Complexity, accuracy and fluency in second language acquisition

Housen, A.; Kuiken, F.

Link to publication

Complexity, Accuracy and Fluency in Second Language Acquisition

[1]ALEX HOUSEN AND [2]FOLKERT KUIKEN

[1]Vrije Universiteit Brussel, [2]Universiteit van Amsterdam

INTRODUCTION

This special issue addresses a general question that is at the heart of much research in applied linguistics and second language acquisition (SLA): What makes a second or foreign language (L2) user, or a native speaker for that matter, a more or less proficient language user?

Many researchers and language practicioners believe that the constructs of L2 performance and L2 proficiency are multi-componential in nature , and that their principal dimensions can be adequately, and comprehensively, captured by the notions of *complexity*, *accuracy* and *fluency* (e.g. Skehan 1998; Ellis 2003, 2008; Ellis and Barkhuizen 2005). As such, complexity, accuracy and fluency (henceforth *CAF*) have figured as major research variables in applied linguistic research. CAF have been used both as performance descriptors for the oral and written assessment of language learners as well as indicators of learners' proficiency underlying their performance; they have also been used for measuring progress in language learning.

A review of the literature suggests that the origins of this triad lie in research on L2 pedagogy where in the 1980s a distinction was made between *fluent* versus *accurate* L2 usage to investigate the development of oral L2 proficiency in classroom contexts. One of the first to use this dichotomy was Brumfit (1984), who distinguished between *fluency*-oriented activities, which foster spontaneous oral L2 production, and *accuracy*-oriented activities, which focus on linguistic form and on the controlled production of grammatically correct linguistic structures in the L2 (cf. also Hammerly 1991).

The third component of the triad, complexity, was added in the 1990s, following Skehan

(1989) who proposed an L2 model which for the first time included CAF as the three principal proficiency dimensions. In the 1990s the three dimensions were also given their traditional working definitions, which are still used today. Complexity has thus been commonly characterized as '[t]he extent to which the language produced in performing a task is elaborate and varied' (Ellis 2003: 340), accuracy as the ability to produce error-free speech, and fluency as the ability to process the L2 with 'native-like rapidity' (Lennon 1990: 390) or 'the extent to which the language produced in performing a task manifests pausing, hesitation, or reformulation' (Ellis 2003: 342).

CAF in SLA research

Since the 1990s these three concepts have appeared predominantly, and prominently, as *dependent variables* in SLA research**.** Examples include studies of the effects on L2 acquisition of age, instruction, individuality features, task type, as well as studies on the effects of learning context (e.g. Bygate 1999; Collentine 2004; Derwing and Rossiter 2003; Skehan and Foster 1999; Freed 1995; Freed, Segalowitz and Dewey 2004; Kuiken and Vedder 2007; Muñoz 2006; Spada and Tomita 2007; Yuan and Ellis 2003). From this diverse body of research, CAF emerge as distinct components of L2 performance and L2 proficiency which can be separately measured and which may be variably manifested under varying conditions of L2 use, and which may be differentially developed by different types of learners under different learning conditions.

From the mid-1990s onwards, inspired by advances in cognitive psychology and psycholinguistics (cf. Anderson 1993; Levelt 1989), CAF have also increasingly figured as the *primary* foci or even as the *independent* variables of investigation in SLA (e.g. Guillot 1999; Hilton, 2008; Housen, Pierrard and Van Daele 2005; Larsen-Freeman 2006; Lennon 2000; Riggenbach 2000; Robinson 2001; Segalowitz 2007; Skehan 1998; Skehan and Foster 2007; Tonkyn 2007; Towell 2007; Towell and Dewaele 2005; Tavakoli and Skehan 2005; Van Daele,

Housen and Pierrard 2007). Here CAF emerge as principal epiphenomena of the psycholinguistic mechanisms and processes underlying the acquisition, representation and processing of L2 knowledge. There is some evidence to suggest that complexity and accuracy are primarily linked to the current state of the learner's (partly declarative, explicit and partly procedural, implicit) interlanguage knowledge (L2 rules and lexico-formulaic knowledge) whereby complexity is viewed as 'the scope of expanding or restructured second language knowledge' and accuracy as 'the conformity of second language knowledge to target language norms' (Wolfe-Quintero *et al*. 1998: 4). Thus, complexity and accuracy are seen as relating primarily to L2 knowledge representation and to the level of analysis of internalized linguistic information. In contrast, fluency is primarily related to learners' control over their linguistic L2 knowledge, as reflected in the speed and ease with which they access relevant L2 information to communicate meanings in real time, with 'control improv[ing] as the learner automatizes the process of gaining access' (Wolfe-Quintero *et al*. 1998: 4).

