

English Development of Chinese EFL Student Writers from Sophomore to Senior Years*

Xiaoling Ji

Shanghai Normal University, China

Second language development is an important topic in second language acquisition research, but as different studies usually focus on different linguistic features, a more comprehensive picture of language development is called for. The present study examines the developmental features in Chinese English majors' writing in the key linguistic features summarized by Wolfe-Quintero, Inagaski and Kim (1998) by comparing 100 learners' compositions from 2003 TEM4 (Test for English Majors Band 4) with their compositions from 2005 TEM8 (Test for English Majors Band 8).

The empirical tests indicate that learners make statistically significant improvement in fluency and lexical features. The minor improvement in syntactic complexity and overall rating of their writings fail to receive statistical support. On the other hand, significant decline is reported for the measure of accuracy. A comparison of learner performance from different universities shows a clear pattern of institutional differences. The results of the study suggest that progress in learners' writing cannot

* An earlier version of the paper was presented at the 6th Symposium on Second Language Writing held at Nagoya Gakuin University from 15-17 September, 2007. The researcher wants to extend her gratitude to the Test Administration Office, National Foreign Language Teaching Advisory Board for the data included in the study. Special thanks go to Prof. Shen Zou at Shanghai International Studies University for her guidance in the research. The researcher is also indebted to Wei Zhang and Xiaoqin Zhang at Shanghai Normal University for their help with the coding and grading of writing samples.

be taken for granted. Of the 100 learners from the four institutions in question, only those from one university show improvement in every index, most of which are statistically significant.

Key words: language development, fluency, syntactic complexity, linguistic accuracy, lexical variation and sophistication, holistic score

INTRODUCTION

Recent years have witnessed an increase in quantitative studies addressing the developmental features of second or foreign language writing, thanks to the various objective measures of linguistic features reviewed in Wolfe-Quintero, Inagaski, and Kim (1998). In China and the Asian context at large, however, this kind of study is not often seen, except Sasaki's (2004) longitudinal study and the cross-sectional study carried out by Wen, Ding and Wang (2003). The majority of survey studies in China suggest that Chinese learners of English are particularly weak at writing (Huang, 1996; Jiang, 1995; Pan, 1992; Yang, 1993; Zou, 1999), but these studies are more like personal reflections whose conclusions have not been put to empirical tests. Besides, as existing studies are all concerned with the writing performance of a group of English learners at a certain point of time, it remains an issue whether the learners' writing proficiency progresses with the improvement of general English proficiency as we would normally assume. To answer this question, we need to examine learners' writings across a period of time.

The present study compares two sets of essays by 100 English majors from four universities in Shanghai (25 from each university) in linguistic features like fluency, syntactic complexity, lexical features, and accuracy as well as holistic rating. The two batches of essays are taken from Test for English Majors Band 4 (TEM4) in 2003 when the participants were sophomores and Test for English Majors Band 8 (TEM8)¹ in 2005 when they were seniors.

¹ TEM4 and TEM8 are nation-wide tests for English majors administered in their

These two tests, though not as influential as College English Test (CET), a test for non-English majors, are high-stakes ones in China. Some institutions associate the passing of TEM4 with the awarding of the bachelor's degree, and having a TEM8 certificate is an asset to a university graduate on the job market.

The research is unique in a few respects. It involves a larger sample than the studies done so far, with 100 essays collected from two time points respectively. There is a longer time lapse—two years between the first and second batches of essays. It is also the only longitudinal study addressing the writings of Chinese EFL learners. The researcher expects the results to contribute to the existing knowledge of development in L2 or foreign language writing.

LITERATURE REVIEW

Many studies have been carried out concerning development in L2 and foreign language writing. Wolfe-Quintero *et al.* (1998) gave a detailed review of the previous studies. They classified all criterial measures into four categories, namely fluency, accuracy, grammatical complexity and lexical features (lexical variation and sophistication). Their work has been widely cited in the literature, inspiring more studies in recent years. The researches reviewed in this section were mostly published after 1998.

