

21

Task-Based Language Learning

Michael H. Long

Jiyong Lee

Kyoko Kobayashi Hillman

21.1 Introduction

Most language learning and teaching is organized around series of linguistic units of various kinds, with grammatical, notional-functional, lexical, and hybrid syllabuses all popular. They use a “synthetic” syllabus, in which the target language is cut up into bite-size pieces and presented and practised one at a time. The learner’s job is to *synthesize* the pieces later for communication. The psychological rationale most commonly invoked for the synthetic syllabus is Skill Acquisition Theory (SAT), an import from first language (L1) psychology not originally developed with language learning in mind. The claim is that instruction first provides students with declarative knowledge of a structure—knowledge *that*, e.g., third person singular present tense verbs in English take a final *-s*. Controlled practice then converts declarative into procedural knowledge, knowledge *how*. Massive practice subsequently automatizes control of the item, until learners can produce *works*, *likes*, etc., effortlessly and fluently. Improvement in listening or speaking depends on practice in the skill concerned.

The dominance of what we will refer to as (in the broadest sense) “grammar-based” approaches and learning language *intentionally* continues despite theoretical and empirical evidence suggesting they are ill-conceived, especially, but not only, if the goal is *implicit* second language (L2) knowledge. It continues despite warnings about SAT’s limited scope:

Skill Acquisition Theory is most easily applicable to what happens in (a) high-aptitude adult learners engaged in (b) the learning of simple structures at (c) fairly early stages of learning in (d) instructional contexts. ([DeKeyser, 2015, p. 101](#))

Task-based language learning (TBLL) and teaching offers an attractive alternative. Task-based approaches are broadly consistent with theory and research findings in second language acquisition (SLA, the process language teaching is supposed to facilitate), and also with lessons learned from the history and philosophy of education. “Task-based” refers to the use of a syllabus whose content is a series, not of linguistic forms, but of pedagogic tasks sequenced in terms of increasing task, not linguistic, complexity, and lessons whose primary focus is communicative use of the L2 to complete those tasks. This is an example of what Wilkins calls an “analytic” syllabus. Much as in child L1A, learners experience gestalt samples of communicative target language use. Their job is to *analyse* the input and induce the underlying rules (or perceive the statistical regularities) and learn the lexis and collocations it contains *incidentally* while performing the tasks. There is no attempt to impose an external overt or covert linguistic syllabus, which would conflict with the internal “learner syllabus”, whose workings are reflected, among other ways, in well-attested common

errors and error types, developmental sequences, U-shaped behaviour, so-called “autonomous syntax”, and gradual approximation to target norms, not sudden categorical acquisition of target structures one at a time, as SAT would imply.

The psycholinguistic rationale proposed for TBLT varies somewhat, but is usually an amalgam of cognitive-interactionist and usage-based theories (see, e.g., Long, 2015a, pp. 30–62; Robinson, 2007, 2015; Skehan, 1998, 2015) developed with *language learning* as the explanandum. When students are adults, whose capacity for purely incidental learning, especially instance learning, is weaker than in young children, a variety of devices is required to enhance incidental learning and thereby speed up the process. The enhancements seek to help learners either *detect* or *notice* new items in the input by increasing their perceptual saliency and by drawing learners’ attention to needed lexis and collocations and grammatical patterns, especially when non-salient forms and form-function or form-meaning relationships are concerned. However, most of the attention-drawing procedures are deployed in response to learner performance, not in advance, as in synthetic approaches.

Grammar-based and task-based approaches are fundamentally different. The main foci of grammar-based language teaching (LT), and of the four-skills Present–Practice–Produce (PPP) pedagogy through which it is usually delivered, are the L2 as object of instruction, intentional learning, and explicit knowledge. The main foci of task-based language teaching (TBLT), conversely, are the L2 as medium of instruction, enhanced incidental learning, and implicit knowledge.

21.2 Four Problems with Grammar-Based LT

The many ways in which use of synthetic linguistic syllabuses conflicts with theory and research findings in SLA have been laid out in detail elsewhere (see, e.g., Long, 2009, 2015a, pp. 16–29; Long & Robinson, 1998). Here, we will briefly summarize four arguments against grammar-based approaches and overuse of the Initiation–Response–Feedback (IRF) exchange structure pervasive in the kind of teacher–student interaction encouraged by PPP pedagogy.

21.2.1 The Problem with Input, Interaction, and Output Quality

Input, interaction, and output in grammar-based PPP are impoverished—limited, repetitive, and mostly a matter of asking and answering display (known information) questions in IRF exchanges designed not for communication, but to manipulate examples of the structure of the day. For example, a teacher may take students through a short reading passage or “dialogue” deliberately crafted by a textbook writer to include numerous examples of the structure, and then ask a series of questions whose purpose is to see that students have “learned” the structure, or in the parlance of SAT, have developed “procedural knowledge” of it.

T: Where does Mary work? (I)

S: She work in a bank. (R)

T: She works in a bank. Works. (F)

She works in a bank. (I)

S: She works in a bank. (R)

T: Good (F)

It is linguistic ping-pong. Utterances are stripped down lexically, and typically have little or no communicative value, even when leaving the well-trodden terrain of information known to everyone from the textbook story, ostensibly to elicit creative examples of the new pattern applied to students' own lives:

T: Where does your father work? (I)

S: He works in an office. (R)

T: Good. (F)

The evaluative feedback move betrays the real purpose of the exchange, and the pseudo-communicative nature of the “conversation”. The lack of implicit knowledge is revealed moments later when the teacher notices that Ana is missing from class and asks if anyone knows where she is. Ana's friend volunteers that “She takes care her sister on Friday.”

Repetitive IRF sequences targeting the structure of the day are more common, especially in “elementary” and “intermediate” classes, than many supporters of grammar-based LT are willing to admit. Such exchanges have constituted a majority of classroom “talk” in many published empirical studies of grammar-based lessons (see, e.g., Long & Sato, 1983; White & Lightbown, 1984), even when the teachers concerned were trained in, and professed support for, some sort of “communicative” approach. Dinsmore (1985) likened the classroom discourse in the lessons he observed to the dialogue typical of the Theatre of the Absurd, such as Beckett's *Waiting for Godot*, where human existence has no purpose, and actors are reduced to meaningless repetitive exchanges using

language full of clichés, circularities, word play, and nonsense. In the L1 literature, the continuous recurrence of IRF sequences across generations of teachers (teach as you were taught?) has been referred to as “the persistence of the recitation” (Hoetker & Ahlbrand, 1969).

SLA research (e.g., Sato, 1990) has shown this is not the way languages are learned. To “teach” isolated grammatical structures first, for communication later, is to put the proverbial cart before the horse. Far from offering a direct path to instant native-like ability, one structure at a time, carefully monitored rehearsal of native-like sentences impedes development and internalization of implicit linguistic knowledge, among other ways because it takes no account of the learner’s internal syllabus, i.e., of what he or she is psycholinguistically ready to learn. Implicit knowledge is what is needed for the instantaneous demands of listening and spontaneous speech, and for functional L2 abilities in general. Really learning a new language entails an inevitably lengthy process of gradual approximation to target norms, replete with errors and interlingual forms along the way, modified and improved mostly by communicative failure and the implicit negative feedback available during negotiation for meaning, which pervades genuine communicative language use but rarely occurs in grammar-based classroom lessons. In the memorable words of Hatch (1978):

... language learning evolves out of learning how to carry on conversations ... One learns how to do conversation, one learns how to interact verbally, and out of this interaction syntactic structures are developed. (p. 404)

21.2.2 The Problem with Learnability

A fixed series of stages in the development of L2 German word order was first identified by the ZIZA group ([Meisel, Clahsen, & Pienemann, 1981](#)). The Wuppertal work came with an *explanation* for the developmental sequence ([Clahsen, 1987](#)), and because this was in terms of universal processing constraints, it could predict as yet unverified sequences, not just in other areas of GSL morphology and syntax, but in other L2s, as well.