Defining CAF

In spite of the long research interest in CAF, none of these three constructs is uncontroversial and many questions remain, including such fundamental questions as how complexity, accuracy and fluency should be defined as constructs. Despite the belief that we share a common definition of CAF as researchers and language teachers, there is evidence that agreement cannot be taken for granted and that various definitions and interpretations coexist. *Accuracy* (or *correctness*) is probably the oldest, most transparent and most consistent construct of the triad, referring to the degree of deviancy from a particular norm (Hammerly 1991; Wolfe-Quintero *et al*. 1998). Deviations from the norm are usually characterized as *errors*. Straightforward though this characterization may seem, it raises the thorny issue of criteria for evaluating accuracy and identifying errors, including whether these criteria should be tuned to prescriptive standard norms

3

(as embodied by an ideal native speaker of the target language) or to non-standard and even non-native usages acceptable in some social contexts or in some communities (Ellis 2008; James 1998; Polio 1997).

There is not the same amount of (relative) denotative congruence in the applied linguistics community with regard to fluency and complexity as there is with regard to accuracy. Historically, and in lay usage, *fluency* typically refers to a person's general language proficiency, particularly as characterized by perceptions of ease, eloquence and 'smoothness' of speech or writing (Chambers 1997; Freed 2000; Guillot 1999; Hilton 2008; Lennon 1990; Koponen and Riggenbach 2000). Language researchers for their part have mainly analyzed oral production data to determine exactly which quantifiable linguistic phenomena contribute to fluency in L2 speech (e.g. Lennon 1990; Kormos and Dénes 2004; Cucchiarini, Strik and Boves 2002; Towell, Hawkins and Bazergui 1996). This research suggests that speech fluency is a multi-componential construct in which different sub-dimensions can be distinguished, such as speed fluency (rate and density of delivery), breakdown fluency (number, length and distribution of pauses in speech) and repair fluency (number of false starts and repetitions) (Tavakoli and Skehan 2005).

As befits the term, *complexity* is the most complex, ambiguous and least understood dimension of the CAF triad. For a start, the term is used in the SLA literature to refer both to properties of language task (*task complexity*) and to properties of L2 performance and proficiency (*L2 complexity*) (e.g., Robinson 2001; Skehan 2001). *L2 complexity* in turn has been interpreted in at least two different ways: as *cognitive complexity* and as *linguistic complexity* (DeKeyser 2008; Housen, Pierrard and Van Daele 2005; Williams and Evans 1998). Both types of complexity in essence refer to properties of language features (items, patterns, structures, rules) or (sub)systems (phonological, morphological, syntactic, lexical) thereof. However, whereas cognitive complexity is defined from the perspective of the L2 learner-user, linguistic complexity is defined from the perspective of the L2 system or the L2 features. *Cognitive complexity* (or

4

difficulty) refers to the relative difficulty with which language features are processed in L2 performance and acquisition. The cognitive complexity of an L2 feature is a variable property which is determined both by subjective, learner-dependent factors (e.g. aptitude, memory span, motivation, L1 background) as well as by more objective factors, such as its input saliency or its inherent *linguistic complexity*. Thus, *cognitive complexity* is a broader notion than *linguistic complexity*, which is one of the (many) factors that may (but need not) contribute to learning or processing difficulty.

