CORRELATIONS OF ENGLISH WRITING COMPLEXITY, ACCURACY, FLUENCY, AND TEST SCORES: A CASE STUDY OF STANDARDIZED WRITING TEST SAMPLES

Ming Huei Lin

English Department, Tamkang University, Taiwan
Phone: +886 2 2621 5656, E-Mail: johnlinminghuei@gmail.com

Abstract. *The Limited Attentional Capacity Model (LACM) (Skehan & Foster, 1997, 2001) is believed to affect the memory processing system, leading to a trade-off model between the levels of complexity, accuracy, and fluency (CAF) in writing output by EFL learners. Time may be one of the critical variables moderating EFL writers' LACM, but few writers, least of all in the context of Taiwan, have studied the possible influence of LACM on the CAF relationship as demonstrated in timed writing,. What has not been properly investigated includes possible CAF interaction varying with Taiwanese EFL writers' proficiency levels. This should also be examined, given the changes in the development of student writers' skills as they learn, that is, mastering greater complexity, accuracy or fluency than before. To shed light on these aspects, the present researcher studied 150 timed writing samples created by 150 EFL test-takers who participated in the General English Proficiency Test High Intermediate. Both inferential and descriptive statistics were used to analyze the data. The research results indicate a possible shifting relationship between complexity and fluency/accuracy, suggesting that LACM impacts on Taiwanese EFL writers' performance, mostly probably subject to the time factor, since no differences of any kind were found between the interrelationship of writing qualities and their CAF.*

Key words: *the Limited Attentional Capacity Model (LACM), EFL writing, writing complexity, accuracy, and fluency*

1. INTRODUCTION

Complexity, accuracy, and fluency (also known as CAF) in writing have recently gained much attention in the field of second language acquisition (SLA), particularly regarding English as a foreign or second language (EFL/ESL) (e.g., Ellis & Yuan, 2004; Lin & Chen, 2015; Lin, Chen, & Chen, 2015; Ong & Zhang, 2010; Polat & Kim, 2013; Robinson & Gilabert, 2007). Many investigators confirm the validity of CAF as effective indicators of language learners' performance (cf. Ellis, 2003; Ellis & Barkhuizen, 2005; Housen, Kuiken, & Veder, 2012; Skehan, 1998, 2009). Others suggest that the more fluently or correctly the writing, the more advanced the English writing/speaking ability (e.g., Fellner & Apple, 2006; Lin, 2012). Some scholars advance this by investigating the interaction between these three linguistic dimensions, specifically as moderated by LACM (the Limited Attentional Capacity Model) (Lin & Chen, 2015; Skehan, 1998, 2001, 2003,

2009; Skehan & Foster, 1999, 2001; Yeh & Lin, 2015), believing that a brain has limited attentional resources. A common assumption about LACM in relation to EFL/ESL writing is that writing under pressure, whether external or internal, consumes much mental attentional capacity, causing the CAF dimensions to compete with each other and consequently form a trade-off relationship (cf. Skehan & Foster, 2001; Housen, Kuiken, & Vedder, 2012). Some examples of external and internal pressure are writing anxiety, motivation to write, language users' linguistic abilities, and task complexity (cf. Housen, Kuiken, & Vedder, 2012 for more discussion).

Interestingly, however, although many researchers still pursue this line of inquiry, they have cannot agree. This is perhaps because they try to examine multiple variables at the same time, for example the variables of time pressure and task complexity. Moreover, they examine different time-frames and writing genres (such as argumentative essays or descriptive writing) (Lin & Chen, 2015). While their approaches are individually valid, they leave the overall effects of each variable rather hard to interpret. Mutually affecting variables may explain why some scholars find that fluency develops at complexity's expense (e.g., Ellis & Yuan, 2004) whereas others observe the trade-off between accuracy and complexity (Skehan, 1996; Skehan & Foster, 1997, 2001). More bleakly, some researchers (e.g. Johnson, Mercado, & Acevedo, 2012) reveal no evidence of any trade-off in students' CAF, even when they wrote in different conditions. This finding would severely undermine the validity of Skehan's (1998) theory of LACM.

