Pertanika J. Soc. Sci. & Hum. 24 (2): 639 – 654 (2016)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Measuring the Written Performance Quality in terms of (CAF) Complexity, Accuracy and Fluency Constructs

Dahmardeh, M.[1]* and Shahmirzadi, N.[2]

[1]*Faculty of Foreign Languages and Literature, University of Tehran, the Islamic Republic of Iran*
[2]*Young Researchers and Elite Club, Robatkarim Branch, Islamic Azad University, RobatKarim, Iran*

ABSTRACT

One of the concerns in language testing is measuring the extent to which the same constructs occur over certain time, across different groups of test takers in an EFL situation. The present study aims to analyse written performance of a number of English language learners over a year in a thick description to provide in-depth empirical data focusing on performance of test takers in terms of CAF. Previously, the majority of language learners' written performance was evaluated based on some scales without taking into account the qualitative performance from the SLA perspective. Therefore, the present study tries to bridge the gap between testing and SLA. As far as the first research question, which is the relative contributions of each of the features, complexity, accuracy and fluency is concerned, the researchers found a noticeable negative correlation between the complexity, accuracy and fluency of learner performance. Concerning the development of participants in writing skill as the second research question, students did not reach the benchmarks of Common European Framework of Reference (CEFR) levels. Moreover, it was observed that forms may not develop linearly among these three dimensions, and a learner's improvement in repeating a task does not guarantee that he/she will get the most from a target-language perspective unless it raises his/her awareness.

Keywords: Complexity; accuracy; fluency; writing

ARTICLE INFO

Article history:
Received: 9 January 2015
Accepted: 11 June 2015

E-mail addresses:
dahmardeh@ut.ac.ir (Dahmardeh, M.),
niloufar_shahmirzadi83@yahoo.com (Shahmirzadi, N.)
* Corresponding author

ISSN: 0128-7702

INTRODUCTION

Applied linguists have identified three components of Second Language (L2) performance, namely, complexity, accuracy and fluency (CAF) to measure L2 development. They could be used in productive skills to provide evidence to

analyse CAF constructs. Moreover, there has also been a call in the field in terms of writing skill for more longitudinal studies with well-chosen measurements (Ortega & Iberri-Shea, 2005; Verspoor, Lowie, & Van Dijk 2008; Norris & Ortega, 2009). In this case, Larsen-Freeman (2009) and Larsen-Freeman and Long (1991) believe that a longitudinal design allows the process of language learning to be followed over time to see the stability of written performance in triad of CAF. Meanwhile, theorising in CAF has focused on one of these three aspects, and relatively less empirical attention has been paid to the characteristics of other features. In line with that, some researchers believe "learners cannot attend to all areas of CAF performance; for example, as for accuracy; different topics seem to encourage different forms, especially in demanding tasks in the field" (Verrcelloti, 2012, p. 37; De Jong & Vercellotti, 2011).

CAF IN SLA - RESEARCH

The aim of the study was to investigate the growth of EFL on written performance with respect to complexity, accuracy and fluency at different levels of language proficiency. To analyse EFL rate of success or failure in writing skill, a model is needed, particularly in this study, for measuring CAF in accordance with new methods of evaluation and teaching (LaBerge & Samuel, 1974; Anderson, 1992; Anderson & Fincham, 1994; Dekeyser, 2001). Although a number of measurements were used, some criteria were needed in line with the present study. In what follows, classification of the three facets (see Ellis & Barkhuizen, 2005) and the method of estimating these criteria are presented.

"... concerning complexity, "*percentage of the total number of separate clauses divided by the total number of T-units*" (Foster & Skehan, 1996); "*percentage of the total number of different verb form used*" (Yuan & Ellis, 2003); "*percentage of the type-token ration: the total number of different words used (types) divided by the total number of words in the text (token)*" (Robinson, 1995) were analysed.

As for measuring accuracy, "*percentage of error free clauses divided by the total number of independent clauses, sub-clausal units and subordination clauses multiplied by 100*" (Foster & Skehan, 1996); "*percentage of error per 100 words: the number of errors divided by the total number of words produced divided by 100*" (Mehnert 1998); "*percentage of target-like use of plurals: the number of correctly used plurals divided by the number of obligatory occasions for plurals multiplied by 100*" (Crooks, 1989) were counted.

To analyse fluency of written performance, "*percentage of the average number of words per text*", "*percentage of the average number of T-units per text*" and "*percentage of the average number of clauses per text*" (Storch & Wigglesworth, 2007) were employed" (pp. 150–154).