In the ESL context, Shaw and Liu (1998) examined indicators of development in L2 writings by international students about to start their undergraduate or graduate study in the UK. The participants, coming from 15 different language backgrounds, wrote a 30-minute essay on either of the two topics assigned prior to their full time summer course in English for Academic Purposes. At the end of the summer course they wrote on the same topic again. A comparison of the two sets of compositions indicated that the

sophomore and senior years respectively.

writings of the group became significantly more academic in that they used fewer contraction, fewer first person singular pronouns, fewer cause clauses, more connectors, and more passive voice expressions. On the other hand, the researchers failed to find such expected changes as improvement in linguistic accuracy (measured by fewer errors per T-unit) or improvement in features of academic styles².

Grant and Ginther (2000) were interested in discovering the utility of a computerized tagging program in distinguishing differences in L2 learners' writing proficiency. The essays in question were written by L2 writers and were all rated holistically using a 6-point band rating rubric. Thirty essays with scores of 3, 4 and 5 were randomly selected and tagged by a computer program to analyze linguistic features. They found that as proficiency level improved, L2 writers tended to write longer essays and longer words.

Larsen-Freeman's (2006) research is more theory-driven, presenting a different view as to "what language is and how SLA transpires" (p. 614). Empirically, the study featured five participants who wrote four narratives over a six-month period. The writings were coded for fluency, grammatical complexity, accuracy, and vocabulary complexity. The group averages showed improvement in all four measures. Individual performance, however, exhibited greater variation.

In the EFL context, Wen *et al.*'s (2003) cross-sectional research, partly a replication of Petch-Tyson's (1998) work, discussed the issue of informality in writing. They collected 240 writing samples from four class years, 60 from each class year. Across the four groups of subjects, the writing samples showed a steady improvement in the number of words per sentence and the number of letters per word. The total number of first person and second person pronouns and demonstratives, emphatics, hedges demonstrated a steady drop except in certain subcategories. In the aspect of vocabulary use, across the four class years there was a decline in the use of 0-1000 words

² Shaw and Liu (1998) interpreted academic style features as "increased use of nominalization, reduction of clauses to prepositional or participle phrases, and either increased subordination or changes in the pattern of subordination" (p. 242).

(high-frequency words), and a rise in the use of 1001-2000 words and academic terms. It was therefore concluded that as learners furthered their study their writings improved in the sense that they featured longer sentences and a lesser degree of informality.

Sasaki (2004) examined the changes in 11 Japanese students' English writing over a period of 3.5 years during which data were collected 4 times. During the study period all the participants demonstrated some improvement in general English proficiency and in writing fluency, represented by two indices: the number of words written and the number of words written per minute, but the changes in writing performance lacked a uniform pattern.

Part of de Haan and van Esch's work (2005) examined the argumentative essays in English as a foreign language by Dutch EFL learners. 47 students wrote on a topic in 2002 and one year later, 31 of them wrote on the same topic. The results showed that from 2002 to 2003 the students demonstrated improvement in fluency in that they produced longer texts, longer sentences and longer words³. However, they fared worse in the measure of the type/token ratio, an indicator of lexical density.

The studies so far, longitudinal or cross-sectional, with large or small samples, have lent support for the development in L2 learner's writing in certain aspects, especially fluency, but not much is known about the other three aspects, since except Larsen-Freeman (2006) no study has actually examined all the four aspects of linguistic features summarized by Wolfe-Quintero *et al.* (1998).

The present study aims to provide a comprehensive picture of English development among Chinese EFL learners in terms of fluency, syntactic complexity, lexical variation and sophistication, linguistic accuracy as well as the holistic rating, and to explore the possible institutional differences. To be more specific, I try to answer the following research questions:

- (1) Does learners' English develop in such linguistic measures as fluency,

³ However, in this aspect it is not the best group that performed the best.

- syntactic complexity, lexical features and accuracy across two years?
- (2) Does learners' English writing proficiency improve across two years?
 - (3) Is there any difference in the performance of learners from different universities?

RESEARCH METHODOLOGY

Participants

100 English majors from four universities in Shanghai (25 from each) were randomly selected for the present study. We selected four universities in the same city with an eye to minimizing the possible regional differences that might affect the findings. The four institutions included a comprehensive university (CU), a foreign language university (FU), a normal university (NU), and a polytech university (PU). The first two universities are both key national universities, whereas the latter two are key local universities in Shanghai.

