Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Effects of a Machine Learning-empowered Chinese Character Handwriting Learning Tool on Rectifying Legible Writing in Young Children: A Pilot Study

Lung-Hsiang WONG[a*], Guat Poh AW[a], He SUN[a], Ching-Chiuan YEN[b], Chor Guan TEO[b], & Yun WEN[a]

[a]*National Institute of Education, Nanyang Technological University, Singapore*
[b]*NUS-KEIO CUTE Center, National University of Singapore, Singapore*
*lhwong.acad@gmail.com

Abstract: The logographic nature of Chinese script is a major dissuading factor for learning handwriting. The challenge is the complex psycholinguistic process behind handwriting. Thus, we developed AI-Strokes, a Chinese handwriting learning tool that assists teachers in facilitating students' handwriting practice in various modalities, and provides personalized feedback for the students. By leveraging a trainable Machine Learning back-end framework, the tool diagnoses and scores students' handwriting errors. This paper reports a pilot study in a Singapore primary school with an early prototype of AI-Strokes. Two classes of students went through AI-Strokes-based Chinese handwriting lessons (the experimental group) and conventional lessons (the control group) respectively. Pre- and post-tests were administered, and their handwriting processes were analyzed regarding errors in stroke orders, extra/missing strokes, and errors in stroke directions. The results show that the experimental group has yielded significantly better learning gains than the control group. It is posited that the personalized feedback of AI-Strokes has formed a feedback loop to support students' trial-and-error process in improving their handwriting skills. The multimodal handwriting task design may have also fostered their orthographic awareness through the activation of alternative psycholinguistic pathways during their handwriting lessons.

Keywords: AI in education, Computer-assisted language learning, Learning of Chinese handwriting, machine learning

1. Introduction

The logographic nature of Chinese script is a major dissuading factor for novice learners in learning handwriting. Chinese handwriting involves complex visual-perceptual-motoring processing (Haas & Rees, 2010). Acquiring such an integrated processing skill is cognitively demanding for, in particular, young Chinese as a second language learners including most young students in Singapore (Wong et al., 2011). Furthermore, in the context of Singapore primary schools, hampered by big class sizes and limited lesson time, timely formative feedback by the teachers during individual students' learning process is hardly materialized.

This study seeks to address the challenges by developing AI-Strokes （爱写乐）, a web-based Chinese handwriting learning tool that assists teachers in facilitating handwriting lessons with various types of prompts, and providing personalized feedback to the students. By leveraging a trainable Machine Learning back-end framework, AI-Strokes could automatically diagnose and score handwriting errors of individual students.

This paper reports on a pilot study in a Singapore school with an early prototype of AI-Strokes. The quasi-experimental study involved two Primary 2 classes who went through AI-Strokes-based Chinese handwriting lessons (the experimental group) and conventional lessons (the control group) respectively. Pre- and post-tests were administered to both groups, and their handwriting processes were analyzed in terms of errors in stroke orders, extra/missing strokes, and errors in stroke directions. The research question is as follows, "Would the incorporation of AI-Strokes into the Chinese as a second language handwriting lessons in lower primary school levels help reduce students' Chinese handwriting errors, as compared to conventional lessons without personalized ICT support?"

2. Literature Review

2.1 Chinese character handwriting

The Chinese scripts are a principled and rule-based system. Each Chinese character comprises one or more components spatially arranged with certain principles (Liang, 2004). Each character is composed of strokes, the basic and smallest unit of a character without any semantic connotation (Lam & McBride, 2018), with fixed, codified stroke orders for identical components in different characters. Handwriting processes involve multifactorial pathways that connect phonology, orthography, and semantics (Yin et al., 2005). The process of writing a Chinese character can be summarized into three steps as follows,

Step 1: Retrieving the orthographic representation (or, the mental image) of a previously learnt character from long-term memory and store in working memory – specifically, the speech sound of the character activates the phonological representation of the character, which activates the lexical representation related by meaning via the lexical semantic pathway (Lam & McBride, 2018).

Step 2: Processing of the mental image in working memory with the stroke orders.