Linguistic complexity, in turn, has been thought of in at least two different ways: as a dynamic property of the learner's interlanguage system at large and as a more stable property of the individual linguistic elements that make up the interlanguage system. Accordingly, when considered at the level of the learner's interlanguage system, linguistic complexity has been commonly interpreted as the *size*, *elaborateness, richness* and *diversity* of the learner's linguistic L2 system. When considered at the level of the individual features themselves, one could speak of *structural complexity*, which itself can be further broken down into the *formal* and the *functional* complexity of an L2 feature (DeKeyser 1998; Williams and Evans 1988; Housen, Pierrard and Van Daele 2005).

Operationalizing and measuring CAF

Clearly, then, accuracy and particularly fluency and complexity are multifaceted and multidimensional concepts. Related to the problems of constructed validity discussed above (i.e. the fact that CAF lack appropriate definitions supported by theories of linguistics and language learning), there are also problems concerning their operationalization, that is, how CAF can be validly, reliably and efficiently measured. CAF have been evaluated across various language domains by means of a wide variety of tools, ranging from holistic and subjective ratings by lay or expert judges, to quantifiable measures (frequencies, ratios, formulas) of general or specific

5

linguistic properties of L2 production so as to obtain more precise and objective accounts of an L2 learner's level within each (sub-)dimension of proficiency (e.g. range of word types and proportion of subordinate clauses for lexical and syntactic complexity, number and type of errors for accuracy, number of syllables and pauses for fluency; for inventories of CAF measures, see Ellis and Barkhuizen 2005; Iwashita, Brown, McNamara and O'Hagan 2008; Polio 2001; Wolfe-Quintero *et al*. 1998). However, critical surveys of the available tools and metrics for gauging CAF have revealed various problems, both in terms of the analytic challenges which they present and in terms of their reliability, validity and sensitivity (Norris and Ortega 2003; Ortega 2003; Polio 1997, 2001; Wolfe-Quintero *et al*. 1998). Also the (cor)relation between holistic and objective measures of CAF, and between general and more specific, developmentally-motivated measures, does not appear to be straightforward (e.g. Halleck 1995; Skehan 2003; Robinson and N. Ellis 2008).

Interaction of CAF components

Another point of discussion concerns the question to what extent these three dimensions are in(ter)dependent in L2 performance and L2 development (Ellis 1994, 2008; Skehan 1998; Robinson 2001; Towell 2007). For instance, according to Ellis, increase in fluency in L2 acquisition may occur at the expense of development of accuracy and complexity due to the differential development of knowledge analysis and knowledge automatization in L2 acquisition and the ways in which different forms of implicit and explicit knowledge influence the acquisition process. The differential evolution of fluency, accuracy and complexity would furthermore be caused by the fact that 'the psycholinguistic processes involved in *using* L2 knowledge are distinct from *acquiring* new knowledge. To acquire the learner must attend consciously to the input and, perhaps also, make efforts to monitor output, but doing so may interfere with fluent reception and production' (Ellis 1994: 107). Researchers who subscribe to

the view that the human attention mechanism and processing capacity are limited (e.g. Bygate 1999; Skehan 1998; Skehan and Foster 1999) also see fluency as an aspect of L2 production which competes for attentional resources with accuracy, while accuracy in turn competes with complexity. Learners may focus (consciously or subconsciously) on one of the three dimensions to the detriment of the other two. A different view is proposed by Robinson (2001, 2003) who claims that learners can simultaneously access multiple and non-competitional attentional pools; as a result manipulating task complexity by increasing the cognitive demands of a task can lead to simultaneous improvement of complexity and accuracy.

OVERVIEW OF THE VOLUME

As the above discussion demonstrates, many challenges remain in attempting to understand the nature and role of CAF in L2 use, L2 acquisition and in L2 research. But despite these challenges, complexity, accuracy and fluency are concepts that are still widely used to evaluate L2 learners, both in SLA research as in L2 education contexts. We therefore thought it timely to take stock of what L2 research on CAF has brought us so far and in which directions future research could or should develop. With this broad goal in mind, four central articles were invited (by Rod Ellis; Peter Skehan; John Norris and Lourdes Ortega; Peter Robinson, Teresa Cadierno and Yasuhiro Shirai), and two commentary articles were commissioned (by Diane Larsen-Freeman and Gabriele Pallotti).