Sample for Analysis

The 2003 TEM4 compositions and the 2005 TEM 8 compositions of the 100 participants were used as samples for analysis. The writing section in 2003 TEM4 required test takers to write a 150-word essay on the importance of keeping a good mood in 35 minutes, and in 2005 TEM8 test takers finished a 400-word essay on job interview in 45 minutes. The rubrics of the two writing tasks are presented in Appendix I.

The researcher is aware of the fact that the two tasks are qualitatively different, which makes the comparison of the two seem less proper. However, we believe they can best reflect test takers' writing proficiency at two different stages. First, the two tests are nation-wide tests for English majors designed in accordance with the English Teaching Curriculum for College

English Majors, which requires that sophomores should be able to write a 150- to 200-word essay in 30 minutes and seniors compose a 300- to 400-word essay in 45 minutes. Second, since the two compositions were timed essays completed in high-stakes exams, the best performance on the part of the participants can be expected.

The 200 essays were digitalized by the researcher and the hard copies were used for coding and rating.

Linguistic Measures of the Writings

Two coders each coded the 200 compositions for T-unit, clause, and errors. The researcher was one of the coders, and the other coder was a Lecturer of English and doctoral student in Language Testing. Lexical features, on the other hand, were examined with the online computer program *VocabProfile*⁴.

Fluency Measure

For the fluency measure, the present study employs two of the three best indices concluded by Wolfe-Quintero *et al.* (1998), namely, *T-unit length* (total number of words divided by total number of T-units), and *clause length* (total number of words divided by total number of clauses). The other measure, *error-free T-unit length* (total number of words in error-free T-units divided by total number of error-free T-units), is excluded because it has more to do with accuracy than fluency.

The present study follows Polio (1997) in defining T-unit as “an independent clause with all the dependent clauses” (p. 138). The guidelines for the coding of T-unit and clause are also drawn from Polio (1997).

Grammatical/Syntactic Complexity Measure

To examine the syntactic complexity of students' writing, *T-unit complexity*

⁴ Available at: <http://132.208.224.131/vp>

ratio (clauses per T-unit) is used, a measure recommended by Wolfe-Quintero *et al.* (1998) and also employed by Larsen-Freeman (2006).

Measures of Lexical Features

Wolfe-Quintero *et al.* (1998) concluded two measures of lexical features on the basis of their literature review: lexical variation and lexical sophistication, which are concerned with the range and size of a learner's productive vocabulary, or how varied and how sophisticated the words are. The best measure of lexical variation is found to be the *word type variation* measure ($WT/\sqrt{2W}$, total number of different word types divided by the square root of two times the total number of words), and the best index of lexical sophistication is the *sophisticated word type ratio* (total number of sophisticated word types divided by total numbers).

The two measures are calculated with the help of online program *Vocabprofile*. For each piece of writing, the program generates information including word type, word token and divides the words into four categories by frequency: 0-1000 (the most frequent 1000 words of English), 1001-2000 (the second most frequent 1000 words), academic words and off-list words⁵. A word here is defined as "a base form with its inflected and derived form" (Laufer & Nation 1995, p. 312), or word family. Two indices of lexical sophistication are examined: the percentage of 1000+ words and 2000+ words⁶.

Following the practice of Laufer and Nation (1995), before running the software *VocabProfile*, the researcher made a few changes with the data. First, all the spelling errors were corrected. Second, the proper names (like Jack or Shanghai), acronyms (like TEM or CET), and misused words (for example "applier" when the writer meant "applicant") were excluded, for they would

⁵ Off-list words refer to the words other than the first and second 1000 words and the academic words.

⁶ 1000+ words refer to those other than the 1st 1000 most frequently used words. 2000+ words are those academic words and off-list words.

otherwise have been categorized in the group of “off-list words”, thus inflating the values of the word sophistication measure.

Accuracy Measure

Wolfe-Quintero *et al.* (1998) suggested 3 accuracy measures, namely: *number of error-free T-units*, *error-free T-units per T-unit*, and *errors per T-units*. The first measure is excluded from the present study to minimize the length effect of the writing samples, as the two writing tasks in question have vastly different length requirement. The latter two measures are employed: *error-free T-units per T-unit*, and *errors per T-units*. Errors are coded following Polio’s Error Classification System (1997).

Holistic Rating of the Writings

Apart from examining the objective linguistic features of the writings, the present study also evaluates the writings holistically. The samples for the study, 100 compositions from 2003 TEM4 and 2005 TEM8 respectively, had already been rated by the time the study was done. However, since they were graded at different time periods (in 2003 and 2005 respectively shortly after the administration of the two tests) by different raters with different expectations⁷, it is problematic to compare the two scores.

