

University of Groningen

The Dynamics of English Writing Development in Advanced Chinese Learners

Hou, Junping

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Hou, J. (2017). *The Dynamics of English Writing Development in Advanced Chinese Learners*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 7

General discussion and Conclusions

The aim of this research was to gain better knowledge of English proficiency development at Chinese universities. At Chinese universities, English is regarded more as a subject one needs to pass tests for rather than a language one would like to use. If students pass the College English Tests (CET) bands 4 and 6, they will have better chances to be employed and be successful in their careers. Actually, most university students are able to pass the CET-4 before they graduate, but still they do complain that they feel their English does not improve much at university despite spending a great deal of time on it. This complaint has led to the main research question of this thesis: Do university students improve their English proficiency during their time at university? This question inspired a series of five separate, consecutive studies.

The series started off with a survey to see to what extent students indeed feel they have improved in English and what their main motivation was in studying English. Then a study was conducted to compare three groups of students—those at the end of high school, those studying for CET-4 and those studying for CET-6 students—that showed that especially students studying for CET-6 seemed to have stagnated. Because it was difficult to believe that these students did not improve at all despite 18 months of studying English, the assumption was that these more advanced students might have reached a ceiling effect in common measures of linguistic complexity and sophistication and improved in other, more subtle areas of language use such as idiomaticity and text quality. These more subtle areas were investigated in detail in the studies presented in Chapters 4 and 5 respectively. The final study zooms in on two individuals in the advanced class, one who clearly made overall progress in English proficiency and one who did not. The question was if their individual trajectories could shed light on their respective developmental processes. Each of these studies is briefly summarized below, after which overall conclusions are drawn.

7.1 Summaries of the studies and overall conclusion

Chapter 2 reported on the results of a survey study at a highly selective university in Xián to find out how students felt about English, how they had scored on previous standardized tests, what motivated them to study English, and whether they felt they had improved their English during their university

courses. A detailed questionnaire was filled in by 418 bachelor students majoring in physics, mathematics and chemistry in their first, second and third year of their college education. By year four more than 90% of the respondents had passed the CET-4, but only 55% had passed the CET-6, indicating that CET-4 is relatively easy, but CET-6 is very difficult to pass.

Many responses had a 60%-40% division. About 60% of the students did not really like English and were not studying English because they were interested but because they had to pass the tests. More than 60% of the respondents felt social pressure to pass the tests and said they spent an inordinate amount of time on studying English. Most felt content with their teachers and classes, but only 40% of the students sought extra English exposure outside class in the form of movies or internet. About half of the students felt they had improved in English but the other half did not.

To find the main predictors for “test-driven motivation” and “perceived improvement” two regression analyses were conducted. The main predictors for test-driven motivation were the amount of social pressure felt, the number of times the CET had been taken, the amount of time and effort spent on studying English, and the degree of appreciation of formal instruction. The main predictors for perceived improvement were positive attitude, the amount of time and effort on learning English, previous academic test scores and out-of-class English exposure.

Finally, an interesting observation was that the students in the lower years had higher NMET scores, suggesting that the student population was changing. The study on the whole confirmed the impression that more than half of the students felt they spent a lot of time on English but did not improve much. To investigate whether we could find evidence confirming their complaints, we conducted a series of studies in the subsequent chapters.

Chapter 3 presented the first of this series of studies and constituted an exploratory study comparing potential gains in L2 writing development in three student groups at different levels: a senior high school group (23 students) with 30 months’ of instruction, and two university groups of two different proficiency levels with 18 months of instruction. The lower level students (n=8) were preparing for the CET-4 and the higher level students (n=18) for the CET-6. All students wrote 12 to 20 texts as part of their class assignments. For calculation the different measures and comparing pretest

and posttest measures, the averages of several measures of the first two and last two texts written in their classes were calculated. These measures included both holistic and analytic measures for which the following procedure was used; All pre- and posttest texts (196) were scored holistically on proficiency by means of a rubric specifically designed for this study asking raters to consider the degrees of linguistic complexity, accuracy, fluency, idiomaticity and coherence (CAFIC). In addition, the texts were coded analytically on 47 complexity measures with Synlex Analyzer (Lu, 2012). The results showed that in holistic scores the high school group showed improvement in complexity and a trend towards improvement in idiomaticity; the lower level university group showed a trend towards improvement in fluency. Interestingly, the higher level university group showed no gains at all. As far as analytical measures was concerned, the high school group and the lower level university group both showed increases in various complexity measures, but different ones per group (see Table 3-7 for details). The highest level university group seemed to develop only in some isolated, subtle lexical measures and non-finite constructions. The findings in this study showed that progress was made in high school, some progress was made at the lower level in university, but hardly any (if any) at the higher level in university in the selected measures. This last finding seemed to substantiate the complaints that were found in the first study reported on in Chapter 2. Still, it was difficult to believe that no progress was made at all at this advanced level despite 18 months of instruction and so much time and effort expended by the students. These findings led to the hypothesis that students at this advanced level no longer progressed in purely linguistic complexity measures as they might have reached a ceiling level, but had developed in more subtle areas of linguistic and textual aspects, such as idiomaticity and coherence. Therefore, we developed two pre-post group studies to explore their chunk use (representing idiomaticity) and their ability to write cohesively and coherently as tested by both a topic-based coherence and a Coh-Metrix-based cohesion.