Step 3: Actual handwriting actions which involve the following sensorimotor performance components, all activated almost simultaneously (Klein, et al., 2011):- visual perception, visual skills, fine motor, and visual motor.

There are at least two levels of objectives for learning to handwrite Chinese characters: "legible writing" and "orthographic retrieval". "Legible writing" (Tsai et al., 2012) focus on Steps 2 and 3 - to handwrite "legible" characters with correct strokes, stroke orders, spacing, etc. Typical learning designs are "copying tasks", e.g., to display a full character for students to copy. Conversely, "orthographic retrieval" (Qu & Damian, 2015) tackles Step 1 where students handwrite characters without visual references of "model characters". The activities could be, (a) written picture naming task (e.g., to show an image of a tree, and the student writes 树); (b) dictation task (e.g., a sound clip utters, "生，生气的生", and the student writes 生); (c) translation task (for second language students; e.g., to show the English word "flower" and the student translates it mentally and write 花).

This study focuses on assisting teachers in improving young students' competencies in performing Steps 1 and 2 of the Chinese character handwriting process with AI-Strokes.

2.2 Automation of diagnosis of students' handwriting

Traditional technology has been used to facilitate the recognition and scoring of Chinese handwriting (Hsiao et al., 2015) according to established systems and rules. Using neural networks for recognition, compared to classical methods, a system may present an advanced alternative to human-based scoring. The ability to learn and establish parameters during the training process means no manual hardcoding is needed. This makes the system scalable and sustainable, and resilient to changes in handwriting styles.

The use of machine learning for recognizing handwritten numbers and English characters has been explored to a great extent, with early works on neural networks recognizing numbers with an error rate of 0.7% (LeCun et al., 1998). The use of Convolutional Neural Networks for recognizing handwritten English characters peaked in 2011, when Ciresan and team achieved an error rate of just 0.27% (Ciresan et al., 2011), making it comparable to human-like performance.

Diagnosis of Chinese handwriting, however, has not been explored in depth, with some having moderate success at handwriting recognition (Bai et al., 2014; Zhang et al., 2019; Zou et al., 2019). These attempts are empowered by large datasets or corpuses of written characters, whilst the system being developed performs machine learning on the process of the writing itself, thereby giving it more data to work with while doing diagnosis. AI-Strokes can also point out the errors made during the writing process (e.g., wrong directions of strokes), which the systems reported in the aforementioned publications cannot.

3. System Design of AI-Strokes

AI-Strokes is a web-based teaching aid that helps teachers manage and record the process of each student's Chinese handwriting and provide personalized diagnosis on their strengths and weaknesses, empowered by the advanced Machine Learning backend. With more student handwriting data being collected, the system can self-learn to analyze and score handwritten characters that it has not been trained with. The web-based tool should be accessed by touchscreen devices for a more natural handwriting experience. The system design is intended for use in classroom lessons (e.g., learning new characters in the textbook). The system user interface (UI) comprises teacher's console and students' client.

Before each lesson, the teacher creates a new lesson plan with the teacher's console UI via a laptop or a tablet computer by specifying the characters to practice. At the lesson, the teacher selects the pre-stored lesson plan and launches it, with the teacher's console UI being projected to the screen. Each student who is assigned a tablet may then log in and join the lesson with the student UI. Every time when the teacher administers a character for the entire class to practice, (s)he may provide the prompt through teacher's console in one of the following modalities: (1) the character itself (copying task); (2) an image or an animation that depicts the meaning of the character (picture naming task); (3) the equivalent English word (translation task); (4) a sound clip that pronounces the character with disambiguation (dictation task). The copying task corresponds to "legible writing" while the three other tasks correspond to "orthographic awareness" (see section 2.1).