Defining and Measuring CAF

Complexity: Ellis and Barkhuizen (2005) describe complexity as:

> *...elaborated language; relative to proficiency, as language that is at the upper limit of the student's interlanguage system connecting with a wider repertoire, which is not fully internalised or automatised by the learner (p. 139).*

Also, it is the most complex, difficult, ambiguous and least understood dimension of the CAF triad in language production (Skehan & Foster, 1997) that could be indicated in terms of 'development or proficiency' (Pallotti, 2009). 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)' (Robinson, 2001), in which performance and proficiency are the concern of the present study.

Accuracy is probably the oldest, most transparent and consistent construct of the triad, referring to the degree of deviancy from a particular norm which is characterised as errors (Hammerly, 1991; Wolfe-Quintero *et al*., 1998; Polio, 1997; James, 1998; Ellis, 2008). It is often measured by 'the learner's suppliance of the specific form in obligatory context, suited for the focused task' (Ellis & Barkhuizen, 2005, p. 151). Schachter and Celce-Murcia (1977) point out that:

> *...about the difficulty in classifying an identified error. Often, an ungrammatical sentence can be corrected in more than one way because the coder does not definitively know what the intention would be (p. 59).*

Fluency: in lay usage, it refers to a person's general language proficiency, particularly as characterised by perceptions of ease, eloquence and smoothness of speech or writing (Lennon, 1990; Chambers, 1997; Guillot, 1999; Freed, 2000; Koponen & Riggenbach, 2000; Hilton, 2008). Successful language performance in writing has often been characterised as writing at a normal rate without interruption resulting in fluency. 'This linear performance in fluency could be reversed in complexity and accuracy (form), that is attention can be directed toward using less advanced or challenging language (complexity), but having greater accuracy' (Skehan 1998).

RESEARCH QUESTIONS

This article attempts to measure the written performance of language learners over CAF constructs during a year to find any outperformance or deterioration. Therefore, in more specific terms, the two research questions for this study based on the aforementioned objective are as follows:

RQ1: What are the relative contributions of each of the features, complexity, accuracy and fluency in the written performance of the students' samples?

RQ2: To what degree does the English writing ability of students in this EFL programme develop with respect to complexity, accuracy and fluency, as a function of time?

MATERIAL AND METHODS

In this study, the researchers sought to define and justify every single procedural step taken throughout the different stages of the study and discuss their rationale and justifications. Thus, every detail of the method and procedure employed in this study is exposed to the readers' scrutiny.

The study was also based on various models for measuring the level of complexity, accuracy and fluency of participants in writing skill. In order to verify the hypotheses of the study, a series of measures were taken, as described below. Therefore, the focal design of the present study is descriptive and qualitative in nature since it aims at describing and reporting the written performance in terms of CAF, which are assigned in meaningful numerical values.

Participants

The data for this study were limited by the population of English language learners, attending classes over time between early April 2012 and late March 2013. Researchers analysed the written performance of 10 random sampled test takers in order to have thick description of who were studying English as a foreign language. However, as is typical of longitudinal study, a steep decline in numbers was observed. This was typically due to transportation difficulties, family relocation, time conflicts or programme completion, final examinations at the university, which were not maintained throughout the period. The students were around 1 to 3 years of studying at the language centre, and between 18 to 30 years of age; female (n=8) and male (n=2). Furthermore, since the progress and quality of students' written performance in terms of CAF was concerned, the participants were included in different levels that are specified in (Table 1). Also, because of the students' proficiency distribution, as well as required threshold levels of abilities for competent performance (Cumming, 2013), the CEFR (Council of Europe, 2001) levels were applied. Moreover, the criteria provided by CEFR were used to verify whether they were in their appropriate level. Finally, the learners were designated by IDs to have more in-depth and accurate comparison. It should be noted that all language learners were required to take a placement test at the beginning of their entrance to the language centre.

Procedure

Data Collection

The corpus of the study was longitudinally collected, which composed of the final written performance of 10 English language learners. During a year, composing of four terms, the writings of the participants were

analysed. Tests were administered in June 2012, September 2012, December 2012 and March 2013. The participants varied limitedly from basic to advance levels. It is worth noting that the participants suffered from poor writing skill; therefore, the researchers tried to analyse the participants' final performance to see their level of improvement and whether they were in their appropriate (proficiency) level or not. Moreover, to observe the rubrics of the standard writing examinations, some topics were chosen for all the levels. Considering task complexity in different levels, the length and time of the written performance were specified accordingly. The levels and topics of the writing examination were also parallel with the course book. Meanwhile to ensure consistency of the topics used, the same topics were administered during a year. Course book sources were World English (Marlin, 2011) and Top Notch (Saslow & Ascher, 2006).

