Annual Review of Applied Linguistics, **46** (2026), pp. 259–267.
© Cambridge University Press, 2017
doi: 10.1017/S0267190517000010

Theory, empiricism and practice: Commentary on TBLT in ARAL 2016

MARTIN BYGATE
Lancaster University

SUSAN M. GASS
Michigan State University

ALISON MACKEY
Georgetown University

RHONDA OLIVER
Curtin University

PETER ROBINSON
Aoyama Gakuin University

ABSTRACT

This commentary includes consideration of theory, opinion, empirical work, evaluative work and practice in a subset of papers that appeared in ARAL (2016) on task-based language teaching. Addressing the wider logic, theoretical underpinnings and instructional implications of TBLT is a serious applied linguistics challenge and one that the TBLT area is increasingly in a position to engage with in the search for evidence-based solutions to problems in designing, implementing and assessing the effects of TBLT. The contributors take a sometimes approving, occasionally critical and always forward looking perspective on the contributions of ARAL 2016 and the directions and challenges facing the field.

Introduction, Alison Mackey:

The 2016 issue of the Annual Review of Applied Linguistics focused on task-based language teaching (TBLT) from a range of perspectives. The journal has recently moved from a model of publishing only reviews to a new format that includes opinion pieces and empirical work as well, enabling a forward looking approach to augment the traditional one of surveying the field. In comparison with the past four decades since ARAL was established, these days things move more quickly in publishing. An increasing number of journals are making papers available on Firstview, or the equivalent, before they appear in print. In addition, the submission-publication process has accelerated with the advent of electronic submission. The journal's new format reflects this faster pace, and the current paper is the first in another new move. From 2017 onwards, in each new issue of

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

ARAL, there will be a reaction paper that responds to the prior issue with the aim of promoting engagement. This reaction will be either singly or multiply authored. Responses were peer-reviewed, as is the case for all papers in ARAL.

The authors of the current paper were each invited to make a short commentary on the 2016 TBLT issue. Each author selected a topic and paper(s) of their choosing. Each author (Bygate, Gass, Oliver, and Robinson) contributed one section, with the editor (Mackey) providing the overview and connections.

The topic of the issue, TBLT, has its own book series (John Benjamins) in addition to monographs and edited collections published in other series (for example, Bygate, Skehan and Swain, 2001, Longman). It has an international association (IATBLT) which holds an international bi-annual conference (TBLT is in Barcelona, Spain in 2017). Hundreds of papers appear in journals, and are given at colloquia or as presentations at conferences. The body of research is such that in recent years, meta-analyses and syntheses of work in the field have emerged (see for example, Bryfonski and McKay's forthcoming meta-analysis for which they assessed 52 studies of TBLT implementations). The 2016 issue of ARAL aimed to capture a small subset of this rich diversity, and with a brief to keep their 'short commentaries' short, contributors had to select and focus on an even smaller subset. Nevertheless, this commentary examines theory and opinion, empiricism and synthesis, and particularly focuses on application and practice, in ways that are sometimes approving, occasionally critical, and always forward looking.

To begin then, Gass points out that is no better way to move the field on than to bring a multiplicity of approaches together in a single volume. In what follows, she comments on one of each type of article from ARAL (2016): an overview, an empirical study, and a meta-analysis.

Theory and empiricism, Susan M. Gass:

TBLT 2016 includes a valuable set of papers that range from thought pieces and review articles to experimental studies and even includes meta-analyses. Starting with thought pieces, Long's article, *In Defense of Tasks and TBLT: Nonissues and Real Issues* gives us pause as we think about criticisms of TBLT. Quite clearly, we need to have a coherent definition of what is meant by any construct before we can consider criticisms and evaluate their validity (or not) as criticisms. Long takes us through a range of criticisms, debunking them where they should be debunked, pointing out the numerous misinterpretations (willful or not) of existing SLA literature. Despite the major emphasis in Long's article on countering false and/or unsubstantiated claims leveled against TBLT, there is a particularly interesting conclusion that helps put things into perspective. As he acknowledges, TBLT is not a panacea; many of the issues that arise from this approach to language teaching are similar to any language teaching method, but at least we have an understanding of the 'true' problems and the bogus ones—no alternative facts here.