Controversial issues

The following issues were offered to the contributors as guidelines for reflection and discussion:

1. *The constructs of CAF: definition, theoretical base and scope*

Exactly what is meant by complexity, accuracy and fluency, i.e. how can they be defined as constructs? To what extent do CAF adequately and exhaustively capture all relevant aspects and

7

dimensions of L2 performance and L2 proficiency? To what extent are the three constructs themselves multi-componential? How do they manifest themselves in the various domains of language (e.g. phonology and prosody, lexis, morphology, syntax)? How do they relate to theoretical models of L2 competence, L2 proficiency and L2 processing? And how do CAF relate to L2 development (i.e. are CAF valid indicators of language development)?

2. *Operationalization and measurement of CAF*

How can the three constructs best be operationalized as components of L2 performance and L2 proficiency in a straightforward, objective and non-intuitive way in empirical research designs? How can they be most adequately (i.e. validly, reliably and practically) measured?

3. *Interdependency of the CAF component*s

To what extent are the three CAF components independent of one another in either L2 performance, L2 proficiency and L2 development? To what extent can they be measured separately?

4. *Underlying correlates of CAF*

What are the underlying linguistic, cognitive and psycholinguistic correlates of CAF? How do the three constructs relate to a learner's knowledge bases (e.g. implicit-explicit, declarative-procedural), memory stores (working, short-term or long-term), and processing mechanisms and learning processes (e.g. attention, automatization, proceduralization)?

5. *External factors that influence CAF*

Which external factors can influence the manifestation and development of CAF in L2 learning and use, such as, for example characteristics of language tasks (e.g. type and amount of planning), personality and socio-psychological features of the L2 learner (e.g. degree of extraversion, language anxiety, motivation, language aptitude), and features of pedagogic intervention (e.g. what types of instruction are effective for developing each of these dimensions within a classroom context?)

The contributions to this special issue all explicitly focus on either one, two or all three of the CAF constructs in relation to one or several of the five issues listed above, which in some cases are illustrated with new empirical research. We will now present a short overview of the topics and questions that are raised by the authors in the four central articles and in the two commentaries.

Ellis

The first article by Rod Ellis addresses the role and effects of one type of external factor, *planning*, on CAF in L2 performance and L2 acquisition. Ellis first presents a state-of-art/comprehensive survey of the research on planning. Three types of planning seem to be relevant with respect to CAF: rehearsal, strategic planning and within-task planning. Ellis concludes that all three types of planning have a beneficial effect on fluency, but the results for complexity and accuracy are more mixed, reflecting both the type of planning and also the mediating role of various other external factors, including task design, implementation variables and individual difference factors.

Ellis then provides a theoretical account for the role of planning in L2 performance in terms of Levelt's (1989) model of speech production and the distinction between implicit and explicit L2 knowledge. Rehearsal provides an opportunity for learners to attend to all three components in Levelt's model – conceptualization, formulation and articulation – and thus benefits all three dimensions of L2 production. According to the author, strategic planning assists conceptualization in particular and thus contributes to greater message complexity and also to enhanced fluency. Unpressured within-task planning eases formulation and also affords time for monitoring, that is, for using explicit L2 knowledge; in this way accuracy increases.

Skehan

The second article, by Peter Shehan, addresses the issue of operationalization and measurement of CAF. Skehan claims that fluency needs to be rethought if it is to be measured effectively. In addition he argues that CAF measures need to be supplemented by measures of lexical use. Not only because empirical evidence suggests that the latter is a separate aspect of overall performance, but also because lexical access and retrieval figure prominently in all models of speech production. Skehan also points to the lack of native speaker data in CAF research. Such data are of crucial importance, as they constitute a baseline along which L2 learners can be compared. Skehan presents a number of empirical studies in which, for identical tasks and similar task conditions, both native and non-native participants are involved, and for which measures of complexity, accuracy (for non-native speakers only), fluency, and lexis were obtained. Results suggest that the difference between native and non-native performance on tasks is related more to aspects of fluency and lexis than to the grammatical complexity of the language produced. Regarding fluency, the major difference between the two groups is the pattern of pause locations, in that native speakers use end-of-clause points for more effective, listener-friendly pausing, pausing there slightly more often albeit for shorter periods, while non-natives pause more mid-clause. Lexical performance is noticeably different between the two groups, both in terms of lexical density and of lexical variety (i.e. the use of less frequent words). Especially interesting is the difference in disruptiveness for fluency of the use of less frequent words, as non-natives are derailed in speech planning when they are pushed to use such words more because of task demands.

