Annual Review of Applied Linguistics (2005) **25**, 26–45. Printed in the USA.
Copyright © 2005 Cambridge University Press 0267-1905/05 $12.00

2. LONGITUDINAL RESEARCH IN SECOND LANGUAGE ACQUISITION: RECENT TRENDS AND FUTURE DIRECTIONS

Lourdes Ortega and Gina Iberri-Shea

Both common sense and expert knowledge tell us that learning a language other than the mother tongue is a complex process that happens through and over time. Time, indeed, is a construct implicated in many of the problems that second language acquisition researchers investigate. The purpose of this chapter is to survey longitudinal SLA research published in the last three years and to offer a critical reflection of best current longitudinal practices and desirable directions for future longitudinal SLA research. We highlight recent trends in longitudinal SLA research, paying particular attention to broad design choices and foci of research organized around four trends in SLA longitudinal investigation, and we describe key exemplary studies under each trend. We close by reflecting on some of the challenges and opportunities that await these longitudinal research programs in the future.

Both common sense and expert knowledge tell us that learning a language other than the mother tongue (a second, foreign, or heritage language; henceforth, an L2) is a complex process that happens through and over time. Indeed, it can be argued that many, if not all, fundamental problems about L2 learning that SLA researchers investigate are in part problems about "time," and that any claims about "learning" (or development, progress, improvement, change, gains, and so on) can be most meaningfully interpreted only within a full longitudinal perspective. It is, therefore, unfortunate that the bulk of disciplinary discussions within the field favors a cross-sectional view of language learning and, as a consequence, discussions about longitudinal research are scarce. For example, in Hatch and Lazaraton (1991), readers will find four pages devoted to time-series research, followed by an incidental remark that this is a special type of longitudinal design. No other specific discussion of longitudinal issues can be found in this 628-page compendium of applied linguistics research methods, now out of print. An inspection of more recent research methods textbooks confirms this state of affairs. SLA researchers who would like to pursue longitudinal research programs can find little guidance about questions such as: What problems about the development of L2 competencies have

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

SLA researchers investigated longitudinally? and What unique challenges and strengths does the longitudinal investigation of L2 learning entail?

This chapter seeks to begin to address these questions. We survey longitudinal SLA research published in recent years and offer a critical reflection of best current longitudinal practices and desirable directions for future longitudinal SLA research.[1] In addressing the general purpose of *ARAL*, we hope this discussion of trends alerts specialist readers to recent cutting-edge longitudinal studies of L2 learning. Given the dearth of methodological discussions about longitudinal research in applied linguistics, another goal we set ourselves is to offer useful considerations for SLA researchers and graduate students who plan to conduct, or are in the process of reporting on, longitudinal studies, and to suggest to them critical areas in need of attention.

The Centrality of Time in SLA

Why is longitudinal research essential to advancement of knowledge in the field of SLA? The simple but uncontestable answer is that many questions concerning second language learning are fundamentally questions of time and timing. A cursory look at some central SLA research areas can easily show this.

For one, fundamental questions related to time concern biological time at the onset of learning, including not only "at what age" but also "for how long" L2 learning should occur, in contexts where near-native success is the desired outcome. These time-related questions have inspired basic research about maturational constraints for L2 learning and the critical periods hypothesis (Hyltenstam & Abrahamsson, 2003; Scovel, 1988). They have also been central in evaluating the research base for educational policy decisions regarding early or late exit from transitional bilingual education programs in the United States (Collier, 1989), and in evaluations of early versus late start of French immersion school education in Canada (Harley & Hart, 1997).

Even in SLA domains where the preponderance of studies favors cross-sectional practices, longitudinal claims are often implicitly posited. Task-based language learning proposals, for example (see Ellis, 2003; Robinson, 2001; Skehan, 2002), make theoretical inferences about competence-related changes that are posited to deploy over time, but they do so on the empirical foundation of static snapshots of learner's capacity for action in the L2 at a given point in time. The same is true of many other areas of instructed SLA research. For example, time issues are central in the empirical determination of ideal lengths and intensities of instruction in a given curriculum, once again, if the goal is to foster advanced L2 capacities: How much instructional time is optimal in a given school curriculum, overall, and how intensively packed into the institutional schedule should it be (Collins, Halter, Lightbown, & Spada, 1999)? Questions about the timing of instruction and how instructional intensity moderates L2 learning outcomes are implicated as well, if perhaps at a more abstract level, in the oft-lamented problem of articulation from high school to college in foreign language education in the United States (Lally,

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

2001). They also underlie characterizations of the crisis of FL collegiate education in North American universities, which many blame at least in part on the dichotomization of programs into lower-division language courses, where time is devoted to the basics of language learning alone, and upper-division content courses, where students experience discontinuity at a turning point when instructional time is suddenly devoted to the exclusive learning of literature and cultural studies (Byrnes, 2002).

Many broader SLA questions could be posed that require answers mustered in longitudinal evidence: What do we know about the pace and pattern of development in second language and literacy, throughout the lifetime of L2 learners? What critical transition points in L2 development need to be taken into account when planning and evaluating educational policy and practice for specific L2 learner populations? Ultimately, it is through cumulative longitudinal findings that the SLA research community would be able to contribute meaningful characterizations of the gradual process of attaining advanced second language and literacy competencies across various contexts. Given the centrality of time in SLA research programs, then, more attention to longitudinal research practices and to findings gleaned from longitudinal studies is desirable.

Methodology for the Present Review

Although we make no claim to exhaustion or completeness in this review, the themes we present emerged from strategies for systematic synthesis of primary research domains (Cooper & Hedges, 1994; Light & Pillemer, 1984). We conducted electronic searches to uncover longitudinal SLA studies published during 2002 and 2003 across 20 major applied linguistics journals. This was supplemented with hand searches of 2004 journal issues that were available at the time of this writing. The search yielded 38 longitudinal studies that focused on L2 learning (including second, foreign, and heritage language learning). A little fewer than half of these studies investigated some aspect of L2 learning by young adult students enrolled in institutions of higher education, whereas about 20% investigated high school or middle school contexts. Even fewer studies addressed adult learners in nonuniversity contexts or preschool or elementary-school-aged children. Broadly speaking, then, recent longitudinal research in SLA is consistent with the focus on college-level populations that is typical of applied linguistics research in general.

