of sensitivity to initial conditions and of equifinality—the notion that a given state or outcome can be reached through multiple pathways. Here we intentionally focused our coding on these categories because many of these contributions are distinguishing features of CDST that other theories do not account for or even investigate.

We were, of course, interested in what practical contributions CDST studies have made to the field. Such contributions are the subject of recent work (e.g., Levine, 2020) and, because they are sought after by many, perceptions that such applications are not readily accessible may act as a curb on wider uptake of CDST in the field (Dewaele, 2019). Table 7 shows that practical contributions in reviewed studies were not few in number. Contributions ranged widely from studies offering direct pedagogical insights (34 studies)

TABLE 6. Empirical contributions to second language development

	<i>k</i>	%	<i>Sample studies</i>
Evidence of dynamic regularities in development	31	19.6%	(Baten & Hakansson, 2015; Lenzing, 2015; Nelson, 2011; Pham, 2011)
Evidence of system interconnectedness/interaction	29	18.4%	(Jessner et al., 2015; Plat et al., 2018; Serafini, 2017)
Methodological innovation/advances	22	13.9%	(Caspi & Lowie, 2013; Huang, 2010; Murakami, 2014; Poupore, 2018)
Evidence of nonlinear predictors/growth	21	13.3%	(Churchill, 2007; de Leeuw et al., 2013; Spoelman & Verspoor, 2010)
Evidence of the role of context in development	20	12.7%	(Fogal, 2015; Hepford, 2017; Kostoulas, 2015; Pomerantz & Bell, 2007)
Evidence of emergent patterns	19	12.0%	(Baba & Nitta, 2014; King, 2013; Larsen-Freeman, 2006; Van Horn, 2017)
Compatibility of CDST with other paradigms	18	11.4%	(Feryok, 2010; Göpferich, 2013; Viswanathan, 2016)
Evidence of interindividual variability	14	8.8%	(Fukuda, 2014; Kopečková et al., 2016; Larsen-Freeman, 2006)
Evidence of system adaptation/self-organization	13	8.2%	(Bragg, 2018; Larsen-Freeman, 2006; Reigel, 2008; Roehr-Brackin, 2014)
Evidence of intraindividual variability	12	7.6%	(Kim et al., 2017; Lowie et al., 2017; Opitz, 2013; Polat & Kim, 2013)
Evidence of nested systems	10	6.3%	(Ebert et al., 2014; Liu & Chao, 2018; Mercer, 2011; Pham, 2016)
Evidence of sensitivity to initial conditions	9	5.6%	(Finch, 2010; Perales Escudero, 2011; Waninge et al., 2014)
Evidence of equifinality	9	5.6%	(Murakami, 2014; Smyk, 2012; Souza, 2013)
Complementarity of group and individual levels	3	1.9%	(Lowie & Verspoor, 2019; Xiao, 2016)
Evidence of fractal nature of development	1	0.6%	(Evans, 2019)
Evidence of multicausality	1	0.6%	(Morel, 2019)

*Note:*  $k = 158$ . Number of theoretical/empirical contributions may not sum to the total number of studies due to some studies including multiple contributions. Full references to the studies cited in this table are listed in the online supplementary material.

and explicit discussion of a fuller, more multidimensional understanding of the phenomena under investigation (24 studies), to the explanatory power of contextual factors in developmental over and above other explanans (23 studies), and confirmation of the particularities of individuals and intraindividual variation (13 studies). Another contribution was the emergence of new previously undiscovered or unapplied criteria for existing issues (10 studies)—for example, using notions of system adaptation from CDST in understanding the development and maintenance of multicompetence, and drawing insights from both CDST and evidence regarding maturational constraints in relation to L1 attrition during L2 acquisition. Other practical contributions related closely to applications for research across these heterogeneous topics. This includes studies that applied a novel perspective that helped uncover new insights into the phenomena under investigation (23 studies), studies that shifted attention to new aspects of existing phenomena (13 studies), or those that showed the limitations of existing perspectives (9 studies). Still others made contributions by integrating multiple complementary data sources (17 studies), developing new conceptual tools for the topics being studied (10 studies), tapping

