Watching Ladders or the Weather: How Should Second Language Teachers Understand and Evaluate Their Students' Writing Progress?

Linda Bilton

Department of English Language and Literature, University of Bahrain P.O. Box 32038, Kingdom of Bahrain

Abstract

This paper describes the emergence of one teacher-researcher's understanding of second language development and the dilemma it poses. Two extant metaphors of language acquisition are applied to a piece of action research involving the assessment of journal writing at the University of Bahrain. The results raise questions about the usefulness of second language acquisition research in the present climate of higher education.

Keywords: Second language writing, journal writing; language development, measurement and evaluation, acquisition research.

1.   Introduction

Theories of Second Language Acquisition (SLA) continue to be influenced by research into child development and learning, which may be viewed either as a 'developmental ladder' or as 'emergent' (Fischer et al., 2003), rather like rainfall from a weather system. 'Emergentism' derives from complexity theory, which describes phenomena in terms of complex, dynamic systems rather than individual elements. A system is not self-existent but emerges from its components. Examples would be a flock of birds, a weather system or the ecosystem of a desert. It has no identifiable cause and its behaviour is unpredictable, but it can be described as it happens and explained retrospectively. It is unstable by virtue of the continual interaction of its components and also external influences - an aeroplane might hit the flock, the weather system might pass over a warm ocean current, or there may be an oil spill. It is complex because each component may in turn be a complex system, revealing a nesting of interacting and interdependent sub-systems. Thus language learning is seen as a complex system involving an individual's brain system, the individual learner, the context of learning, his/her socio-cultural group and the speech community. All of these subsystems should be accounted for in a description of language learning as they interact (Larsen-Freeman and Cameron 2008).

Applied to SLA, the two metaphors encompass opposing sets of postulates and assumptions which are comprehensively expounded by Larsen-Freeman (2006).and are

summarised as follows. Researchers working with the ladder metaphor assume fixed and homogenous native and Target Languages (TLs) and view second language acquisition as increasing conformity to TL norms. This approximation proceeds linearly through discrete stages and all learners move consistently, although not necessarily at the same speed. Language is exclusively a cognitive resource and progress of acquisition can be adequately monitored along one dimension, for example, fluency, by examining one subsystem, for example, lexis. In contrast, 'emergentists' claim that languages are not fixed, but dynamic systems which change with use. Acquisition is not linear and there are no discrete stages of acquisition, only periods where certain language forms dominate temporarily. Language is both a cognitive and a social resource since it is used communicatively, and context always affects performance. Therefore, progress can only be monitored by considering the interaction of multiple dynamic systems at different levels moving at different speeds (Larsen-Freeman, 1997).

It should be noted that both schools assume acquisition: either that exposure to the TL plus a need to learn it (with or without teacher intervention) will produce acquisition, or that using a learner's language resources will change those resources positively over time. These assumptions are questionable. Furthermore, both approaches entail problems of measurement and evaluation. Progress is usually ascertained either by comparing an initial state with an end state, as deduced from entrance and exit tests, or by examining a series of performances between those two states for evidence of sustained change (the time-series method). Either way involves measurement, which Bachman defines as 'the process of quantifying the characteristics of persons according to explicit procedures and rules' (Bachman, 1990:18). The data for measuring language acquisition is performance. The ladder metaphor assumes a fixed, direct link between that and competence, which is disproved by the present study. However, if every performance is contingent and its language products possibly ephemeral, qua 'emergentist' theory, it matters little whether evaluation uses two or multiple tests, for acquisition as achievement remains elusive and evaluation shaky.

Teachers, however, need certainties. Most language teaching takes place in educational institutions, where particular programmes are introduced, altered or discarded according to some idea of usefulness. They are constantly being evaluated, whether formally or informally. Evaluation, the 'collection and perusal of information in order to make decisions about people' (Bachman, 1990:22), is a necessary part of teaching. Recognising the persistent instability of complex systems (Percival, 1993) may encourage teachers to tolerate learners' Inter-Language (IL), but they are forced to look for improvement in performance as evidence of learning in order to justify changes

in their teaching. Some of the tensions between examining the process of acquisition and seeking its products are illustrated in the following report written by a teacher-researcher.