Chapter 4 was a pre-posttest study on chunk use and development in the advanced group. From a usage-based perspective, ‘chunk’ refers to a conventionalized way of expressing a certain concept with sequences consisting of two or more orthographic words, some of which might include

variable slots. We classified chunks into grammatical and lexical types with 7 sub-categories as in Verspoor et al. (2012). Each chunk was hand-coded for type and length in number of words. Several calculations were done to see if there were correlations with the holistic scores and if there were any differences between the pretest and posttest writings. The findings were that ‘chunk coverage’, the total number of chunk words divided by the total number of words, correlated with several holistic proficiency scores, and increased significantly over time. Average chunk length did not show any correlations with proficiency scores nor did it increase over time. As far as type of chunks was concerned, it was found that more proficient writers used relatively more chunks, specifically collocations (one of the lexical chunks, containing at least two words in it, e.g., ‘pay attention’, ‘bright future’) and the use of especially collocations was found to increase from pre to posttest.

Chapter 5 was another pre-posttest study in this advanced group to explore the development of text quality, operationalized as coherence and cohesion. Coherence refers to the general semantic relationships between sentences that are necessary for the reader to make sense of the entire text. Coherence was operationalized as topic-based coherence in which the number of related concepts are counted and the density of moves (DoM), operationalized as the average number of moves in 10 T-units, and strength of moves (SoM), operationalized as the average distance between these concepts, is calculated. Cohesion was operationalized as the explicit mention of cohesive devices (local cohesion) and the semantic overlap between sentences and paragraphs (global cohesion). As far as coherence is concerned, the statistical analysis revealed that SoM showed very few correlations with the CAFIC proficiency measures and showed no significant difference between the pre and posttest. However, DoM correlated with almost all CAFIC proficiency measures, not only when all the data were taken together but also with the pretest and posttest data separately. Moreover, there was a significant difference between the pre and posttest. We may conclude that DoM is a good measure of text quality and that the students improved significantly in coherence, operationalized as making fewer moves per 10 T-units. These findings on the relationship between topic-based coherence and holistic proficiency are consistent with Watson Todd et al.’s (2004) findings that the strength of relationships between key

concepts does not correlate with the teachers' score for coherence, but density does. As this measure is relatively easy to establish as only T-units and key concepts have to be established, it may be a useful measure in assessing coherence. As far as cohesive devices are concerned, counted automatically with the Coh-Metrix tool (Crossley et al., 2016), only three measures, two global and a local one, correlated with human judgments of overall text quality as measured with the CAFIC rubric.

The findings might mean that DoM, which was based on the subjective identification of key concepts, partially tapped into the same construct as the 3 content word overlap measures and 3 LSA overlap measures, indicating the text stayed more on topic and therefore had more focus at the end of the 18 months' course in this advanced groups' writings as compared to their work written at the beginning of the course.

Combining these results with the ones we have found in chapters 3, 4 and 5, we conclude that the advanced learners in our project made no progress in any typical CAFIC measures, nor in any syntactic measures, but they did improve in some more subtle lexical areas and idiomaticity and text quality as well, both of which may be argued to be higher language skills. All studies discussed thus far are group studies, which can explore how a number of variables in the group have changed over time. However, group studies cannot show how different variables develop over time and interact with each other; in other words, group studies cannot tap into the actual process of development. In Chapter 3, we found that the group as a whole did not make progress in either the holistic scores nor in the analytical scores, but still some students did and some did not. To explore potential developmental differences between strong and weak learners, one strong and one weak learner were selected from the advanced group and traced over the course of 18 months.