Out of the ZIZA work grew Processability Theory (PT; [Pienemann, 1998](#); [Pienemann & Kessler, 2012](#)). PT has motivated numerous studies (continuing to this day) of developmental sequences in a variety of typologically unrelated languages. The findings have been broadly consistent with PT predictions. PT-motivated work (e.g., [Bonilla, 2014](#); [Ellis, 1989](#)) conducted to determine whether instruction can alter the sequences has found them robust, and the same is true in naturalistic and instructed language learning, i.e., unchanged by textbook sequences or grammar-based instruction that typically seeks to present and practice production of the final NS stage in a sequence from the get-go. Aside from the processing constraints, attempting to teach mastery of one structure at a time ignores the fact that few are free-standing, even at the elementary level, but are instead interdependent, requiring facility with others that go to make them up. English negative utterances, for example, require knowledge of basic word order, tense, number, and auxiliaries (*Mary doesn't like apples, John can't play the guitar, Peter didn't read the book*, etc.). The conclusion is that passage through developmental sequences can be sped up, but stages cannot be skipped.

Such results led Pienemann (1984, and elsewhere) to formulate his *Processability, Learnability, and Teachability Hypotheses*: what learners can process determines what they can learn, and what they can learn determines what teachers can teach. This presents a fundamental challenge to synthetic, grammar-based instruction, the implicit assumption underlying which is the opposite: teachers can teach what they want (the structure of the day), when they want (subjunctives on Tuesday), to whomever they want (all students, regardless of their psycholinguistic readiness).

Despite the robust research findings, the dominant approach to LT worldwide remains the same: the linguistic *plate du jour* delivered via PPP, seasoned with grammar to taste (i.e., focus on forms). This is not the fault of ISLA research, which has done its job well. Rather, it reflects (1) the immense power of commercial textbook publishing, an industry worth billions of dollars each year (billions with a b) and (2) the influence of washback from standardized discrete-point language testing, another industry worth billions of dollars each year. On the positive side, however, in addition to revealing the limitations of synthetic syllabi and PPP, the ISLA research findings constitute one of several rationales for the analytic syllabus and for TBLT (*task-based*, not *task-supported*, LT).

21.2.3 The Problem with Quantity of Practice

In courses intended to develop listening and speaking abilities, PPP and the IRF exchange structure have a negative effect, not just on the quality, but also the quantity of speaking opportunities for students. Using round numbers for ease of

exposition, and recognizing that class size, among other factors, can vary in different parts of the world, assume three sixty-minute lessons per week for a class of thirty students. Roughly fifty per cent of class time in a school year is taken up by reading, writing, classroom management, and testing, leaving an average of ninety minutes per week for listening and speaking. Since teachers generally control the initiation and feedback moves, students (if still paying attention) respond in the thirty minutes remaining for them. However, since about fifty per cent of aural–oral work (dictation, drills of one kind or another, etc.) is handled “lockstep”, in whole-class format, only about fifteen minutes is left for individual student oral production. Split among thirty students, this equates to an average of roughly thirty seconds per student per week—enough time for, say, four short utterances each. Practice in pairs and small groups can alleviate the problem somewhat. However, assuming thirty weeks in a school year, students are effectively asked to learn to speak individually in thirty seconds per week x thirty weeks, i.e., a total of about fifteen minutes each per year. This is a tall order for any students not blessed with unusually high language aptitude.

21.2.4.. The Problem with Purely Intentional Learning

Learning a new language as an adult is an immense task, especially if the learner’s L1 and L2 are typologically unrelated. For example, it has been estimated that non-native speakers (NNSs) require approximately 9,000 word families to read newspapers and novels, and about 6,000 to watch videos

(Nation, 2006). The time available in a typical foreign language course is hopelessly inadequate for treating each of those word families explicitly, and even more inadequate for the far greater number of collocations involved, in addition to L2 phonology, grammar, and pragmatics. Explicit instruction and intentional learning are too slow, an L2 too big, and time too short.

Explicit instruction, moreover, induces intentional learning, a conscious operation in which learners attend to aspects of a stimulus array in a search for underlying patterns or structure. When intentional learning works, it results in *explicit knowledge*: people know something, and know they know. For a functional command of a language, however, usable fluently in spontaneous communication, what students need is *implicit knowledge*, accessible instantly and without thought. Implicit knowledge is knowledge learners have but are unaware they have. It is the result of incidental, not intentional, learning.

Incidental learning is learning *without intention* to learn at least part of what is learned, i.e., while the learner's *attention* is focused on something else (task completion). If, and only if, the learner remains *unaware* of what is learned, the incidental learning process is also a case of *implicit learning*, i.e., learning *without intention and without awareness* of the end product, *implicit knowledge*. Almost all of native speakers' (NSs) knowledge of their L1 is implicit.

Implicit learning is more basic and *more important* than explicit learning, and *superior* (Whong, Gil, & Marsden, 2014). It is automatic and fast; it is what underlies listening comprehension, spontaneous speech, and fluency. It is the result of deeper processing and so more durable, and it obviates the need for explicit knowledge, freeing up attentional resources to focus on message content.

For many of their academic, occupational, and everyday social survival tasks—everything from listening to academic lectures, answering a teacher's question, and participating in a classroom discussion, through to demonstrating a new product, delivering a sales report, or recommending a course of medical treatment to a patient, to following street directions, buying a train ticket, or opening a bank account—adults will depend first and foremost on their *implicit knowledge*. To a lesser degree, the same is true of reading.

Since implicit L2 learning is the primary goal, students need opportunities for incidental learning. There are three problems, however. First, like intentional and explicit learning, incidental and implicit learning require more time, and a lot more input (rich, not impoverished, input) than is available in typical foreign language courses. Second, implicit learning may be limited to contingent items, or items either adjacent or brought into adjacency through some other means—by attention or by the use of a meaningful context in which they occur (Williams, 2009). Third, although laboratory studies of implicit learning of rules in artificial language grammars (see, e.g., Aslin & Newport, 2012) and field studies of naturalistic SLA (e.g., Granena & Long, 2013) demonstrate that adults remain capable of incidental and implicit learning, there are equally well-documented maturational constraints on language learning. The latter show that their capacity for incidental and implicit language learning is weaker than in young children. This is especially true of *instance learning*, i.e., learning of arbitrary, non-rule-governed items, such as noun gender, lexis, and collocations. For all these reasons, the size of the learning task, the limited time available, the priority accorded implicit knowledge, the possible limits to purely implicit language

learning, and the negative effects of increasing age, what will be needed is *enhanced incidental learning*, both to speed up acquisition and to improve ultimate attainment (Long, 2017).

Enhancements can take many forms, depending on the kinds of learners and, especially, the kinds of target linguistic features involved. For example, with task performance the primary focus, some analytically oriented learners may not detect perceptually non-salient items in the input they experience while doing the task, especially if the items are communicatively redundant (Ellis, 2017).