We located and read these studies and discussed main methodological features. We also recorded the extent to which studies defined and documented "change over time" (quantitatively, qualitatively, or both), and we evaluated the strengths and weaknesses of each study, noting any longitudinal challenges explicitly discussed in the primary reports. This process compelled us to refine and broaden our working definition of "longitudinal" SLA research, as we realized the study of second language and literacy development over time has been undertaken from a number of distinct epistemological traditions, including quantitative descriptive studies of linguistic features and qualitative interpretive studies of sociocognitive and sociocultural dimensions of multilingual development. In the remainder of this

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

chapter, we characterize four broad trends in recent longitudinal SLA research. We conclude with a methodological discussion that we hope is useful for SLA researchers and graduate students who plan to conduct, or are in the process of reporting on, longitudinal studies.

Descriptive-Quantitative Longitudinal Studies of L2 Development

Many longitudinal SLA studies published in the last three years have adopted a descriptive (that is, nonexperimental), quantitative design. Such studies focus on quantifiable variables, but without researcher's manipulation and without random sampling or random assignment of participants to conditions or to contexts for learning. They are characterized by multiwave data collection from the same individuals over a relatively long period of time, which typically spans anywhere between four months and four years. Although the longitudinal data in these studies are quantified, the number of participants is far too small to warrant the use of inferential statistics. For this reason, descriptive statistics displayed in the form of frequencies, percentages, and proportions, and other analytical tools, such as visual displays and implicational scaling, are favored.

The research focus is often strictly linguistic, following up on the tradition of early SLA longitudinal studies that have become classics (e.g., Huebner, 1983; Sato, 1990; Schumann, 1978; but see Schmidt, 1983, for a broader linguistic focus). Although the early interlanguage work involved intense case studies of usually one or two participants only, more recent studies increasingly tend to involve multiple participants, although still typically small groups. For example, Perdue, Benazzo, and Giuliano (2002) charted the development of finiteness in the spoken production of five adult Italian and Spanish immigrant workers in Europe during 30 months of naturalistic exposure to English or French, and Bardovi-Harlig (2002) investigated the development of future tense as evinced in the written production by 16 college ESL learners during 7 to 17 months of study in a university's English Language Program (this is the same sample and longitudinal corpus previously studied in Bardovi-Harlig, 1994). Well established descriptive-quantitative longitudinal lines of SLA research to the present day concentrate on L2 morphology, as the studies just mentioned illustrate, and on L2 phonology (e.g., Abrahamsson, 2003; Hansen, 2004).

Descriptive-quantitative longitudinal SLA research draws on linguistic theories, often of a functional bent, and seldom includes nonlinguistic variables. These stable trends notwithstanding, it is interesting to note that in other recent SLA studies employing this type of design there has been a broadening not only of the linguistic focus but also of the epistemological approach to language development. This is the case of a longitudinal study conducted within a Vygotskian sociocultural framework by Belz and Kinginger (2002). The researchers documented critical incidents that contributed to the learning of indexical politeness (specifically, the use of *tu*/*vous* and *Du*/*Sein*) by two fourth-semester foreign language students at a U.S. university. The participants were involved in a telecollaboration project that afforded them the opportunity to establish a personal relationship with a native-speaking peer in France and Germany, respectively. The critical incidents were defined by the

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

researchers a priori as instances in which a student chose the formal *you* during an e-mail exchange in the L2, causing the peer interlocutor to focalize overtly on the inappropriateness of this choice. The researchers argue that solidarity (for the French student) and even intimacy (for the German student) developed between cross-site partners, and that this personal investment infused the choice of term of address with special significance. In sum, the personal communicative investment enabled by the telecollaboration assignment outside of the classroom made choice of address term functionally salient for these students, whereas it also afforded them fine-tuned and timely assistance from their native-speaking peer.

The descriptive-quantitative type of design discussed here corresponds closely with traditional definitions of "longitudinal" research in a number of fields. Its main strength, according to the canon of statisticians (Johnson, 2001), is that it provides evidence for time relationships among variables, thus helping not only to describe phenomena and association among variables, but also to begin to explain causes and effects by looking at antecedent-consequent relationships (e.g., the co-occurrence of critical incidents and systematic shifts in the use of indexical politeness documented by Belz and Kinginger, 2002). In the social sciences, the longitudinal tradition of life-course studies is based on this descriptive, quantitative approach, and has contributed massive panel studies as well as more developmentally oriented studies, in both cases involving hundreds or thousands of participants followed over what is some times several decades of their lives (Phelps, Furstenberg, & Colby, 2002). However, the longitudinal SLA tradition, with its small samples and short-span lengths of investigation, is much closer to the spirit of longitudinal research in the feeder field of first language acquisition (e.g. Karmiloff-Smith & Karmiloff-Smith, 2001). Here, detailed case studies are the norm, and the chronological frame of several months up to five or six years is well established as sufficient for any given study, because this is the biological time scale naturally needed to capture the object of inquiry in child first language acquisition.

Longitudinal Research on L2 Program Outcomes

A very different kind of longitudinal design has been adopted more rarely in recent SLA studies, which we would like to call, in lieu of a better term, "programmatic" longitudinal research. By comparison to the descriptive-quantitative longitudinal design discussed in the previous section, the programmatic longitudinal design is used for evaluation of L2 curricular options, employs relatively large samples, and focuses on longer periods of observation scaled on institutional time (typically four to six years), coupled with fewer data collection waves spaced along wider time gaps (there are often one or two collection points per year). Although this longitudinal design has been infrequently employed in SLA, it addresses questions of utmost importance for policies and practices in L2 programs, such that an increase of longitudinal investigations of this kind would be a positive development in future instructed SLA research. In this section we discuss two representative studies: Lightbown, Halter, White, and Horst (2002) and Klapper and Rees (2003).