Skehan also considers the issue of interdependency between CAF measures; in particular between accuracy and complexity, since positive correlations between these two aspects have been less common in the literature. In order to account for these correlations. Skehan explores rival claims from his own Trade-off Hypothesis and Robinson's Cognition Hypothesis. Skehan argues that such joint raised performance in accuracy and complexity is not a function of task

difficulty (as Robinson's Cognition Hypothesis would predict) but, rather, that it reflects the joint operation of separate task and task condition factors. Like Ellis, Skehan tries to link the research findings to Levelt's (1989) model of speaking.

Robinson, Cadierno and Shirai

The article by Peter Robinson, Teresa Cadierno and Yasuhiro Shirai exemplifies a particularly prolific strand of empirical research on CAF, namely research on the impact of task properties on learners' L2 performance. The authors present results of two studies that measure the effects of increasing the complexity of task demands in two conceptual domains (time and motion) using specific rather than general measures of the accuracy and complexity of L2 speech production. The studies are carried out within the theoretical framework of Robinson's Cognition Hypothesis. This hypothesis claims that pedagogic tasks should be sequenced for learners in an order of increasing cognitive complexity, and that along resource-directing dimensions of task demands increasing effort at conceptualization promotes more complex and more grammaticized L2 speech production.

The specific measures used are motivated by research into the development of tense-aspect morphology for reference time, and by typological, cross-linguistic research into the use of lexicalization patterns for reference to motion. Results show that there is more complex, developmentally advanced use of tense-aspect morphology on conceptually demanding tasks compared to less demanding tasks, and a trend to more accurate, target-like use of lexicalization patterns for referring to motion on complex tasks. By using specific measures of complexity and accuracy (alongside general measures), these authors address the issue of measurement of CAF in their contribution. They contrast the effectiveness of these conceptually specific metrics with the general metrics for assessing task-based language production used in previous studies, and argue for the use of both. In addition, Robinson, Cadierno and Shirai also argue for a higher sensitivity

of the specific measures which are used in order to gauge cognitive processing effects on L2 speech production along selected dimensions of task complexity.

Norris and Ortega

The article by John Norris and Lourdes Ortega addresses the crucial issue of the operationalization and measurement of CAF. They critically examine current practices in the measurement of complexity, accuracy, and fluency in L2 production to illustrate the need for what they call more *organic* and *sustainable* measurement practices. Building from the case of syntactic complexity, they point to impoverished operationalizations of multi-dimensional CAF constructs and the lack of attention to CAF as a dynamic and inter-related set of constantly changing sub-systems. They observe a disjuncture among the theoretical claims researchers make, the definition of the constructs that they attempt to measure, and the grain size and focus of the operationalizations via which measurement happens. Furthermore they question current reasoning, under which a linear or co-linear trajectory of greater accuracy, fluency, and complexity is expected. Instead they want to consider measurement demands that stem from a dynamic, variable, and non-linear view of L2 development. They therefore call for a closer relation between theory and measurement and argue for a more central role for multi-dimensionality, dynamicity, variability, and non-linearity in future CAF research.

This overview of the four central articles in this volume shows that the authors approach CAF from various perspectives, focus on different issues and investigate distinct research topics. What they share is their desire to build further on the results to date. This is where the commentaries by Diane Larsen-Freeman and Gabriele Pallotti come in.