Given such conflicts, recent researchers (e.g., Yeh & Lin, 2015; Lin & Chen, 2015; Lin, Chen, & Chen, 2015) have suggested further research to plug this gap, in particular using a stand-alone variable. Time is suggested as crucial among the many candidate variables (Lin & Chen, 2015; Yeh & Lin, 2015) mostly because it has been widely examined along with other variables, but never alone. This omission is even striking in Taiwan's EFL/ESL writing context. Lin and Chen (2015) successfully identified a twofold trade-off model in the writing samples produced in different time-frames by Taiwanese EFL students: a macro model of complexity versus accuracy/fluency and a micro model of accuracy versus fluency. On the one hand, this may synthesize the disagreement among previous CAF studies (e.g., Ellis & Yuan, 2004); Skehan & Foster, 1997, 2001) but, on the other, it emphasizes the need to investigate the exclusive effect of time on CAF. The findings of Lin and Chen were drawn from a small-scale case study of fewer than 50 writing samples; this only intensifies the above need.

An equally important aspect so far ignored in CAF studies in Taiwan is the English proficiency of EFL student writers. This need not entail investigating blends of variables (time and writing abilities). Rather, the present researcher believes that articles by writers of different proficiency should not only be studied as a whole, to gauge the norm of Taiwanese EFL writers' CAF performance, but they should each be treated in accordance with its proficiency level, to reflect the fine interactions (if any) of CAF as they vary, subject to LACM, among writers with different skills (Lin & Chen, 2015). Such treatment is essential for tracing how the development of student writers' skills changes as they learn (cf. Hunt, 1965).

Clearly, although writing CAF has often been studied, more investigation is still needed, in particular in the EFL writing context of Taiwan. Hence, the present project aims to plug the gap by considering a batch of timed writing samples collected from the General English Proficiency Test (GEPT) High Intermediate Level. The results of this project are intended to   answer the two sets of research questions below:

1-1. Does LACM play a role in the trade-off between the CAF in Taiwanese GEPT High Intermediate test-takers' performance in timed writing

1-2. If so, what is the trade-off model like between the CAF?

2-1. Does LACM play a role in the trade-off between the CAF in Taiwanese GEPT High Intermediate test-takers' performance in timed writing when writing quality is considered?

2-2. If so, what are the trade-off models like between the CAF in these writing samples of varying quality?

2. DEFINITIONS OF CAF

2.1. Writing complexity

Writing complexity is arguably the most 'complex' feature to define. In the current literature, many find that writing complexity has two major aspects: the total number of linguistic units involved and the variety of linguistic patterns used. However, because language systems have a great many different linguistic features, measuring the two dimensions seems time-consuming, if not unfeasible. Among the many alternative methods of assessment, indicating the complexity levels of textual outputs by the number of T-units (devised by Hunt in 1996) is common (e.g., Biber, Gray, & Poonpon, 2011; Ellis & Yuan, 2004; Lin & Chen, 2015; Wolfe-Quintero, Inagaki, & Kim, 1998; Yang, Lu, & Weigle, 2015), 5), mostly combined with the number of sentences or clauses used in the same text (e.g., Ishikawa, 1995; Lin & Chen, 2015; Lu, 2010; Yang, W., Lu, X., & Weigle, S. C. (2015)). Such practices greatly increase the feasibility of assessing writing complexity and also enhance the validity of measuring complexity, given the close correlation found between T-unit-related measures and advanced writing samples (Hunt, 1965). The present study chose the widespread measurement of writing complexity by counting the number of T-units per sentence (T/S) (cf. Lin, 2015; Lin & Chen, 2015; Wolfe-Quintero, Inagaki, & Kim, 1998).