For our purpose, these essays were rated independently by two raters. The researcher of the study was one of the raters, and the other was a colleague of the researcher. The two raters each gave a holistic score on a scale of 1-10, and the two scores were then averaged as the final score for analysis.

⁷ As the raters know the program level of the test takers (sophomores or seniors), they tend to have different expectations. In other words, a highly-rated TEM4 composition might not be an excellent piece of writing to a rater of TEM8 compositions.

RESULTS AND DISCUSSION

For the linguistic measures of the writings, the two coders each coded the 200 compositions for T-units and clauses, error-free T-units and errors per T-unit, and the average was taken as the final number. Table 1 reports the inter-coder reliability of these four variables. The samples were also rated holistically. The distribution of the scores and the inter-reliability of the two raters for the two sets of essays are reported in Table 2.

TABLE 1
Inter-Coder Reliability of Linguistic Measures

		Inter-coder reliability	Sig.
TEM4	Number of clauses	0.944	0.000
	Number of T-units	0.97	0.000
	Error-free T-units/T-unit	0.82	0.000
	Errors/T-unit	0.924	0.000
TEM8	Number of clauses	0.989	0.000
	Number of T-units	0.99	0.000
	Error-free T-units/T-unit	0.934	0.000
	Errors/T-unit	0.957	0.000

TABLE 2
Distribution of Holistic Scores and Inter-Rater Reliability

	N	Minimum	Maximum	Mean	Inter-rater Reliability	Sig.
Rater1 TEM4	100	4	8	6.54	0.917	0.000
Rater2 TEM4	100	4	9.5	6.77		
Rater1 TEM8	100	5	10	6.7	0.934	0.000
Rater2 TEM8	100	4.5	9.5	6.89		

Table 1 shows the inter-coder reliability of the four variables for the two batches of papers has all been found to be statistically significant. However, with regard to the coding of accuracy measures, the high reliability

coefficient should be interpreted with caution. Despite a high consistency between the two coders in the *number* of errors, there is a possibility that the two coders did not agree on what the errors actually were in a piece of writing, an issue pointed out by Polio (1997) as well.

The inter-rater reliability for the holistic rating of TEM4 and TEM8 compositions reported in Table 2 stands at 0.917 and 0.934 respectively, and both are statistically significant.

The linguistic measures and holistic ratings of the two sets of compositions are presented in Table 3. With two years intervening between the two tests, significant improvement in all the variables is expected.

TABLE 3
Linguistic Features and Holistic Ratings of TEM4 and TEM8 Compositions

		TEM4 Mean	TEM8 Mean	T	Sig.
Fluency	Clause length	9.58	10.22	3.911	0.000
	T-unit length	14.56	15.99	4.707	0.000
Syntactic Complexity	T-unit complexity ratio	1.53	1.57	1.27	0.207
Word Variation	Word type variation	5.17	6.31	17.516	0.000
Word Sophistication	1000+ words	18.28%	22.97%	8.657	0.000
	2000+ words	10.96%	15.43%	10.261	0.000
Linguistic Accuracy	Error-free T-unit ratio	67.13%	58.24%	-5.56	0.000
	Errors per T-unit	0.43	0.583	5.617	0.000
Writing Proficiency	Holistic rating	6.65	6.79	-1.667	0.099

Linguistic Measures

Fluency

In the studies reviewed by Wolfe-Quintero *et al.* (1998) clause length ranged from 5.2 words to 10.83. The participants here compose clauses with an average of 9.6 words in TEM4 compositions and 10.2 in TEM8

compositions, a statistically significant change.

The learners write an average of 14.6 words and 16 words per T-unit in the two essays respectively, with a statistically significant increase of 9.7%. However, compared with the highest of 23 words for advanced learners reported in Wolfe-Quintero *et al.* (1998), 16.1 words per T-unit is by no means a high value.