Li, Ellis, and Zhu take us down an interesting path highlighting the always vexed question of elicitation types and the interpretation that stems from looking at the results based on one elicitation type or another. In the case of the Li et al. study, we

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

have one measure which purports to elicit explicit knowledge and the other which in their words elicits 'automated' knowledge. All participants performed a dictogloss with treatment groups in which there were different attentional emphases. It is nice when results converge and to some extent they did. In both elicitation measures (untimed grammaticality judgment and elicitation imitation), attention to form through explicit instruction and feedback was the best facilitator of learning. Unexpected was the result that differentiated participants by their degree of prior knowledge. This was the case with the results from the elicited imitation task where those with prior knowledge benefited more from explicit instruction coupled within-task feedback.

The idea of needing prerequisite knowledge before new information is useful goes back in SLA at least to work by Andersen (1983) and more recently to Carroll (2012). In the case of the former, the topic was transfer and in the case of the latter, the topic was salience. Carroll, in a discussion of the acquisition of target language intonation, claims that salience is relevant "after they [L2 learners] have acquired the relevant knowledge of the target language intonation and its role in information structure" (p. 63). What this suggests is that something becomes salient as a result of learning and in this sense salience prepares a learner to acquire new knowledge. In the case of Li et al., prior knowledge allows new information to become automated.

Finally, the article by Plonsky and Kim uses meta-analysis to investigate a number of critical issues in task-based research, focusing on both learning outcomes and learner performance. A number of meta-analyses have been carried out in the field, many by the first author himself and in general by those that support 'methodological reform' (Byrnes, 2013). What is disturbing in general, but perhaps not shocking, are the comments that indicate that studies that investigated methodological issues "do not generally paint a very favorable picture of L2 research methods" (p. 78). This coupled with comments that authors do not often "provide any justification (e.g., TOEFL score) for their assessment of proficiency" (p. 82) is somewhat disheartening especially given that the studies reported on are from the past 10 years. Thomas's plea (1994) for more accurate measures of proficiency has not been heard. This article is important as it points the way forward with some easy steps for conducting research in SLA ranging from suggestions for design, analysis, and reporting. The relevance and stability of the field depend on our successes in this arena.

This commentary by Gass and together with the piece she is commenting about by Plonsky and Kim (2016), both make an important point about methodology. A current move in the field, indeed in the social and psychological sciences as a whole, is to emphasize the need for more work on replication (e.g. Porte, 2012). Journals are increasingly beginning to recognize the importance of publishing enough detail to allow replication, and the free, searchable, downloadable IRIS database is facilitating the ease of this sort of research by encouraging researchers to upload instruments from all published second language research (for more information, see Marsden and Borowczyk's blog post on this topic, published on the

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

journal's website). Researchability is something that Peter Robinson also takes up but before that, we return to theory with a piece by Rhonda Oliver who points out that ARAL's new mission includes "state-of-the-art overviews of the field … position pieces" noting that, as ever, no one is left in any doubt about Michael H. Long's position.

Theory and practice in the Australian context,
Rhonda Oliver:

In his 2016 ARAL paper, Michael H. Long provides a definitive statement on what tasks are (and are not) and what TBLT really means. And, as the originator of TBLT, so he should. The eloquent arguments, including a succinct list of methodological principles, and his wicked turns of phrase had me both nodding agreement, but also chortling at his word choices. It is, perhaps, my favorite 'Long' read yet. Using his encyclopedic knowledge of the field as ammunition and with the precision of a sharp shooter, Long picks off the arguments of the TBLT detractors. If you are going to criticize the underpinnings of TBLT you should get your facts straight. Long's adroit arguments leave no margin for error. However, as much as I enjoyed the cut and thrust of the academic argument, it was his word choice that really engaged me ("a linguistic caboose," p. 8, 28). This on its own was enjoyable, but when used to reflect on the type of classroom instruction I have sadly and recently witnessed in Aboriginal schools (for more information, go to Google and search for *Direct Instruction*) describing it as "mind-numbing monotony" (of forms focused instruction) I was ready to order the t-shirt.

Hence, I was persuaded not only by the very strong theoretical underpinnings of this article, but also because this piece captured and reflected my own pedagogical experiences. As an ESL teacher of children – when I was doing TBLT before I knew the term – I would look at the mainstream classes where my students would finally move to, what the mainstream curriculum required of them, what real life language and activities they would need to support success and, accordingly, design pedagogic tasks (a term which is also clearly elucidated in the current missive). In more recent times I have undertaken task-based needs analysis – the very foundation of TBLT (negating Swan's "straw man" argument) - to determine the real world activities that Australian Aboriginal vocationally-oriented students could be performing and then the real world language they needed to be able to do these, or, as Long suggests, I found "the real-world communicative uses to which learners will put the L2 beyond the classroom" (p. 6). The fact that my work has now been translated into appropriate classroom strategies by teachers is further, albeit anecdotal evidence, for Long's claim.