Examples (differentially relevant as L2 proficiency increases, and depending on the L1 and L2 involved) might include noun gender, Romance clitic pronouns, adverb placement, and inflectional morphology, such as English third person singular *-s* and *-ed*, whose function can be independently understood from context or other linguistics devices, such as adverbs (*Peter arrived yesterday*; Sato, 1986). A variety of attention-drawing devices are available, either built in to task-based materials or deployed by teachers in real time, to make such features more salient, so more likely to be *detected*, i.e., perceived without awareness on the student's part (Tomlin & Villa, 1994), or *noticed*, i.e., perceived with awareness at the moment of perception (Schmidt, 1990, and elsewhere).

21.3 Task-Based Language Learning and Teaching

Having identified serious problems with grammar-based approaches, we now turn to a brief description and evaluation of an increasingly popular alternative, task-based language learning and teaching. Task-based approaches are not

without their critics, too. For responses to the major charges, see Long (2016).

Fundamental differences between the two approaches will become clear from the following brief descriptions of six stages in the design, implementation, and evaluation of a full-fledged TBLT course.

21.3.1 Needs Analyses

A properly designed TBLT course begins with a needs analysis (NA), in two parts. The first is the *identification of target tasks* for particular groups of students, i.e., the real-world communicative uses to which learners will put the L2 beyond the classroom, the things they will need to *do* in and through the L2. The NA is carried out using (i) multiple oral and written *sources* of information, domain experts being the most important, (ii) multiple *methods* of gathering the information, such as interviews, questionnaires, and content analysis of training manuals and union contracts, and (iii) *triangulation* of sources and methods. The rationale for an NA is simple: learners differ, as do their target tasks. Occupations vary enormously, for example, and so do the tasks involved, as can easily be seen by spending a few minutes perusing the *Dictionary of Occupational Titles* or its online successor *O*Net*, the Occupational Information Network (both accessible and free of charge). The tasks that define the work of a hotel receptionist, diplomat, or automobile mechanic, for example, differ greatly from one another, and all of them from those performed by students in university or vocational training programmes.

If the learners are, or will be, using the L2 in a second language environment, there will be an auxiliary set of “social survival” tasks. Some may

be relevant for the majority of people working or studying overseas, regardless of the primary reason for their presence there, but others may not. Many learners will need to use public transport, open a bank account, rent an apartment, and visit a doctor, but not all will need to buy a car, enroll a child in school, or attend a job interview. Delivering the same course for everyone without first determining what they need to be able to *do* in the L2 is as unacceptable as it would be for a physician to prescribe the same treatment for all patients without first conducting a medical diagnosis. (For a detailed rationale for NA, examples, and methodological options, see, e.g., [Gilabert, 2005](#); [Jasso-Aguilar, 2005](#); [Long, 2005, 2013, 2015a](#), pp. 85–168; [Serafini, Lake, & Long, 2015](#).)

The second stage in a task-based NA involves gathering genuine samples of target language use by NSs performing the most frequent or most critical target tasks, and subjecting the samples to a flat, data-mining-type *analysis of target discourse* (ATD). Like the tasks themselves, language use can vary considerably in terms of the skills, genres, registers, lexis, and collocations involved, and often also in the meanings and functions of grammatical structures in different discourse domains, e.g., modals in scientific discourse and in doctor-patient interaction. The data for an ATD may consist of recorded speech samples, e.g., of radio and TV weather forecasts, economics lectures, and medical diagnoses, but also written texts, e.g., recipes, application forms, and equipment manuals. The target discourse samples are analysed in preparation for the production of task-based instructional materials. In an ATD, common patterns are sought, idiosyncratic features removed, and the findings distilled to produce

prototypical examples of the way particular target tasks are handled successfully by NSs. The procedure has been applied to such tasks as buying and selling a cup of coffee, obtaining and following street directions, decoding drug labels, and negotiating a police traffic stop. (For a detailed description and examples of the procedures followed in an ATD, as distinct from a comprehensive, generative, hierarchically structured, discourse analysis, see Long, 2015a, pp. 169–204.)

An ATD of a particularly complex task, requiring very advanced L2 proficiency, was conducted by Hillman (2017). Hillman had first used multiple sources, multiple methods, and triangulation of methods and sources to identify sixty-eight target tasks and the frequency with which they were used by Foreign Service Officers (FSOs) at the US embassy in Tokyo and in consulates elsewhere in Japan. An ATD was then conducted of one of the most important tasks—and because of the subtle honorific expressions required in Japanese, one of the most difficult for speakers of English: “Delivering celebration speeches in a formal setting.” Using transcripts of recordings of six celebration speeches, patterns in sequences of sub-tasks were identified, and the analysis validated by having two of the speeches coded independently by a second rater with high inter-rater reliability (Cohen’s kappa of 0.79 and 1.00). Important nouns, verb phrases, collocations, and politeness formulae were then tagged and linked to the sub-tasks. After the overall pattern for celebration speeches had been ascertained, and possible variations in the sequences of sub-tasks captured in a flow diagram, the findings were distilled to produce prototypical celebration speeches.

Modified *elaborated* (not linguistically simplified) versions of the prototypical spoken or written target discourse models are used in the

production of task-based materials—*pedagogic tasks*. Pedagogic tasks (and secondarily, teacher and student speech) constitute the major source of new language for learners, and because based on genuine samples of NS language use during performance of target tasks for those learners, a rich source of *relevant* new language.

2.1.3.2 Syllabus Design

The unit of analysis at every stage of a TBLT course is the *task*. When designing a syllabus, *target tasks* identified by the NA are first classified into more abstract *target task-types*, after which task-based instructional materials are produced in the form of gradually more complex *pedagogic tasks*. Series of pedagogic tasks make up the *task syllabus*. Target tasks are classified into target task-types for two reasons: economy of time and effort, and as one way of dealing, at least in part, with differing needs in heterogeneous groups. For example, making or changing a restaurant, theatre, train, or plane reservation (four target tasks) are all examples, at a more abstract level, of making or changing a reservation (one target task-type). Filling out application forms for a visa, a job, or a driver's license, or to join a club or open a bank account, are all examples of filling out an application form.

In a class with heterogeneous learner needs, presenting an oral or written annual sales report (a target task for a company employee in charge of, say, BMW sales in the UK) and giving a public talk or writing a newspaper article on climate change (target tasks for a climate change activist) share much in common at the level of target task-type, e.g., presenting a report of a series of

events. (For a sequence of six PTs suitable for the target task-type, Delivering a sales report, see Long, 2015a, pp. 287–291). For example, both will include narrative segments describing change over time, requiring reference to a series of past events (*first, next, then, before, afterwards, past* and *past perfect verb tenses*), cause and effect (*because, as a result, consequently, hence*), and some of the same lexis and collocations (*rose, fell, increased, decreased, a rapid/sharp/notable/encouraging/alarming increase/decline/improvement/drop/fall in X, reduced, doubled, halved, shrank, soared, plummeted*). The same target task-type will be relevant for researchers and students reporting empirical studies at scientific conferences, or in journal articles, lectures, and term papers.