2.2. Writing accuracy

Accuracy is easier to define and measure than complexity. Unsurprisingly, writing accuracy concerns the presentation of correct linguistic features. However, since the relative correctness of features is contentious, the common measure adopted to show correctness is the total number of linguistic errors/mistakes found, lower accuracy being shown by more errors/mistakes, and so on. Instead of exclusively using the total number of errors, this type of assessment, like that of complexity, is often associated with other important linguistic characteristics, such as T-units, forming such typical measures as error-free T-units (EFT)—the total number of T-units without errors identified (cf. Lin & Chen, 2015; Polio, 1997). As implied, the greater the number of EFT found, the more accurate the article. EFT is thus used for this project (Lin, 2015).

2.3. Writing fluency

Although not fully focusing on 'writing' performance, Brown (1994, 2001) offers a somewhat fuller observation about fluency in language performance, which focuses on the 'flow' of language, or more precisely, a steady flow of language without correction within a limited timeframe. Brown's definition is widely reflected in contemporary measures of fluency

in literature, in which "words per minute" is probably the most frequent assessment (see Ellis and Yuan (2004), Freed (2000), Housen and Kuiken (2009), Ishikawa (1995), Latif (2013), Lin (2015), and Lin and Chen (2015), for example). Although time is considered an important element in measuring fluency, this study recorded no precise periods for each writing sample; only the total number of words indicated the writing fluency (cf. Lin, 2012) since all the writing samples collected were produced in the same time-frame.

3. METHODS

3.1. The writing samples

The data used for the following analysis come from a batch of 150 writing samples produced by 150 test-takers (58 females and 92 males) for the 2013 GEPT High Intermediate Level writing test. The GEPT contains five levels: elementary, intermediate, high-intermediate, advanced, and superior. Elementary level corresponds to A2 in the Common European Framework of Reference for Languages (CEFR); intermediate, B1; high-intermediate, B2; advanced; C1; and superior, C2. In this writing test, the examinees had 50 minutes to complete two parts: a passage of Chinese-English translation and a guided writing task of 150-80 words. For the present project, only writing samples were collected; they all discussed why many youngsters nowadays enter talent contests (e.g., dancing and singing competitions, etc.). The 150 samples comprise 37 articles graded 2 in this test, 53 graded 3, 54 graded 4, and 6 graded 5 (maximum). Grades 4-5 indicate a pass; higher (lower) grades suggest more (less) advanced writing skills All the scores presented here are the averages of those produced by two anonymous trained raters from the Language Training and Testing Center (LTTC: https://www.lttc.ntu.edu.tw/), who used the 6-level (0 to 6) holistic writing rubric designed by the center. Articles marked 0 are unanswered or contain too little information (fewer than 40 English words) to judge their quality, while those scoring 1 are either irrelevant or cannot be evaluated due to their content or grammar. For these reasons, the analysis ignored levels 0 and 1.

It should be noted that to answer the research questions proposed above, the 150 writing samples had first to be examined as a whole, and then divided into two batches for detailed investigation: Sub-Batch 1 (those marked 2 and 3) and Sub-Batch 2 (those marked 4 and 5). The division was based on the threshold mark because it indicated a decisive qualitative difference in the writing. The division was anticipated to shed light on the fine differences in the CAF interactions between writing samples of varying quality.

3.2. CAF raters and measures

Two raters, not those from the LTTC, were involved in the CAF assessment, one of whom was the researcher and the other an experienced TESOL teacher. Before the raters started the assessment, they had agreed on the fundamental elements to form the measures to assess the complexity and accuracy qualities in this study: T-units per sentence (T/S) for complexity and error-free T-units for accuracy. These formed the criteria for measuring errors and T-units (Hunt, 1965; Young, 1995). To confirm the agreement in the measurements of the two raters, two correlation tests (Pearson's r, two-tailed) were performed. The results indicate statistically significant inter-rater reliability for both items, since statistically significant correlation was found for them both (T-units: $r = .998$, $p = .000$; errors: r .981, $p = .000$), ensuring the validity

of the data analysis afterwards. The average of the raters' scores was then used for further analysis. It should be noted that they needed no inter-rater reliability test because the total number of word tokens (for assessing fluency) and of sentences was automatically generated by computer using software (WordSmith 5.0).