As a teacher educator I, too, have heard the criticism of TBLT: 'yes, but not for my context'. So I applaud the respect and support Long gives to the autonomy of teachers and to the role that peers can play in this process. As his pithy arguments clearly outline, the explicit pedagogical implications of TBLT is that it can be used for teaching all facets of language, and also be done in way that supports diverse learners. The need to cater to diversity in this way is a banner I proudly wave.

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

Australia, like many nations, is now invested in a national testing regime based on 'one size fits all', the washback effect of which is catastrophic for minority groups. Long's article provides a counterpoint, not only to TBLT detractors, but to educators more generally highlighting the importance of addressing the real needs of learners, and doing so in ways that reflect what we know about how languages are learned.

Oliver's comments on theory and her own practice in the Australian context are well taken. In the next section, Peter Robinson carries the discussion forward, commenting on researchability.

Theoretical, feasible, and researchable imperatives,
Peter Robinson:

Theoretically motivated, pedagogically feasible, and empirically researchable proposals for operationalizing aspects of TBLT are essential to a credible, overarching instructional design theory that brings them together and articulates them in ways that are transparent to teachers, and beneficial to language learners in instructed settings. Papers in the 2016 ARAL on TBLT address, to varying extents, these theoretical, feasible, and researchable imperatives, with some leaning towards addressing one imperative, more than the others.

The analysis by Plonsky and Kim leans more to the empirically researchable imperative. What units of analysis are most appropriate for studying the effects of task demands on learner production and interaction? They conclude that a lack of consistency across studies, e.g. of complexity, accuracy and fluency (CAF) measures, impedes efforts to draw stable conclusions about research questions addressed. Of course, research needs to be consistent to be cumulative. But this must not imply a need to be reductive in the measures we take of learner language, or interaction.

There is certainly a need to adopt comparable general measures of CAF when researching what I have called resource-dispersing dimensions of task demand, such as plus or minus planning time, or single versus dual task demands, and I have acknowledged, and borne this in mind in my own research. But on what I have called resource-directing dimensions (which differ in the conceptual domains they direct learner effort to linguistically encode) necessarily, different specific measures will be needed (see Robinson, Cadierno & Shirai, 2009).

The situation is no different in studies of the effects of child conceptual development on language performance, across a variety conceptual domains, such as reference to time, mental states of others, causality and so on. In this line of research (see Bowerman & Levinson, 2001), accuracy and complexity are broad categories of analysis (shared by developmental researchers) that have specific instantiations, and indices appropriate to particular conceptual domains.

Plonsky and Kim also recommend a move towards larger samples and fewer contrasts. In some areas of research this is feasible and so desirable, but it is less so

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

in others. For example, studying the effects of sequences of tasks, over time, in intact classroom settings and groups, limits the number of participants to normal class sizes (in the relevant educational settings), while also multiplying the number of contrasts that need to be made across each temporal task performance by any one individual in any one group, versus others, in order to capture the multiplicative effects of extended task sequences on measures of language learning and performance.

Moving on to the paper by Révész and Gurzynski-Weiss (2016), this addresses the pedagogic feasibility imperative by examining what contributes to teachers' perceptions about the difficulty of tasks for learners. This is an essential issue for pedagogically implementing tasks, sequencing of tasks in an increasing order of complexity, and for understanding whether teachers' perceptions about *in situ* task difficulty match a task designer's intentions to gradually increase their demands on learners. Do teachers perceive learner difficulty during the performance of tasks in ways anticipated by task designers, and can they reliably classify tasks as more or less difficult for learners in ways intended by task designers, drawing on a theoretically motivated metric for increasing the complexity of the cognitive processing and performative demands they make on learners (see also Baralt, Harmath-de Lemmos & Werfelli, 2014, for relevant recent research)? These are issues essential to the feasibility of task design frameworks during classroom activity.