2   Materials and Methods

2.1   Background

This normative study arose from a piece of action research undertaken by the author with her upper-intermediate writing class in the Department of English Language and Literature at the University of Bahrain. In order to increase motivation and, it was hoped, mastery of the written code, weekly journals were introduced over a period of nine consecutive weeks to supplement the course's text-book based compositions, which were products of a cognitive and pragmatic approach to writing. The journals were written at home over the weekend. The subjects to be written on, chosen by the teacher, were of a general and accessible nature such as 'people, gardens, monsters,' but could be treated as the writer wished. For example, she could be reflective, or compose a story or a poem or even a play. Various suggestions were discussed in class at the outset of the programme. The only constraints imposed were that a writer should not spend more than thirty minutes on the task and should write spontaneously where possible. Work would be graded for effort, not achievement, and language would not be corrected because the reader would be focusing on content. Journal writing was presented as a challenge to write interestingly and entertainingly for the reader.

Most of the students (11 out of 17) rose to this challenge and in their final review of their experience and their written texts, all of them commented favourably, even with qualifications, on journal writing. That they were confident enough to express their opinions without fear of being penalised by their teacher testified to the success of the programme, which had been devised in order to develop self-confidence as well as to create enjoyment in writing. However, the author felt that in addition to these personal benefits, it would be helpful when selling this innovation to colleagues, to point to an improvement in mastery of written English.

2.1.2   Selection of Material

Therefore, an objective study was begun of the language used by a sample of students, selected by their course-end achievement scores - which reflected various composition skills - to cover the whole range of language competence in the class. Those who had failed the course were included in the initial survey, but a close examination of the language of their journals soon

revealed a high proportion of plagiarism such as recycling of old school compositions or copying from published sources. From the successful students, those who had scored A to D, seven were selected on the basis of the highest number of submitted journals. Thus only the writing of the more motivated students was examined for evidence of language development. A total of 51 journals was ultimately examined.

2.2   Measures of Language Development

The need for an objective measure of second language development was felt as early as the 1970s, with prominent researchers like Diana Larsen-Freeman calling for 'a developmental yardstick against which global (i.e. not skill or item specific) second language proficiency could be gauged' (1983:287) and which should 'increase uniformly and linearly as learners proceed towards full acquisition of a target language'(1978:440). Whether or not acquisition proceeds uniformly, some index which moves in tandem with it is definitely useful. For example, the effect of a particular teaching context, such as the one described in this study, might be ascertained by measuring differences on a developmental measure. The most popular and effective ones have proved to be measures of fluency, and grammatical and lexical complexity (Wolfe-Quintero et al., 1998), and they have been employed here.

2.2.1   Fluency

'Fluency' is a slippery concept which can include coherence, complexity, appropriacy and creativity. Put simply, it is the ease with which a language user can retrieve the language items that s/he needs. This can vary according to context and individual ability. It has been described as 'the processing of language in real time' (Schmidt, 1982:.358), with a focus on 'the primacy of meaning' (Foster and Skehan, 1996:.304). Fluency is related to the production pressures of trying to communicate a message. The present researcher has chosen rate of production and length of production as units of fluency measures, agreeing with Wolfe-Quintero et al. that 'fluency means that more words and more structures are accessed in a limited time, whereas a lack of fluency means that only a few words or structures are accessed' (1998:.14).

2.2.2   Grammatical Complexity

Grammatical or syntactic complexity manifests in writing primarily as grammatical variation and sophistication, and its development is seen in 'progressively more elaborate language' and 'a greater variety of syntactic patterning' (Foster and Skehan, 1996:303). It reflects linguistic mastery

in so far as a variety of simple and more sophisticated structures is available and accessible to a writer (Wolfe-Quintero et al, 1998). Of course, sophistication may be variously interpreted: complex noun phrases in a simple subject-verb-object sentence are arguably more sophisticated than a bunch of dependent finite clauses. An early observer of first language writers (Hunt, 1965) noted that learners appeared to move from co-ordination to subordination to reduced phrases. This progression seems intuitively plausible and has been noted in studies of second language writing (Munroe 1975; Sharma 1980; Ishikawa 1995); but as Wolfe-Quintero et al. (1998) caution, assuming the ladder metaphor, stages of development would overlap in any sample and certain structures might be temporarily privileged. Furthermore, availability does not presuppose maximum employment, for a sophisticated writer may choose to restrict his/her range of structures for dramatic effect; for safety under test conditions; because they become locked into a pattern as they write; or because they cannot be bothered to display their repertoire. Due regard to performance factors must be paid when matching evidence of grammatical complexity against proposed indicators of language development, although over a longer period such performance anomalies might be ironed out and a trend discerned. Nevertheless, despite these important qualifications when examining data for evidence of language development, grammatical complexity of writing remains a useful measure.