Task-based materials written for heterogeneous groups will obviously not be as effective as materials produced for homogeneous groups. However, they will be more relevant than commercially published, grammar-based, “generic” materials written on the basis of textbook authors’ unreliable intuitions, not genuine target discourse samples, and marketed by publishers as appropriate for everyone, but in reality, for no-one in particular. The step up from target tasks to target task-types can ameliorate (not solve) the heterogeneity problem, as can work by individual students or sub-groups on occupation-specific or discipline-specific materials at certain stages within a module, and do so while retaining the psycholinguistic, methodological, and motivational advantages of TBLL.

Classification of target tasks into target task-types helps address additional problems facing most course designers, teachers, and students:

focusing on only one or two of the target tasks in each target task-type saves time and money. The same reasoning applies when the time comes to test students' task-based abilities. The assumption is that learners will be able to generalize what has been learned when they need to perform other target tasks within the target task-type that were not explicitly covered or tested during the course. That said, the extent to which it is in fact possible to predict the generalizability of task-based abilities is still unknown and an issue on which very little empirical research has been carried out as yet. (For a promising beginning, see [Benson, 2014](#).)

Pedagogic tasks (PTs) are simpler versions of target tasks. They are included in a task syllabus (selection) on the basis of the criticality and frequency of the corresponding target tasks revealed by the NA. As in Hillman's study, data on frequency and/or criticality can be gathered at the same time that target tasks are identified (see [Malicka, Gilabert, & Norris, 2017](#)). After NA findings have informed selection decisions, the next question (for any syllabus, regardless of the unit of analysis) is how items should be sequenced (grading). In TBLT, PTs are sequenced, or graded, according to their complexity—task complexity, not linguistic complexity ([Long, 2015a](#), pp. 227–241; [Robinson, 2009, 2011](#)).

Robinson has proposed making clear distinctions between task complexity, task difficulty, and task conditions. *Task complexity* refers to inherent, fixed, unchanging qualities of tasks. It can depend on the number of components, steps, or dimensions tasks involve, on whether they occur in the here-and-now (context-embedded) or there-and-then (context-reduced,

displaced time and space), or on the reasoning demands required or number of possible solutions to a problem. *Task difficulty* concerns the challenge a task presents to particular learners, depending on the qualities they bring to it. Thus, the difficulty of a task will vary for learners with different levels of L2 proficiency, IQ, language aptitude, relevant content knowledge, and so on. *Task conditions* are the circumstances under which learners perform a task, e.g., with or without planning time, speeded or unspeeded, alone or in groups, holding all or only part of the information required to complete the task, or with or without teacher assistance. Task difficulty can be raised or lowered by manipulating task conditions. Initial PTs in a sequence are simpler (how simple depends on students' current abilities), becoming progressively harder until they reach the full complexity of the target task that motivated their inclusion. The final task in a sequence is the full target task, or a simulation thereof, and serves as the exit task for the module.

There is a large and growing literature documenting the search for reliable, objectively measurable dimensions of task complexity. Many studies are carried out within the simplify–stabilize–automatize–restructure–complexify (SSARC) framework ([Robinson, 2007](#)). Robinson defines *resource-directing* variables as those that increase cognitive demands (requiring more reasoning, involving more steps or components, etc.), forcing learners to devote more attention to linguistic features, e.g., using more modifiers or relative clauses to differentiate persons, places, or things. This stimulates language development and can result in both increased complexity and accuracy. Resource-directing features include +/– here and now, +/– few elements, and +/– spatial, causal,

and intentional reasoning. After a statistical meta-analysis of eighty-six studies, [Malika and Sasayama \(2017\)](#) report that, in general, displaced time and space (there and then) leads to greater syntactic complexity, and more elements or reasoning demands result in greater lexical variety. *Resource-dispersing or depleting* variables are those that require attentional resources to be allocated to more things at once, e.g., talking to a customer while taking written notes about her order. Robinson claims that resource-dispersing or depleting dimensions of a task do not affect learners' attention to linguistic features, but do affect procedural demands and automaticity, and can result in increased fluency, but lower complexity and accuracy. Resource-dispersing or depleting task features include +/- single task, +/- task structure, +/- few steps, and +/- independence of steps.

[Malika and Sasayama \(2017\)](#) found few clear patterns among relationships between task complexity and oral and written production, but some tendencies, nonetheless. Planning, repetition, structure, support, and familiarity made tasks simpler, with positive effects on all four CALF variables—complexity, accuracy, lexis, and fluency—especially complexity and accuracy. They also found task complexity effects *within* categories of variables. Among resource-directing variables, the number of elements or reasoning demands, but not displaced time and space, had a positive effect on lexis. Among resource-dispersing or depleting variables, planning affected syntactic complexity and accuracy more than other variables, but affected fluency less than repetition and structure.

The goal of identifying reliable effects of task complexity on L2 performance has proven elusive. Findings are often non-cumulative because researchers use very different PTs in studies, different dimensions and operational definitions of complexity, and a plethora of outcome measures. In an early statistical meta-analysis, for example, Jackson and Suethanapornkul (2013) found that no fewer than eighty-three different measures had been employed for just one outcome variable, fluency. Clearly, coherent research programmes are needed, rather than one-off studies.

Two additional factors have been problematic. First, few studies have reported NS and NNS data on the same tasks. If the only participants are NNSs, variation in performance could be task-induced and/or simply a result of interlanguages being more unstable than NS grammars. Second, Révész (2014) has drawn attention to the fact that researchers often fail to establish that their manipulations of task features actually result in more or less complex versions, simply assuming, instead, that if the CALF variables behaved as predicted, it was due to the intended differences in task complexity. Révész suggested use of measures of cognitive load to eliminate this circularity by obtaining independent evidence of complexity. It has since become customary for researchers to include one or more such measures in their work.

A study by Lee (2018) addressed both issues. First, Lee examined data from NSs to provide a more reliable window on task complexity effects unfiltered by variability in NNS competence, so that any changes in performance could be attributed to complexity alone. Forty-two NSs of English performed three versions—least complex, mid-complex, and most complex—of each of

three oral tasks: a map task; a seating arrangement task; and a car accident task.

Task complexity was operationalized as the number of elements involved.

Second, to obtain an independent measure of cognitive load before linguistic outcomes were considered, Lee used (i) participant ratings of perceived difficulty, mental effort, stress, interest, and willingness to perform a similar task, (ii) prospective duration estimations, and (iii) a dual-task method involving participants having to respond to computer screen colour changes that occurred as they worked on the primary tasks. Results from all the measures showed that increasing task complexity led to the expected changes in cognitive load: greater levels of perceived difficulty, mental effort, stress, and interest; shorter prospective duration estimates; and slower reaction times on the dual-task method. As measured by Guiraud's Index, greater task complexity led to the most lexically diverse speech. Syntactic complexity, however, measured by the number of subordinate clauses per AS-unit, was highest on the mid-complex task, due to a tendency (encountered in several studies), whether intentional or unintentional, for some participants to simplify the most complex versions of tasks, either by failing to notice or simply ignoring added elements, resulting in less complex speech.