Robinson's message, that some of the issues raised by papers in the 2016 ARAL on TBLT point to the need for further discussion and thought with the ultimate goal of theory and research in TBLT resulting in evidence-based solutions to problems in designing, implementing and assessing the effects of TBLT, is also a central theme of the final contribution to this commentary by Martin Bygate who begins by pointing out that ARAL 2016 brings together a representative showcase of developments in TBLT, notably in the growing interest in pedagogical issues (e.g. Li, Ellis & Zhu (2016), Philp & Duchesne (2016), Révész & Gurzynski-Weiss (2016), and Van den Branden (2016). Bygate's review helpfully concludes this commentary, by suggesting the practice area warrants future focus.

Tasks and classroom practice, Martin Bygate:

Rather than explore the papers in this issue of ARAL, in what follows I will take them as starting point to consider future directions. It seems to me that there is increasing need to engage with the classroom and with key stakeholders, and with this in mind there are four ways in which TBLT could usefully reposition itself:

(a) By exploring more fully what learners can be expected to learn when working on tasks;
(b) By strengthening the relationship between task-based programs and the language curriculum, including between tasks and their potential for extending learners' language;
(c) By investigating more widely the repertoire of ways in which learning can be promoted by various types of complementary activities around tasks; and finally

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

(d) By researching ways in which TBLT programs can be effectively structured together with the nature of their impact.

The first issue is a central paradox at the heart of the SLA approach to TBLT. The paradox is that on the one hand tasks are proposed because they enable the integrated learning of everything needed to successfully meet particular real world demands; but that on the other hand, the overriding concern of SLA studies is how and how far language can be acquired within the context of tasks. There is little concern (if any) in the research literature for the question, 'what do learners learn through doing those particular tasks that enables them to handle them better?' Or put another way, if particular 'tasks' are the units of learning and teaching, what do learners learn by performing them? Expressing the TBLT enterprise in this way takes us beyond a focus on lexico-grammatical features, or learners' relative complexity, accuracy or fluency of language, and towards a broader applied linguistic perspective.

Learners must also develop discourse capacities, task completion strategies, socio-cognitive pathways for accomplishing the task, and presumably various organizing schemata which they can take from the experience for future use. So the question I think we should be moving towards is: what is it "to learn to successfully perform a task"?

Language is, however, plainly important, and here there is the second paradox: the SLA literature may be defining the relationship between task and language too narrowly. There appears to be a consensus that denies the possibility of any language curriculum other than that defined by learners' real world needs and their internal syllabi (e.g. Long, 2016).

This position seems to risk becoming a shackle: teachers, learners and other stakeholders need to know TBLT will enable learners to explore, and appropriate or acquire the language curriculum as a whole. Task-based approaches in other areas of education (as in, for example the relationship between tasks and the content curriculum such as math, physics, and so on) as not problematic, but as central in justifying the approach, In primary/elementary level education, for example, an approach known as 'Mantle of the Expert' (see Heathcote & Bolton, 1995) requires children to imagine they are part of a group of experts such as scientists, archaeologists or librarians as they work to learn about their jobs and responsibilities). In other words, other areas of education see tasks as a gateway to the curriculum. In adopting a similar stance, TBLT would recognize the language curriculum as the terrain to be explored through tasks, even if without pre-conditions regarding the sequence of the various domains of language, or the way they will be 'dressed' for particular groups of learners.

Such an approach does not deny the decades of research into SLA, but is upfront about the relationship between the task-based program and the target language. Pointing only to a dossier of language-neutral real world tasks (Long, 2016; Van den Branden, 2016), to the learners' internal syllabus (Long, 2016) and to the potential of tasks to develop the fluency, accuracy and complexity of learners' language (Skehan, 2016) seems unnecessarily limiting and likely to do a disservice to our

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

rich understanding of language built up over the past 50 years. We should be able to share more openly the ways in which the use of tasks can open up access to a whole language.

Moving to the third point, the repertoire of ways in which learning can be promoted between and around task-based work, I would argue that it is time that we went beyond the hitherto limited number of theoretically-driven teaching procedures surrounding the use of tasks. Long's approach to TBLT favors the use of focus on form procedures, which although an important potential element, needs backing up with other varieties of implicit and explicit language work.

In contrast, Li, Ellis and Zhu's welcome exploration of one type of task-supported language teaching (TSLT), unfortunately seems (wrongly in my view) to imply that upfront explication of language is the only TSLT option, whereas in fact it is possible to provide implicit focused pre-teaching prior to a task (familiar to anyone trained in the International House approach to contextualized PPP since the 1970's – Samuda (personal communication). Alternatively, it is perfectly possible to conduct a TSLT lesson by starting the lesson with the task. So it seems to me that while researchers may not intend this, the SLA assumptions underpinning our research designs (here to test out the hypotheses of the single or dual models) could be leading us into limiting the pedagogical options available to teachers. There are far more options than real time focus on form within tasks, or straight explicit presentation of structures before the task than some readers might conclude from the current literature.