3.3. Data analysis

In order to observe whether there was a trade-off between the CAF of the overall timed writing samples in Sub-Batch 1 and Sub-Batch 2, both inferential statistics and descriptive statistics were used. First, Pearson's *r* examined the interaction between the CAF; descriptive statistics entailing charts were then used to convey further the nature of the interaction (cf. Lin & Chen, 2015).

4. RESULTS

All illustrations (figures, photographs, line drawings, graphs) should be numbered in series and all legends should be included at the bottom of each illustration. All figures, photographs, line drawings and graphs, prepared in electronic form, should be converted in TIFF or JPG (max quality) files, in 300 dpi resolution, for superior reproduction. Figures, line drawings and graphs prepared using elements of MS Drawing or MS Graph must be converted in form of pictures and unchangeable. All illustrations should be planned in advance so as to allow reduction to 12.75 cm in column width. Please review all illustrations to ensure that they are readable.

4.1. Interaction between the CAF: Writing samples as a whole

Table 1 presents the basic information of the writing samples and Table 2 shows rather complex results: a statistically significant positive correlation between accuracy and fluency (Pearson's $r = .585$, $p = .000$), a statistically significant yet weak negative correlation between complexity and accuracy (Pearson's $r = -.189$ at $p = .021$), and a non-significant correlation between complexity and fluency (Pearson's $r = -.066$ at $p = .425$). These results suggest that in these samples, lengthier articles tend to contain more accurate linguistic features. Alternatively, the results also mean that student writers who write more accurately may write faster or more. In contrast, such accuracy may also be developed at the cost of complexity, or vice versa, because a statistically significant negative correlation was found here. However, no significant interaction was found between writing fluency and complexity.

Table 1 Basic information on the writing samples ($n = 150$)

Items	Minimum	Maximum	Mean	SD
Sentences	3.00	25.00	12.27	4.44
T-units	4.00	26.00	13.60	4.58
Errors	.50	45.50	15.95	9.96
T/S (Complexity)	.85	3.00	1.14	.25
EFT (Accuracy)	.00	15.50	5.54	3.81
Tokens (Fluency)	61.00	353.00	205.61	62.39

Table 2 Correlation tests (Pearson's r) for CAF interactions

	Complexity	Accuracy	Fluency
Complexity	1		
Accuracy	-.189[*]	1	
Fluency	-.066	.585[***]	1

Figure 1 provides further evidence that LACM affects the participants' writing samples. To be specific, despite some exceptions, the *complexity* curves remained similar in shape, slope, and height, complementing the inferential finding that a very small negative correlation coefficient was found between complexity and the rest. However, both *fluency* and *accuracy* curves changed drastically and seemingly rose or dropped together, suggesting corresponding interactions between them. This further confirms a positive correlation between accuracy and fluency, above.
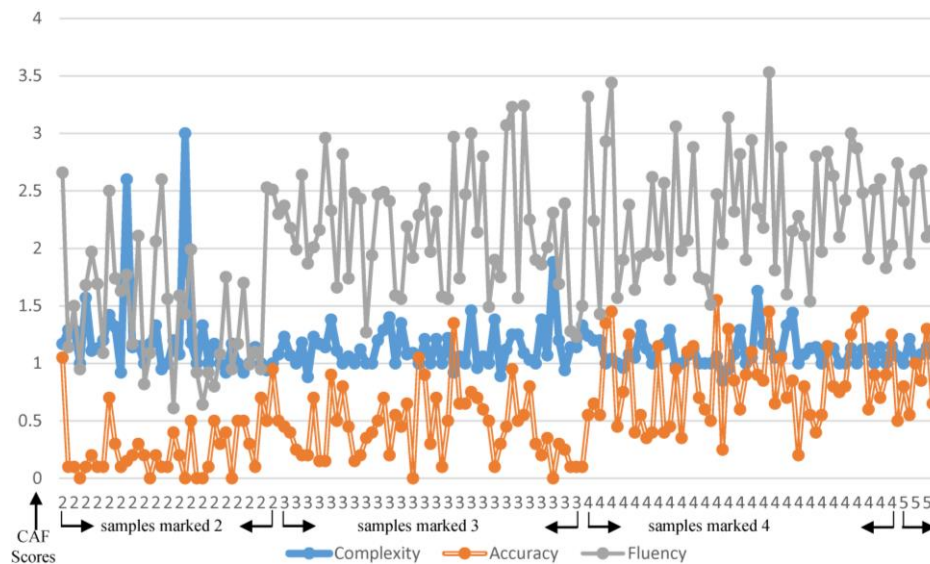