Test administration task

Each term, 20 sessions running for 90 minutes each for two months and a half, 6 units (out of 12) of World English (the chosen course book series by the language centre published by Heinle to be taught to the language learners) and 5 units (out of 10) of Top Notch (A chosen course book series by the language centre published by Pearson Education to be taught to the language learners) were used to teach the different classes. The materials that the centre used were the two course book series selected based on the quality as well as public popularity in terms of market and other language centres across the country. Both require the learners to provide some writing activities and tasks; the two books were also categorised according to the same CEFR level. As for the final examinations, the participants were supposed to have one written section per term in line with the aforementioned materials.

Test administration procedures

As mentioned earlier, the tests were administered 4 times. On each testing occasion, the subject group and other students took the test simultaneously at the exam venue. For consistent tests administration, the test were organised on one day to supervise the uniformity and security of the testing procedure.

Meanwhile, the analysis procedures are presented here chronologically, in the order they occurred. First, since the aim of study was a longitudinal evaluation of the participants, it took a year to collect the writing samples in the institutional final examinations. Then, the four writing performances of each 10 sample were analysed accordingly.

After manually coding the selected writings of the participants with respect to the theoretical framework of the study, collected data related to complexity, accuracy and fluency were analysed exactly based on the framework as mentioned in section 2 (CAF in SLA research) to accommodate the needs of the present study.

Finally, after coding and combining the different aspects of complexity, accuracy and fluency separately (Appendix A), the mean score of each individual was calculated. Then, to measure the performance of language learners during a year, T–Kendall, a non-parametric correlation, was used.

RESULTS

After analysing the selected writings of the participants who attended four terms continuously, the collected data related to accuracy, fluency and complexity were analysed based on a written performance scale to accommodate the needs of the present study. The results are reported in percentage terms. Also, based on the theoretical framework of the study provided in CAF in SLA research, written sentences of each paper were analysed and coded several times to measure complexity, accuracy and fluency - in detail (a sample is provided in Appendix A). Meanwhile, accuracy, fluency and complexity were measured by error free clauses, errors and target-like use of plurals, number of words, T-units and clauses, and type-token ration, clauses and verb form, respectively. Table 1 shows the mean score of the participants by consulting three facets over the four terms. It is important to note that sample analysis was peer reviewed to check reliability of analysis along with the framework, while the difficulty level of the test was increased based on course book from test 1 to test 4. From the vertical perspective, the difficulty level of tests from level A to level to C_1^+ was also considered in the pilot study of test in a step wise fashion.

TABLE 1
Descriptive statistics of accuracy, fluency and complexity

Candidates (cardinal numbers)/ Levels	Mean Scores over Four Terms		
	Accuracy %	Fluency %	*Complexity %*
1 A A_1 A_1^+ A_2	44.23	41.99	51.28
2 A A_1 A_1^+ A_2	47.75	58.49	63.37
3 A A_1 A_1^+ A_2	49.33	76.10	46.68

Table 1 (continue)

4 A A_1 A_1^+ A_2	23.88	72.49	57.47
5 A_1 A_1^+ A_2 A_2^+	40.26	55.33	42.34
6 A_1 A_1^+ A_2 A_2^+	29.54	63.72	63.67
7 A_1 A_1^+ A_2 A_2^+	25.18	59.19	67.31
8 A_1^+ A_2 A_2^+ B	41.08	66.36	44.36
9 B_2 B_2^+ C_1 C_1^+	30.15	74.83	52.03
10 B_2 B_2^+ C_1 C_1^+	60.03	84.45	71
	39.14	**65.3**	**55.95**

As shown in Table 1, almost all of the participants received low scores on accuracy and complexity although they received high score on fluency. This could be justified as some might employ stream of consciousness without using the usual method of writing. It is also important to consider the slight differences on accuracy and complexity. As they often skipped on how to use grammatical structures properly, the scores did not reach the peak. This issue could be justified by the fact that the participants of the study did not experience continuous error correction in order to be able to have some repairs like restructuring, wisely. Surprisingly enough,

since complexity is the most sophisticated feature among the three, outperformance could be considered in comparison to accuracy. The point is that the participants might apply some prefabricated patterns to increase the complexity. A more in-depth analysis of grammatical complexity, based on the total number of subordinate clauses and phrases divided by the sum of the T-units (all clauses and phrases) (Foster & Skehan, 1996), the reversed results were shown. For example, a person who did well on separate clauses did badly on using different verb forms. As for the present study, Table 2 displays trade off in the mean scores.