Grammatical/Syntactic Complexity

For T-unit complexity ratio, Wolfe-Quintero *et al.* (1998) concluded that this measure not only has a high correlation with program or school level, but is also correlated with other proficiency measures. The T-unit complexity measure from their review ranged from 1.07 clauses per T-unit to 2.17, with the latter coming from the most advanced learners. The results here indicate that on average the *clause/T-unit ratio* in TEM4 compositions stands at 1.53 (in other words, in every T-unit there are about 1.53 clauses), rising to 1.57 in TEM8 composition, but the minor improvement fails to find statistical support.

The insignificant improvement, we believe, is not totally unexpected, as Ortega (2003) already suggested that syntactic complexity metrics should not be regarded as “absolute developmental indices or as direct indices of language ability” (p. 494). Besides, as high school English course in China covers all the key grammatical points, by the time learners take TEM4 in their sophomore year, they should have been able to compose sentences complex enough, and another two years of study will probably not result in any big changes.

Lexical Features

For lexical variation, Table 3 shows that the 100 subjects as a group make significant improvement, suggesting that the words they use in TEM8 demonstrate a higher degree of variation.

Lexical sophistication also increases significantly with program level for this group of learners, whether it is operationalized as 1000+ words or 2000+ words. On average in TEM4 essays, 1000+ and 2000+ words account for 18.28% and 10.96% of the vocabulary respectively, and the ratios rise to 22.97% and 15.43% in TEM8 writing.

Accuracy

Regarding accuracy measure, Wolfe-Quintero *et al.* (1998) warned that none of the widely applied measures of accuracy were “clearly related to program or school level” as relevant studies had produced mixed results (p. 62). In the studies reviewed by them, error-free T-unit/T-unit ranged from 13.3% to 85.2% and errors per T-unit from 0.35 for advanced learners to 1.51.

The present study shows that as the learners in question continue their English study, the percentage of grammatically correct T-units actually declines, from an average of 67.13% in TEM4 essays to an average of 58.24% in TEM8 essays. The T test indicates that the drop is statistically significant. The participants also fare worse in the aspect of errors per T-unit, making 0.43 errors every T-unit in their TEM4 compositions, and 0.58 errors per T-unit in their TEM8 compositions, a significant increase of 35%.

In a word, the learners demonstrate a lower degree of accuracy in their writing as they further their study, a result contrary to the expectation but consistent with findings from relevant studies. Casanave (1994) and Shaw and Liu (1998) arrived at a similar conclusion, and the former explained this away with the following remark, “I have observed that as students relax into their writing and as they write more fluently thoughtfully, and insightfully, the grammatical accuracy of writing of some of them decreases” (Casanave 1994, p. 197).

The explanation makes sense in her case—a study of learners’ journals, but it does not seem to apply to the present study. As TEM4 and TEM8 are important national tests, we have every reason to believe that test takers want to demonstrate their best writing proficiency and are therefore more likely to

be risk averse. The researcher believes that the significant downward movement in accuracy results from a couple of factors.

First, the courses these participants took in the 3rd and 4th years after they had taken TEM4 were no longer language skill oriented but more related to their major, linguistics and literature, and half of the participants did not have any writing instruction in the two years. With their focus of study diverted from language, a decline in the accuracy of their written products is not unexpected.

The second factor concerns the vast differences between the two writing tasks in question. TEM4 requires test takers to finish a 150-word composition in 35 minutes while TEM8 expects a 400-word essay in 45 minutes. It can be safely concluded that TEM8 places a more stringent requirement on writing proficiency. For TEM4, given the limited length requirement, test takers can avoid using certain structures or expressions they are not sure about, and very likely they have enough time to monitor or to proofread their writings. On the other hand, in TEM8 when they must finish a 400- word essay in 45 minutes, the length and time constraint may pose a big difficulty for them⁸. As they have to produce a long piece of writing, the avoidance strategy may not work very well. When they are doing their utmost to write a 400-word composition, they will certainly have difficulty monitoring their writing. The time constraint also means that they will not have enough time to proofread their writings carefully. It is therefore not totally unexpected that the learners make more errors per T-unit and produce fewer error-free T-units in TEM8 compositions.

In summary, of the four linguistic measures, the 100 participants in the present study have made significant improvement in fluency, lexical variation and sophistication, slight but statistically insignificant improvement in syntactic complexity and significant regression in accuracy.

⁸ Of the 100 participants only 28 actually produced an essay of 400 words or above. The longest essay was 540 words long, and the shortest 190 words, with an average of 358 words.