Chapter 6 traced a strong learner—one who had made progress in the holistic scores--and a weak learner—one who had stagnated--from a Complex Dynamic System Theory (CDST) perspective. Each had written 12 texts over the course of their 18 month course. We took our complexity, accuracy, fluency, idiomaticity and coherence (CAFIC) rubric as a starting point and traced both the holistic sub-scores and a number of analytical scores that had proven to be robust predictors of development in advanced writers from the previous studies. For linguistic complexity measures we

used average word length (AWL) and finite verb ratio (FVR) as these had proven to be robust in other longitudinal studies at the advanced level. In addition, the CAFIC measures and their sub-constructs, text length (for fluency), collocation ratio (Coll) for idiomaticity, and density of moves (DoM) for coherence were calculated. The results showed that the stronger learner developed most of the measures (especially the holistic ones) rather synchronously, suggesting there was coordination in the sub-CAFIC variables over time. In addition, her total CAFIC scores, text length and FVR developed synchronously and increased over time, confirming the idea that there was coordinated development in various measures. She also developed a significant peak in AWL, meaning she developed in using longer words, which are usually less frequent and more academic. The weak learner, on the other hand, showed a rather scattered pattern in the development of the CAFIC measures, and especially in the FVR, which represents syntactic complexity. The weak learner was also significantly more variable than the strong one. Towards the end there was stagnant development in CAFIC measures, accuracy and AWL. The differences between the two learners was further confirmed by the significant differences in the degrees of variability between the two individuals in text length and AWL over time.

To conclude, the questionnaire report in Chapter 2 seemed to substantiate the complaints among university students that they do not improve much in English during university. Most pass the CET-4 rather easily, but the CET-6 especially takes a lot of effort. Despite 18 months of instruction, the advanced group (studying for the CET-6) showed no clear indication of improvement. However, when we focused on the development of idiomatic language and text quality, we saw that the group did make significant improvement. The main reason may be that as far as common complexity measures is concerned, the advanced group had reached a ceiling level and they were able to develop in more subtle areas of language. Our last study on two individuals, one who did progress and one who did not, also shows that the strong learner had a rather automated and coordinated linguistic system and the weak learner did not. We might say that the university students' complaints of making little improvement in the English language learning is partially substantiated. There are indeed learners, like our weak one, who seem to be struggling and who stagnate, but there are

also some like our strong learner who progressed quite a bit. If we then reason back to our first study, we may carefully conclude that good learners probably already have higher English proficiency (NMET scores) to begin with, have a positive attitude towards English and seek extra English exposure in the media. One interesting observation is that the younger students in our study (in year 1) have higher NMET scores when they enter university, suggesting that they may have less trouble passing the CET-6 at university.

7.2 Implications

7.2.1 Implications for L2 development

Our studies explored differences and developmental patterns in group studies examining a number of linguistic subsystems that might retroactively explain growth or stagnation and found that for different groups at different proficiency levels, the development took place in different measures: lower level groups developed more than higher level groups in both holistic measures and complexity measures. We explain this from a dynamic perspective. According to CDST, L2 development is a complex dynamic process that is composed of many intertwining, interconnected elements that affect each other. As Verspoor et al. (2012) argued, different sub-systems develop at different stages or phases. In our first pre-posttest study (Chapter 3), the lower level students developed significantly in various linguistic complexity measures, but the higher level students did not. Further studies showed that the advanced students developed significantly in chunk coverage and collocations, suggesting that they used more idiomatic language overall. The final study (Chapter 6) suggested that the strong learner who had progressed in holistic scores in the course of her 18 month study had an automated linguistic system in which most holistic measures were coordinated, which allowed her to make further progress. Our studies give credence to the idea that at different phases of development, the whole system reorganizes and that would mean that different variables will change at different stages.

However, the developmental variables that worked well at the group level were not visible in the individual trajectories of development. For the developmental patterns of individual learners, we found that the holistic measures (CAFIC sub-scores) were better than the analytical ones, such as

collocation ratio or density of moves. The reason could be explained as follows. A human looks at a particular construct holistically. For example, human judgement can take into account that coherence is more than the sum of its parts: it has not only to do with staying on topic (density of moves) and referencing (keeping track of who does what), but also with referencing in a consistent manner and not repeating the same information. An analytical measure can tap only into one of the sub-constructs at a time and not a single one may be strong enough to represent the interaction of the various sub-systems. In group studies, the analytical measures gain strength through numbers, but in individual studies, the numbers of specific sub-constructs may be too variable to be used meaningfully. These overall findings confirm the CDST perspective that the L2 developmental trajectory is not linear but rather variable (Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor et al., 2008; Penris & Verspoor, 2017).

7.2.2 Implications for research

In addition to the theoretical implications addressed above, our choice and comparison of methodologies and measures can give other researchers information on what measures and methodologies to use how and when.