TABLE 7. Practical contributions to the field of language learning

	<i>k</i>	%	<i>Sample studies</i>
Pedagogical insights	34	21.5%	(Fukuda, 2014; Mercer, 2011; Yashima et al., 2018)
Multidimensional understanding of issues	24	15.2%	(Baba & Nitta, 2014; de Leeuw et al., 2013; Hepford, 2017; Nelson, 2011)
Importance of context	23	14.5%	(Braga, 2013; Thompson, 2017; Waninge et al., 2014)
Applying a novel perspective to topic under investigation	23	14.5%	(Jessner et al., 2015; Polat & Kim, 2013; Xiao, 2016)
Integrating multiple complimentary data sources	17	10.8%	(Kikuchi, 2017; Lasagabaster, 2017; Pham, 2016)
Attention to unexplored aspects of existing phenomena	13	8.2%	(Churchill, 2007; Lowie et al., 2016)
Learner differences/variability as source of information	13	8.2%	(Lowie & Verspoor, 2019; Rosmawati, 2014; Spoelman & Verspoor, 2010)
Emergence of new insights on existing topic	10	6.3%	(Larsen-Freeman, 2006; Serafini, 2017; Wang, 2016)
Increased ecological validity	10	6.3%	(Bragg, 2018; Cameron & Deignan, 2006; Scholz, 2017)
Developing new conceptual tools	10	6.3%	(King, 2013; Link, 2015; Plat et al., 2018)
Evidence of the limitations of existing perspectives	9	5.6%	(Baten & Hakansson, 2015; Kounatidis, 2016; Larsen-Freeman, 2006)
Capturing phenomenological reality	8	5.0%	(de Bot, 2015; Onitsuka, 2018; Zheng, 2016)
Demonstrating feasibility of a transdisciplinary approach	1	0.6%	(Göpferich, 2013)

*Note:*  $k=158$ . Number of practical contributions may not sum to the total number of studies due to multiple contributions coded in some studies and none in others. Full references to the studies cited in this table are listed in the online supplementary material.

into greater phenomenological reality in the issues under investigation (8 studies), and achieving superior ecological validity (10 studies).

### STUDY QUALITY

Our third and final research question relates to methodological rigor and what areas, if any, were apparent for improving CDST research going forward. To this end, we examined apparent limitations of study design (Table 8) in our review pool. Note that these were design limitations we explicitly coded as such and not those listed by authors as limitations of their studies.

Some of the most prevalent design issues we identified were studies relying on data or analyses that were seemingly inappropriate for investigating change and development (41 studies; 26%), and studies relying on data or analyses that were poorly suited to investigating connections in context (22 studies; 14%). For instance, it is not hard to appreciate why studies drawing on a single round of interviews or cross-sectional test data at one or two time points would struggle to shed light on such issues. This result was also incongruent with the strong evidence in this pool that phenomena of interest or constructs under study were complex and dynamic (see Table 6). We return to this unanticipated finding further in the text that follows and reflect on the extent to which these studies were indeed informed by CDST in their design.

TABLE 8. Study limitations in CDST research

	<i>k</i>	%
data/analyses incongruent with investigating dynamic change	41	26.0%
limited scope of data	31	19.6%
data/analyses incongruent with investigating connections in context	22	14%
sample selection bias	21	13.3%
findings with limited transferability/applicability	15	9.5%
ambiguity in applying CDST concepts	15	9.5%
contextual factors underspecified	14	8.8%
exclusive metaphorical application of CDST	12	7.6%
analytical strategy underspecified	10	6.3%
ecological fallacy	8	5.0%
participant information underspecified	6	3.8%
statistical assumptions violated	4	2.5%
abstractness of the analytical strategy	3	1.9%

Other design limitations we observed included the limited scope of data many studies drew conclusions from (31 studies; 19.6%) and sample selection bias (21 studies; 13.3%), evident, for example, in studies with no sampling frame or a nonpurposive sample. Evident here too was the limited transferability or generalizability of a handful of accompanying conclusions to similar samples or contexts (15 studies; 9.5%), due to inattention to external validity. It is rarely generalizability in its conventional sense that CDST scholars are chasing (Hiver & Al-Hoorie, 2020b; Larsen-Freeman, 2017). However, especially when considering the lack of detail in specifying contextual factors (Tables 2, 3, and 8) and data analysis techniques adopted (Tables 5 and 7) that was apparent in some studies, this finding was not entirely unanticipated.