General Writing Proficiency

A comparison of the holistic ratings of the two sets of essays shows that on average the participants in question write better compositions as they further their study. The mean scores for TEM4 and TEM8 are 6.65 and 6.79 respectively, but this improvement fails to find statistical significance.

It is difficult understand why two years of study does not lead to a significant improvement in general writing proficiency. We believe it has a lot to do with a lack of writing practice for most participants. As discussed above, half of the participants did not have writing course in the two years intervening between TEM4 and TEM8. Brauer (2000) highlighted the significance of writing practice; writing can be a “powerful tool” for learning. “Learning to write” and “writing to learn” are equally important in EFL classes (p. 17). A lack of writing instruction and practice certainly does not help learners’ writing proficiency.

Institutional Differences

This section examines the performance of the participants from the four different institutions to determine if the group performance holds for each institution as well, as Casanave (1994) and Larsen-Freeman (2006) have warned us of the inconsistency between group average and individual performance. Table 4 reports the institutional performances in all the variables.

TABLE 4
Institutional Differences in the Variables

		CU		FU		NU		PU	
		TEM4	TEM8	TEM4	TEM8	TEM4	TEM8	TEM4	TEM8
Fluency	clause length	10.17	11.03	9.78	10.24	8.64	9.56*	9.73	10.03
			*						
	T-unit length	15.3	17.14	15.13	15.96	13.45	15.07*	14.37	15.78*
			*						
Syn. sophistication	clauses/T-unit	1.517	1.558	1.557	1.561	1.577	1.579	1.482	1.577

English Development of Chinese EFL Student Writers from Sophomore to Senior Years

Word variation	Word variation	5.49	6.58*	5.26	6.42*	4.72	5.84*	5.22	6.40*
Word sophistication	1k+ words, %	21.10	24.7*	19.2	24.3*	15.40	20.7*	17.40	22.2*
	2k+ words, %	13.10	17.7*	12.20	17.1*	8.80	12.5*	9.80	14.4*
Accuracy	Error-free T-unit	69.80	72.40	71.70	56.3*	66.60	54.1*	60.50	50.2*
	Errors/T-unit	0.39	0.343	0.363	0.594	0.417	0.646*	0.542	0.752*
Writing proficiency	Holistic rating	7.14	7.81*	6.92	6.84	6.35	6.3	6.19	6.22

(* indicates that the changes from TEM4 to TEM8 are significant at 0.05 level.)

It can be seen that the institutional performance in most variables is, generally speaking, in line with the group performance. For the fluency measures, two institutions show significant improvement in clause length and three in T-unit length. In syntactic complexity, all the four institutions make slight yet insignificant improvement, which is consistent with the results when the 100 participants are examined as a whole group. Also in accordance with the group performance, all the four institutions demonstrate significant progress in word variation and sophistication.

The biggest difference between the group and institutional performance lies in linguistic accuracy and the holistic rating of the compositions. For the former, participants from the CU (the comprehensive university) actually demonstrated some improvement, though statistically insignificant, whereas those from the other three institutions show significant regression. In the holistic rating, it was again the participants from the CU that exhibit significant progress, at 0.05 level, as opposed to the minor changes by those in the other three institutes. In a sense, the improvement in the holistic rating as shown in Table 3 is mostly driven by learners from the CU.

Many variables might have contributed to the differences in institutional performance, such as institutional factors related to curriculum, instructors' proficiency level and teaching methodology as well as the individual factors of the learners in question. An examination of all these differences is beyond

the scope of the present study. However, some variables seem to stand out as plausible reasons and thus call for some discussion.

The first factor is related to curriculum. As mentioned above, since TEM8 composition features a more stringent requirement than TEM4 composition, without regular writing instructions or enough writing practice, it would be difficult to compose a TEM8 essay satisfactorily. Of the four institutions here, the CU is the only one that offers three years of writing instruction, from sophomore year to senior year, while the other three institutions have regular writing courses for one or two years only. The other possible reason has something to do with the expansion of university enrollment. The consequences of university enrollment expansion have aroused a lot of concern and discussion in China these days. For the English Department at many universities, what comes with the expansion of enrollment is not just the growing student body, but more importantly, an increase in class size. Of the four institutions here, the CU, however, is the only one that has not expanded their enrollment, with a class size of 16 or 17 students. For the rest three, the class size is in the 20s or 30s. This difference might also contribute to the good performance of learners from that university.