This project explored L2 proficiency development from both a general view focusing on group comparisons and a more detailed view focusing on individual developmental trajectories. Additionally, the project focused on both holistic ratings as well as analytic measures. To address the different sub-questions, we applied quite a few different methodologies, some of them newly created by the authors themselves and others established and well-known measures adopted from previous research on related topics. Some of these measures were manually coded and others were automatically calculated by programs already available. Each methodological approach was useful and fruitful in its own right, but the comparison of this variety of measures is rather interesting.

We started this project with a questionnaire designed to get the confirmation of the university students' complaints. The questionnaire consisted of 57 questions, all of which were designed to test aspects related to learning English. The internet WeChat forum was used to collect the questionnaire, which was very convenient for both the researcher and the

participants and saved time at this step. Altogether 418 out of 434 effective response sheets were received and analyzed in the study.

We created a rubric to assess essay quality; it included complexity, accuracy, fluency, idiomaticity and coherence (CAFIC). We compared the relationships within the rubric measures with each other and with other analytic measures. High correlations between all CAFIC measurements across all of our studies showed the CAFIC measures to be a better operationalization of the constructs in both the group as well as the individual studies than the analytic measures (e.g., Coll and DoM only proved good measures in Chapters 4 & 5). Previous studies used various classifications of chunks. For our purposes, we decided to build our own chunk corpus with a classification of grammatical and lexical types with 7 sub-categories in line with the purpose of our study. In Chapter 4, we hand-coded chunks in all pre- and posttest texts in the study presented and in Chapter 6 in all 12 texts written by the two individuals produced over 18 months. In Chapters 5 and 6, we hand-coded topic-based coherence in terms of SoM (the strength of moves) and DoM (the density of moves) following Watson Todd et al. (2004). The results showed that the advanced learners used more chunks in their writings after 18 months of instruction, specifically collocations. However, we would argue in line with Smiskova-Gustafsson (2013) that chunk coverage may be the best measure to operationalize idiomatic language as it does not rely on categorizing specific types of chunks nor on the length of each chunk. Moreover, it avoids disentangling embedded chunks. As far as coherence was concerned, the advanced learners developed significantly in DoM, which also correlated significantly with CAFIC measures. For the longitudinal study we used the collocation ratio and DoM to represent idiomaticity and coherence, but as mentioned above these measures showed no clear patterns of development in the two individual learners (Chapter 6). This suggests that the holistic scores are better used for constructs such as idiomaticity and coherence. However, it would be interesting to see if chunk coverage—which is very labor intensive to calculate—would have been an appropriate measure in a longitudinal study.

In addition to holistic measures and hand-counted analytical measures, various automated tools were applied in the studies. There are quite a few automatic linguistic tools available on the Web and choosing the right one in

a particular study is not straightforward. In Chapter 2, Synlex Analyzer (Lu, 2012) was used for an analysis of specific complexity measures and it was found that some isolated lexical complexity measures increased significantly in the advanced group, even though no differences had been found in the CAFIC measures. Web VP-Classic (v.4) was used for some automatic counting and calculations on tokens, ratios and frequencies in Chapters 1 and 6. In Chapter 5, the automated computer Coh-metrix tool was used to code for coherence and cohesion following Crossley et al. (2016). The standard deviation of LSA and the number of causal connectives significantly increased in the advanced group's writing data. However, each tool produces numerous measures, so one needs to be weary of pure data mining practices as the findings can be both overwhelming and erratic. Therefore, the user needs to carefully consider to what extent 'statistical' differences in isolated measures are meaningful or not meaningful as it is not known yet to which extent these measures tap into the constructs they aim to represent.

Finally, in Chapter 6, typical CDST methodology was applied: in addition to the visualization of developmental patterns in variables separately and in combination in terms of raw numbers, LOESS smoothed and normalized (0-1) data were used to be able to see beyond the day to day variability. All these were done in Excel. Monte Carlo analyses were run to test for developmental peaks or dips and differences in degrees of variability with Pop tools.

As in other scientific studies, SPSS was applied in most of the calculations and analyses on correlations and differences within or among variables and individuals. In the questionnaire study, a factor analysis and multi-linear regression analyses were conducted to see which factors influenced perceived proficiency and test-driven motivation. In most studies, we applied Pearson and Spearman correlation analyses, different *t*-tests, and non-parametric tests (if the data were not normally distributed) in calculating the relations and differences among or across the groups at pre-posttest levels, between measures within the advanced group, and between measures within or among individuals.