Several other limitations in study design highlighted through our coding include the presence of some ambiguity in the application of CDST concepts and terminology (15 studies; 9.5%). This may be partly due to our inclusion criteria which selected for studies self-labeled as CDST. In several studies, for example, readers are presented with direct claims about the importance of CDST for the research but based on the questions explored in the study and the design and methods used; it was unclear how CDST had informed the study. In several other studies that were terminology heavy, it was unclear in lay terms what the “system” being discussed by the researchers was, what precisely made it “adaptive,” “self-organizing,” or “nonlinear” in nature, or what patterns were “emergent.” This limitation links to another, regarding the exclusive metaphorical application of CDST applying only its terms or concepts (12 studies; 7.6%)—these were distributed nearly equally across report type and year of publication. Larsen-Freeman and Cameron (2008) propose that CDST is a necessary metaphor that can “push the field towards radical theoretical change” (p. 11) but they are equally clear that CDST is much more than metaphor when it is “literalized into field-specific theory, research, and practice” (p. 15). We agree, and discuss below how future applications of CDST might extend beyond its value as metaphor.

Other less frequent study limitations we observed included underspecified participant information and analytical techniques (6 and 10 studies respectively), the ecological fallacy—assuming that relationships observed for groups apply equally for individuals

and vice versa (8 studies) (Lowie & Verspoor, 2019), and violation of basic statistical assumptions (4 studies) (see e.g., Al-Hoorie & Vitta, 2019). Taken together these limitations point to some clear implications regarding areas for improving CDST study design going forward.

## DISCUSSION

This scoping review looked first at the methodological characteristics of CDST research, at the contributions this body of research has made to the field, and finally at CDST study quality. Our review pointed to clear trends in how the field has investigated complex and dynamic phenomena of interest and—based on this body of research—what shared concerns and issues in the field we now think of differently.

First, this body of work clearly supports the claims that have been made in the theoretical literature that language, language use, and language development/learning are complex and dynamic—these are all notions, our review suggests, that are now undisputed. The two most prominent contributions that studies in our review made are in fact related to the existence of dynamic regularities in development and the complex, interconnected, and interactive nature of the topics and constructs under investigation (Table 6). As mentioned earlier, scholars have previously highlighted several core objectives of CDST research in applied linguistics (Hiver & Al-Hoorie, 2016). It is clear from our review that the field has made particularly strong advances relating to the first two of these objectives (i.e., describing various complex systems and identifying various patterns of dynamic change in context), and has begun work on the third objective (i.e., modeling complex mechanisms and dynamic patterns), but—despite more than 50% of studies collecting data from an instructed L2 setting—has left the remaining objective largely aside (i.e., understanding how to intervene or influence systems' behavior). Applied linguists arguably aim to go further than mere description and enact certain forms of complex praxis in social contexts (Al-Hoorie et al., 2021; Larsen-Freeman, 2016a). Application of a field's scientific findings and insights is one of the most important modes of social science research. By consequence, with more than two and a half decades of thinking and research on the matter, continued work with descriptive findings limited to insights such as “phenomenon X is complex in its make-up” or “process Y is nonlinear in its development” is unlikely to push the field forward in a substantive way at this stage as such claims are now already established.

The contribution of CDST work going forward will be to offer more robust explanatory conclusions at increasingly relevant timescales and levels of resolution. Given the shift of perspective that accompanies a familiarity with CDST, there is a need for greater work on systemic interventions (Byrne & Callaghan, 2014). While the field has been quick to amass evidence that many phenomena are relational, nonmechanistic, and indeterminate in their development, as an applied field we have yet to do the necessary work to understand whether and how to intervene in and influence the complex dynamic realities of the phenomena under investigation. Here, by intervene in and influence systems we mean intentionally generate positive change that is complex, situated, iterative, time-scaled, and reciprocal in nature (see e.g., Steenbeek & van Geert, 2015; van Geert & Steenbeek, 2014, for similar arguments). Criteria are also needed for developing and evaluating these systemic interventions that are sensitive to features of context-

dependence, multiplicity, and interactions. Complex interventions will be those designed to respond adaptively to a number of relational components in context, when various levels of the system (e.g., individual, group, or organizational levels) are targeted by the intervention, managing a number of anticipated and surprising behaviors manifested by those involved, and leading to variability in outcomes.<sup>6</sup> As our finding that more than 82% of studies had an exploratory and descriptive aim suggests, we have much to do to think in CDST terms about deliberate intervention and to develop research tools for this (Osberg & Biesta, 2010).