Larsen-Freeman

Larsen-Freeman starts by reminding us of the fact that, historically, CAF research has come out

of the search for an L2 developmental index. The big challenge has always been how to

operationalize CAF. According to Larsen-Freeman the measures we have been using to date may

be too blunt and not suitable because we may not have been looking at the right things in the

right places. She therefore seconds Robinson, Cadierno and Shirai's suggestion not to stick to

general measures, but to use more specific measures and to look at more detailed aspects of

performance. She further points out that the operationalization and measurement issue is

complicated by the interdependency of the CAF components. As mentioned by some of the

authors in this volume, there is an increasing amount of evidence, that complexity, accuracy and

fluency do not operate in complete independence form each other, and that findings obtained by

CAF measures depend on the participants involved and on the context in which the data have

been collected. For those reasons Larsen-Freeman does not expect much from studying the CAF

components one by one to see what effect they have on learner performance in a linear causal

way. In her view such a reductionist approach does little to advance our understanding, as we risk

ignoring their mutual interaction. Instead, we should try to capture the development of multiple

sub-systems over time, and in relation to each other. With reference to Wolfe-Quintero *et al*.

(1998) who have demonstrated that many, if not all, aspects of language development are non-

linear, Larsen-Freeman calls for a broader conceptual framework and for more longitudinal and

non-linear research, in which difference and variation occupy a central role. She considers a

dynamic or complex systems theory, in which more socially-oriented measures of development

are employed as the best candidates for such a framework.

Pallotti

Pallotti starts by signaling some definitional and operationalizational problems of CAF

constructs. As an example of an unresolved question in this area he opposes Skehan – who

doubts whether lexical and syntactic complexity are 'different aspects of the same performance area' or two separate areas – to Norris and Ortega, who consider syntactic complexity to be a multi-dimensional construct with several sub-constructs. Pallotti considers CAF to be a good starting point for describing linguistic performance, but they do not constitute a theory or a research program in themselves. He emphasizes that a clear distinction should be made between on the one hand CAF, referring to the properties of language performance as a product, and linguistic development on the other, referring to a process, with its sub-dimensions such as route and rate.

In line with Larsen-Freeman, and with specific reference to the contributions by Norris and Ortega and Robinson *et al.*, Pallotti welcomes the use of specific measures in addition to the more general ones, as one cannot expect that '*all* sorts of task complexification lead to higher complexity of *any* linguistic feature.' He questions, however, what the use of specific measures may contribute to theorizing about CAF. Although by using specific measures the relationship between task difficulty and linguistic complexity may become more reliable, 'discovering such relationships looks more like validating the tasks as elicitation procedures for specific linguistic features than like confirmations of general theories about speech production.'

Pallotti agrees with Larsen-Freeman's call for a more central role of non-linearity in L2 acquisition. He illustrates this by referring to Norris and Ortega's example that syntactic complexity as measured by means of a subordination ratio may not always increase linearly, but that syntactic complexity may grow in other ways, for example by phrasal and clausal complexification. And also for accuracy it is not always the case that 'more is better'. He does not, however, embrace Larsen-Freeman's idea that variation should move to the front of CAF research. This is what he calls 'the necessary variation fallacy': research should not only be concerned with variations and differences, but also with constants and similarities. Instead he argues that *adequacy* be included as a separate dimension of L2 production and proficiency,

alongside complexity, accuracy and fluency. In this respect, he points to an interesting paradox, namely that most studies have assessed CAF within the contexts of communicative tasks, but very few discuss how the communication unfolded and whether it was successful in achieving its goals. Adequacy, the appropriateness to communicative goals and situations, should be seen as both an independent construct based on task success and as a way of interpreting CAF measures.

The debate continues

The work presented in this special issue on CAF thus presents new perspectives on the empirical study of CAF in SLA, as well as raising important theoretical and methodological questions. Crucial to these questions is the need to further refine testing instruments and measures and better define the constructs to be measured as well as learner external and internal factors surrounding, affecting and perhaps impeding the development or manifestation of CAF in L2 performance. These are all issues for further exploration. We hope that the articles in this issue will contribute to further debate on CAF, shedding light on existing theoretical and methodological issues in the field as well as opening up new areas of inquiry.

REFERENCES

Anderson, J. R. 1993. *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum.

Brumfit, C. J. 1984. *Communicative Methodology in Language Teaching.* Cambridge University Press.

Bygate, M. 1999. 'Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks.' *Language Teaching Research* 3: 185–214.