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

The participants in Lightbown et al. (2002) were a cohort of 73 French-speaking school students in the Canadian province of Nouveau-Brunswick who had received all their English L2 instruction via a self-access, comprehension-only experimental curriculum (that is, the program was entirely based on self-paced reading and listening activities). This cohort's overall proficiency in English was mapped during six years of comprehension-based instruction, with data points on the students' first year of English study (in grade 3), their last year of primary school (in grade 6), their first year of secondary school (in grade 7), and a year later (in grade 8). In addition, the cohort's L2 competencies were also compared to a group of 27 students from the same province who studied English under the traditional (audiolingual) curriculum and—for grades 6, 7, and 8 only—with another group of 14 traditional program students. The fact that these learners were followed for six years makes this one of the longest periods of observation investigated in any longitudinal SLA study focusing on group, rather than individual, performance. The analyses based on second-wave observations in grade 6 (the last year of primary school) yielded data that supported a no-difference interpretation. An inspection of the evidence over the longer period of six years, however, revealed a similar-sized improvement in reading and listening abilities for the experimental and the traditional curriculum samples, but a clear advantage in productive skills (particularly L2 writing, including the use of verbal morphology) for the traditional curriculum groups.

Klapper and Rees (2003) undertook a four-year comparison between two groups of university students (a total of 57 learners) in the same German department in the United Kingdom, who experienced a traditional grammar instruction curriculum (mostly grammar and grammar-translation courses for 2 to 3 hours a week in a major degree concentration) or a focus-on-form curriculum (mostly content- and task-based courses for 2 to 3 hours a week in a minor degree concentration). The two groups began their programs with comparable scores and received near identical amounts of instructional time during the first two years of study. Both groups' programs were enriched with a one-year study abroad component on the third year of the four-year college curriculum. Klapper and Rees examined changes on a grammar test and a test of general academic language proficiency, both administered in three waves, yearly, for years 1 through 3 of the program. The comparisons also included one-time indicators of academic ability and success (an IQ test at the beginning of the program; final exams and degree marks at completion of the program in year 4). Klapper and Rees concluded that the pace of development of the focus-on-forms and focus-on-form cohorts was different, but the final outcomes were comparable. Specifically, the focus-on-forms group showed large increases in grammar and proficiency test scores during the first two years of the program, whereas the focus-on-form group showed stagnation during the same period but similarly large progress after the one-year study abroad experience on the third year. In essence, then, the two groups had started equal and finished equal as well, but they exhibited interesting differences in the pace of development within the four-year observation period. Notably, these differences were eventually equalized by the turning-point experience of study abroad.

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

Despite rather large differences in L2 learning context and in instructional focus, these two programmatic longitudinal studies share strong design similarities. They are similar in the quantitative orientation of the analyses, the sampling of rather large groups, the comparison of broad curricular programmatic options, and the unusual length of the observation span, which was scaled on institutional time, coupled with infrequent, year-spaced data collection waves. Perhaps these shared features are responsible in part for the fact that in both studies the findings changed radically when the length of observation was doubled, from three to six years in Lightbown et al. and from two to four years in Klapper and Rees. Had the research teams in either study decided that three or two years was "long enough," this choice would have led them to the conclusion, erroneous in a fuller longitudinal perspective but consistent with the first available waves of data, that there was no difference between the two programs compared, in one case (Lightbown et al.), or that there was a difference, in the other (Klapper & Rees). Thus, these two programmatic SLA studies attest to the tremendous value of taking a very long view on L2 development in SLA research.

The Longitudinal Investigation of L2 Instructional Effectiveness

A third recent, noteworthy trend is the gradual emergence of what could eventually give rise to a longitudinal design tradition in effectiveness of L2 instruction research. For one, there is a pronounced tendency in recent quasi-experimental studies of instruction to feature longer-lasting interventions (eight weeks seems to be a favored choice) and to include delayed posttests (often only a month after instruction, some times up to three months later). This, for example, can be found in the designs adopted by Han (2002) and Lyster (2004). This is a valuable improvement over the prevalent practices in this research domain up until 1998 reported by Norris and Ortega (2000). Of the 49 studies included in that meta-analysis, the majority (30 or 61%) featured short duration treatments of a maximum of two hours, and only 22 (or 45%) included a delayed posttest. Although the presence of long-lasting interventions and delayed post-tests does not amount to longitudinal research per se, when data are collected and analyzed across multiple waves to document gains on several posttests, some times even repeatedly during the treatment period as well, the process of learning is presented in longitudinal perspective to some extent.

Moreover, some recent studies have implemented true longitudinal intervention designs, although they cannot be characterized as fully quasi-experimental because no control or comparison group was employed. A representative case in this category is an investigation of instruction by Kim and Hall (2002). These researchers, like Belz and Kinginger (2002), sought to document L2 pragmatic learning through the study of Vygotskian microgenetic change, taking a close look at the process of change throughout the history of a learning event. More specifically, their purpose was to investigate the benefits of interactive book sessions that took place between the teacher-researcher and four third-grade Korean students who had been learning English for a year at the time of the study. The interactive book reading sessions (the instructional 'treatment') took place over four months in

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

30-minute biweekly meetings, during which seven age-appropriate English books were read. One book was read per session during the first month, and then the same seven-book cycle was repeated each successive month. After 20 minutes of interactive reading, the last 10 minutes in each session were typically devoted to the children's role playing of some aspect of the story. This role play was identified as the ideal "testing" context, from the viewpoint of microgenesis. Pragmatic L2 development was operationalized as increases during these role play events in mean number of words, mean number of utterances, average use of moves that served to manage talk (that is, initiations, elaborations, conclusions, and formulaic expressions), and average use of moves that served to manage meaning (that is, self- and other-corrections). The researchers found clear improvement in the four children's EFL pragmatic competencies in terms of all dependent variables but one (the meaning management moves) over the four-month period.