Fig. 1 The CAF interactions in the overall writing samples
(*Note.* The dots represent the scores for each linguistic feature.)

4.2. Interaction between the CAF: Sub-Batch 1

Table 3 presents the basic information about the writing samples of Sub-Batch 1 and Table 4 shows a statistically significant positive correlation between accuracy and fluency (Pearson's r = .517, p = .000) but no significant correlation between complexity and accuracy (Pearson's r = -.197 at p = .063) or between complexity and fluency (Pearson's r = -.017 at p = .874). These results, mostly resembling those of the writing samples overall, suggest that despite their relatively poor quality, lengthier articles in this batch still tend to contain more accurate linguistic features, or that student writers who write more accurately probably develop fluency too. Although fluency and accuracy clearly benefit each other, no clear evidence shows the way that complexity development entered this relationship.

Figure 2 provides further evidence that LACM has an effect on test-takers' writing performance. Despite some exceptions, the shape and slope of the *complexity* curve clearly remained much the same from right to left, which may again be taken to support the inference that complexity and the other two features are uncorrelated. However, both *fluency* and *accuracy* curves changed drastically and seemingly rose or dropped together, implying a corresponding interaction with one another. This further confirms the positive correlation between accuracy and fluency shown above, notwithstanding the relatively low writing quality of Sub-Batch 1.

Table 3 Basic information on Sub-Batch 1

Items	N	Minimum	Maximum	Mean	SD
Sentences	90	3.00	24.00	11.59	4.58
T-units	90	4.00	25.50	13.04	4.80
Errors	90	2.00	45.50	19.46	9.53
T/S (Complexity)	90	.88	3.00	1.17	.30
EFT (Accuracy)	90	.00	13.50	3.77	2.92
Tokens (Fluency)	90	61.00	324.00	187.96	63.03

Table 4 Correlation tests (Pearson's *r*) for CAF interactions in Sub-Batch 1

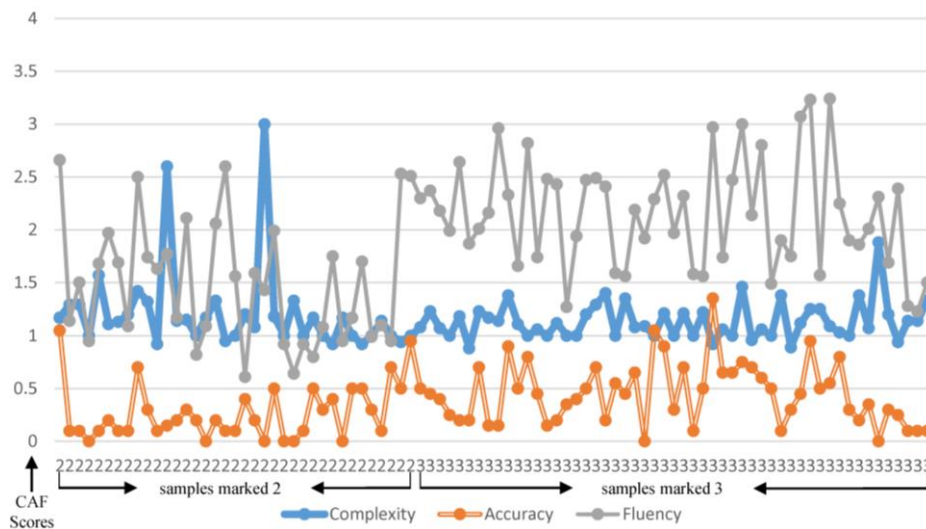|	Complexity	Accuracy	Fluency
Complexity	1		
Accuracy	-.197	1	
Fluency	-.017	.517***	1

Fig. 2 The CAF interactions in the Sub-Batch 1
(*Note.* The dots represent the scores for each linguistic feature.)

