Impacts of Microsoft Pinyin Input Software on CFL Essay Writing: A Pilot Study

Yu Zhu

Overseas Education College

Xiamen University

Xiamen, China

e-mail: zhuyu@xmu.edu.cn

Abstract—**Collecting data based on experiment, questionnaire, and interview, this study compared Chinese as a foreign language (henceforth, CFL) learners' attitude to, experience about, and scores on handwritten and computer-processed compositions. It was found that the majority of the subjects considered composition via handwriting and MS pinyin input software as different processes, had higher self efficacy on computer-delivered writing, performed significantly better when writing through word processor than using paper-and-pencil. Moreover, inter-rater reliability was found to be higher for compositions done on computer than on paper. Based on such evidence, the authors called for word-processor with pinyin input softwre rather than pencil-and-paper as writing tool for high-stake and/or large scale CFL writing assessment.**

Keywords-software; pinyin input method; computer assisted learning; writing assessment

I. INTRODUCTION

Writing is one of the two most crucial productive communication tools. There exists numerous literature on writing, yet few of them investigated the relationship between writing medium and writing process and/or quality. Although it is almost an consensus that "the days of students sifting through piles of file cards, producing detailed outlines, and handwriting drafts have slipped into the distant past" [1], what remains unclear is whether or not writing medium as word processor with pinyin input software or paper-and-pencil would have different impacts on writing itself, and if so, would the impacts be negative or positive.

The current study sets out to examine the impacts of writing tools as paper-and-pencil or word processor with pinyin input software have on writing assessment of CFL learners.

II. RESEARCH DESIGN

A. Participants

Participants of this study were studying abroad students in one of the keystone universities of south China. Over 200 foreign students were studying Chinese in the university. They were mainly from Southeast Asia and Europe or North America. They had diverse educational background and their age difference was huge.

The 17 volunteers were all taking the 2-credit core course "intermediate Chinese" in the spring semester of 2010. Five of them were male and the rest were female. Four of them were from European countries, and the rest were from seven Southeast Asian countries. Most of them aged between 22 and 35, although two were over sixties. By the time of the experiment started, they had been learning Chinese for one to nine years, with a mean of three years; their former experience of living in China ranged from 2 months to 2 years, with a mean value of one year or so. Mean of self-reported average duration of typing in the latest month was 90 minutes, with an average of 43 minutes spent on typing Chinese per day. Their self reported average duration of handwriting in the latest month had a mean of 80 minutes, with an average of 72 minutes contributed to Chinese character writing. Furthermore, they unanimously preferred Pinyin input methods than any other Chinese input software.

B. Instruments

1) Writing assessment

The experiment asked each participant to independently complete two essays. They are computer-delivered composition "My Favorite Female" and hand-written composition "My Favorite Male". Both essays should be around 300 Chinese characters, identify participant's favorite persons, and explain why he/she like these persons. The writing should be clean and characters should be legible with correct usage of punctuations.

2) Questionnaire

Questionnaire was intended to collect participant's demographic information and their experience and feelings of different writing tools. The individual information section asked participants to fill in their name (optional), gender, age, nationality, first language, duration for learning Chinese, time living in China, preferred Chinese input software, and time spent on typing and writing Chinese in the latest month. Questionnaire on experience and feelings about writing medium included 10 Likert scale items. The higher a participant was scored on item one, three, and five, the stronger tendency he/she would prefer word-processor as writing medium; the higher scores a participant got on

item two, four, and nine, the higher self-efficacy in computer-delivered composition. Item six and seven intended to compare participant's frequencies of finishing an essay with pencil or computer. Participant's response would be considered invalid if he/she scored higher on both items, which would reflect unserious considerations on these questions. Item nine and item ten respectively explored participant's perceptions of processes and qualities of writing with different tools.