Chambers, F. 1997. 'What do we mean by oral fluency?' *System* 25: 535-544.

Collentine, J. 2004. 'The effects of learning contexts on morphosyntactic and lexical development.' *Studies in Second Language Acquisition 26/2*: 227-248.

Cucchiarini, C., H. Strik and **L. Boves.** 2002. 'Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech.' *The Journal of the Acoustical Society of America* 111/6: 2862-2873.

DeKeyser, R. 2008. 'The complexities of defining complexity'. Paper presented at AAAL 2008, April 1st 2008, Washington D.C.

Derwing, M. and **J. Rossiter.** 2003. 'The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech.' *Applied Language Learning* 13: 1–18.

Doughty, C. and **M. Long** (eds.). 2003. *The Handbook of Second Language Acquisition*. Oxford: Blackwell.

Ellis, R. 1994. 'A theory of instructed second language acquisition' in Ellis, N. (ed.): *Implicit and Explicit Learning of Language*. London: Academic Press.

Ellis, R. 2003. *Task-based Language Learning and Teaching.* Oxford University Press.

Ellis, R. 2008. *The Study of Second Language Acquisition (2ⁿᵈ edition).* Oxford University Press.

Ellis, R. and **G. Barkhuizen**. 2005. *Analysing Learner Language*. Oxford University Press.

Freed, B. 1995. 'What makes us think that students who study abroad become fluent?' in Freed, B. (ed.): *Second Language Acquisition in a Study Abroad Context*. Amsterdam: Benjamins.

Freed, B. 2000. 'Is fluency, like beauty, the eyes, of the beholder?' in Riggenbach (ed.).

Freed, B., **N. Segalowitz** and **D. Dewey**. 2004. 'Context of learning and second language fluency in French.' *Studies in Second Language Acquisition* 26/2: 275-301.

Guillot, M.-N. 1999. *Fluency and its Teaching*. Multilingual Matters: Clevedon.

Halleck, G. B. 1995. 'Assessing oral proficiency: A comparison of holistic and objective measures.' *Modern Language Journal* 79/2: 223-234.

Hilton, H. 2008. 'The link between vocabulary knowledge and spoken L2 fluency.' *Language Learning Journal* 36/2: 153-166.

Housen, A., S. Van Daele and **M. Pierrard**. 2005. 'Rule complexity and the effectiveness of explicit grammar instruction' in Housen, A. and M. Pierrard (eds.): *Investigations in Instructed Second Language Acquistion*. Berlin: Mouton de Gruyter.

Iwashita, N., A. Brown, T. McNamara and **S. O'Hagan**. 2008. 'Assessed Levels of Second Language Speaking Proficiency: How Distinct?' *Applied Linguistics* 29: 24-49.

James, C. 1998. *Errors in Language Learning and Use: Exploring* Error Analysis. London: Longman.

Koponen, M. and **H. Riggenbach**. 2000. 'Overview: Varying Perspectives on Fluency', in Riggenbach (ed.).

Kormos, J. and **M. Dénes**. 2004. 'Exploring measures and perceptions of fluency in the speech of second language learners.' *System* 32/2: 145 - 164.

Kuiken, F. and **I. Vedder**. 2007. 'Task complexity and measures of linguistic performance in L2 writing.' *International Review of Applied Linguistics in Language Teaching* 45/3: 261-284.

Larsen-Freeman, D. 2006. 'The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English.' *Applied Linguistics* 27/4: 590-619.

Lennon, P. 1990. 'Investigating fluency in EFL: A quantative approach.' *Language Learning* 40: 387-417.

Lennon, P. 2000. 'The lexical element in spoken second language fluency' in Riggenbach (ed.).

Levelt, W. 1989. *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Muñoz, C. (ed.). 2006. Age *and the Rate of Foreign Language Learning*. Clevedon: Multilingual Matters.

Norris, J. and **L. Ortega.** 2003. 'Defining and measuring SLA' in Doughty, C. and M. Long (eds).

O'Brien, I., N. Segalowitz, B. Freed and **J. Collentine**. 2007. 'Phonological memory predicts second language oral fluency gains in adults.' *Studies in Second Language Acquisition* 29: 557-581.