TABLE 2
Mean score over four terms

Separate Clauses	Verb Forms
30.51	74.88

According to Richards and Schmidt (2002), the closer the type-token ratio is to "1", the greater the lexical richness will be. To prove the stated point, it might be justified in the study that the average scores of the type-token ratio were between 50 and 70, which was from 1. Therefore, extending vocabulary repertoire was neglected. Table 3 presents the poor lexical richness of the participants in their writing skill.

TABLE 3
Descriptive statistics of lexical richness

Candidates (cardinal numbers)/ Levels	Mean Scores over Four Terms
1	
A	
A_1	0.83
A_1^+	
A_2	
2	
A	
A_1	0.73
A_1^+	
A_2	
3	
A	
A_1	0.61
A_1^+	
A_2	
4	
A	
A_1	0.61
A_1^+	
A_2	
5	
A_1	
A_1^+	0.65
A_2	
A_2^+	

Table 3 (continue)

6 A_1 A_1^+ A_2 A_2^+	0.65
7 A_1 A_1^+ A_2 A_2^+	0.77
8 A_1^+ A_2 A_2^+ B	0.58
9 B_2 B_2^+ C_1 C_1^+	0.53
10 B_2 B_2^+ C_1 C_1^+	0.65
	0.66

According to Table 3, the mean score of each participant's lexical richness is below 1, and the total mean score of all is 0.66. As a result, the learners were found to suffer from lack of accuracy and complexity in their written performance. More importantly, the high score on fluency might not guarantee learner's competency. Likewise, the notion that all CAF constructs grow, although not necessarily by the same route in L2 writing (Larsen-Freeman, 2006), supports the findings of this study.

In addition, as a non-parametric statistics, T–Kendall was estimated to see the relationship between accuracy, fluency; accuracy, complexity, and complexity, and fluency. Table 4 depicts the findings measured by the SPSS programme.

TABLE 4
Non-parametric correlations

	Accuracy–Fluency	Accuracy–Complexity	Complexity–Fluency
Sig. (2–tailed)	0.53	0.78	0.53
Kendall's tau–b	–0.15	–0.67	0.15

As shown in Table 4, although each of three pairs is significantly meaningful, that is more than 0.5, no relationship between the first two pairs was observed since the scores were –0.15, –0.67. As for complexity and fluency, the relationship was found to be too weak, that is 0.15.

DISCUSSION AND CONCLUSIONS

In order to achieve a thorough understanding of the CAF traits and to capture the complex, dynamic and developmental nature of the CAF phenomena, the researchers attempted to find markedly steeper gains with saturated sample through thick description in development or trade-off effects between the constructs of language performance.

In the ultimate analysis of CAF, it is always measured for particular purposes in particular settings and with particular developmental targets in mind. Likewise, in an organic and sustainable approach to instructed SLA research, raising CAF is probably not going to be sufficient. According to Norris and Ortega (2009), researchers do not ignore other phenomena that are essential to a more complete understanding of second language learning besides defining the CAF triads' effect.

In relation to the present study and to tackle the first research question, there is no linear correlation between complexity and accuracy with fluency in each and all learners' performance, respectively. As shown, almost all the participants had higher fluency at the expense of both complexity and accuracy. Considering the correlation between fluency and lexical variety, especially at lower levels, negative correlation was observed. Table 5 supports the above mentioned statement for all the participants and lower level, which is –0.48 and –0.65, respectively (based on CEFR: lower levels or basic users are A, A_1, A_1^+, A_2, A_2^+).

TABLE 5
Non-parametric correlations

	Fluency–Lexical Variety (All Levels)	Fluency–Lexical Variety (Lower Levels)
Sig. (2–tailed)	0.56	0.46
Kendall's tau–b	–0.48	–0.65

This can be justified since lexical variety may cause fluency breakdown, except for proficient ones. Surprisingly, there is a downward trend between accuracy and complexity. As for grammatical complexity and lexical richness, a negative correlation was predicted between grammatical complexity and lexical variety, following Larsen-Freeman (2006).