2.2.3 Lexical Complexity

A similar vagueness of terminology weakens descriptions of vocabulary acquisition. Hyltenstam (1988:.71) says, uncontroversially, that 'a reasonably large lexicon' is crucial for effective communication, Proficiency necessarily includes mastery of a vocabulary resource, and the richness of a learner's lexicon reflects the stage of his/her language development. Lexical complexity is revealed by the variation (range) and sophistication (size) of a writer's productive vocabulary. Put developmentally, lexical complexity means that a wide variety of basic and sophisticated words is available and can be accessed quickly, whereas a lack of complexity means that only a narrow range of basic words is both available and accessible. Therefore, with due regard for the cautions about a writer's choice expressed in 2. 2. 2, it can be accepted that 'Learners who have more productive vocabulary items available, are able to vary their word choices more freely; consequently, larger ratios in variation and sophistication measures should reveal greater lexical proficiency' (Wolfe-Quintero et al., 1998:101).

2.2.4   Choice of Measure

In their wide-ranging meta-analysis of developmental measures, Wolfe-Quintero et al. (op. cit.) looked for measures which worked consistently with different subjects. They were all ratio measures. For fluency they were: Words per T-unit (W/T), Words per clause (W/C) and Words per Error-Free T-unit (W/EFT). For grammatical complexity, the best measures were Clauses per T-unit (C/T), Words per Clause (W/C), and Dependent Clause per Clause (DC/C). They found that the most useful measure of word variation which took into account both the length of the text and the number of word types was Arthur's (1979) formula, which divides the number of lexical word types by the square root of double the number of Lexical Word Tokens (LWT/$\sqrt{2LW}$). 'This measure captures changes in word variation from the first half of a writing class [weeks 1-4] to the second [weeks 5-8]' (Wolfe-Quintero et al., 1998.104). Lexical sophistication or size of vocabulary has been measured by dividing the number of Sophisticated Word Types by the number of Word Types (SWT/WT). This lexical complexity measure was found to be similarly significantly related to short-term change.

The present analysis uses the clause as its production unit since it is more readily recognisable and generally acceptable. Hence, fluency was measured by counting the number of Words in a Clause (W/C) and grammatical complexity by dividing the number of dependent clauses by the number of main or independent clauses (DC/C). Lexical variation was examined using Arthur's formula, but there was no time to study lexical sophistication. Wolfe-Quintero *et al.* warn that type/token ratios should only be used where there is a time or conceptual limit on production. In this study, such a limit was set by the task, 'Spend half an hour writing freely on this subject.' This is not as casual as it seems since few of our students would spend more time on a minimally graded task. Moreover, even under timed test conditions, writers may finish early or not at all.

2.3   Applying the Tools

2.3.1   Fluency

Since the chosen fluency measure was the number of Words per Finite Clause **(W/C)**, many phrases were omitted from the textual analysis. Chiefly, these were

(a)   Discourse markers, which do not belong to the sentence, for example: *Well, no doubt, in contrast, also, i.e.;*

(b)   Fragments such as *According to their agenda;*

6

(c)   Apposition: Music - *a beautiful song full of passion, a rhythm you can dance to and a harmonic passage - takes you to the world of dreams;*

(d)   Lists

(e)   Apostrophized names: *"Dakota, I think…"*

Certain problems of categorizing were encountered. Direct speech was counted separately, rather than as a clause or clauses dependent on the reporting verb. Thus:

"It will help us finish our research," (Main Clause) *she whispered* (Main Clause).

This was justified on the grounds that each sentence belonged to different levels of discourse and that it simplified the analysis. Of course, the reporting verb might have been excluded, but this would have involved exclusion of any dependent adverbial phrases such as (hypothetically): *she whispered, putting down her binoculars*, which is definitely part of the narrative.

Co-ordination also turned out not to be straightforward. Ellipsis where the same verb was left as understood, as in *Some left their traces and others a memory*, was treated as two main clauses; but in the case of different verbs with a shared object such as *We could feel and see* (X), the whole string was treated as a single clause, likewise co-ordinated complements *It was exciting and …sometimes even dangerous*. A more difficult decision involved sentences such as *I want to be the beauty girl and live a comfortable life*. Assigning two main clauses would have involved adding the verb phrase I want to, whereas just inferring a *'to'* makes this a single finite clause containing two co-ordinated verb phrases which seem to encapsulate one idea. In contrast, three parallel qualifying but elided clauses were counted separately:

If you don't water it with hope (DC), *let the sun of friendship shine through it* (DC)
and let pure love nourish it (DC), *it will dry and die* (MC)

However controversial, these decisions were applied consistently throughout the analysis.