4.3. Interaction between the CAF: Sub-Batch 2

Table 5 presents the basic information of Sub-Batch 2; Table 6 shows the inferential results, finding a statistically significant positive correlation between accuracy and fluency (Pearson's r = .501, p = .000). However, no statistically significant correlation was detected between complexity and accuracy (Pearson's r = -.020 at p = .882) or between complexity and fluency (Pearson's r = -.047 at p = .758). These results also suggest that, in this test, lengthier articles tend to contain more correct grammar, or that skilful student writers who write more correctly may also write faster or more in the given time. However, developing either accuracy or fluency had no obvious interrelationship with greater complexity.

Table 5 Basic information of Sub-Batch 2

Items	N	Minimum	Maximum	Mean	SD
Sentences	60	5.00	25.00	13.28	4.06
T-units	60	6.00	26.00	14.43	4.13
Errors	60	.50	41.00	10.68	8.16
T/S (Complexity)	60	.85	1.63	1.10	.13
EFT (Accuracy)	60	2.00	15.50	8.19	3.45
Tokens (Fluency)	60	143.00	353.00	232.07	51.49

Table 6 Correlation tests (Pearson's r) for CAF interactions in Sub-Batch 2

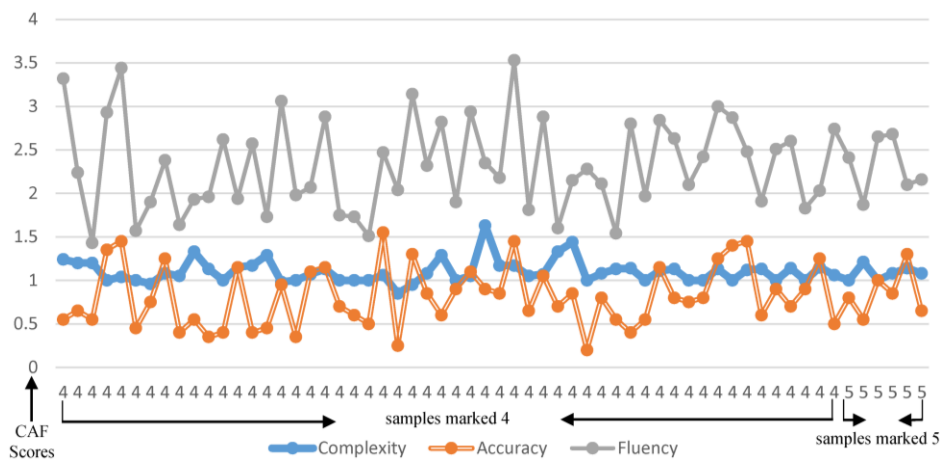|	Complexity	Accuracy	Fluency
Complexity	1		
Accuracy	-.020	1	
Fluency	-.041	.501***	1

Fig. 3 The CAF interactions in the Sub-Batch 2
(*Note.* The dots represent the scores for each linguistic feature.)

Figure 3 also shows a clear interaction between accuracy and fluency, which may suggest that LACM has an effect on test-takers' writing performance. This is evidenced by

the way that both *fluency* and *accuracy* curves changed drastically and seemingly rose or dropped together. However, like those in Figures 1 and 2, the shape and slope of the *complexity* curve remained mostly flat from right to left, which may again signify support for the inferential finding that complexity and the other two features are not correlated. Clearly, the relatively high writing quality of Sub-Batch 2 contained no different CAF interaction from that shown in low quality writing.