Further, assessing accuracy over a time may be complicated since the learners may attempt to use new lexical items

and grammatical forms. As such, Norris and Ortega (2003, pp. 717-61) caution that 'accuracy of specific forms may not develop linearly but rather curvilinearly'. The downward trend, which is also in the complexity, can be explained on the basis of cognitive factors. The fact may applicable not just in oral but in written in fluency as well.

Seemingly according to Wendel (2007, pp. 13-41), 'participants with higher initial proficiency were expected to have different performance in comparison to lower proficient ones as they had more room for improvement'. This can also be explained in terms of accuracy and lexical variety, which is shown in Table 6 to compare minute difference in accuracy of the participants between the first and last tests, which is 36.96 and 37.66, respectively. Surprisingly, the participants could not manage to have richness in lexical variety during the given time span. This could perhaps be attributed to the fact that the subjects had not attained a sufficient level of proficiency to handle all the three facets at a time.

TABLE 6
Mean score and standard deviation of the first and last tests

	First Test		Last Test	
	Mean	SD	Mean	SD
Accuracy	36.96	8.15	37.66	24.17
Lexical Variety	70.78	7.78	60.39	9.30

These findings can be interpreted within the framework of Skehan's trade-off hypothesis. Skehan (1998) proposes that speaker performance be examined in terms of an initial contrast between meaning and form, with form further distinguished with regard to 'control' and 'restructuring'. In this model, meaning is reflected in fluency, while form is manifested in either accuracy (if control is prioritised) or complexity (if opportunities for restructuring arise because of the learner's willingness to take risks). Skehan's model can also be generalised to written performance where accuracy can be considered as more meaning-oriented and complexity is regarded as more form-oriented. Therefore, we can conclude that in the present study, the form-oriented measure of performance (i.e., complexity) is slightly related to the meaning-oriented components of performance (i.e., accuracy). Skehan (1998) contends that in selecting or designating linguistic tasks, there is a trade-off between cognitive processing and focus on form. Hence, we can expect that in attempting cognitively demanding tasks, the amount of attention language learners can devote to the formal features of the language is reduced, something that is deemed to be necessary for producing grammatically complex structures. This can partly explain the point that in the present study, learners with the highest fluency performed very badly in terms of

accuracy and complexity and the ones with the lowest lexical richness did very well in terms of complexity.

As for the second research question, analysis of the written performance in terms of CAF described on the CEFR scales for languages: learning, teaching, assessment. It should be taken into consideration that students' writings did not attain to the level of proficiency that is comparative to common European reference levels: self assessment grid. Below are some samples to introduce the Common Reference Levels in writing skill developed and validated for CEFR in the research project, Council of Europe (2001).

"… A_1: I can write a short, for example sending holiday greetings.

A_2: I can write short, simple notes and messages relating to matters in areas of immediate need.

B_1: I can write simple connected text on topics which are familiar or of personal interest.

B_2: I can write clear, detailed text on a wide range of subjects related to my interests.

C_1: I can express myself in clear, well-structured, text, expressing points of view at some length.

C_2: I can write clear, smoothly flowing text in an appropriate style" (p. 26).

Hence, nearly all the learners did not stand in their appropriate levels compared to the CEFR, and no correlation was observed between the CAF constructs in pairs. Therefore, the instability and multi-componential nature of CAF constructs were materialised (Skehan, 1998; Ellis, 2003, 2008; Ellis & Barkhuizen, 2005; Ortega & Byrnes, 2008).

All in all, it is recommended for further research to raise the awareness of language practitioners, especially writing skill in an EFL situation. Feedback is also considered essential in the process of writing tasks to integrate knowledge gained and make progress (LaBerge & Samuels, 1974). Meanwhile, according to Housen *et al*. (2012, p. 16), it is also suggested to take into account 'the interacting effects of both learner-internal and learner-external factors, which are to contribute to a fuller understanding of L2 knowledge, use and development'. Finally, it is suggested to narrow the study down to only C_1 and C_2 (proficient users) to determine the final step of proficiency which is automacity in utilising complexity, accuracy and fluency of quality written performance. In addition, it is recommended to evaluate both speaking and writing skills of a large number of participants in - one shot study based on CAF, and then compare the results of these two productive skills.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of the University of Tehran for this research under grant number 27320/1/2.

REFERENCES

Anderson, J. R. (1992). Automaticity and the ACT theory. *American Journal of Psychology, 105*(2), 165–180.

Anderson, J. R. & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20* (6), 1322–1340.