2.3.2   Grammatical Complexity

Measuring grammatical complexity by dividing the number of dependent clauses by the number of independent or main clauses (DC/MC) proved fairly straightforward after the above decisions. A main clause was taken to include all gerund and participial phrases:

Mr L *decided to build a wall, separating himself from his neighbour (MC)).*

Occasionally, a more difficult problem of embedded clauses presented itself as in:

They brought their essential tools, knowing that the less they bring, the better they go.

Strictly taken, the whole sentence after *knowing* should be included as part of the non-embedded adverbial phrase of reason and only one unit counted, that is, one main clause. However, this

analysis was felt to misrepresent the syntactic complexity of the writer's sentence; so, a compromise was made, counting the two embedded noun clauses separately, whilst including the gerund *knowing* in the word count of the main clause.

At this juncture, it is worth recalling the caveats raised earlier concerning measuring grammatical complexity. It can be argued that a simple main clause containing complex noun and verb phrases is syntactically more complex than one main clause with several dependent clauses. In the present analysis, this sophistication is captured in the fluency count, which suggests that the two measures of fluency and grammatical complexity are more instructive when taken together.

2.3.3 Lexical Complexity

For the measure of lexical variety/richness it was necessary to establish the status of a lexical word or lexeme. All grammatical words, such as prepositions, articles, pronouns and possessive adjectives were excluded as were expletives. Also excluded were closed sets like adverbs of frequency and time *(today, last week, now, before, previously)*, intensifiers and quantifiers. In addition, sentence fragments were also disregarded to maintain consistency with the measures of fluency and grammatical complexity, but misused lexemes were included on the grounds that the writer showed knowledge of the word without mastery of its meaning. When recording types and tokens, the lexemes were categorised by word class (Noun, Verb, Adjective and Adverb) to seek patterns of use.

2.4 The Analysis

Fluency and grammatical complexity were thoroughly analysed before work was begun on lexical complexity. To begin with, the first and last journals of six students were examined. In four cases, these were Journals 1 and 9. However, data for two students began from Journal 4 only. One student (S6) was omitted from the initial analysis as her data started from Journal 5. The analysis was recursive in that every fresh problem of categorisation necessitated a review of earlier analyses. Some consistency was maintained by copying working notes from each student's analysis sheet into a dossier which was kept for later reference, but no statistical test of intra-rater reliability was made. The analyses were reviewed after18 months. It had been hoped to assess language development by comparing early and late writing; however, results from a tentative analysis of median journals cast doubt on this enterprise. So it was decided to analyse all the students' journals (51) for fluency and complexity. This final analysis was conducted 18 months after the review. The whole enterprise thus took three years, delays being due to full-time teaching commitments.

The analysis of the journals' lexical richness followed a similar procedure. The first and last journals of four students were piloted, then their median journals to check for monotonic trends. Twenty two months later, this was extended to the other students' journals, and five months thereafter the analysis of all 51 was completed.

3 Results

3.1 Fluency (W/C)

Table 1: Measures of fluency (J = journal; 164 = word count; E= essay; S = story)

Student	J1	J2	J3	J4	J5	J6	J7	J8	J9
1	5.86	8.12	7.96	6.62	7.39	6.90	6.77	7.60	6.24
	164	552	446	139	244	662	548	441	249
	E	S	S	E	S	S	S	S	E
2	6.05	7.48	9.21	5.38	6.00	6.57	6.12	7.31	7.33
	127	217	175	457	174	217	208	190	285
	E	S	S	S	E	S	E	E	S
3				6.52	7.59	8.46	6.23	8.04	6.34
				150	167	237	187	193	222
				E	S	E	E	S	E
4	6.74	6.41	5.78	5.85	6.12	6.24	6.77	7.00	5.58
	182	141	145	234	153	362	176	146	162
	E	S	S	E	E	S	E	E	E
5				6.30	7.32	7.35	8.76	7.21	6.82
				206	227	228	184	137	148
				S	S	S	E	S	S
6					9.14	7.18	6.40	7.00	6.20
					201	158	192	182	230
					E	E	E	E	S
7	7.87	7.02	5.51	5.87			6.89	6.66	7.00
	181	323	204	223			131	193	140
	E	S	S	S			E	E	E
Subject	People	Crime	Memory	Music	Garden	Sea	Mistake	Gift	Animals

Table 1 shows the ratio of words per finite clause for every journal submitted by the selected students, with word count and genre added.