In important work comparing independent measures of cognitive load, [Révész, Michel, and Gilabert \(2015\)](#) attempted to validate task complexity using the dual-task method, participant ratings of overall task difficulty and mental effort required, and expert judgments. Forty-eight English NSs and forty-eight English as a second language (ESL) speakers performed simple and complex versions of a picture-narration, map, and decision-making task. As in [Lee's](#)

(2018) study, the dual-task method involved screen colour changes to which participants had to respond while performing the primary oral task. Using the same questionnaire, expert judgments of the complexity of the different versions were collected from sixty-one ESL teachers, who also provided the reasons for their answers. Results for all three methods showed that the complex versions of tasks imposed a greater cognitive load on participants. In the dual-task method, NNS accuracy on the secondary task was higher on the simple version, and the data on both ESL learners' and teachers' mental effort and task difficulty showed higher ratings for the more complex versions. With the objective (dual-task method) and subjective (self-ratings and expert judgments) measures all corroborating their predictions, the authors noted the practical advantages of the ease of use and high face validity of subjective measures for future studies.

The progress made notwithstanding, the vast majority of studies to date have focused on relationships between task complexity and various dimensions of L2 *performance*, usually one or more CALF variables in oral L2 production. There has been little work on the more important issue of L2 *acquisition*. This is a second important area for future research on task-based language learning and teaching.

21.3.3 Task-Based Materials

Modules of task-based instructional materials take the form of a series of pedagogic tasks, receptive, productive, or both. PTs are sequenced, or graded, according to their complexity as tasks, not linguistic complexity, using which would entail speculation and signal a return to grammar-based LT. The value of a

thorough NA becomes doubly clear at this stage, for the results remove the guesswork from deciding what learners should work on and mean that genuine samples of the language they need, already linked to relevant tasks for them, are available for manipulation by the materials writer.

Task-based materials are not linguistically controlled in the way grammar-based materials are. Authors of textbooks for grammar-based LT are customarily restricted to use of specified structures and verb tenses, and lexical items drawn from specific vocabulary frequency ranges, everything supposedly appropriate for a particular band on one of the ever-increasing number of “proficiency scales” (ILR, ACTFL, CEFR, IELTS, Pearson, etc.) so popular with publishers when marketing their teaching materials and achievement tests.

Aside from the earlier discussed lack of psycholinguistic credibility, linguistically graded spoken or written texts often come across as stilted, as inauthentic (which they are), and as “not the way NSs speak”. Even when whole stories or “dialogues” are not inundated with examples of the verb tense or structure of the day, lexical items and collocations NSs would use naturally are replaced by items from a higher frequency range, thereby failing to present learners with realistic models. Collocations and idiomatic usage suffer similar fates. Comprehensibility at a specified “proficiency level” is maintained, but by removing from the input the very items students need to learn.

In the design of task-based materials, it is tasks that are simplified, not the linguistic input, and gradually increasing task complexity is the basis for the ordering of PTs within sequences. (For detailed examples of PT sequences, see, e.g., [Long, 2015a](#), pp. 259–297; [O'Connell, 2014](#).) The elaboration process

preserves the realism of the genuine NS speech samples collected during stage 2 of the NA. If particular lexical items or collocations are appropriate for a task, but unknown to students, they are retained in the input, made comprehensible by building redundancy into the text, e.g., by paraphrasing or the addition of synonyms. Empirical studies (e.g., Oh, 2001; Yano et al., 1994) have confirmed that the level of comprehensibility of elaborated spoken or written text is usually close, and sometimes comparable, to that achieved by linguistic simplification, without the latter's toll on acquisition. Also, of considerable importance for such programmes as immersion, CLIL, and TBLT, where the L2 is not the only thing being learned, research has found input simplification, but not elaboration, to dilute curricular content (see, e.g., Long et al., 2017; Long & Ross, 1993).

21.3.4 Methodology and Pedagogy

The way a task syllabus is delivered differs fundamentally from the PPP procedure typically employed with grammar-based approaches. It is informed by ten *methodological principles* (MPs) for LT. Each is motivated by theory and research findings in SLA (Long, 2009, 2015a, pp. 300–328) and independently, in most cases, by the radical tradition in the philosophy of education represented by such thinkers as William Godwin, Leo Tolstoy, Peter Kropotkin, John Dewey, Paul Goodman, and Colin Ward, and by the rich history of free schools established by Sebastien Faure, Madeleine Vernet, Paul Robin, and Francisco Ferrer y Guardia, among many others. (For historical reviews, see, e.g., Avrich, 1980; Long, 2015a, pp. 63–83; Smith, 1983; Suissa, 2006). There are ten MPs: MP1: use task, not text, as the unit of analysis; MP2: promote learning by doing;

MP3: elaborate input; MP4: provide rich input; MP5: encourage inductive “chunk” learning; MP6: focus on form; MP7: provide negative feedback; MP8: respect learner syllabi and developmental processes; MP9: promote cooperative collaborative learning; and MP10: individualize instruction. MP1, MP3, and MP6 are original to TBLT.

MPs are realized at the local classroom level by a wide range of context-sensitive *pedagogic procedures* (PPs). There are no “best” PPs; rather, teachers, the experts on their local classroom situation, decide which are appropriate. This often requires moment-by-moment decision-making as a lesson unfolds, taking into account such factors as learner age, proficiency, literacy, and language aptitude, and whether a linguistic feature is marked or unmarked, perceptually salient or non-salient, and learnable by the student(s) at that time. By way of illustration, consider MP7: provide negative feedback. A teacher may decide it is necessary to provide negative feedback to address persistent errors with inversion, e.g., **John asked why did Neymar leave the best team*. Because the error is non-salient (string-internal, not affecting communication), the PP chosen might need to be drawn from the explicit end of the spectrum, e.g., a prompt or (if the learners concerned are thought capable of benefiting) a pedagogical rule of thumb. For a perceptually salient adverb-placement error, on the other hand, e.g., **I watch every week the Barcelona game*, an implicit PP for providing negative feedback, such as a recast or a clarification request, may suffice.

Reflecting the MPs listed above, interaction in TBLT classrooms (assuming an NA has shown listening and speaking skills to be relevant for the learners concerned) is very different from the IRF structure prevalent in

grammar-based PPP lessons. Instead of language as object, the primary focus is on communication. When working on a two-way PT (MP1: use task, not text, as the unit of analysis; MP2: promote learning by doing), for example, new information must move in both directions between teacher and student(s), and among students, for the task to be completed successfully. Their talk will have a purpose. Due to the learners' linguistic limitations, teachers and students will inevitably need to negotiate for meaning (as in all communication, even among NSs), especially, although not only, when communicative trouble arises. Instead of the focus in grammar-based LT on rehearsal of particular sentence patterns, interaction in task-based lessons is motivated by the need to understand or send messages, to solve interesting problems posed by PTs, to express feelings and opinions, and more. Being vested in interesting conversation motivates students, who are more likely to attend to the input, as a result—a requirement for acquisition. (For examples of tasks usable in these ways even at the most basic proficiency level, see [Long, 2015a](#), pp. 259–274; [Shintani, 2016](#).)

Many other positive consequences for classroom interaction follow from the basic focus on task accomplishment and communication. To begin with, turns at talk and conversational roles are more evenly distributed. Because students often possess information and opinions unknown to the teacher or to other students with whom they are working on PTs, they must collaborate, initiating exchanges, responding to initiations by others, and providing feedback, e.g., to confirm or check their understanding of what the teacher or another student has said, or to request clarification (MP9: promote cooperative collaborative learning). Exchanges depart from the repetitive IRF pattern limiting students to

passive responses to display questions, followed by feedback on grammatical accuracy, not the message. Now, topic-initiating moves, whether by teachers or students, often trigger extended sequences of reacting moves. Classroom talk becomes more like talk outside classroom. Input is richer and both the quality and quantity of student input and output are greatly improved (MP3: elaborate input; MP4: provide rich input). New language is more readily incorporated into students' grammars and lexicons, in part because it was intended for them (MP5: encourage inductive "chunk" learning; MP10: individualize instruction). They are aware of its usefulness at the moment it is encountered; because they need it to complete a task, they are attending (MP5: encourage inductive "chunk" learning). Not surprisingly, studies show students learn better from the acquisition-rich environment created by task-based interaction (see, e.g., the statistical meta-analysis by [Mackey & Goo, 2007](#)).