Ortega, L. 2003. 'Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing.' *Applied Linguistics* 24: 492-518.

Polio, C. 1997. 'Measures of linguistic accuracy in second language writing research.' *Language Learning* 47: 101-143.

Polio, C. 2001. 'Research methodology in second language writing research: The case of text-based studies' in Silva, T. and P. K. Matsuda (eds.): *On Second Language Writing*. Mahwah, NJ: Lawrence Erlbaum.

Riggenbach, H. (ed.). 2000. *Perspectives on Fluency.* Ann Arbor: The University of Michigan Press.

Robinson, P. 2001. 'Task complexity, task difficulty, and task production: exploring interactions in a componential framework.' *Applied Linguistics* 22: 27-57.

Robinson, P. 2003. 'Attention and memory during SLA' in Doughty, C. and M. Long (eds).

Robinson, P. and **N. C. Ellis**. 2008. 'Conclusion: Cognitive linguistics, second language acquisition and instruction: Issues for research' in Robinson, P. and N. C. Ellis (eds.): *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge.

Segalowitz, N. 2007. 'Access fluidity, attention control, and the acquisition of fluency in a second language.' *TESOL Quarterly* 41: 181-186.

Skehan, P. 1989. *Individual Differences in Second Language Learning*. London: Edward Arnold.

Skehan, P. 2001. 'Tasks and language performance' in Bygate M., P. Skehan and M. Swain (eds.): *Research Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. London: Longman.

Skehan, **P**. 1998. *A Cognitive Approach to Language Learning*. Oxford University Press.

Skehan, P. 2003. 'Task based instruction.' *Language Teaching* 36: 1-14.

Skehan, P. and **P. Foster**. 1999. 'The influence of task structure and processing conditions on narrative retellings.' *Language Learning* 49: 93-120.

Skehan, P. and **P. Foster**. 2007. 'Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing Research' in Van Daele *et al*. (eds.).

Spada, N. and **Y. Tomita**. 2007. 'The complexities of selecting complex (& simple) forms in instructed SLA research' in Van Daele *et al.* (eds.).

Tavakoli, P. and **Skehan, P.** 2005. 'Strategic planning, task structure, and performance testing' in R. Ellis (ed.): *Planning and Task Performance in a Second Language*. Amsterdam: Benjamins.

Tonkyn, A. 2007. 'Short-term changes in complexity, accuracy and fluency: Developing progress-sensitive proficiency tests' in Van Daele *et al*. (eds.).

Towell, R. 2007. 'Complexity, accuracy and fluency in second language acquisition research' in Van Daele et al. (eds.).

Towell, R., R. Hawkins and **N. Bazergui**. 1996. 'The development of fluency in advanced learners of French.' *Applied Linguistics* 17: 84-119.

Towell, R. and **J.-M. Dewaele**. 2005. 'The role of psycholinguistic factors in the development of fluency amongst advanced learners of French.' in Dewaele, J.-M. (ed.): *Focus on French as a Foreign Language: Multidisciplinary Approaches.* Clevedon: Multilingual Matters.

Van Daele, S., A. Housen and **M. Pierrard**. 2007. 'Psycholinguistic mechanisms underlying the manifestation and development of 2nd language complexity, accuracy and fluency' in Van Daele *et al*. (eds.).

Van Daele, S., **A. Housen**, **F. Kuiken**, **M Pierrard** and **I. Vedder**. (eds). 2007. *Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching*. Brussels: Contactforum.

Vermeer, A. 2000. 'Coming to grips with lexical richness in spontaneous speech data.' *Language Testing* 17/1: 65-83.

Williams, J. and **J. Evans**. (1998). 'What kind of focus and on which forms?' in Doughty, C. and J. Williams (eds.): *Focus on Form in Classroom Second LanguageAacquisition*. Cambridge University Press.

Wolfe-Quintero, K., S. Inagaki and **H.-Y. Kim**. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Yuan, F. and **R. Ellis**. 2003. 'The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 oral production.' *Applied Linguistics* 24: 1-27.