In both modes, students are given immediate feedback (Figure 1) upon submission of their handwriting via the automated scoring system powered by the Machine Learning backend. The Machine Learning backend scores each student's writing with how well it matches the correct answer in terms of number of strokes, order of strokes, directions of strokes, and speed. The aspects of handwriting to be diagnosed is a subset of the validated Tseng's Handwriting Problem Checklist (THPC) (Tseng, 1993). According to THPC, there are 24 common Chinese character handwriting problems among young students. Only 16 of them may be automatically diagnosed by typical AI-empowered web-based software. For this early prototype of AI-Strokes, automated diagnoses on 8 of the problems are implemented, lumped into 4 categories as listed in the score card in Figure 1. In this early prototype, a submitted handwritten character may be awarded a star for each of the achievements of correct number of strokes, order of strokes and directions of strokes. Thus, the full score is 3 stars.

In teacher's console, there is an additional classroom view. After administering a character for the students to practice, the teacher may switch to the classroom view with all students' writing are displayed in real-time. The teacher may get a glimpse of how each student has done, or project it on the screen for class wide comparison and discussion.

4. Research Design of the School-based Pilot Study

Fifty (50) Primary 2 students from a public school in Singapore participated in the study, including 28 from the experimental group/class and 22 from the control group/class. The experiment comprised four one-hour lessons that spanned across 1.5 months. During the experiment, the teachers of both classes tapped on the regular Chinese lessons to conduct handwriting activities with the same sets of characters. However, AI-Strokes was employed in the experimental class while the prevailing handwriting instructions were carried out in the control class. In the control class, the teacher first demonstrated writing a character on the whiteboard. The students then practiced writing individual characters in the air with fingers or on their mini whiteboards and displayed them to the teacher. Eventually, all students were drilled in handwriting in their exercise books by copying the given characters or by dictation.

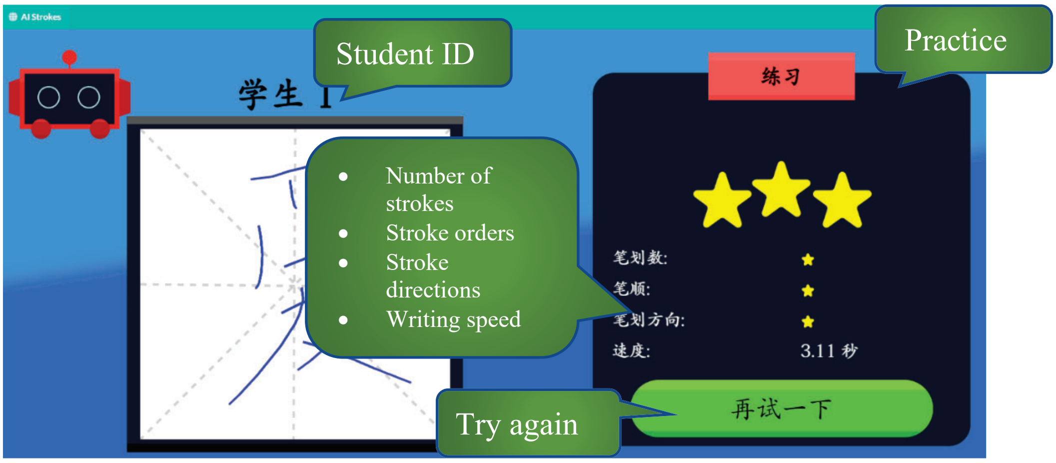

Figure 1. The student's UI of AI-Strokes with the handwriting field on the left and the score card on the right

The pre-test and the post-test took place before the first lesson and after the final lesson respectively and were administered as copying tasks. At each test, students used AI-Strokes to handwrite 10 characters – 瓜，亲，哭，汤，能，爱，起，笔，着，极. The system captured the students' writing processes (stroke orders) and products (the final appearance of each character) for analysis. The characters were selected from 30 characters covered in the four lessons with a maximum variety in the character components.

Stroke-by-stroke images of the writing process of each character were captured by the system for analysis. The process was assessed in: (a) Errors in stroke orders ("s-orders" hereafter): The number of strokes written in wrong orders; (b) Number of extra/omitted strokes ("eo-strokes" hereafter): The sum of the absolute number of extra strokes and the absolute number of omitted strokes – including cases where the student breaks one stroke into two or more strokes (e.g., writing the third stroke of 瓜 in two strokes – counted as 1 extra stroke), or write two or more strokes in one stroke (e.g., writing the three-stroke component 口 in one stroke or two strokes – counted as 2 or 1 omitted stroke(s)); (c) Errors in stroke directions ("s-dir" hereafter) (e.g., writing 横 (a horizontal stroke from left to right) as 撇 (a titled stroke from right to lower left) is considered an error). Thus, for all three indicators, the lower the numbers are, the better the student's performance is.