Although the research trends in effectiveness of instruction discussed so far are clearly encouraging and take a longer view of instructed L2 development, little progress seems to have occurred regarding the use of time-series designs in SLA research. Time-series design constitutes the single best formal strategy for investigating effects of instruction longitudinally. The principles for this design were clearly and forcefully laid out as a promising longitudinal experimental research option in SLA by Mellow, Reeder, and Forster (1996). After nearly a decade since the publication of that article, we were able to uncover only one study of L2 instruction adopting a time-series longitudinal experimental design, Ishida (2004).

Ishida investigated whether a regime of intensive recasts on the aspectual Japanese form *–te i-(ru)* delivered over four 30-minute, biweekly tutorial sessions would have lasting effects on the oral accuracy exhibited on this form by four second- and third-year semester students. The time-series design was operationalized as two waves of data collection that established a baseline for all four participants (on week 1), followed by four more waves in which the treatment was delivered and immediate effects were documented (on weeks 2 and 3), and two final waves that served as posttest (on week 4) plus a final wave of data collection seven weeks later (on week 11 of the study) that provided delayed posttest observations. Ishida found that the intensive recasting treatment exerted an immediate effect, as shown in learners' more accurate use of *–te i-(ru)* immediately after the onset of the treatment. The accuracy gains were sustained on each successive wave of observation up until week 11. These results, together with those reported by Han (2002), offer longitudinal evidence that intensive, focused recasting is likely to be causally associated with higher levels of accuracy. When these findings are compared to those reported by Lyster (2004), the converging evidence suggests that recasts under tutoring conditions, but not in larger classroom settings, can be an effective means to promote long-term formal accuracy in a second language.

A longitudinal view on L2 instructional effectiveness is not only useful to establish the durability of effects, but it also helps elucidate effects that deploy unevenly and nonlinearly over time (cf. Mellow et al., 1996). A recent meta-analysis by Keck, Iberri-Shea, Tracy, and Wa-Mbaleka (forthcoming) yielded findings that

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

underscore this point. These researchers synthesized 14 quasi-experimental studies (published between 1994 and 2003) that tested the direct link between interaction and acquisition of L2 forms. When they aggregated and compared effect sizes on immediate versus delayed posttests, they uncovered a systematic association between the timing of the posttests and the magnitude of effects for the two design features of task-essentialness and task usefulness (see Loschky & Bley-Vroman, 1993). Namely, aggregated effect sizes yielded by immediate posttests suggested no difference between task essential and task useful interaction treatments, as both were equally effective; but when effect sizes from delayed posttests were inspected, a trustworthy (that is, statistically significant) and large difference was noted, as groups who had experienced L2 forms through task essential treatments, rather than task useful treatments, were able to demonstrate more knowledge of the given forms on delayed post-tests (d=1.66 versus d=.76, respectively). Keck et al. were able to capture this important effect for task-essentialness, which was theoretically postulated first by Loschky and Bley-Vroman (1993) but had not been investigated in any single primary study, thanks to the meta-analysis methodology, but also because 10 of the 14 interaction studies synthesized featured delayed post-tests, scheduled as soon as seven days and as late as sixty days after the treatment.

In sum, we wish to reiterate the importance of this overall move towards a more longitudinal view on instructional effectiveness in SLA research. Even though quasi-experimental, cross-sectional designs have been predominant in this research domain, issues of intensity, timeliness, duration, and durability are hotly debated (see Doughty & Williams, 1998; Mellow et al.,1996; Norris & Ortega, 2000; Spada, 1997). Because the theoretical and educational claims implied in L2 instructional studies are tantamount with claims about change that unfolds gradually (but not linearly) over time and that is long-lasting, they will ultimately demand longitudinal evidence from researchers working in this SLA domain.

Recent Qualitative Longitudinal SLA Research

A fourth noticeable trend is the gradual accumulation of longitudinal research that employs primarily qualitative methodologies. These qualitative longitudinal studies are usually framed within Vygotskian sociocultural theory (Lantolf & Thorne, forthcoming), or within a socialization theory perspective (Kramsch, 2002; Watson-Gegeo, 2004), and they some times draw from both. Accumulation is only gradual, however, because the focus in these longitudinal studies is far from homogeneous.

Sociocultural SLA Longitudinal Studies

Recent qualitative longitudinal studies framed within a Vygostkian sociocultural perspective range from microanalysis of paralinguistic aspects of interaction (e.g., in the 15-week study of gesture development by a Taiwanese EFL student reported in McCafferty, 2002), to qualitative observation of classroom life and the contingencies of motivation (e.g., in the account of a summer experience with Chinese as a foreign language classroom learning presented by Lantolf &

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

Genung, 2002), to autobiographical longitudinal accounts of the relationship between pronunciation and identity in foreign language learning (Marx, 2002). We will examine in some detail this last study, because of its unique and promising approach to mapping longitudinal change qualitatively.

Marx (2002) offers a detailed autobiographical chronicle of the development of identity that she underwent over three years of living in Germany, and as her accent in her second foreign language (German, which she had learned after French) became a site of identity (trans)formation. She frames this chronicle theoretically by positing six longitudinal stages originally proposed by Pavlenko (1998). When Marx arrived in Germany and was identified as "American" because of her English-accented German, she felt displaced and defined herself for the first time as "Canadian," in opposition to the American identity (understood as "from the United States") that her German interlocutors attributed to her. In a second stage after four months in the target environment, she developed a French accent in her German. This accent, which implied a loss of her Canadian identity, nevertheless allowed her to be identified by German interlocutors as "anything but American." After one year immersed in the L2 setting, stage three emerged as Marx consciously attempted to develop a native-like accent that would allow her to pass for a German native speaker. After two years, stage four was reached: the construction of her L2 identity, and with it her native-like accent, had consolidated, while her L1 began to experience signs of loss through the emergence of a British accent that in fact was truly "foreign" for her. After three years in Germany, Marx reentered the L1 context. This fifth stage lasted for about three months and was characterized by a strong British or German accented English. This foreign accent allowed her to position herself as "not just Canadian," that is, as someone who has been transformed by the experience of being a foreigner and a nonnative speaker. After three months, however, the final sixth stage of learner identity formation emerged. That is, her Canadian identity was finally reconstructed and her foreign accent in English subsided.