5. DISCUSSION AND CONCLUSION

The present study investigated whether or not LACM might affect the trade-off relationship between the CAF in Taiwanese GEPT High Intermediate test-takers' writing performance, considering samples of varying quality as a whole. The data were first examined as a whole and then individually according to their writing quality (two batches of different levels). This was done in the hope of portraying not only the interaction between the CAF subject to LACM, but also the fine interrelationship in CAF attributed to the influence of writing proficiency. Not completely in line with previous, the results of this study indicate a potential trade-off model; that is, student writers, regardless of proficiency level, apparently focus on developing both accuracy and fluency, probably at the expense of complexity, an interrelationship meriting discussion.

To begin with, although no statistically significant correlations were found between complexity and fluency/accuracy, the sharp contrast between the consistent *negative* interrelationships of complexity versus fluency/accuracy and the *positive* interrelationship of fluency and accuracy highlights the possibility that accuracy and fluency are developed to the cost of complexity. The reason that such correlation was not fully evidenced in statistical form may be the sample sizes adopted in this project. This interpretation gains strength from reflecting that only in the overall samples (150 articles) was complexity seen in a significant (although weak) negative correlation with accuracy, never in Sub-Batch 1 (90 articles) or 2 (60 articles). The decreasing possible correlations between fluency and accuracy from the data as a whole ($r = .587$) through Sub-Batches 1 (.517) and 2 (.501) lend further support to this reasoning. Hence, future researchers may choose relatively large samples in re-testing the trade-off model claimed in this study.

Second, confirming the LACM theory, the results also echo those by Skehan (1996) and Skehan and Foster (1997, 2001), that accuracy competes with complexity and also those by Ellis and Yuan (2004), who find fluency developed at the expense of complexity (e.g., Ellis & Yuan, 2004). Such interpretations may seem in conflict, but to some extent, they resemble those of Lin and Chen (2015). The present results may actually synthesize those by Skehan (1996), Skehan and Foster (1997, 2001), and Ellis and Yuan (2004), supporting part of the formulation by Lin and Chen (2015) of a macro trade-off model (complexity versus accuracy/fluency). In contrast to Lin and Chen, however, while such a model entails the micro-interrelationship of accuracy versus fluency, the micro-interrelationship found in this study benefits both parties.

One possible reason why the formulation of this study partly differs from that by Lin and Chen (2015) in terms of the micro model is the different degrees of time pressure imposed on the EFL writers. While all the writers in the present study were given the same amount of time to write, in Lin and Chen's study the EFL student writers were required to compose different articles in different time-frames. In other words, in addition to

examining timed writing, Lin and Chen also looked at the influence of different time pressures on Taiwanese EFL student writers' performance; sometimes greater time pressure clearly affects the balance between accuracy and fluency, causing mutual conflict. Given this, it seems valid at this point to conclude that not only does the present study ensure a micro model between fluency and accuracy; it also indicates time as one of the most influential factors affecting its balance, in turn justifying the need to investigating time as a sheer variable in studies of this type. This interpretation is further strengthened by this: the results of the present study also echo the early study where Kuiken and Vedder (2008) report that language proficiency levels have no effect on the interrelationship between complexity and accuracy.