3.1.1 Text length

The number of words actually written in each text has been included for convenience in this table. It serves as a rough indicator of motivation. S1 averaged 385 words per text, S2, 228 and S4, 167. The lack of data for the other students may have skewed their averages, which cluster around 200 words per text written in about 30 minutes. Of the four students who submitted their first three journals, three of them (S1, S2 and S7) show an increase then a decrease in the length of text. Since all chose to write stories for their second and third journals, genre seems not to have been a factor; and there is no means of determining whether the topic or an extraneous factor was in play. What is clear, however, is that the second half of the semester produced shorter texts from six of the seven students studied. This may have been due to a growing pressure on their study time - some journals were submitted late - caused by an increase in assessment of various forms such as second midterm tests, projects, presentations; or it may reflect a loss of interest in free writing. However, only one student (S4) admitted this, and the quality of the others' writing did not deteriorate.

3.1.2 Fluency

There is no overall increase in fluency for the period of this study: students tend to finish at about the same level of fluency with which they began. The general impression is of fluctuation. However, there are noticeable differences between students and possible patterns within each student's performance. Clearly, individual writers behave differently in regard to topic, treatment and position in the course. It should be remembered that fluency is glossed as the length of finite clauses, not ease of expression.

3.1.3 Genre

On this measure, it appears that three students (S1, S2 and S5) wrote stories more fluently than they did essays; two students (S4 and S7) wrote essays more fluently; and two students (S2 and S6) showed no link between fluency and their chosen way of writing. More data, for example, a full return of journals or writing spanning two semesters, might present a clearer picture. Meanwhile, no link between fluency and genre can be established.

3.2   Grammatical Complexity

Table 2 shows the measures of grammatical complexity for all students during the period of this study. Again, genre (Essay/Story) has been added.

Table 2: Measures of grammatical complexity with genre (E = essay; S = story)

Student	J1	J2	J3	J4	J5	J6	J7	J8	J9
1	1.55 E	0.74 S	0.33 S	1.10 E	0.38 S	0.78 S	0.37 S	0.45 S	1.50 E
2	0.91 E	0.60 S	0.19 S	0.35 S	1.07 E	0.43 S	1.27 E	0.37 E	0.44 S
3				0.77 E	0.47 S	0.65 E	0.88 E	1.67 S	1.06 E
4	0.87 E	0.38 S	0.60 S	0.48 S	0.79 E	0.57 S	0.63 E	0.31 E	0.53 E
5				0.65 E	0.41 S	0.82 S	0.75 E	0.27 S	0.83 E
6					0.38 E	0.83 E	1.00 E	1.60 E	0.12 S
7	0.77 E	0.53 S	1.06 S	0.41 S			0.46 E	0.61 E	1.22 E

There is no overall increase in grammatical complexity. The three students who returned a complete set of journals exhibit marked fluctuations between each journal but a general decrease in grammatical complexity after their first journals. This pattern, such as it is, seems to be mirrored in the incomplete sets of the other writers. Perhaps students wished to show their mettle initially but relaxed thereafter.

Although variation seems to be random and may in part have been caused by temporary factors such as enthusiasm or fatigue, treatment of topic (genre) does appear to influence complexity, with stories being on the whole syntactically simpler than the discursive or reflective essays.

3.3   Lexical Complexity

The complete findings of the analysis of the writers' lexical richness/variety using Arthur's formula are displayed in Table 3.

Table 3: Measures of lexical richness (J - journal)

Student	JI	J2	J3	J4	J5	J6	J7	J8	J9
1	4.12	7.20	6.19	4.19	5.86	6.69	6.89	7.21	4.76
2	3.62	5.61	5.40	6.64	5.03	5.03	4.40	4.99	6.66
3				4.01	4.53	5.35	4.24	4.14	4.41
4	3.87	4.61	3.64	5.19	4.47	6.25	3.92	3.86	3.67
5				4.22	4.24	4.82	4.04	3.66	3.75
6					4.32	3.98	4.21	3.75	4.79
7	3.83	5.96	3.38	4.66			3.27	3.47	3.56

Sought was a general increase in lexical richness over the nine-week period of journal writing, and any increase within that, on the assumption that an increase over two or more weeks should indicate greater confidence in writing and improved, or rediscovered, mastery of the language. Also sought were possible links between high/low performance scores and factors such as external conditions and the internal constraint of genre - how a writer chooses to treat her subject, whether discursively or as narrative. Subject itself was not considered a factor in this study as topics were very broad, writers were free to treat them as they wished and, on appeal, an alternative topic could be given.