The autobiographical methodology that Marx (2002) adopted, together with her theoretical framework for identity transformation through participation in and out of L1 and L2 communities, makes this study unique as a three-year longitudinal account. It provides valuable theoretical and psychological insights into learner agency and identity negotiation, in ways that highlight, from the privileged vantage point of the agent, the dynamic development of multicompetence (Cook, 2002) as a learner transits through chronological and social time.

Ethnographies of L2 Learners

Socialization theory has also begun to contribute increasingly large numbers of longitudinal studies that chart the development of academic literacies in a second language. These studies tend to adopt a case study methodology and employ a rich mixture of qualitative approaches, including microanalysis of discourse and ethnographic observation and interviews. An illustrative case is Kobayashi (2003), who studied two undergraduate students from Japan during their preparation outside

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

of class for their first group oral presentation in their first semester of study abroad, in a Canadian university. Although the preparation for the academic oral assignment was confined to a short cycle of five days, Kobayashi draws his insights from wider ethnographic data of the entirety of this Japan-Canada exchange program, which involved one academic year and a total of eighty undergraduate Japanese students. He uncovers a preparation process comprised of three stages spread over five days of meetings and group work. During the first stage, the students negotiated task content and assignment expectations, often using their shared L1 (Japanese) and drawing on their collaborative interpretation of written class documents. At a second stage, they collaborated to generate the language of the oral presentation. This stage was characterized by L1-L2 code-switching and by reliance on multiple modalities (speech and writing), supported by the use of technological literacy involving Powerpoint. A strong sense of audience and genre was also captured in the second-stage group data. The third and final preparation stage was devoted to rehearsal of the presentation, and favored the use of the L2 (English) and peer coaching, often through theatrical appropriation of the teacher voice by members of the group who acted as critics of each member's rehearsed contribution. Kobayashi invokes Rogoff's (1995) notion of "participatory appropriation" to acknowledge that true longitudinal data would be necessary to ascertain whether the preparation practices uncovered eventually contributed to the students' development of oral presentation skills in their future academic endeavors.

It is indeed surprising that many longitudinal qualitative studies adopting a case study or ethnographic approach end up reporting on only a small subset of the richer longitudinal data, making readers wait on a promised larger study that may take years (and several other shorter-spanned reports drawn out of the larger study) to appear in print. This trend is no doubt in part explained by the academic pressure to produce publications at a rate that works against multiple-year engagement with the same site and participants (this same pressure may also discourage dissertators from choosing longitudinal methodologies). However, even when considering ethnographic studies, a more fundamental issue emerges as well. Namely, we repeatedly encountered cases where an ethnographic report spanned long enough to be considered longitudinal in its own right, yet it was unclear whether it should be considered longitudinal because the results were presented with little attention to illuminating change over time. For example, Gebhard (2002) collected data from third and fourth graders at a magnet school in California for over two years. She motivated her focus on these grades during elementary school as "important turning points" in the academic trajectories of children at this age. She contextualized her findings against a four-year longitudinal perspective on the changing demographic, curricular, and material landscape of the school. Jackson (2002) drew her insights about classroom oral participation from a three-year ethnography of Chinese undergraduate students majoring in business at a Hong Kong university. She carefully triangulated a survey approach with systematic and in-depth observations and interviews over a 14-week semester. Similarly, Leki (2001) documented six English L2 students' semester-long experiences during group projects across university classes, but drew on a fuller longitudinal perspective and even richer data on the same informants over their total university life experience, which lasted for

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

five years. In light of such features, all three ethnographic studies could be considered indeed "longitudinal." Yet, all three turn out to present a thick but relatively static description of the problems they sought to understand. The goal of these researchers in each case seems to have been, rather than to document and understand change over time, to capture continuity and to achieve deep understanding of roles, interrelationships, and intentionalities in ecologies of second language learning that are viewed as stable over time. We will return to this observation in our discussion in the next section.

Future Directions

The studies reviewed in this chapter, together with many other longitudinal studies of L2 learning and learners that have accumulated over the years since the inception of SLA, point at many implicit observations and questions regarding the unique benefits and the special challenges of investigating second language and literacy development over time, from a truly longitudinal perspective. One generalization that should serve SLA researchers well is that longitudinal research can be as diverse as the epistemological approaches that inspire it. This diversity is likely to continue in the future, and we hope it will also encourage longitudinal research that capitalizes on the strengths of mixed methods designs (see Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 2002). The following are some of the more persistent issues that we saw emerging across studies and authors.

How Long is "Longitudinal"? Time Scales and Turning Points

A main definitional feature of longitudinal research is its long time span. Yet, little is known about the optimal length of observation for the longitudinal study of bilingual and biliterate development. When the focus is on linguistic development, there are often as many arguments to call for studies that span a few months to capture the phenomenon of interest (e.g., an observation period of about a year was enough for the development of English past perfect to be mapped in 10 of the 16 learners in Bardovi-Harlig, 1994), as there are to posit the entire life time of a learner as the right time frame for longitudinal SLA investigations (e.g., a 16-year observation span may not be sufficient evidence in the case of fossilization; Han, 2004). Yet few SLA researchers have acknowledged this challenge or discussed related methodological choices explicitly in their longitudinal reports. Most recent longitudinal SLA studies span anywhere between three or four months and up to six years. This seems related to two factors: (a) the influence of child first language acquisition studies, where longitudinal research is scaled on biological time; and (b) the convenient scaling of other studies on institutional time, particularly the one-semester and four-year periods widely meaningful in tertiary education, where many SLA studies are conducted. Thus, decisions about how long is long enough for the longitudinal study of L2 development are implicitly made in SLA research by recourse to biological and institutional time scales.