Negative feedback is plentiful in task-based interaction (MP7: provide negative feedback). Further, as learners negotiate for meaning, more of the feedback they receive is tuned to their current level of development (MP8: respect learner syllabi and developmental processes). The utterance they hear immediately following their own, be it a *confirmation* that what they said was understood or a *comprehension check* or *clarification request* of some kind, will often constitute a slight reformulation of their output. Because the feedback is contingent on, and similar to, what they just said, its meaning will usually already be clear, so their attentional resources will be freed up to devote to the form of the response, in context, and to hold it and what they said in short-term memory long enough to notice any differences (MP6: focus on form, not forms).

In other words, the learnability issue is at least partly addressed, because more of the input is triggered by learners' attempts to use the L2 communicatively, with the parties' primary focus on what is said, not how it is said:

Student 1: Green triangle beside blue square?

Student 2: The green triangle beside the blue square, yes

(confirmation/recast)

Recasts have been found to be the most frequent form of negative feedback in all sorts of L2 classrooms, from immersion to foreign language lessons, and over sixty studies have shown they produce substantial and durable learning gains in grammar and vocabulary. (For excellent narrative reviews, see [Goo & Mackey, 2013](#); [Yilmaz, 2016](#); and for two statistical meta-analyses, [Li, 2010](#), and [Mackey & Goo, 2007](#).) They do so without the disruption to communication caused by overt, on-record "error correction". Moreover, due to the primary (not exclusive) focus on meaning and task completion, much of the language learning that takes place is incidental, making implicit knowledge a more likely end-product.

Task-based interaction also succeeds when students work with fellow students in pairs or small groups, not least because individual output opportunities are far more numerous than during lockstep classroom interaction, and because input is individually tailored. As part of an eight-week pre-test, post-test, delayed post-test study of young adult learners of German receiving twelve hours of instruction per week, [Eckerth \(2008\)](#) conducted detailed analyses of fifty hours of audio-recorded peer-peer interaction. He found little empirical support for the idea that learners reinforce one another's errors. Rather, thirty-eight per cent of non-target-like hypotheses were replaced

by correct forms and integrated into the learners' IL grammars over the ten-week period, with the majority of changes traceable to learning opportunities arising from the task-based interactions. Moreover, the collaborative dialogue led to learning gains after task completion in twenty-two per cent of cases when, at the outset, neither learner possessed the relevant L2 knowledge. Eckerth's verdict was that the task-work by pairs of learners indicated "significant gains in the short and medium term. Thus, it can be concluded that learners are able to provide each other with feedback rich in acquisitional potential" (Eckerth, 2008, p. 133).

21.3.5 Task-Based Assessment

Student abilities are assessed using *task-based, criterion-referenced performance tests* (Brown & Hudson, 2002; Long, 2015a, pp. 329–341; Norris, 2009; Van Corp & Deygers, 2013). This is a relatively straightforward process in the many cases, such as attending an economics lecture, conducting a visa interview, setting up laboratory equipment, or following street directions, where the ability to perform target tasks can be assessed directly and discretely, with unambiguous behavioural outcomes as achievement measures. For example, in a university economics lecture, can students identify at least eighty per cent of the information bits designated important by a domain expert (the lecturer), as shown by their scores on a multiple-choice test? Task-based assessment can be problematic, however, and involve choices among less desirable options, including whether or not to add a linguistic caboose when comprehension or simply getting a message across may not suffice for acceptable task performance

(Nielson et al., 2009). For example, as noted earlier, a diplomat's after-dinner speech in Japanese must be delivered using the correct register, with appropriate formulae and grammatical particles to mark politeness (Hillman, 2017).

21.3.6 Programme Evaluation

Studies comparing grammar-based PPP and TBLT have produced the same general findings: TBLT produced comparable results to grammar-based PPP on forms-focused or discrete-point outcome measures, and superior results on communicative measures. Shintani (2011) employed a quasi-experimental, pre-test-post-test-delayed post-test design to compare the learning of twenty-four new nouns by thirty-six young Japanese children (ages 6–8), in six intact groups of six, over a twelve-week course. Two output-focused grammar-based PPP classes employed five drill-like games. Language learning was intentional, the children having been told at the outset that the goal was for them to learn the twenty-four new words. The two TBLT classes were not told about the goal, and focused on input-based, game-like, listen-and-do tasks. There was a case of incidental learning. The last two groups, functioning as controls, received TPR instruction, English songs, practice writing the alphabet, and no exposure to the target words. Recordings showed there were 200 IRF sequences and no negotiation for meaning in the grammar-based PPP lessons, with teacher-initiated tokens of the target vocabulary items dominating. That compared with only twenty-five IRF sequences in the task-based lessons, where student-initiated tokens dominated. Overall exposure to the target words was roughly

the same in the PPP and TBLT lessons, but target-word production was 444 in the output-based PPP, and only 144 in the input-based TBLT (3:1). There were two listening and two speaking measures, one of each discrete-point, the other communicative and task-based. Both the TBLT and PPP groups outperformed the control groups on all four measures. The TBLT groups performed as well as the PPP groups on the discrete-point listening and speaking measures (despite fewer production opportunities in the input-based TBLT lessons), comparably on the task-based speaking measure (again despite the more limited production opportunities in the input-based lessons), and significantly better than the PPP group on the measure of task-based listening.

In a second study, Shintani (2013) randomly assigned forty-five six-year-old Japanese child beginners to form three classes of fifteen: TBLT/focus on form; PPP/focus on forms; and control. Nine lessons over the next five weeks focused on twenty-four nouns and twelve adjectives. Outcomes were measured by a combination of discrete-point and task-based and communicative production measures. The classroom process data again showed that only the TBLT/focus on form condition featured contextualized input, student initiation, and negotiation for meaning. The TBLT/focus on form groups did as well as the PPP/focus on forms groups for nouns on both discrete-point and task-based/communicative production tests, and better than the PPP/focus on forms groups for adjectives on both tests. (For detailed discussions of Shintani's studies, see Long, 2015a, pp. 357–361, 2015b, pp. 66–67.)

In a study with young adults, Borro (2017) randomly assigned two intact classes of Chinese students at an Italian university to a TBLT ($n=11$) or PPP

(n=10) condition. Each group received two three-hour lessons based on genuine samples of NS Italian used when opening a bank account and changing a mobile phone contract in a shop, each having been identified by an NA as target tasks for the Chinese students. The PPP condition involved intentional learning. Lesson content began with a focus on vocabulary before reading a linguistically simplified version of the NS speech data, along with comprehension questions, explicit grammar instruction, fill-in-the-blank grammar exercises, and a final role-play or text-writing and more output. The linguistic target was a high frequency structure in the input for both tasks, third person clitic pronouns, both direct and indirect object (*Le fragole costano poco, allora *la* compro *Strawberries are cheap, so I buy *it*). The PPP group was exposed to twenty pronouns in the bank text and ten in the mobile text. They completed two fill-in-the-blank grammar exercises per lesson, with about twelve items each, for a total of seventy-eight tokens.