5. Findings

We employed SPSS 28 to analyze the pre- and post-test data (Table 1). To determine whether parametric or non-parametric Analysis of Covariance (ANCOVA) should be performed on the three indicators respectively (Corder & Foreman, 2009), the normal distribution tests of Shapiro-Wilk were executed. The results showed that there were normal distributions in both the pre- and post-test scores of s-orders ($p=.324>.05$ and $p=.258>.05$ respectively), while there was normal distribution in the pre-test score of eo-strokes ($p=.153>.05$) but no normal distribution in the post-test score of the same indicator ($p<.05$), and there were no normal distribution in both the pre- and post-test scores of s-dir ($p<.001$ and $p<.001$ respectively). Therefore, parametric ANCOVA was performed to compare the pre- and post-test scores of s-orders between the two groups, while Quade's non-parametric ANCOVA was executed to compare the eo-strokes and s-dir scores respectively.

The results showed that the handwriting performance of the experimental group was improved significantly as compared to the control group. By comparing the pre- and post-test means of the control group, there were even dips in their performance in s-orders and eo-strokes. Notwithstanding, the dips are insignificant, according to our supplementary paired samples *t*-tests (s-orders: p=.079>.05; eo-strokes: p=.353>.05; s-dir: p=.164>.05). On the contrary, the experimental group exhibited significant improvements in all indicators, as seen in the results of paired-sample *t*-tests (s-orders: p<.001; eo-strokes: p<.001; s-dir: p=<.05).

Table 1. *Summary of descriptive statistics and* results *of parametric or non-parametric ANCOVA on the three indicators of students' handwriting performance*

Indicator	Group	Pre-test Mean (SD)	Post-test Mean (SD)	*F*	*p*
s-orders (parametric ANCOVA) [1]	Experimental	77.43 (4.23)	70.93 (4.94)	11.447	.001
	Control	73.86 (3.91)	75.77 (3.89)		
eo-strokes (Quade's ANCOVA)	Experimental	3.96 (2.43)	1.96 (2.73)	28.587	.000
	Control	5.55 (3.14)	5.86 (1.71)		
s-dir (Quade's ANCOVA)	Experimental	1.75 (1.84)	0.89 (1.69)	8.812	.005
	Control	2.14 (2.05)	1.55 (1.54)		

[1] Levene's Test of Equality of Error Variances: *F*=.252, *p*=.618>.05.

High error rates of stroke orders in both tests and across both groups are observed in the statistics. The correct total number of strokes of the 10 tested characters is 88. Thus, the error rates according to the four mean scores of s-orders is ranging from 80.5% to 88.0%. The roots of the phenomenon can probably be found in the student demography and the teachers' instructional approaches, which will be discussed in the next section.

6. Discussion and Conclusion

This study found that the experimental group which experienced the AI-Strokes-empowered Chinese handwriting lessons has yielded significantly better learning gains in the three measured indicators, as compared to the control group. It is posited that the personalized feedback feature of AI-Strokes has formed a feedback loop to trigger and support students' trial-and-error process to improve their handwriting skills. The multimodal handwriting tasks have also played the role of fostering students' orthographic awareness through activation of alternative cognitive and psycholinguistic pathways during their handwriting lessons.

Yet why were there high error rates in stroke orders even at the post-test? Scholars (Lam & McBride, 2018; Law et al., 1998) reiterated the need to teach general stroke-order rules. When students have internalized these rules, they could handwrite any novel character with correct order. However, during the limited lesson time, teachers tend to focus on demonstrating stroke orders of individual characters, and only introduce scattered stroke order rules by chance. Students might ignore these rules but just memorize the stroke order of each character. In paper-based assignments, the correctness of stroke orders cannot be assessed as only the completed handwriting is submitted for grading. Thus, most students are not motivated to learn correct stroke orders as long as their "final products" "look right".