Second, in both empirical and applied terms, the important role of context in understanding development is clearly apparent. It has almost become a truism for studies to conclude that the spatiotemporal context plays an integral role in affecting development. The fact that “outcomes and change not only emerge in context, they are also mediated and adapted by contextual factors” (Hiver & Al-Hoorie, 2016, p. 746) is an integral part of necessary design considerations for CDST research. This conclusion, however, must also be juxtaposed with the somewhat surprising number of studies reviewed in which the research setting or the instructional context were either *underspecified* or *unspecified* (see Table 2). This is especially bewildering given the large number of dissertations in the pool that purport to draw on ecological frameworks (e.g., informed by the work of van Lier, 2004) that presuppose detailed descriptions of context. It is important to be able to develop evidentiary accounts and explanations that go beyond the unique instance (Byrne & Ragin, 2009), and one way of doing so is to use contextual information to specify the range of applicability of developmental mechanisms, without essentializing context. Context, which itself changes, is much more than background variables and should be understood as more than a constellation of such macrofactors. Going forward, instead of token, perfunctory mentions that context is influential, CDST research must articulate explicitly what contextual factors are being taken into account and how context informs study design. This way, information about the role of particular contextual factors in particular causal mechanisms will come to be incorporated more clearly and more concretely in evidentiary accounts and explanations in the field (see also Kaplan et al., 2020).

Third, as a research community too, the field has developed new ways of operating that are accompanied by and that “require a different framing” (Larsen-Freeman, 2020, p. 202). The methodological characteristics of this body of CDST research have certainly made the case that idiographic research is not only valid, but also necessary and important. It has taken some time for this notion to gain traction, yet judging by the nearly 10% of studies in the pool with an *N* of 1, and a full 45% of studies—regardless of sample size—adopting the individual as the unit of analysis, this is an understanding that has gained wider acceptance. There is also significant value in the field’s growing recognition of the importance of innovating with new modes of data elicitation and dynamic analytical strategies, whether case-based or variable-based (Table 5). Expanding the methodological repertoire beyond conventional methods and developing expertise in new designs and analytical techniques are key initiatives that the field should continue to pursue (see Hiver et al., 2021; MacIntyre et al., 2017). One indication of the importance of this relates to our finding (Table 8) that many studies reviewed relied on data or analyses that were seemingly inappropriate for investigating change and development or were poorly suited to investigating connections in context. Our report pool contained studies claiming evidence for dynamic development that did not draw on data with a temporal aspect in

a way that would allow for such an interpretation. Other reports argued for evidence of intraindividual variability while looking at data in an insufficiently individual way. Form must follow function: the choice to adopt certain methods of data elicitation and analysis should be driven by the aim(s), unit(s) of analysis, and the outcome(s) or process(es) under investigation.

Other findings also indicate the need for increased transparency and rigor in methodological designs and in reporting relevant choices—issues also articulated in other subdomains of the field (see e.g., Hu & Plonsky, 2021; Marsden et al., 2018; Paquot & Plonsky, 2017). For instance, the large number of CDST studies in which the general approach to data collection and the length of study was unspecified, or the data analysis technique unclear, is cause for concern. This finding may also be linked to the large number of studies in which CDST concepts were applied ambiguously, in an exclusively metaphorical way, or due to their exploratory nature. CDST is not merely a useful set of metaphors for conceptualizing second language development phenomena: complexity is an empirical reality. As such, CDST research must move beyond the exclusively metaphorical application that describes findings with a language borrowed from CDST (Hiver & Al-Hoorie, 2020b). Metaphors may be adequate if we wish to conceptualize phenomena (Larsen-Freeman & Cameron, 2008); however, the field must move forward to operationalize and validate these phenomena and investigate them empirically (see also Brown et al., 2018; Nicklin & Plonsky, 2020; Vitta & Al-Hoorie, 2021). These findings suggest the importance of greater transparency and rigor in the design choices of future CDST research, and also underscore the need for study designs to clarify the ways in which they are informed by CDST (see Gass et al., 2021, for a detailed discussion of study quality).