The TBLT condition involved incidental learning. There was no decontextualized vocabulary instruction before students were exposed to genuine and elaborated oral and written versions of the input and written transcriptions of dialogues containing the target structures enhanced (in bold). The students worked on pedagogic tasks, e.g., matching dialogues with correct fliers among five or six describing special offers, spotting differences between different kinds of bank accounts, and a final role-play. There was no explicit grammar instruction unless in reaction to student questions. The TBLT group encountered eighteen instances of the pronouns in the spoken dialogues, twenty-

two pronouns in the bank task, and seventeen in the mobile task, for a total of fifty-seven tokens.

Borro administered four versions of the same unspeeded grammaticality judgment test (UGJT) and a moving-window self-paced reading test (SPRT) as pre- and post-tests, assessing explicit and implicit knowledge of the target structure, respectively. The SPRT comprised twelve sentences containing the target structure, and twenty-four fillers. Each item occurred in grammatical (semantically consistent), ungrammatical (semantically inconsistent), masculine, and feminine versions. Scores on the outcome measures showed that the TBLT and PPP groups both improved statistically significantly from pre- to post- on the UGJT, but did not differ significantly from one another. Conversely, performance on the SPRT, which assessed implicit knowledge, was superior for those in the TBLT condition. Reaction times (RTs) to grammatical semantically congruent sentences in the TBLT group decreased statistically significantly after the instruction and were significantly shorter than those in the PPP group. RTs to grammatical semantically incongruent sentences decreased slightly, but not significantly. RTs in the PPP group barely decreased from pre- to post- on any sentence types. It is noticeable that learning of the target grammatical structure was superior in the incidental learning TBLT condition, despite the lack of overt focus on it there, and despite the higher number of tokens (seventy-eight versus fifty-seven) and explicit treatment of the structure in the grammar-based PPP lessons.

These explicit comparisons of grammar-based PPP and TBLT are small-scale and limited in several ways, but each suggests advantages for TBLT at both

the process and product levels. Their findings support the claims discussed earlier (Section 3.4) concerning the superior quality, quantity, and structure of task-based classroom interaction for L2 acquisition, and its advantages for incidental learning. They are in line with results from other evaluations of TBLT lessons, courses, and whole programmes. Reports of the first twenty years of successful implementation of TBLT in over 200 state primary and secondary schools in Flanders appeared in 2006 (Van den Branden, 2006). Additional formative and summative evaluations of TBLT for children and adults have been published since, all favourable to a greater or lesser extent. (For details and a review of studies, see [Long, 2015a](#), pp. 341–366). A statistical meta-analysis of fifty-two evaluations of programme-level implementations of TBLT in real classroom settings in many parts of the world ([Bryfonski & McKay, 2017](#)) revealed an overall positive and strong effect ($d = 0.93$) for TBLT implementation, compared with grammar-based PPP, on a variety of learning outcomes, and positive stakeholder perceptions towards the TBLT programmes. However, while results so far are positive, more evaluations of TBLT are needed, especially of complete programmes *in situ*. The small number of such studies (already far larger than for any other approach) is mostly due to the advent of TBLL and TBLT having been relatively recent.

21.4 Conclusion

Language teaching in most parts of the world is grammar-based. However, research has shown that, among other major problems, externally timed discrete linguistic items have little credibility as units of acquisition. This is one reason

for the growing interest in task as an alternative organizational unit around which to organize learning. Fundamental differences between grammar-based and task-based approaches to language learning and teaching are summarized in [Table 21.1](#).

Begin Table 21.1

Table 21.1

Grammar-based and task-based language learning and teaching.

Grammar-based LT	TBLT
Language as object	Language as communication
Language for (future) use	Language through use
No needs analysis	Needs analysis
Language for nebulous purposes	Language for specific purposes
Intentional learning	Enhanced incidental learning
Noticing	Noticing or detection
Explicit knowledge	Implicit knowledge
Structure as the unit of analysis	Task as the unit of analysis
Generic materials	Needs-driven materials
Restricted input	Rich input
PPP	Ten MPs and numerous PPs
Lessons often boring	Lessons usually stimulating
Textbook in control	Teacher in control
Norm-referenced assessment	Criterion-referenced assessment
No programme evaluations	Some programme evaluations
Vanishingly little research support	Considerable research support

End Table 21.1

Tasks and task-based approaches are attractive, due to their greater psycholinguistic credibility from a language-learning perspective, and because the incidental and implicit learning they encourage can result in the implicit knowledge needed by learners requiring a functional command of the L2. Moreover, the ability to perform tasks is a more tangible way of identifying what learners need a new language for, and then of evaluating their ability to do what

they need to do. *Learners know a language, but they do not do a language; they do tasks through a language.*

Over the past thirty years, a substantial body of research has developed concerning several dimensions of tasks and TBLL, including task criticality, frequency, complexity, and difficulty, and the conditions under which tasks are performed. Experience has also grown with practical issues in the design and implementation of TBLT, and with ways of dealing with problems that can arise in the field. As several commentators have noted, despite its comparatively recent appearance, the hundreds of studies in the published literature have already made task-based language learning and teaching by far the most extensively researched approach to language teaching the field has ever seen.

chapter-references

References

- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170–176.
- Avrich, P. (1980). *The modern school movement. Anarchism and education in the United States*. Princeton, NJ: Princeton University Press.
- Benson, S. (2014). Task-based language teaching: An empirical study of task transfer. *Language Teaching Research*, 20(3), 341–365.
- Bonilla, C. (2014). From number agreement to the subjunctive: Evidence for Processability Theory in L2 Spanish. *Language Teaching Research*, 31(1), 53–74.

Borro, I. (2017). Comparing the effectiveness of TBTL and PPP on L2 grammar learning. A self-paced-reading study with Chinese students of Italian L2. Ms. University of Portsmouth.

Bryfonski, L., & McKay, T. (2017). TBTL implementation and evaluation: A meta-analysis. *Language Teaching Research*.

Clahsen, H. (1987). Connecting theories of language processing and (second) language acquisition. In C. Pfaff (ed.), *First and second language acquisition processes* (pp. 103–116). Cambridge, MA: Newbury House.

DeKeyser, R. D. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (eds.), *Theories in second language acquisition. An introduction* (2nd edn.) (pp. 94–113). Mahwah, NJ: Lawrence Erlbaum.

Dinsmore, D. (1985). Waiting for Godot in the EFL classroom. *ELT Journal*, 39(4), 225–234.

Eckerth, J. (2008). Investigating consciousness-raising tasks: Pedagogically targeted and non-targeted learning gains. *International Journal of Applied Linguistics*, 19(3), 119–145.

Ellis, N. (2017). Salience in usage-based SLA. In S. Gass, P. Spinner, & J. Behney (eds.), *Salience in second language acquisition* (pp. 21–40). London: Routledge.

Ellis, R. (1989). Are classroom and naturalistic acquisition the same? A study of the classroom acquisition of German word order rules. *Studies in Second Language Acquisition*, 11(3), 305–328.