CONCLUSION

A comparison between the TEM4 and TEM8 compositions of 100 Chinese EFL learners aiming at finding features of language development has produced some mixed results. The participants sampled for the present study demonstrated 1) significant improvement in fluency and lexical variation and sophistication, 2) slight and yet insignificant improvement in syntactic complexity and in the overall quality of the writings, and 3) significant regression in linguistic accuracy.

The results concerning measures of accuracy coincide with other studies using the same measures. A lack of improvement in this aspect may be due to several factors. First, the courses offered in the 3rd and 4th years were less

language skill oriented, and half of the participants did not even have a writing course. Second, the writing part in TEM8 is a demanding task as it requires test takers to finish a 400-word essay in 45 minutes, a requirement that many participants failed to meet.

The holistic rating in the present study is used to reflect learners' general writings proficiency. Despite the improvement in most of the linguistic features examined, the general quality of TEM8 essays does not improve significantly over TEM4 ones. Both sets of compositions receive a comparatively low rating: 6.65 and 6.79. This seems to echo the concerns of many researchers in English education in China. However, an examination of the institutional differences shows that participants from some university do make significant improvement. This is also the university that offers three years of regular writing course and the university that features small class size. More quantitative studies are therefore called for to test the effect of curriculum design and class size.

The present study has a few limitations. As the two writing tasks are not highly comparable, the tasks *per se* might have affected the results. Future studies should try to examine essays of more or less similar nature, as done by Zijlstra (2007), cited in Verspoor, Lowie and van Dijk (2008). The holistic rating employed by the present study may not tell too much about the writing quality. It would be ideal to use analytic rating, which will present a clearer picture of the strengths and weaknesses of the learners' writings. Finally, due to the scope of the present study, the causes behind the institutional differences are not probed into in greater detail, which could have shed more light on the instruction of EFL writing.

THE AUTHOR

Xiaoling Ji is associate professor in the Dept. of English, Foreign Languages College of Shanghai Normal University. Her current research interests cover EFL writing and language testing. Her recent publications

include *Development in EFL Writing of Chinese English Majors* (2007) and several papers at various journals in China.

Email: jxl_sh_cn@hotmail.com, xlji@shnu.edu.cn

REFERENCES

- Bräuer, G. (2000). Product, process, and the writer within: history of a paradigm shift. In G. Bräuer (Ed.), *Writing across languages* (pp. 15-22). Stamford, Connecticut: Ablex Publishing.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179-201.
- de Haan, P., & van Esch, K. (2005). The development of writing in English and Spanish as foreign languages. *Assessing Writing*, 10(2), 100-116.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123-145.
- Huang, W. X. (1996). Cong yici xiezuo lianxi kan dangqian yingyu zhuan ye xuesheng de "yingyang buliang" [Problems with English majors' writing from a writing exercise]. *Xiandai Waiyu [Modern Foreign Languages]*, 74(4), 38-42.
- Jiang, J. P. (1995). Nuli tigao xuesheng de yingyu xiezuo nengli [Greater efforts are needed to improve learners' English writing proficiency]. *Waiyujie [Foreign Language World]*, 60(4), 23-26.
- 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.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
- Pan, S. Z. (1992). Daxue yingyu xiezuo zhong de wenti yu duice [Problems and solutions in college English writing]. *Waiyujie [Foreign Language World]*, 45(1), 23-26.
- Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107-118). Longman: London and New York.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research.

- Language Learning*, 47(1), 101-143.
- Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning*, 54(3), 525-582.
- Shaw, P., & Liu, E. T. K. (1998). What develops in the development of second-language writing? *Applied Linguistics*, 19(2), 225-254.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231.
- Wen, Q. F., Ding, Y. R., & Wang, W. Y. (2003). Zhongguo daxuesheng yingyu shumianyu zhong de kouyuhua qingxiang [The tendency toward colloquial English in written English of Chinese university students]. *Waiyu Jiaoxue Yu Yanjiu [Foreign Language Education and Research]*, 35(4), 268-274.
- Wolfe-Quintero, K., Inagaski, K. S., & Kim, H. Y. (1998). *Second language Development in writing: Measures of fluency, accuracy, and complexity*. Second Language Teaching and Curriculum Center: University of Hawaii at Manoa.
- Yang, X. R. (1993). Cong dajuan kan jiaoxue—TEM4-93 zuowen yuejuan shouji [A look at the writing instruction from students' writing performance—reflections of rating compositions of TEM4 1993]. *Waiyu Jiaoxue Yu Yanjiu [Foreign Language Education and Research]*, 96(4), 70-72.
- Zijlstra, D. (2007). *Variation and development in L2 writing*. Unpublished master's thesis. Department of Applied Linguistics. University of Groningen, The Netherlands.
- Zou, S. (1999). TEM8 xiezuo nengli pinggu—yaoqiu, wenti ji duice [The evaluation of writing proficiency in TEM8—requirements, problems and solutions]. *Waiyujie [Foreign Language World]*, 76(4), 56-60.