As our inclusion criteria show, our search cast a wide net by including all self-labeled CDST studies in the report pool. However, our analysis highlighted the fact that this self-labeling may not always be robust or that reports did not always warrant a CDST label. Many studies in this pool appeared to operate within a CDST perspective but did not unambiguously articulate how, or only called attention to the fact indirectly or fairly late. Some studies were not substantively conceived of or designed as CDST research in any major sense of what might be expected (i.e., a focus on relational and dynamic phenomena in context). Specifying that studies explicitly identify themselves as adopting a CDST perspective or design added clarity to our report pool, but many studies went no further. What is, therefore, unclear from our review, and rarely transparent from reports themselves, is whether studies in our pool approached the phenomena of interest in an exploratory fashion and discovered that CDST principles fit their data and accounted for these phenomena well, or if studies were in fact looking for evidence of such principles in their data and so applied these *ex ante*. By not discussing how CDST informs the design and methods, studies like these run the risk of spurious assumptions of complex phenomena from a dataset that may not support these claims. This limitation points to the need for CDST research to take up preregistration and other open science initiatives in research methods designed to increase study quality (see Hiver & Al-Hoorie, 2020a).

Future applications of CDST research must be transparent about the reasons for choosing to adopt the CDST metatheory and specify why situating a study within this perspective is a sound theoretical and empirical choice (Larsen-Freeman, 2017). Articulating how CDST informs their approach to research explicitly can help researchers situate the design of their study, their research questions, data analyses, and the results and

TABLE 9. Nine tenets for CDST research

Tenet	Purpose
<i>For an individual study</i>	
1. Provide a rationale for why adopting a CDST research perspective is a sound choice	Helps guard against overly loose and opportunistic applications of CDST (i.e., applications that are purely semantic or metaphorical)
2. Articulate how CDST informs the design and methods	Helps to establish how a study substantively draws from CDST in its conception and design Helps avoid spurious assumptions of complex, dynamic phenomena
3. Specify the aim(s), unit(s) of analysis, and the outcome(s) or process(es) under investigation	Helps to increase transparency Helps to leverage an integrative design (see Tenet 8)
4. Adopt methods of data elicitation and analysis that are driven by the aim(s), unit(s) of analysis, and the outcome(s) or process(es) under investigation	Helps to ensure that methods adopted are suited to investigating connections in context and are appropriate for investigating change and development
5. Specify information about the role of particular contextual factors in particular processes or outcomes	Helps to incorporate contextual detail more clearly and more concretely in evidentiary accounts and explanations Helps to develop an understanding of contextual influences that go beyond the unique instance
<i>For a program of research</i>	
6. Identify areas for complex interventions	Allows researchers to focus on influencing, intervening in, and generating positive change (e.g., in systems) that is complex, situated, and adaptive Helps to build robust explanatory conclusions of complex, systemic change
7. Develop criteria for designing and evaluating these systemic interventions	Helps to account for objectives targeted by systemic interventions Helps to appropriately frame and assess the efficacy of adaptive interventions for various levels of systems (e.g., individual, group, or organizational levels)
8. Adopt more integrative designs	Allows researchers to integrate exploratory × falsificatory aims, individual × group analyses, and qualitative × quantitative methods Helps drive ongoing methodological innovation
9. Become comfortable with a more problem-based, transdisciplinary orientation	Helps to avoid rigid, paradigm-driven research Allows CDST research to address issues in socially useful and participant-relevant ways Allows researchers to work in transdisciplinary ways and teams

discussion more clearly within this perspective (Lowie, 2017). This can also guard against using CDST too loosely—in the sense that anything with multiple interacting parts can be construed as CDST research—and in an opportunistic, post hoc manner.

## CONCLUSION

Even though it has been a quarter of a century since it was introduced to the field, CDST is still a relatively new paradigm. The limitations we reviewed in this article are therefore a

natural part of its growth and more mainstream acceptance of this meta-theory. Yet, as is also apparent, methodological advances and applications now exist that point the way forward for the field—particularly those allowing researchers to tap into the system of within-person dynamics and draw inferences about the underlying patterns of language development (e.g., Kliesch & Pfenninger, 2021; Murakami, 2016, 2020; Pfenninger, 2020; Verspoor et al., 2021). We acknowledge that the insights and guidelines CDST offers can be overwhelming, and this can slow the progress of the field. We have therefore synthesized the methodological lessons we obtained in this review, and refer to them here, as the “nine tenets” of CDST research. Table 9 presents these tenets and the purpose of each.