My final point, researching ways in which TBLT programs can be effectively structured together with the nature of their impact, is to highlight the lack of a time-wise longitudinal perspective on tasks within the literature as a whole. My own interest in task repetition originated from wanting to understand what learners took from an initial engagement with a task. But beyond this is the need for a perspective on the ways in which tasks can be used to structure whole programs, and by implication, the ways in which task-based learning might occur through and across different levels of proficiency. If TBLT is to mature and capture the commitment of mainstream language teaching, we need to be able to picture how whole courses might unfold. We can do this of course for traditional language teaching, but we seem to be some way off being able to do it for TBLT.

In conclusion then, Bygate notes that addressing the wider logic and implications of TBLT for stakeholders is a serious applied linguistics challenge that the TBLT area is increasingly in a position to engage with. This commentary has included consideration of theory, opinion, empirical work, evaluative work, and practice as these areas were represented in ARAL 2016.

REFERENCES

Andersen, R. (1983). Transfer to somewhere. In S. Gass, & L. Selinker (Eds.), *Language transfer in language learning* (pp. 177–201). Rowley, MA: Newbury House.

Baralt, M., Harmath-de Lemmos, S., & Werfelli, S. (2014). Teachers' application of the Cognition Hypothesis when lesson planning: A case study. In Baralt, M., Gilabert, R., & Robinson, P. (2014). (Eds.), *Task sequencing and instructed second language learning*. London: Bloomsbury Academic.

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press

Bowerman, M., & Levinson, S. (2001) (Eds.), *Language acquisition and conceptual development*. Cambridge, UK: Cambridge University Press.
Bryfonski, L., & McKay, T. H. (in preparation). TBLT Implementation and Evaluation: A Meta-Analysis. *Language Teaching Research*.
Bygate, M., P. Skehan, & Swain, M. (2001). *Researching pedagogic tasks: second language learning, teaching, and testing*. Harlow, UK: Longman.
Byrnes, H. (2013). Notes from the editor. *Modern Language Journal, 97*, 825–827.
Carroll, S. E. (2012). When is input salient? An exploratory study of sentence location and word length effects on input processing. *International Review of Applied Linguistics in Language Teaching, 50*, 39–67.
Heathcote, D. & Bolton, G. M. (1995). Drama for learning: Dorothy Heathcote's mantle of the expert approach to education. Portmouth, NH: Heinemann.
Li, S., Ellis, R. and Zhu, Y. (2016). Task-Based Versus Task-Supported Language Instruction: An Experimental Study, *Annual Review of Applied Linguistics*, *36*, 205–229.
Long, M.H. (2016). In Defence of Tasks and TBLT: Nonissues and Real Issues, *Annual Review of Applied Linguistics*, *36*, 5–33.
Mackey, A. (2016). New directions for the Annual Review of Applied Linguistics, *Annual Review of Applied Linguistics*, *36*, 1–4.
Philp, J., & Duchesne, S. (2016). Exploring Engagement in Tasks in the Language Classroom, *Annual Review of Applied Linguistics*, *36*, 50–72.
Plonsky, L., & Kim, Y. (2016). Task-Based Learner Production: A Substantive and Methodological Review, *Annual Review of Applied Linguistics*, *36*, 73–97.
Porte, G. K. (2012). *Replication research in applied linguistics*. New York: Cambridge University Press.
Révész, A., & Gurzynski-Weiss, L. (2016). Teachers' Perspectives on Second Language Task Difficulty: Insights from Think-Alouds and Eye Tracking, *Annual Review of Applied Linguistics*, *36*, 182–204.
Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics, 30*, 533–554.
Skehan, P. (2016). Tasks Versus Conditions: Two Perspectives on Task Research and Their Implications for Pedagogy, *Annual Review of Applied Linguistics*, *36*, 34–49.
Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Learning, 44*, 307–336.
Van den Branden, K. (2016). The Role of Teachers in Task-Based Language Education, *Annual Review of Applied Linguistics*, *36*, 164–181.

https://doi.org/10.1017/S0267190517000010 Published online by Cambridge University Press