3.3.1 Trends

There is no general increase of lexical complexity, but a noticeable decrease of lexical richness for the later journals, which corresponds to the beginning of the second half of the semester after the week-long mid-semester break.

Six of the seven students show one or more upward trends over different parts of this period. (A trend was regarded as an observable pattern over at least three consecutive assignments.) These same students all show an increase somewhere in the first half -semester (Journals 1-6), which suggests that there was an improvement in lexical mastery after all. Surprisingly, however, five of the seven students start the second half of the semester with scores lower than those attained before the break, although four of these subsequently show an increasing richness in their writing. This would support the contention that regular free writing promotes development of vocabulary.

3.3.2 Performance and External Conditions

From the results presented in Table 3, it is clear that the mid-semester break coincided with a change in most students' writing behaviour. Only S1 and S6 continued to write with the same degree of complexity as before; the others reverted to their initial level. Moreover, although most did show weekly improvement, it was not as marked as it had been earlier in the semester. This 'underperformance' may be related to the increased pressure of assessment - first and second mid-semester tests plus major assignment deadlines - which took time and, perhaps, interest away from the less motivated students.

3.3.3 Genre and Lexical Richness

A glance back at Table 3 suggests a possible link between the type of writing chosen by students and the depth of lexical complexity or richness exhibited. Broadly, 'essay' (E) was any writing in which the author addressed her reader directly and expressed opinions. In one case, an essay included a poem. 'Story' (S) covered any piece of writing which consisted of more than 50% narrative, for example, when a writer illustrated an opinion with a personal anecdote or referred to a news item at

Table 4: Lexical richness and genre

Student		Essay				Story	
	Highest		Lowest		Highest		Lowest
1	4.76		4.19		7.21		5.86
Range		0.57				1.35	
2	5.03		4.40		6.66		5.03
Range		0.63				1.63	
3	4.41		4.01		5.55		4.14
Range		0.40				1.41	
4	4.47		3.67		6.25		3.64
Range		0.80				2.61	
5	4.22		3.75		4.82		3.66
Range		0.47				1.16	
6	4.32		3.75		4.79		-----
Range		0.57				-----	
7	3.83		3.27		5.96		3.38
Range		0.56				2.58	

length. In order to compare the lexical richness of these two types of writing, each student's highest and lowest scores for essay and story were tabulated (Table 4)

It can be seen that for every student the richest story is lexically richer than the richest essay, in most cases, considerably richer. Similarly, the poorest essay was generally poorer than the poorest story, although two students (S4, S5) behaved differently. Overall, however, students' stories were lexically richer than their essays. Students seemed to be using safe, simple vocabulary to express their opinions, whereas stories appeared to give them more freedom to explore and utilize their stock of words, probably because stories lend themselves to description, action and change of scene.

4   Discussion of Findings

4.1   Language Development

From the start of the 'free writing' project to its end, a matter of nine consecutive weeks, there appears to be no overall increase in language development in terms of fluency, grammatical complexity or lexical richness. Rather, there is marked fluctuation from one assignment to the next. There are also noticeable differences between the performances of individual writers, with detectable short-term patterns within each writer's progression. Nevertheless, lexical richness increased over the first half of the semester (Journal 1 to Journal 6) before the midterm break, which supports the finding of Wolf-Quintero et al. (2003:140) that lexical complexity is significantly related to short-term change.

4.2   The Influence of Genre

Also, although fluency was not affected by genre (type of writing), grammatical complexity and lexical richness were, with stories being syntactically simpler but lexically richer than essays. The investigation of lexical richness suggests that genre influences the activation of a learner's lexicon. For the intermediate and upper-intermediate students of this study, writing stories proved a more useful exercise for language development than essays.

4.3   Language Processing

The findings of this study show that fluency and grammatical complexity relate to each other in an intriguing way whilst at the same time reflecting a common pattern of development. In most students' journals, fluency appears to be inversely related to grammatical complexity: in other words, the longer the clauses, the fewer clauses per sentence, which suggests that these intermediate or upper-intermediate students were more comfortable with basic sentences and

coordination when writing freely. This is, perhaps, not surprising. According to 'emergentism,' learners have limited resources of memory, attention span and willingness to spend time on a task, so that 'at one point in time, higher performance on one dimension of proficiency, say, accuracy, can seemingly detract from performance in others, say fluency and complexity' (Larsen-Freeman 2006:.593). The apparent inverse relationship between fluency and complexity in the present study would support the notion of competition for attention.