7.2.3 Implications for teaching

Every year about 10 million Chinese students take the NMET or CET-4

exam. Higher scores enable students to enter a better university or obtain better employment. Even though most students are relatively positive about English (Xu, 2010), most study English mainly to pass the tests, which we have referred to as ‘test-driven motivation’. This is unfortunate as pure test-driven motivation showed to be the worst predictor for success (Chapter 2). Two of the best predictors for success in passing the CET-6 and ‘perceived improvement’ was the score on the NMET and out of class exposure. Our findings may have some implications for English teaching at both senior high school level and university.

On the whole, we have argued that the tests that the students take are communicatively oriented, well-constructed and fair, so we do not assume that the tests have a negative washback effect on teaching. The classroom observations at the selective university where the studies took place have professional teachers who speak English in class and spend time on communicative activities in addition to explaining grammar. The main problem is thus neither the exams nor the teachers, but the fact that learning any foreign language--especially one that is typologically very distant like Chinese and English are--takes a great deal of time (which the students do spend) and unless the level of exposure is maintained, students tend to regress in skills, especially in reading as Xu (2010) has shown.

In our study, we found that the role of out-of-class exposure, such as watching English movies, rather than appreciation of formal instruction positively correlated with the positive attitude towards the language, which in turn was a positive predictor for higher perceived improvement, suggesting that in addition to the limited exposure in the classroom, the positive effects and importance of out-of-class exposure should be stressed by teachers. In other words, our hard-working colleagues who teach English to university students should encourage their students to watch English movies, use the internet and practice writing more in English to develop target-like English proficiency. If the students were exposed in such out-of-class English environment quite often, they would not only feel they get better at English and pass the CETs, but they would probably also enjoy it more than listening to standardized listening exercises.

If we take a usage based view to language (Ellis, 2009; Schmid, 2015; Verspoor, in press), it is frequency of exposure to the target language within a meaningful setting that is the key to learning the language. In other words,

more meaningful exposure to the target language, which in turn may motivate the students to learn the language for its own sake, may be the key to improving the levels of English of Chinese students. There are several ways to achieve this in our modern age with access to internet with movies for all ages, games, and lots of information on almost any subject. Sun (2015) has already shown that extra exposure outside the class was one of the stronger predictors of L2 proficiency in children between the ages of 2 and 5. Both in senior high school and at university movies could be used to serve as input (Verspoor & Hong, 2013), which will motivate learners to learn English for the sake of English (rather than for taking the test) and perhaps at university, teachers could use texts that are relevant to the learners' field of study. If the student realizes that there are meaningful texts in English that contribute to his or her knowledge of his or her field of study, students may be more likely to learn English for communicative purposes.

One interesting finding from our questionnaire study was that already 25% of the students are seeking extra English exposure through the media and that these were also the students who were most motivated to study English. Another interesting finding was that the younger students at university (first year students) had higher NMET scores to begin with. At this point we do not know if this reflects stricter admission criteria or an overall rise in scores, but if the latter then this would point to more exposure to the English language at a younger age already taking place.

7.3 Limitations and directions for further research

The dissertation project was exploratory to begin with and new questions and studies emerged from the findings of the previous ones. Throughout the project, while analyzing the data in different ways, several limitations to the studies became apparent.

Our questionnaire in Chapter 2 was originally created on the basis of the basic research questions, but later we realized that other items about learning strategies and aptitude might have been useful. It would also have been better to question in detail the cohorts whose data was collected longitudinally. Furthermore, on the whole, for all the detailed studies the numbers were small and from highly selective institutions in a large urban area, so the results cannot be generalized to other contexts. There are also

limitations in the types of measures that were used. The findings in Chapter 3 suggested that no single complexity measure was robust for all proficiency levels and that for the highest levels, other metrics tapping into inter-clausal complexity could have been added. In the study into chunk use (Chapter 4) subjectivity may have played a strong role in identifying chunks, even though we tried to keep it as consistent as possible. However, due to the elusive construct of chunks, there may be better ways to identify, categorize and operationalize them as we used a number of categories that might not be a chunk in every one's view. Different coding schemas might have yielded different results. The same was true for our hand-counted measures in topic based coherence. In addition, the types of topics our participants wrote about were limited in genre, and therefore we should not generalize beyond this genre. In Chapter 6, we traced two individuals over the course of 18 months. Unfortunately we had only 12 texts. More texts and more learners over time would have been preferable. Also the variables that were decided upon might not have been really representative of the subsystems of the whole linguistic system.

Finally, we collected longitudinal free-response data from the three groups at different levels during their English years, but only the data from the advanced group was studied longitudinally. We hope to conduct further studies involving the two lower level groups.