3) Interview

The study applied structured interview. Thirteen participants agreed to take part in the interview. Each interview was tape recorded upon interviewee's approval. Major goal of the interview was to understand participant's feeling toward writing experiences using different medium. More specifically, following questions were asked: whether or not participant considered the two essays' topics equivalent in difficulty; which medium allowed participant more time on idea-thinking; which medium made them took more time on producing characters; which writing tool forced participant to spend more time on correcting spelling and/or grammatical mistakes; which medium offered them more time on polishing the composition; and their confidence levels on essays produced in the two medium.

C. Material and Investigators

The whole process of data collection happened in a multimedia speech laboratory in the university mentioned above. There were one central control machine and forty-eight terminals arranged in six rows and eight lines. A passage in the middle of the laboratory separated the terminals into left and right area.

The first author of this paper directed the data collection process. Three master's student majored in education or psychology received short-training in advance. They administered the writing assessments and interviewed the participants.

D. Procedure

Before the experiment started, experimenters announced the confidential terms, data collection procedures, and requirements for the participants. The participants were randomly divided into two groups sitting in front of terminals on each side of the passage. The writing assessment experiment applied an AB-BA design.

While participants in one group handwriting the assigned topic for 35 minutes, the other group typed their essays for the other topic. Five-minute break was arranged right after the first essay writing session. Then the two groups switched writing medium and finished their assigned topics in the second 35 minutes.

After the writing assessment finished, participants were asked to respond to the questionnaire for five minutes, followed with a short interview with the investigators. Participants picked their souvenir and left when the interview was done.

III. DATA ANALYSES AND FINDINGS

Data analyses included deciding a score for each essay, calculating inter-rater reliability, and comparing the mean scores for essays done with different writing medium.

A. Deciding Essay's Score

Adapting analytic scoring criterions as described in [2], the study focused on three aspects of an essay, namely idea and content, linguistic expressions, and cohesion and coherence. Each aspect was evaluated in terms of qualitative and quantitative quality. Thus the final scoring table had six scoring aspects. Each participant's score on each scoring aspect ranged from one to five points. In other words, raw score for each essay would be in between 6 to 30 points.

Scoring of influential writing assessment programs conventionally adopted grading mechanism instead of giving a raw score for each essay so that only raw score difference leading to grade difference would be counted as a meaningful difference. Borrowing such experience, the study also transformed raw scores into grade scores. Table I showed each grade point and its corresponding raw scores.

TABLE I.   RAW SCORES AND CORRESPONDING GRADE POINT

Raw Score	Grade Point
6~10	1
10~15	2
15~20	3
20~25	4
25~30	5

Two scholars having experience of teaching CFL over three years were invited to be the chief raters of this study. One of them was a female professor, the other one was a male associate professor. Independently, they took anonymous review and rated a raw score for each essay. The Pearson correlation coefficient for their raw scores was 0.52 for handwritten essays and 0.63 for computer-delivered essays. Such finding was consistent with results of many previous studies, for example [3], which implied that essays done with word-processor might achieve more homogeneous scorings among different raters.

Since the two inter-rater reliability were both lower than 0.70, a third rater were invited, who had rich experience in teaching writing courses in the department of the participants. Based on the convention of large-scale writing assessment, mean of the two grade points given by the chief raters was assigned as the final score for an essay when the difference of the two grades was less than one; otherwise, the median of the grade points given by all the three raters was the final score of that essay.

B. Analysis of Questionnaire Data

Valid responses on the tow items about the participants' frequently used writing tool revealed that about half of the

participants usually used both tools, another one third or so participants often used computer to deliver a composition. Students who used paper-and-pencil more frequently and those who had no clear tendency each accounted for 8% of the sample.

Results for the items regarding the feeling of the writing process and quality were as following: (1) about three fourth of the participants considered wrting with computer or pencil as two different processes; nearly 19% of them had no idea whehter the two processes were different or not; and only 6% participants thought the processes were the same no matter which medium was used to deliver a composition. (2) about 59% participants thought the quality of essays were different when they were produced with different writing tools; 29% of them were not sure if the quality were same; and only 12% participants denied a quality difference would be existed in between their essays produced in paper-and-pencil and on a computer.