4.4   Limitations of an Objective Study

The findings of this objective study do not harmonize with the writers' own judgment (Journal 10) that their writing had improved over the nine weeks of journal writing. Several explanations for this mismatch spring to mind. One is that the language development of intermediate or upper-intermediate students may perhaps become discernible along objective dimensions only after a longer writing period, unlike Arthur's (1979) finding of rapid development with elementary writers. Another is that a valid analysis of improvement should consider pragmatic as well as linguistic competence, since the students' journals were socially situated. Larsen-Freeman (2006), for example, would consider learners' goals and intentions, the particular tasks given by their teacher and each performance on that task, using multivariate analysis to describe and measure development. Yet a third explanation is that improvement may have been primarily in the affective domain, with writers remembering increasing mastery of their medium as they re-read their journals, whilst the teacher/researcher took a normative and cognitive stance, imposing a ladder of language development on the phenomena and seeking extant evidence of this in forms of expression. What emerges from this difficulty is the lesson that no single method of evaluating second language writing can be privileged, although it should be pointed out that in this study performance along the dimension of lexical richness/variety mirrored the students' ranking in their final achievement test, which assessed many aspects of writing such as content, rhetoric, and grammatical/ mechanical accuracy.

Nevertheless, the limitations of every approach need to be kept in mind when making a preferably inclusive judgement of language development.

4.5   Implications

The conduct and results of this objective study of language development bear implications for three actors in the language teaching enterprise: second language acquisition researchers, teachers-as-researchers, and teachers themselves.

4.5.1 Second Language Acquisition Researchers

The results of the present study would seem to endorse Larsen-Freeman's 'emergent' metaphor of language acquisition as 'a complex process of dynamic construction within multiple ranges in multiple directions' (2006:591), with a learner's language changing as it is used and showing no discrete stages with invariant performance, but only periods of fluctuation. At the same time, with the partial exception of increased lexical complexity, the analysis offers scant evidence of construction, This, of course, may be a feature of the limited time frame (nine weeks) and suggests that any meaningful study of development needs to be extended over a year or more. However, researchers ought also to countenance the possibility of fossilisation and investigate likely causes in each learning context. Here, a complex systems approach would be fruitful.

4.5.2 Teachers-as-Researchers

A teacher is well placed to observe the social context of student writing and can record language development along many dimensions, thus providing the detailed account of contingent behaviour which emergentist research calls for. However, she cannot encompass the full range of levels or timescales from neural processing to life-long learning and even when she makes a considered selection, her analysis cannot ignore the putative impact of this excluded data. Given the fact that widely separated events may be more relevant to meaningful behaviour than closer events (Larsen-Freeman and Cameron, 2008), there will always hang a mist of uncertainty over each noted change of behaviour and any retrospective explanation, even if students provide historical information.

On a practical level, a teacher/researcher will be hampered in her rich data analysis by the concurrent demands of teaching. Specifically, in university departments which follow a modular programme, she will probably be unable to conduct a longitudinal study beyond one semester because the context of the research will change. Given the slowness of foreign language development beyond the elementary level of proficiency as suggested by this study, evidence of acquisition will therefore be hard to find, and the value of the requisite research effort extremely problematic. At the same time, the teacher/researcher is also a participant in her study with a separate agenda, whose short-term evaluative needs cannot, it seems, be met by instructed second language acquisition research. Thus the tension between research and teaching may increase.

4.5.3   Teachers

This particular teacher/researcher had introduced journal writing into a pragmatically oriented composition course to increase confidence and bring enjoyment into the writing process. However, in order to justify her innovation to colleagues, she later sought evidence of language development by measuring the fluency, grammatical accuracy and lexical complexity of the journals of a selection of students. Unfortunately, no sustained increase along those measures was observed; rather, considerable variation was found amongst students, and within their journal writing 'careers.' In most cases, narrative appeared to promote fluency, grammatical complexity and range of vocabulary more than discursive writing. The context of writing - its position within the semester, greater pressure from testing in other subjects in the later weeks - also seemed to influence performance. These findings may be helpful to teachers who are considering introducing free writing to similar university courses. Nevertheless, the final picture is of variability and fluctuation, with no way of determining which interim forms will predominate and no sure indications of language development. Therefore, it cannot be affirmed that the students' mastery of the written code was improved by writing journals.