As well as contributing to the understanding of LACM in timed writing produced by GEPT High Intermediate test-takers, the research focus and methods of this investigation open up some opportunities for further studies, although the design of this study is individually valid. First, while time-frames were not considered in assessing fluency in this study, future researchers may, like many previous investigators, incorporate this element into fluency measurements (cf. Ellis & Yuan, 2004; Freed, 2000; Housen & Kuiken, 2009; Ishikawa, 1995; Latif, 2013; Lin & Chen, 2015), thus shedding greater light on the fine interaction between CAF, more precisely between fluency and accuracy, given their positive interaction. More researchers, too, may examine whether patterns found in this study may resemble those in the CAF interactions observed in the writing samples at other GEPT levels, such as advanced or beginning. Following this line of inquiry would portray the CAF trade-off models even more fully. Similarly, an interesting topic to develop further is a closer look at the fine interaction between complexity and accuracy/fluency. While this study temporarily demonstrated the existence of such a relationship, it remains unsolved what, or more specifically which linguistic feature in complexity, is traded off when test-takers focus on accuracy/fluency. To complement the results of this study or any future studies as suggested here, researchers in this field may consider qualitative inquiries, a rare practice in this field. The exact CAF interaction in Taiwanese GEPT test-takers' writing samples would be helpful to discover, and so would the exact reason(s) for students to develop one (or two) linguistic dimension(s) over another. In a related area, future researchers may adopt new perspectives in observing how CAF may differ in writing samples of varying quality, an idea inspired by the result in this study that the writing quality of Sub-Batch 1 was found slightly more complex than that of Sub-Batch 2 (see Tables 3 and 5)—a counter-intuitive phenomenon against the general assumption that advanced writers would write with more complexity than less skillful writers. As this aspect does falls outside the research goal of this study, too little data were generated to tackle it, leaving some room for future studies to fill. Last, future researchers may extend the understanding of CAF interactions by examining how such interrelationships may serve as predictors of students' writing scores. It would be helpful to perform this task using logistic regression and the findings might alter our perspectives on the meaning of CAF interactions, in turn justifying the need for more CAF studies in future.

REFERENCES

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59-84.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59-84.

Evans, N. W., Hartshorn, K. J., Cox, T. L., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing*, *24*, 33-50.

Groom, N., & Littlemore, J. (2011). Doing applied linguistics: A guide for students. London and New York: Routledge.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. In A. Housen, F. Kuiken, & I. Vedder (Eds.), Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA (Vol. 32) (pp. 1-20). Amsterdam: John Benjamins Publishing Company.

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

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, *21*, 264-282.

Latif, M. M. M. A. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, *34*(1), 99-10.

Lin, M. H. (2012). Blog assisted language learning in the EFL writing classroom: An experimental study. Unpublished Doctoral dissertation. University of Birmingham, Birmingham, UK.

Lin, M. H. (2014). Effects of Classroom Blogging on ESL Student Writers: An Empirical Reassessment. *The Asia-Pacific Education Researcher*, *23*(3), 577-590.

Lin, M. H. (2015). Writing complexity, accuracy, and fluency: Validity as indicators for Taiwanese EFL writers? Paper presented at *The 1st Interschool Academic Conference on Linguistics, Literature, and Language Teaching*. Providence University, Taichung, Taiwan.

Lin, M. H., & Chen, I.-T. (2015). Time factors in writing complexity, accuracy, and fluency: A preliminary trade-off model found in Taiwanese EFL students' compositions. In the Proceedings of 2015 International Conference and Workshop on TEFL & Applied Linguistics (pp. 161-169). Taipei: Crane Publishing.

Lin, M. H., Chen, I.-T., & Chen, H.-K. (June, 2015). What does time buy? A preliminary investigation on Taiwanese EFL students' writing performance. Paper presented at the 11th Annual English Conference. New Taipei City, Taiwan.

Lin, M. H., Li, J. J., Hung, P. Y., & Huang, H. W. (2014) Blogging a Journal: Changing Students' Writing Skillsand Perceptions. *ELT Journal*, *68*(4), 422-431.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496.

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218-233.

Polat, B., & Kim, Y. (2013). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied linguistics*, *35*(2), 184-207.

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing. *Language Learning*, *47*, 101-143.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 161-176.

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

Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.). Researching pedagogic tasks: Second language learning, teaching and testing (pp. 167-185). Harlow: Pearson Education.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1-14.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532.

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

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retelling. *Language Learning*, *49*(1), 93-100.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Ernst Klett Sprachen.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53-67.

Yeh, S.-Y., & Lin, M. H. (May, 2015). The trade-off relationship between English writing fluency, accuracy, and complexity: The time factor in college students' English writing performance. Paper presented at *the 32nd International Conference on English Teaching & Learning*, Taipei, Taiwan.

Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning*, *45*(1), 3-42.