For item about the preference of writing medium, around 70% preferred to word-processor delivered essay, another one fourth of them tended to like paper-and-pencil writing, and the rest showed no preference to either one of the two writing tools.

The item investigating self-efficacy in writing with word processor software reflected that about 63% participants had high self-efficacy in writing with computer, 31% had low self-efficacy, and the rest had no idea about their self-efficacy.

C. Comparison of the Essay Scores

An important assumption for comparing the two essays delivered by different writing medium was that the perceived difficulty levels were equivalent for the two essays. Although the participants' instructor had confirmed in advance that the topics of the two essays were about the same difficult for the participants, whether or not the participants actually agreed with their teacher's judgment can only be made clear through interview.

Thirteen of the seventeen participants accepted the interview. Two of the interviewees made it very clear that one of the two topics was more difficult; two thought writing with paper-and-pencil was more difficult and two thought writing with computer was not easier as using pencil. The rest seven considered the two topics were about the same in term of difficulty. Based on this information, only these seven participants' data was valid for comparison of the two essays' scores.

Research literature had contradictory conclusions for direction of the difference between the paired sample means of essays written with computer and pencil. However, the descriptive statistics (see Table II for more details) showed that as a whole, mean of the scores of computer delivered essays was higher than that of essays written with paper-and-pencil. Furthermore, the questionnaire data indicated that majority of the participants had a high

efficacy in computer-delivered essay. Therefore the null hypothesis of the study was: mean of the participants' word-processor delivered essay scores was higher than that written with paper-and-pencil. In other word, the statistical significance test was one-tailed.

TABLE II.   DESCRIPTIVE STATISTICS OF PAPER-AND PENCIL VERSUS COMPUTER-DELIVERED ESSAY SCORES

Types of Essays	n	Mean	Std. Dev.	Min. Grade	Max. Grade
Paper-And-Pencil Essays	7	2.57	0.53	2	3
Computer-Delivered Essays	7	3.57	1.27	1	5

Because of the sample size was quite small, and the essays' final score are multinomial ordinal data, nonparametric test of marginal homogeneity was applied in data analysis. The result was shown in Table III.

TABLE III.   TEST OF MARGINAL HOMOGENEITY FOR PAIRED SUBJECTS' SCORES OF ESSAYS VIA PAPER-AND-PENCIL AND WORD PROCESSOR WITH PINYIN INPUT SOFTWARE

	Paper-And-Pencil Group & Pinyin Input Software Group
Distinct Values	5
Cases in Off-Diagonal	7
Observed MH Statistics	18.000
Mean of MH Statistics	21.500
Standard Deviation of MH Statistics	1.800
Standardized MH Statistics	-1.940
2-Tailed Asymptotic Significance	0.026

Result of table III indicated that final mean score of computer-delivered essays was significantly higher than that of handwritten essays. Such a difference had statistical significance (p=0.026, n=7).

D. Analysis of the Interview Data

Thirteen participants accepted the interview. Reults showed that regardless of their medium preference, almost all participants thought that the speed of essay writing using a word-processor with pinyin input software was faster, and its accuracy was higher.

Some interviewees volunteerily reported that pinyin input software allowed them to correctly produce some characters for which they did not have a clear memory or even forgot. For example, some interviewees described the difference of the two essay writing processes as "hand writing Chinese characters depends on your memory. You have to get the memory of a character in order to produce it correct. In typing Chinese characters, you simply relied on

pinyin of a character. It doesnot matter if you forgot any part of a character as long as you still knew its pinyin. You typed the pinyin, and then can get all the homophones on your screen. And all you need to do is to select based on your blurred memory which is not that hard at all."

Ovearall, analysis of the interview data cross-validates the analyses of questionnaire data and experiment data. The majority of the participants had a preference of word-processor with MS pinyin input software as writing medium and a higher self-efficacy in computer-delivered compositions. Moreover, interview data provided richer explanatory information for the above findings.