Gilabert, R. (2005). Evaluating the use of multiple sources and multiple methods

in needs analysis: A case study of journalists in the Autonomous

Community of Catalonia (Spain). In M. H. Long (ed.), *Second language*

needs analysis (pp. 182–199). Cambridge: Cambridge University Press.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in*

Second Language Acquisition, 35(1), 127–165.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language

aptitude, and ultimate L2 attainment in three linguistic domains. *Second*

Language Research, 29(3), 311–343.

Hatch, E. M. (1978). Discourse analysis and second language acquisition. In E. M.

Hatch (ed.), *Second language acquisition: A book of readings* (pp. 402–

35). Rowley, MA: Newbury House.

Hillman, K. K. (2017). Target tasks for US Foreign Service Officers: The challenge

for TBLT of the Japanese celebration speech. Scholarly Paper, Ph.D. in SLA

Program, University of Maryland.

Hoetker, J., & Ahlbrand, W. P. (1969). The persistence of the recitation. *American*

Educational Research Journal, 6(1), 145–167.

Jackson, D. O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A

synthesis and meta-analysis of research on second language task

complexity. *Language Learning*, 63(2), 330–367.

Jasso-Aguilar, R. (2005). Sources, methods and triangulation in needs analysis: A

critical perspective in a case study of Waikiki hotel maids. In M. H. Long

(ed.), *Second language needs analysis* (pp. 127–158). Cambridge:

Cambridge University Press.

Lee, J. (2018). Task complexity, cognitive load, and L1 speech. *Applied Linguistics*.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365.

Long, M. H. (ed.) (2005). *Second language needs analysis*. Cambridge: Cambridge University Press.

Long, M. H. (2009). Methodological principles for language teaching. In M. H. Long & C. J. Doughty (eds.), *Handbook of language teaching* (pp. 373–394). Oxford: Blackwell.

Long, M. H. (2013). Needs analysis. In C. Chapelle (ed.), *The encyclopedia of applied linguistics*. Oxford: Wiley-Blackwell.

Long, M. H. (2015a). *Second language acquisition and Task-Based Language Teaching*. Malden, MA: Wiley.

Long, M. H. (2015b). Experimental perspectives on classroom interaction. In N. Markee (ed.), *Handbook of classroom discourse and interaction* (pp. 60–73). Oxford: Wiley-Blackwell.

Long, M. H. (2016). In defense of tasks and TBLT: Non-issues and real issues. *Annual Review of Applied Linguistics*, 36, 5–33.

Long, M. H. (2017). Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition*, 1(1), 7–44.

Long, M. H., Al-Thowaini, A., Al-Thowaini, B., Lee, J., & Vafaee, P. (to appear). A micro process-product study of a CLIL lesson: Linguistic modifications, content dilution, and vocabulary knowledge. *Instructed Second Language Acquisition* 1(2), 2017.

Long, M. H. & Robinson (1998). Focus on form: Theory, research and practice. In C. Doughty & J. Williams (eds.), *Focus on form in second language acquisition* (pp. 15–41). Cambridge: Cambridge University Press.

Long, M. H., & Ross, S. (1993). Modifications that preserve language and content. In M. Tickoo (ed.), *Simplification: Theory and application* (pp. 29–52). Singapore: SEAMEO Regional Language Centre.

Long, M. H., & Sato, C. J. (1983). Classroom foreigner talk discourse: Forms and functions of teachers' questions. In H. W. Seliger & M. H. Long (eds.), *Classroom-oriented research on second language acquisition* (pp. 268–285). Rowley, MA: Newbury House.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (ed.), *Conversational interaction in second language acquisition* (pp. 407–452). Oxford: Oxford University Press.

Malicka, A., Gilabert Guerrero, R., & Norris, J. M. (2017). From needs analysis to task design: Insights from an English for specific purposes context. *Language Teaching Research*.

Malika, A., & Sasayama, S. (2017). Cognitive task complexity: A research synthesis. Paper presented at the Seventh International Conference on TBLT, University of Barcelona, 18–21 April.

Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3(1), 109–135.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63, 59–82.

Nielson, K. B., Masters, M. C., Rhoades, E., & Freynik, S. (2009). *Prototype implementation of an online Chinese course: An analysis of course implementation and learner performance*. (TTO 82131). College Park, MD: University of Maryland Center for Advanced Study of Language.

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (eds.), *Handbook of language teaching* (pp. 578–594). Oxford: Blackwell.

O'Connell, S. (2014). A task-based language teaching approach to the police traffic stop. *TESL Canada Journal*, 31(Special issue 8), 116–131.

Oh, S-Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, 35(1), 69–96.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6(2), 186–214.

Pienemann, M. (1998). *Language processing and second language development*.

Processability theory. Amsterdam/Philadelphia, PA: John Benjamins.

Pienemann, M., & Kessler, J.-U. (2012). Processability Theory. In S. M. Gass & A. Mackey (eds.), *The Routledge handbook of second language acquisition* (pp. 228–246). New York: Routledge.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, 35(1), 87–92.

Révész, A., Michel, M., & Gilabert, R. (2015). Measuring cognitive task demands using dual task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*, 28(4), 1–35.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia-Mayo (ed.), *Investigating tasks in formal language learning* (pp. 7–26). Bristol: Multilingual Matters.

Robinson, P. (2009). Syllabus design. In M. H. Long & C. J. Doughty (eds.), *Handbook of language teaching* (pp. 294–310). Oxford: Blackwell.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (ed.), *Second language task complexity. Researching the Cognition Hypothesis of language learning and performance* (pp. 3–37). Amsterdam: John Benjamins.

Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 87–121). Amsterdam: John Benjamins.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.

Sato, C. J. (1986). Conversation and interlanguage development: Rethinking the connection. In R. R. Day (ed.), *"Talking to learn": Conversation in second language acquisition* (pp. 23–45). Rowley, MA: Newbury House.

Sato, C. J. (1990). *The syntax of conversation in interlanguage development*. Tubingen: Gunter Narr.

Serafini, E. J., Lake, J. B., & Long, M. H. (2015). Methodological improvements in identifying specialized learner needs. *English for Specific Purposes*, 40, 11–26.

Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research*, 15, 137–158.

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginner learners. *TESOL Quarterly*, 47(1), 36–62.

Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. Amsterdam/Philadelphia, PA: John Benjamins.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 123–155). Amsterdam: John Benjamins.

Smith, M. P. (1983). *The libertarians and education*. London: George Allen & Unwin.

Suisse, J. (2006). *Anarchism and education. A philosophical perspective*. London: Routledge.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition*, 16, 183–204.

Van Gorp, K., & Deygers, B. (2013). Task-based language assessment. In A. Kunan (ed.), *The companion to language assessment. Vol. 2: Approaches and development* (pp. 578–593). Oxford: Wiley-Blackwell.

White, J., & Lightbown, P. M. (1984). Asking and answering in foreign language classes. *Canadian Modern Language Review*, 40, 228–244.

Whong, M., Gil, H.-G., & Marsden, E. (2014). Beyond paradigm: The “what” and the “how” of classroom research. *Second Language Research*, 30(4), 551–568.

Williams, J. N. (2009). Implicit learning. In W. C. Ritchie & T. K. Bhatia (eds.), *The new handbook of second language acquisition* (pp. 319–353). Bingley: Emerald Group Publishing.

Yilmaz, Y. (2016). The linguistic environment, interaction and negative feedback. *Brill Research Perspectives on Multilingualism and Second Language Acquisition*, 1(1), 45–86.