This conclusion must disappoint any teacher who wishes to justify introducing journal writing into a product-based writing course, especially if the language programme is subject to the demands of quality assurance with its need for measurable outcomes. A positive response in a post programme questionnaire, albeit gratifying, carries little weight in such an ethos, which is regrettably taking over many university departments as they scramble for accountability to ensure their continuance.

A further disappointment arises at a practical level. For the relatively short-term evaluation of (upper) intermediate level students enrolled in semester-long modules, time-series measurement proves unsuitable and may even be subversive; for by showing that a final performance in a series is not necessarily the best, it casts doubt on the validity of exit tests. However, teachers need to make course-end evaluations when they work in institutions which move cohorts of students through semester-long modules; so for them 'last is best,' as promised by a ladder metaphor of acquisition, may be a necessary fiction.

5   Conclusion

Although 'emergentism' offers a more credible account of language acquisition than the ladder metaphor, it poses problems for diagnosis and engineering - two responsibilities of teachers

17

- because it eschews cause-effect relationships and prediction. If language acquisition is a self-organising complex system, then teachers cannot interfere benevolently in the process to speed it up: any imposed change (such as more interaction or free writing) merely introduces a new element into the complex system which may or may not produce a discernible effect. Teachers can neither control nor measure what is happening; nor can they apply the results of one investigation to future, similar teaching/learning situations because everything is flux. At best they can hope to provide a favourable, stress-free context for acquisition. This, of course, is not possible in educational institutions influenced by the standards movement where teaching has been squeezed into an explicit Objectives-Outcomes-Test-Review mould and all endeavour is predicated on the ladder metaphor of learning.

In a recent review of second language acquisition research, Larsen-Freeman (2012) seems to pull back from championing the complex systems theory of language acquisition, accepting diverse explanations and seeing "diversity in unity". This writer does not share her optimism. The SLA community needs consensus, rather than competition, if it is to promote better understanding of how languages are learned, particularly to educators and those who manage institutions of higher education. So long as these pursue policies that are inimical to language acquisition, prospects remain grim for research, informed teaching and optimal learning.

REFERENCES

1. Arthur B (1979) Short-term changes in EFL composition skills In: Yorio C, Perkins K, and Schachter, J (eds) *On TESOL '79: The Learner in Focus*: 330-342. Washington, D.C.: TESOL

2. Bachman L F (1990) *Fundamental Considerations in Language Testing* Oxford: Oxford University Press

3. Fischer K, Yan Z and Stewart J (2003) Adult cognitive development: Dynamics in the developmental web. In: Valsiner J and Conolly KJ (eds) *Handbook of developmental psychology*. London: Sage

4. Foster P and Skehan P (1996) The influence of planning and task type on second language performance. *Studies in Second Language Acquisition* **18**: 299-323

5. Hunt KW (1965) *Grammatical Structures Written at Three Grade Levels*. Urbana, Ill.: The National Council of Teachers of English

6. Hyltenstam K (1988) Lexical characteristics of non-native second language learners of Swedish. *Journal of Multilingual and Multicultural Development* **9**: 67-84

7.   Ishikawa S (1995) Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing* **4**:51-70

8.   Larsen-Freeman D (1978) An ESL index of development. TESOL *Quarterly* **12:** 439-448

9.   Larsen-Freeman D (1983) Assessing global second language proficiency. In Seliger HW and Long M (eds) *Classroom-oriented Research in Second Language Acquisition* 287-304. Rowley, MA: Newbury House

10.   Larsen-Freeman D (1997) Chaos/complexity science and second language acquisition. *Applied Linguistics* **18**:141-165

11.   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

12.   Larsen-Freeman D (2012) Fro_unity to diversity…to diversity within unity. *English Teaching Forum* **50** (2): 22-27

13.   Larsen-Freeman D and L. Cameron (2008) Research methodology on language development from a complex systems perspective. *Modern Language Journal* **92** (2):200-213

14.   Monroe JH (1975) Measuring and enhancing syntactic fluency in French. *The French Review* **48**:1023-1031

15.   Percival I (1993) Chaos: A science for the real world. In Hall N (ed.) *Exploring Chaos: A Guide to the New Science of Disorder.* New York: Norton & Company

16.   Sharma A (1980) Syntactic maturity: Assessing writing proficiency in a second language. In: Silverstein R (ed.) *Occasional Papers in Linguistics* **6**:318-325. Carbondale: Southern Illinois University

17.   Schmidt R (1992) Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* **14**:357-385

18.   Wolfe-Quintero K, Inagaki Sand Kim H-Y (1998) *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity* Hawaii: University of Hawaii Press