We might think of CDST research in the field as now being at a crossroads. As CDST research assesses how far it has come, with one eye to the future, it is important not to simply scrutinize and critique without also offering alternatives. We hope to have done both in this paper, and our results have shown that there is robust empirical evidence as well as ample methodological guidance on which future work can build. We hope that future CDST research will draw on these lessons and continue to offer substantive insights to the field of language learning and development.

#### COMPETING INTERESTS

At the time this paper was initially submitted for review, Ali Al-Hoorie had not yet taken up duties on the *SSLA* editorial board.

#### SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <http://doi.org/10.1017/S0272263121000553>.

#### NOTES

<sup>1</sup>We use this term to mean a frame of reference for thinking that provides guiding notions for methods of scientific inquiry.

<sup>2</sup>While this was necessary for obvious reasons, the more than 70 conceptual articles are additional testament to the robustness of the field.

<sup>3</sup>While this was necessary for obvious reasons, the more than 70 conceptual articles are additional testament to the robustness of the field.

<sup>4</sup>As one reviewer pointed out, prior to the fairly recent adoption of the term “CDST,” the field used “CT” or “DST” and even “chaos theory,” though not always as entirely interchangeable concepts.

<sup>5</sup>Of these 74 studies, 25 were from the article subpool and 49 were from the dissertations subpool.

<sup>6</sup>An example from a parallel field might be *psychotherapy* in which the content of each consultation is tailored to the individual needs of patients, where each client responds in different ways to treatment, and the treatment is adapted as the program of consultations unfolds.

## REFERENCES

- Al-Hoorie, A. H., Hiver, P., Kim, T.-Y., & De Costa, P. I. (2021). The identity crisis in language motivation research. *Journal of Language and Social Psychology, 40*, 1–18.
- Al-Hoorie, A. H., & Vitta, J. P. (2019). The seven sins of L2 research: A review of 30 journals' statistical quality and their CiteScore, SJR, SNIP, JCR Impact Factors. *Language Teaching Research, 23*, 727–744.
- Alexander, P. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research, 90*, 6–23.
- Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics, 40*, 134–142.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*, 19–32.
- Bastardas-Boada, A. (2013). Language policy and planning as an interdisciplinary field: Towards a complexity approach. *Current Issues in Language Planning, 14*, 363–381.
- Blommaert, J. (2014). From mobility to complexity in sociolinguistic theory and method. *Tilburg Papers in Culture Studies, 103*, 1–24.
- Brown, A. V., Plonsky, L., & Teimouri, Y. (2018). The use of course grades as metrics in L2 Research: A systematic review. *Foreign Language Annals, 51*, 763–778.
- Brutus, S., Gill, H., & Duniewicz, K. (2010). State-of-science in industrial and organizational psychology: A review of self-reported limitations. *Personnel Psychology, 63*, 907–936.
- Byrne, D., & Callaghan, G. (2014). *Complexity theory and the social sciences: The state of the art*. Routledge.
- Byrne, D., & Ragin, C. C. (Eds.) (2009). *The SAGE handbook of case-based methods*. SAGE.
- Cowley, S. J. (Ed.) (2011). *Distributed language*. John Benjamins.
- de Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal, 92*, 166–178.
- Dewaele, J.-M. (2019). The vital need for ontological, epistemological and methodological diversity in applied linguistics. In C. Wright, L. Harvey, & J. Simpson (Eds.), *Voices and practices in applied linguistics: Diversifying a discipline* (pp. 71–88). White Rose University Press.
- Ellis, N. C., & Larsen-Freeman, D. (Eds.) (2009). *Language as a complex adaptive system*. Wiley-Blackwell.
- Eskildsen, S. (2009). Constructing another language—Usage-based linguistics in second language acquisition. *Applied Linguistics, 30*, 335–357.
- Evans, R. (2019). *Bifurcations, fractals, and non-linearity in second language development: A complex dynamic systems perspective* (Unpublished doctoral dissertation). University at Buffalo, New York.
- Gass, S., Loewen, S., & Plonsky, L. (2021). Coming of age: the past, present, and future of quantitative SLA research. *Language Teaching, 54*, 245–258.
- Han, Z.-H. (2020). Usage-based instruction, systems thinking, and the role of Language Mining in second language development. *Language Teaching*. Advance online publication. <https://doi.org/10.1017/S0261444820000282>
- Han, Z.-H. (Ed.) (2019). *Profiling learner language as a dynamic system*. Multilingual Matters.
- Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism*. Multilingual Matters.
- Hilpert, J., & Marchand, G. (2018). Complex systems research in educational psychology: Aligning theory and method. *Educational Psychologist, 53*, 185–202.
- Hiver, P., & Al-Hoorie, A. H. (2016). A dynamic ensemble for second language research: Putting complexity theory into practice. *The Modern Language Journal, 100*, 741–756.
- Hiver, P., & Al-Hoorie, A. H. (2020a). Reexamining the role of vision in second language motivation: A preregistered conceptual replication of You, Dörnyei, and Csizér (2016). *Language Learning, 70*, 48–102.
- Hiver, P., & Al-Hoorie, A. H. (2020b). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.
- Hiver, P., Al-Hoorie, A. H., & Larsen-Freeman, D. (2021). Toward a transdisciplinary integration of research purposes and methods for Complex Dynamic Systems Theory: Beyond the quantitative–qualitative divide. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. <https://doi.org/10.1515/iral-2021-0022>.
- Hu, Y., & Plonsky, L. (2021). Statistical assumptions in L2 research: A systematic review. *Second Language Research, 37*, 171–184.