IV. DISCUSSION

Results of statistical analyses showed that in general, the participants had a higher mean score for essays written with word-processor than with paper-and-pencil. And such a mean score difference reached statistical significance level. This finding implied that in CFL writing assessment, compared to the conventional writing tool (i.e. paper-and-pencil), taking computer input software as writing medium could better reveal examiniees' writing potentials, allow the examinees to have a generally better score, and promote validity of the writing tests. Although this finding was inconsistent with some previous literature (e.g., [4], [5], [6], [7], [8], etc.), the authors regarded the descripency as caused by the difference between the Chinese and English input software. Specifically, most CFL learners use pinyin to type Chinese on a computer, in other words, a pinyin input software can bridge the gap between alphabetic input and character output. Compared to handwriting, pinyin input software usually can produce Chinese characters in a faster and much more legible manner (please also referred to the section of analysis of interview data). While for typing English on a computer, regardless of one's native langauges, each participant has to input and output alphabets, which does not provide any "bonus" effect of input software.

The current study also investigated particpants' accustomed writing medium, preferred writing tool, the writing tool s/he feels competent, experience and self-efficacy about the writing processes through questionnaire and interview questions. Results showed that most participants got used to writing in either interested medium. Quite some of them used computer to write more frequently, and only few were more accustomed to handwriting an essay. Many considered the processes of handwriting or typing an essay different, and had a higher self-efficacy on computer-delivered writing.

CFL leanrers preference of computer-delivered writing and higher self-efficacy in it could be summarized as faster speed, more convenient operation (as usually, the writer only need to type pinyin and then make a selection), and more desirable output (correct and neat).

It is widely acknowledged that the focus of any writing assessment is the writing ability per se of the examinees. Correctness of spelling or character writing as well as the legibility issue are no doubt unignorable aspects of scoring rubrics, however, sophisticated raters would pay much more attention to features such as the idea and logic development, orgnization of the linguistic units at different levels, and effectiveness of communication. Interview as well as experience told us that in CFL writing tests, foreign learners, examinees from non-Chinese-characer cultural sphere in particular, frequently encountered situations of partial or total memory loss of some once familiarized characters. Many of them struggled with producing legible characters. Consequently, they spent quite some time in character writing while time they could have spent on writing itself was substantively reduced. Furthermore, in essay grading, a large number of CFL learners lost points for causing antipathy of raters with messly produced or hard to recognize essays. As a result, validity of any CFL writing assessments specficying paper-and-pencil as the sole writing tool would have to be doubted.

In conclusion of the discussion, authors of this paper advocate pinyin software sun as MS pinyin input method as writing tool for high-stake and/or large-scale assessments of writing competence of CFL learners. More prudently, whereever it is possible, exmaniees should be allowed to choose character input tool for writing assessments.

REFERENCES

[1] P. C. Kuriloff, "Rescuing writing instruction: how to save time & money with technology," Liberal Education, 2004: 90(4), 36-41.

[2] D. Nie, "A summary of the development of grading criteria of HSK writing assessment," Journal of Yunnan Normal University, 2009: 7(6), 15-20.

[3] H. K. Lee, "A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test," Assessing Writing, 2004: 9(1), 4-26.

[4] C. O. Sweedler-Brown, "Computers and assessment: the effect of typing versus handwriting on the holistic scoring of essays," Research & Teaching in Developmental Education, 1991: 8(1), 5-14.

[5] D. E. Powers, M. E. Fowles, M. Farnum, and P. Ramsey, "Will they think less of my handwritten essay If others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays," Journal of Educational Measurement, 1994: 31(3), 220-233.

[6] M. D. Roblyer, "Technology and the Oops effect: finding a bias against word processing," Learning and Leading with Technology, 1997: 24(7), 14-16.

[7] E. W. Wolfe, and J. R. Manalo, "Composition medium comparability in a direct writing assessment of non-native English speakers," Language Learning & Technology, 2004: 8(1), 53-65.

[8] H. Breland, Y.-W. Lee, and E. Muraki, "Comparability of TOEFL CBT essay prompts: response-mode analyses," Educational and Psychological Measurement, 2005: 65(4), 577-595.