Chambers, F. (1997). What do we mean by fluency? *System, 25*(4), 535–544.

Council of Europe. (2001). *Common European Framework for Languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

Crookes, G. (1989). Planning and interlanguage variability. *Studies in Second Language Acquisition, 11*(4), 367–83.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and Perils. *Language Assessment Quarterly, 10*(1), 1–8.

DeKeyser, R. M. (2001). Automaticity and Automatisation. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125–151). Cambridge: Cambridge University Press.

De Jong, N., & Vercellotti, M. L. (2011). *Norming picture story prompts for second language production research: Fluency, linguistic items, and speakers' Perceptions.* Paper Presented at the American Association for Applied Linguistics conference (AAAL), Chicago, IL.

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

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

Ellis, R. & Barkhuizen, G. (2005). *Analysing Learner Language*. New York: Oxford University Press.

Foster, P. & P. Skehan. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299–323.

Freed, B. (2000). Is fluency, like beauty, the eyes, of the beholder? In H. Riggenbach (ed.), *Perspectives on Fluency* (pp. 243–265). Michigan: The University of Michigan Press.

Guillot, M. N. (1999). *Fluency and its Teaching*. Philadelphia: Multilingual Matters.

Hammerly, H. (1991). Fluency and Accuracy: Toward Balance in Language Teaching and Learning. *Multilingual Matters*, *15*(2), 267–269.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 Fluency. *Language Learning Journal, 36*(2), 153–66.Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.Koponen, M. & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach, *Perspectives on Fluency* (pp. 5–24). Ann Arbor, MI: University of Michigan Press.

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

LaBerge, D. & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

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.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 579–589.

Larsen-Freeman, D. & Long, M. H. (1991). *An Introduction to Second Language Acquisition Research*. Harlow: Longman Group.

Marlin, M. (2011). *World English. Student Book*. Heinle: Cengage Learning.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second language Acquisition, 20*(1), 83–108.

Norris, J. & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty and M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 717–61). Malden, MA: Blackwell.

Norris, J. M. & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578.

Ortega, L. & Byrnes, H. (2008). Theorising advancedness, setting up the longitudinal research agenda. In L. Ortega and H. Byrnes (Eds.), *The Longitudinal Study of Advanced L2 Capacities* (pp. 281–299). New York: Routledge: Taylor & Francis.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. Annual Review of Applied Linguistics, *25*, 26-45.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590–601.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research." *Language Learning, 47*(1), 101–143.

Richards, J.C. & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd Ed.). London: Longman Pearson Education.

Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning, 45*(2), 283–331.

Robinson, P. (2001). Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.

Saslow, J. & Ascher, A. (2006). *Top Notch.* Pearson Education, Inc.

Schachter, J., & Celce-Murcia, M. (1977). Some reservations concerning error analysis. *TESOL Quarterly, 11*, 441–451.

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

Skehan, P. (2001). Tasks and language performance assessment. In M.P. Bygate. P. Skehan, and M. Swain (Eds.), *Researching pedagogic tasks* (pp. 167–185). International Handbook of English language Teaching: Springer.

Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185–211.

Storch, N. & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M. Del Pilar Garcia-Mayo (Ed.), *Investigating tasks in formal language settings* (pp. 157–177). Clevedon: Multilingual Matters.

Vercellotti, M.L. (2012). *Complexity, Accuracy, and Fluency as Properties of Language Performance: The Development of the Multiple Subsystems over Time and in Relation to Each Other*. (Doctoral Dissertation). University of Pittsburgh.

Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal, 92*(2), 214–231.

Wendel, J. N. (2007). An assessment of English language learner writing. *Journal of Bunkyo Gakuin University Department of Foreign Languages and Bunkyo Gakuin College,* 13–41.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu, HI: Second Language Teaching and Curriculum Centre, University of Hawaii at Manoa.

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

APPENDIX A

An Example of a Part of Coded Sample

Number of T-Unit (TU) : 8
Number of verb form (VF) : 6
Number of Types (Ty) : 26
Number of Token (TO) : 33
Number of words (W) : 33
Number of clauses (C) : 4
Number of correct use of plural (P) : 1
Number of error free clauses (EFC) : 2

(TU) I'm (VF) a doctor. (C) (TU) I treat (VF) this sick people (C) (TU) (EFC) after I talk (VF) with them (TU) about their aches (P). (C) (TU) (EFC) I examine (VF) them (C) (TU) before I surgery them (TU) until help (VF) them. (TU) My job is (VF) interesting and good.