A main caveat to consider is that SLA researchers lack the certainty afforded by a biological scale such as that intrinsically given in child first language

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

acquisition. Institutional schedules, on the other hand, may be at odds with social and developmental chronologies experienced by the L2 learners (for a useful discussion of such cases in educational longitudinal research, see Entwisle, Alexander, and Olson, 2002). We would like to suggest that, whether biological or institutional scales are chosen for longitudinal SLA research, they can be better motivated when key events and turning points in the social or institutional context investigated are considered. Indeed, such crucial turning points can be essential in motivating the appropriate length of observation.

When it comes to charting the developmental evolution (arguably, from emergence to mastery) of a given L2 linguistic feature, a strategic option would be to center on defining the time scale of the specific learning phenomenon under inquiry, rather than on the biological chronology of the subjects. Such strategy lies behind Belz and Kinginger's (2002) operationalization of the Vygostkian concept of microgenesis, which offers guidance for establishing the organic length of observation needed to map out the history of a learning event. Namely, these researchers collected a large longitudinal corpus of electronic communication transcripts (contextualized with a wealth of quantitative and qualitative data collected during the implementation of their three-year project). They then undertook "a thorough review of the interactions in the database, arranged chronologically, to select cases" (p. 196) of critical incidents involving the use of *tu*/*vous* and *Du*/*Sein* and decided to go as far before and after the incident in the data as was needed to trace the history of the phenomenon longitudinally. Aided by this inspection process, Belz and Kinginger determined an observation window around the critical events of 60 days (for the French student) and 49 days (for the German student) and they sampled data at five and nine observation points, respectively, within that window. Although rather short by traditional longitudinal methodological standards, this was precisely the organic length of observation needed to map out the history of the two critical incidents.

Notably, Belz and Kinginger (2002) had engaged in a wider three-year longitudinal study during which they had collected a substantial longitudinal corpus representative of the full length of the telecollaboration project. The availability of such a rich (and longer spanning) longitudinal corpus was therefore crucial in enabling the microgenetic method to study critical incidents, and allowed the researchers to identify the appropriate length of observation. In future studies using this microgenetic approach, an improvement could be to further inspect data recorded for the same participants in the full longitudinal corpus. This would allow researchers to trace evidence for the long-term impact of the critical incidents on learners' overall development of the L2 aspect under study, in effect enabling them to address Rogoff's (1995) longitudinal construct of "participatory appropriation."

Programmatic longitudinal SLA studies, which often scale study length on institutional time, also attest to the usefulness of considering turning points. For Lightbown et al. (2002), this meant including a long enough period of observation that would allow researchers to look at the transition from elementary to secondary school; for Klapper and Rees (2003), it meant documenting the turning point of the

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

study abroad experience. In the future, it will also be useful to situate individual change over time against institutional and even sociopolitical and historical time. We also hope the field engages in explicit discussion of the relative merits of various criteria and rationales for determining the appropriate length for investigating specific SLA research questions longitudinally.

Multiwave Data Collection: How Many Waves, and How Comparable?

Another constitutive feature of longitudinal research is multiwave data, because data are collected repeatedly and usually more than twice to map change and growth over time. Thus, associated with the question of appropriate overall study length is that of optimal timing (frequency and spacing) of measurements. Our discussion about scales and turning points is equally relevant when making the decision of how often participants need to be observed and data collected, in that the estimated chronological scope of biological and institutional key events should help determine appropriate frequency and total number of observation points (see Entwisle et al., 2002). In addition, there is a likely trade-off in longitudinal SLA research among three factors: the overall length of study, the size of the sample, and the frequency of data collection waves. Finally, we would like to stress that the frequency of data collection is in part a matter of grain size of the phenomenon under investigation. It would certainly be impossible to attain the level of detail needed to chart many L2 developmental phenomena without striving for frequent data collection waves. For example, Bardovi-Harlig (1994) collected all classroom assignment products over six seven-week terms, for a total study length of slightly under one and a half years. She divided the language samples into two-week periods and sampled three texts per learner from each period. Consequently, her analyses of L2 development of past perfect are truly developmental and longitudinal, as they reflect individual growth data consisting of 34 waves in the best individual cases. In this sense, the case-study in depth approach chosen by descriptive-quantitative SLA researchers, which is modeled on first language acquisition research, represents a strength. For other longitudinal SLA research programs, however, it may be more appropriate to choose yearly or biannual data collection waves, as is the case with programmatic SLA longitudinal studies.

A special challenge with multiple data collection points that is likely to arise in any longitudinal SLA designs is the comparability of observations. The problem is that when different tasks and topics are employed over the length of a longitudinal study, it is difficult to disentangle time-induced and topic- or task-induced variability. Such topic-specific effects were found, for example, by Kim and Hall (2002), who reported large amounts of book-induced variability, as some books elicited much briefer role plays from the children than other books did, regardless of the time period at which the data were collected. This problem could be minimized logically by keeping elicitation techniques and content as similar as possible across waves. On the other hand, however, this may not always be the best strategy. For one, if SLA researchers were to repeat exactly the same tasks and the same topics over long periods of longitudinal study, the potential for diminished interest (and even demotivation or boredom), as well as practice effects, among participants would

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

be a clear danger to the validity of the data. Further, it is possible that certain tasks are inappropriate at early points of development (because they require knowledge of the L2 that is not in place yet) or for the initial waves of data collection (because the participants, for example, are too young in the beginning of the study to be able to understand the task instructions). Such constraints were likely to be a consideration in Lightbown et al. (2002), for example, because what third graders and eighth graders can do with and in their L2 (and their L1) looks rather different. Thus, enhancing the equatability of data collection points is a formidable challenge in SLA longitudinal research.