- Hulstijn, J. (2020). Proximate and ultimate explanations of individual differences in language use and language acquisition. *Dutch Journal of Applied Linguistics*, 9, 21–37.
- Hult, F. (2010). The complexity turn in educational linguistics. *Language, Culture and Curriculum*, 23, 173–177.
- Kang, E. Y., Sok, S., & Han, Z. (2018). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*, 23, 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, 285–288.
- Ke, J., & Holland, J. H. (2006). Language origin from an emergentist perspective. *Applied Linguistics*, 27, 691–716.
- Kliesch, M., Pfenninger, S. E. (2021). Cognitive and socio-affective predictors of L2 micro-development in late adulthood: A longitudinal intervention study. *The Modern Language Journal*, 105, 237–266.
- Koopmans, M. (2020). Education is a complex dynamical system: Challenges for research. *The Journal of Experimental Education*, 88, 358–374.
- Kramsch, C., & Whiteside, A. (2008). Language ecology in multilingual settings: Towards a theory of symbolic competence. *Applied Linguistics*, 29, 645–671.
- Kretzschmar, W. (2015). *Language and complex systems*. Cambridge University Press.
- Larsen-Freeman, D. (1994, October). *On the parallels between chaos theory and second language acquisition* [Paper presentation]. Second Language Research Forum, McGill University, Montreal.
- Larsen-Freeman, D. (2007). On the complementarity of chaos/complexity theory and dynamic systems theory in understanding the second language acquisition process. *Bilingualism: Language and Cognition*, 10, 35–37.
- Larsen-Freeman, D. (2015). Ten “lessons” from complex dynamic systems theory: What is on offer. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 11–19). Multilingual Matters.
- Larsen-Freeman, D. (2016a). Classroom-oriented research from a complex systems perspective. *Studies in Second Language Learning and Teaching*, 6, 377–393.
- Larsen-Freeman, D. (2016b). Thoughts on the launching of a new journal: A complex dynamic systems perspective. *Journal for the Psychology of Language Learning*, 1, 67–82.
- Larsen-Freeman, D. (2017). Complexity theory: The lessons continue. In L. Ortega & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 11–50). John Benjamins.
- Larsen-Freeman, D. (2018). Resonances: Second language development and language planning and policy from a complexity theory perspective. In F. Hult, T. Kupisch, & M. Siiner (Eds.), *Bridging language acquisition and language policy* (pp. 203–217). Springer.
- Larsen-Freeman, D. (2020). Complexity theory: Relational systems in interaction and in interlocutor differences in second language development. In L. Gurzynski-Weiss (Ed.), *Cross-theoretical explorations of interlocutors and their individual differences* (pp. 189–208). John Benjamins.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.
- Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65, 127–159.
- Levac, D., Colquhoun, H., & O’Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5, 69.
- Levine, G. S. (2020). A human ecological language pedagogy [Monograph issue]. *The Modern Language Journal*, 104, 1–130.
- Lowie, W. (2017). Lost in state space? Methodological considerations in complex dynamic theory approaches to second language development research. In L. Ortega & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 123–141). John Benjamins.
- Lowie, W., & Verspoor, M. (2019). Individual differences and the ergodicity problem. *Language Learning*, 69, 184–206.
- Lowie, W., Verspoor, M., & de Bot, K. (2010). A dynamic view of second language development across the lifespan. In K. de Bot & R.W. Schrauf (Eds.), *Language development over the lifespan* (pp. 125–146). Routledge.
- MacIntyre, P. D., MacKay, E., Ross, J., & Abel, E. (2017). The emerging need for methods appropriate to study dynamic systems: Individual differences in motivational dynamics. In L. Ortega & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 97–122). John Benjamins.

- Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*, 861–904.
- Massip-Bonet, À., Bel-Enguix, G., & Bastardas-Boada, A. (Eds.) (2019). *Complexity applications in language and communication sciences*. Springer.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*, 276–282.
- Molenaar, P. C. (2015). On the relation between person-oriented and subject-specific approaches. *Journal for Person-Oriented Research*, *1*, 34–41.
- Molenaar, P. C., & Campbell, C. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*, 112–117.
- Mufwene, S. S., Coupé, C., & Pellegrino, F. (Eds.) (2017). *Complexity in language: Developmental and evolutionary perspectives*. Cambridge University Press.
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*, 143.
- Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, *66*, 834–871.
- Murakami, A. (2020). On the sample size required to identify the longitudinal L2 development of complexity and accuracy indices. In W. Lowie, M. Michel, A. Rousse-Malpat, M. Keijzer, & R. Steinkrauss (Eds.), *Usage-based dynamics in second language development* (pp. 20–49). Multilingual Matters.
- Nicklin, C., & Plonsky, L. (2020). Outliers in L2 research: A synthesis and data re-analysis from self-paced reading. *Annual Review of Applied Linguistics*, *40*, 26–55.
- Osberg, D., & Biesta, G. (2010). The end/s of school: Complexity and the conundrum of the inclusive educational curriculum. *International Journal of Inclusive Education*, *14*, 593–607.
- Overton, W. F. (2007). A coherent metatheory for dynamic systems: Relational organicism-contextualism. *Human Development*, *50*, 154–159.
- Overton, W. F., & Lerner, R. M. (2014). Fundamental concepts and methods in developmental science: A relational perspective. *Research in Human Development*, *11*, 63–73.
- Overton, W. F. (2015). Taking conceptual analyses seriously. *Research in Human Development*, *12*, 163–171.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The British Medical Journal*, *372*, n71. <https://doi.org/10.1136/bmj.n71>.
- Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, *3*, 61–94.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences*. Blackwell.
- Pfenninger, S. E. (2020). The dynamic multicausality of age of first bilingual language exposure: Evidence from a longitudinal CLIL study with dense time serial measurement. *The Modern Language Journal*, *104*, 662–686.
- Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, *5*, 371–385.
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*, 655–687.
- Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal*, *98*, 450–470.
- Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 103–125). SAGE.
- Schmid, M. S., Köpcke, B., & de Bot, K. (2013). Language attrition as complex, non-linear development. *International Journal of Bilingualism*, *17*, 675–683.
- Siddaway, A. P., Wood, A. M., Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*, 1–24.
- 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.

- Ushioda, E. (2021). Doing complexity research in the language classroom: A commentary. In R. Sampson & R. Pinner (Eds.), *Complexity perspectives on researching language learner and teacher psychology* (pp. 245–256). Multilingual Matters.
- 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.
- van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Kluwer.
- Van Orden, G., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132, 331–350.
- Verspoor, M., & de Bot, K. (2021). Measures of variability in transitional phases in second language development. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. <https://doi.org/10.1515/iral-2021-0026>.
- Verspoor, M., de Bot, K., & Lowie, W. (Eds.) (2011). *A dynamic approach to second language development: Methods and techniques*. John Benjamins.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in L2 development from a dynamic systems perspective. *The Modern Language Journal*, 92, 214–231.
- Verspoor, M., Lowie, W., & Wieling, M. (2021). L2 developmental measures from a dynamic perspective. In B. Le Bruyn & M. Paquot (Eds.), *Learner corpus research meets second language acquisition* (pp. 172–190). Cambridge University Press.
- Vitta, J. P., & Al-Hoorie, A. H. (2020). The flipped classroom in second language learning: A meta-analysis. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/1362168820981403>.
- Vitta, J. P., & Al-Hoorie, A. H. (2021). Measurement and sampling recommendations for L2 flipped learning experiments: A bottom-up methodological synthesis. *The Journal of Asia TEFL*, 18, 682–692.