APPENDIX I

Requirements of TEM4 2003 and TEM8 2005 Compositions

TEM4 2003 composition

People in modern society live under a lot of pressure, from education, career, or family. So it is important for them to keep a good mood under whatever circumstances.

Write on ANSWER SHEET ONE a composition of about 150 words on the following topic:

The Importance of Keeping a Good Mood-

You are to write in three parts.

In the first part, state specifically what you view is.

In the second part, support your view with one or two reasons.

In the last part, bring what you have written to a natural conclusion or a summary.

Marks will be awarded for content, organization, grammar and appropriacy. Failure to follow the instructions may result in a loss of marks.

TEM8 2005 composition

Interview is frequently used by employers as a means to recruit prospective employees. As a result, there have been many arguments for or against the interview as a selection procedure. What is your opinion? Write an essay of about 400 words to state your view.

In the first part of your writing you should state your main argument, and in the second part you should support your argument with appropriate details.

In the last part you should bring what you have written to a natural conclusion or make a summary. You should supply an appropriate title for your essay.

Marks will be awarded for content, organization, grammar and appropriateness. Failure to follow the above instructions may result in a loss of marks. Write your composition on ANSWER SHEET FOUR.

APPENDIX II

Writing Samples

TEM 4 Writing

The Importance of Keeping a Good Mood

You woke up on a nice, sunny morning just to find that you overslept with the alarm not working. As you rushed to the bathroom, you hurt your foot by tripping over the books on the floor. And just when you were about to have your breakfast cereal, there was no milk left. Well, at this moment, you faced with the situation where you could stay grumpy all day long, or you can laugh at yourself with a good sense of humor. A good mood sustains us all the happiness and energy throughout the day.

It is proved by experiments that people with merry moods are more focused and likely to achieve goals than those in bad moods. For example, when two are assigned homeworks, one of them complains and grumbles about how hard it is, and how much time he has to waste on it; while, the other one accept the fact with a bright attitude, and tries to finish the homework by all efforts. The later one is likely to be more successful, of course.

To keep a good mood, we need to cultivate patience, a good sense of humor, and a positive mind. A beautiful life is on our hand to control, once

we adopt the concept of the importance of a good mood.

TEM 8 Writing

The essential of interview

There is no better way to recruit your employee than meeting him up and giving him an interview. One does not simply decide who to hire by just looking plainly at the resume sheets, the documents cannot stand for any man's heart and mind for sure. The following paragraphs are the merits and necessities to hold up an IV for your potential workers.

Like the Hollywood film, "You've Got Mail", two strangers build their relationship on the internet, they do not know each other's real life and personality until they have really met in person. So is the case with recruitment, interview plays the role of a bridge that connects the interactions of the employers and the employees. It gives both of the two sides a chance to know what their future partners would be like. They could get a brief understanding of each other's needs and requirements. Mm serves on a mutual base for the employees and the employers.

For the employee's benefits, mm allows them a chance test the employee's honesty by checking if their abilities really match with the words on the resumes. The companies could as well observes the applicants' personalities and EQ's. Most important of all, to see if the applicants will fit into the position and the team workers.

For the employee's benefits, mm requires them to learn more about what the companies' features and images are; in this way, they would know how well they can fit in for the job. The mm also allows the employees to ask the companies questions and doubts in their minds. And they could get to meet their future boss and colleagues through this opportunity, it gives them a second chance to reconfirm their determination to work for the companies.

To conclude, it is never too troublesome for an mm. An mm only brings benefits to both the employers and the employees. It is the essential step to a good working environment and a better future.