At least for studies in which learners' general cognitive and linguistic abilities are not posited to change dramatically over the course of a lengthy investigation, a good strategy to deal with the multiwave comparability problem is to build in a cyclical elicitation schedule. Although it is unclear whether Kim and Hall (2002) planned this by design, they were able to compare four points of data for each of the seven books employed during the treatment, precisely because each book remained constant for four data collection waves, all four being equidistant in time over four months. In essence, then, in this study the longitudinal data analyses on each new book can be taken as seven cases of four-month long replications in search for reliable evidence of instruction-induced change over time. Even better would be the planning of cyclical elicitation points where similar tasks or content are employed over equally-spaced waves, interspersed with a range of dissimilar tasks, also spaced at equal intervals over the length of the study. This would ensure anchored points of comparison as well as strategic points of more varied performance that would not only improve participants' engagement but also enhance the richness of the longitudinal corpus available for study. Other creative strategies will need to be developed (and this problem discussed more explicitly) by longitudinal researchers so as to preserve the comparability of observations while ensuring participant affective, cognitive, and linguistic investment in the tasks and tests used.

Innovative Use of Analytical Options for Quantitative Longitudinal SLA Research

We found a general methodological shortcoming in current practices among longitudinal SLA researchers who draw on quantitative data, particularly in those studies that focused on program or instruction effects, because in these cases the number of participants and the quasi-experimental designs call for the use of inferential statistical analyses. Namely, none of these studies have benefited from advances in longitudinal quantitative analysis that began in the 1970s and that have been recently updated and presented in a comprehensive fashion by Singer and Willett (2003). Instead, SLA longitudinal data are often analyzed by recourse to the same inferential statistics that are employed in cross-sectional research (t-tests, multivariate analysis, and so on). Moreover, the larger-sized studies are often less about longitudinal change of a group (and much less so of individuals within the group), and more about comparing cohorts at repeated intervals of time. That is, often these multiwave between-groups comparisons are conceptually cross-sectional, and only indirectly longitudinal (e.g., Klapper & Rees, 2003; Lightbown et al., 2002). This tendency to recast longitudinal change through the analytical lens of

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

repeated cross-sectional comparisons partly explains the absence of analysis of magnitude of longitudinal change in longitudinal SLA studies (in the form of repeated-measures effect sizes; see Ortega, 2003). In the future more SLA researchers may seek to map the L2 developmental path of a cohort transiting through curricular and instructional time with key turning points, different phases of stagnation and differently-paced development, and nonlinear progress. If this is the case, and more large-size longitudinal quantitative studies are conducted in SLA, it will be important to train ourselves in the use of statistical analytical options that are available specifically for use with longitudinal designs and data.

Strengthening Longitudinal Qualities of Ethnographic Studies

As mentioned earlier, when it comes to ethnographic qualitative studies of L2 learners, what counts as "longitudinal" is not easily decided. Sheer length of study and multiwave data collection are definitional features of longitudinal research in the social sciences, but so are two research goals: to capture change over time and to establish antecedent and consequent relations (cf. Phelps, Furstenberg, & Colby, 2002). Although the feature of prolonged engagement undeniably makes ethnographies of L2 learners "longitudinal" in one sense, ethnographers of L2 learning do not always appear to enter the multiple-year study with longitudinal research goals explicitly formulated. This does not need to be the case, however, as qualitative and specifically ethnographic studies can provide the most convincing evidence for fine-grained conceptualizations of change and growth and even for richly contextualized antecedent and consequent relations (for how qualitative research can do the latter and address causality in addition to description, see Maxwell, 2004).

Some L2 ethnographies in the past have successfully embodied all definitional features of longitudinal research. For example, Peirce (1995) ethnographically documented shifts and changes in five immigrant women's motivation and investment to learn and use English. From the recently published literature, an excellent example is Vavrus (2002), who reports on an in-depth and longitudinal investigation of high school students' changing views of the benefits of English-medium instruction versus vernacular education in Tanzania. Vavrus collected her ethnographic data over six years and three waves of surveys and focus groups with enrolled students who eventually became graduates from five high schools in the same Kilimanjaro area. Studies such as these reflect the best features of qualitative research drawing on ethnographic techniques and adopting a complex longitudinal view on L2 learners' lives. They strive to produce thick descriptions as much as depictions of change and growth over long periods of time.

Concluding Remarks

Research practices and disciplinary discussions in SLA typically favor a cross-sectional view of language learning and, in spite of much lip service, discussions about longitudinal research are scarce. In this review, we have highlighted recent trends in longitudinal SLA research, paying particular attention to

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

broad design choices and foci of research organized around four trends of SLA longitudinal investigations, with key exemplary studies described for each trend. We have also reflected on some of the challenges and opportunities afforded by these longitudinal research programs.

We hope more explicit discussions will ensue in the future regarding what makes a study "longitudinal" and what the unique benefits and challenges of doing longitudinal SLA research are. We also hope SLA researchers will increasingly seek to look at second language and literacy development longitudinally. Ultimately, longitudinal findings can have a central place in advancing our SLA theories and research programs. We are hopeful that the diversity and accumulation of recent and future longitudinal research will help chart the development of advanced L2 capacities and help us understand the appropriate timing, duration, and content of optimal educational practices for L2 learning across educational settings and multilingual contexts.

Note

1. We are indebted to *ARAL* editor, Mary McGroarty, for her comments on an earlier version of this paper. This article is based on findings from a larger synthesis of longitudinal research in applied linguistics, funded by a National Academy of Education/Spencer Postdoctoral Fellowship to the first author and a research assistantship to the second author. We would like to thank the National Academy of Education and the Spencer Foundation for their generous support of this work over the period of 2003–2005. The views expressed here are not necessarily those of either supporting entity.

REFERENCES

Abrahamsson, N. (2003). Development and recoverability of L2 codas: A longitudinal study of Chinese-Swedish interphonology. *Studies in Second Language Acquisition*, *25*, 313–349.

Bardovi-Harlig, K. (1994). Reverse-order reports and the acquisition of tense: Beyond the principle of chronological order. *Language Learning*, *44*, 243–282.

Bardovi-Harlig, K. (2002). A new starting point? Investigating formulaic use and input in future expressions. *Studies in Second Language Acquisition, 24*, 189–198.

Belz, J. A., & Kinginger, C. (2002). The cross-linguistic development of address form use in telecollaborative language learning: Two case studies. *The Canadian Modern Language Review, 59*, 189–214.

Byrnes, H. (2002). Toward academic-level foreign language abilities: Reconsidering foundational assumptions, expanding pedagogical options. In B. L. Leaver &

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

B. Shektman (Eds.), *Developing professional-level language proficiency* (pp. 34–58). New York: Cambridge University Press.
Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly*, *23*, 509–531.
Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, *33*, 655–680.
Cook, V. (Ed.). (2002). *Portraits of the L2 user*. Clevedon, UK: Multilingual Matters.
Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). New York: Cambridge University Press.
Ellis, R. (2003). *Task-based language learning and teaching*. New York: Oxford University Press.
Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2002). Baltimore Beginning School study in perspective. In E. Phelps, F. F. J. Furstenberg, & A. Colby (Eds.), *Looking at lives: American longitudinal studies of the twentieth century* (pp. 167–193). New York: Russell Sage Foundation.
Gebhard, M. (2002). Fast capitalism, school reform, and second language literacy practices. *The Canadian Modern Language Review*, *59*, 15–52.
Han, Z.-H. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly, 36*, 543–572.
Han, Z.-H. (2004). *Fossilization in adult second language acquisition*. Clevedon, UK: Multilingual Matters.
Hansen, J. G. (2004). Developmental sequences in the acquisition of English L2 syllable codas: A preliminary study. *Studies in Second Language Acquisition, 26*, 85–124.
Harley, B., & Hart, D. (1997). Language aptitude and second-language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*, 379–400.
Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: HarperCollins/Newbury House.
Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma.
Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.
Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form -te i-(ru) by learners of Japanese as a foreign language. *Language Learning*, *54*, 311–394.
Jackson, J. (2002). Reticence in second language case discussions: Anxiety and aspirations. *System, 30*, 65–84.
Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher, 30*(2), 3–13.
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

Karmiloff-Smith, K., & Karmiloff-Smith, A. (2001). *Pathways to language: From foetus to adolescent*. Cambridge, MA: Harvard University Press.

Keck, C., Iberri-Shea, G., Tracy, N., & Wa-Mbaleka, S. (forthcoming). Investigating the empirical link between interaction and acquisition: A quantitative meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching*. Philadelphia, PA: John Benjamins.

Kim, D., & Hall, J. K. (2002). The role of an interactive book reading program in the development of second language pragmatic competence. *The Modern Language Journal, 86*, 332–348.

Klapper, J., & Rees, J. (2003). Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research*, *7*, 285–314.

Kobayashi, M. (2003). The role of peer support in ESL students' accomplishment of oral academic tasks. *The Canadian Modern Language Review, 59*, 337–368.

Kramsch, C. (Ed.) (2002). *Language acquisition and language socialization: Ecological perspectives*. London: Continuum.

Lally, C. G. (Ed.) (2001). *Foreign language program articulation: Current practice and future prospects*. Westport, CT: Bergin & Garvey.

Lantolf, J. P., & Genung, P. B. (2002). "I'd rather switch than fight": An activity-theoretic study of power, success, and failure in a foreign language. In C. Kramsch (Ed.) *Language acquisition and language socialization: Ecological perspectives* (pp. 175–196). London: Continuum.

Lantolf, J. P. & Thorne, S. L. (forthcoming). *Sociocultural theory and the sociogenesis of second language development*. New York: Oxford University Press.

Leki, I. (2001). "A narrow thinking system:" Non-English-speaking students in group projects across the curriculum. *TESOL Quarterly, 35*, 39–67.

Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lightbown, P., Halter, R., White, J., & Horst, M. (2002). Comprehension-based learning: The limits of "do it yourself." *The Canadian Modern Language Review, 58*, 427–464.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning* (pp. 123–167). Clevedon, UK: Multilingual Matters.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.

Marx, N. (2002). Never quite a "native speaker:" Accent and identity in the L2- and the L1. *The Canadian Modern Language Review, 59*, 264–281.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(1), 3–11.

McCafferty, S. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal, 86*, 192–203.

Mellow, D., Reeder, K., & Forster, E. (1996). Using time-series research designs to investigate the effects of instruction on SLA. *Studies in Second Language Acquisition, 18*, 325–350.

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*, 492–518.
Pavlenko, A. (1998). Second language learning by adults: Testimonies of bilingual writers. *Issues in Applied Linguistics, 9*(1), 3–19.
Peirce, B. N. (1995). Social identity, investment, and language learning. *TESOL Quarterly*, *29*, 9–31.
Perdue, C., Benazzo, S., & Giuliano, P. (2002). When finiteness gets marked: The relation between morphosyntactic development and use of scopal items in adult language acquisition. *Linguistics*, *40*, 849–890.
Phelps, E., Furstenberg, F. F. J., & Colby, A. (Eds.). (2002). *Looking at lives: American longitudinal studies of the twentieth century*. New York: Russell Sage Foundation.
Robinson, P. (2001). Task complexity, cognition and second language syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). New York: Cambridge University Press.
Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. D. Rio & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 139–164). New York: Cambridge University Press.
Sato, C. (1990). *The syntax of conversation in interlanguage development*. Tübingen: Gunter Narr.
Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137–174). Rowley, MA: Newbury House.
Schumann, J. H. (1978). *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.
Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Philadelphia, PA: John Benjamins.
Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *29*, 1–15.
Tashakkori, A., & Teddlie, C. (Eds.). (2002). *The handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
Vavrus, J. (2002). Postcoloniality and English: Exploring language policy and the politics of development in Tanzania. *TESOL Quarterly, 36*, 373–398.
Watson-Gegeo, K. A. (2004). Mind, language, and epistemology: Toward a language socialization paradigm for SLA. *The Modern Language Journal*, *88*, 331–350.

https://doi.org/10.1017/S0267190505000024 Published online by Cambridge University Press