In our study, however, the experimental group students have constantly received feedback from AI-Strokes after handwriting each character. If the feedback indicated that their stroke order was not correct, they might try to figure out what went wrong by recalling what the teacher had taught before or asking around their classmates, with the hope of scoring higher in their next writing attempt. Thus, we postulate that AI-Strokes had subtly influenced the experimental group students to start paying attention to the stroke orders.

Despite the significant improvement in stroke orders among the experimental group students, their error rates in the post-test are still relatively high. By comparing their stroke orders between the pre- and post-tests, it is observed that at the post-test they were more likely to handwrite the characters with partially correct stroke orders. Yet most of the time they did not manage to handwrite whole characters with perfectly correct stroke orders. The reasons are two-folded. First, the prevailing issue of the lack of systematic explicit teaching of stroke order rules would have limited the students' development of stroke order "competency". Second, the feedback given by the current AI-Strokes prototype is relatively coarse-grained and is unable to pinpoint the sources of errors; but just alerts the students that "something was wrong" and they need to find out the errors through other means.

To address these limitations, we advocate a holistic revamp in AI-Strokes-based lessons. In the instructional aspect, the teachers should set aside the time to systematically cover the general stroke order rules. When a new rule is introduced, the teachers should let students practice writing multiple characters where this rule is applicable (rather than insisting to practice the characters appearing on the current textbook passage) on AI-Strokes, to help them internalize the rule and later apply it to writing novel characters with the same "stroke configuration". In the technical aspect, AI-Strokes should be upgraded to offer finer-grained feedback. AI-Strokes may play a crucial role in providing personalized formative assessment of handwriting, which is unable to be performed by the teachers.

Given its current machine learning backbone engine, AI-Strokes has the potential to be further upgraded to incorporate the capability of diagnosing additional aspects of THPC including spacing, spatial relationship, size and formation. More young children's handwriting samples need to be collected as training data for this purpose. With an upgraded AI-Strokes system, a longer-term study can be carried out with a full-fledged Chinese handwriting lessons to develop a stronger psycholinguistic and handwriting foundations in the students.

References

Bai, J., Chen, Z., Feng, B., & Xu, B. (2014). Chinese image character recognition using DNN and machine simulated training samples. International Conference on Artificial Neural Networks, Hamburg, Germany.

Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2011). Convolutional neural network committees for handwritten character classification. 2011 International Conference on Document Analysis and Recognition, Beijing, China.

Haas, B., & Rees, G. (2010). Multiple stages of cross-modal integration in visual processing: Comment on crossmodal influences on visual perception: By L. Sham & R. Kim. *Physics of Life Reviews*, *7*, 287-288.

Hsiao, H.-S., Chang, C.-S., Chen, C.-J., Wu, C.-H., & Lin, C.-Y. (2015). The influence of Chinese character handwriting diagnosis and remedial instruction system on learners of Chinese as a foreign language. *Computer Assisted Language Learning*, *28*(4), 306-324.

Lam, S. S.-Y., & McBride, C. (2018). Learning to write: The role of handwriting for Chinese spelling in kindergarten children. *Journal of Educational Psychology*, *110*(7), 917-930.

Law, N., Ki, W. W., Chung, A. L. S., Ko, P. Y., & Lam, H. C. (1998). Children's stroke sequence errors in writing Chinese characters. In C. K. Leong & K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages. Neuropsychology and cognition* (pp. 113-138). Springer.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278 - 2324.

Liang, Y. (2004). The characteristics of Chinese character components and Chinese as a Foreign Language teaching. *Language Education and Research*, *6*, 76-80.

Qu, Q., & Damian, M. F. (2015). Cascadedness in Chinese written word production. *Frontiers in Psychology*, *6*(1271), 1-11.

Tseng, M.-H. (1993). Factorial validity of the Tseng Handwriting Problem Checklist. *Journal of Occupational Therapy Association Republic of China*, *11*, 13-28.

Wong, L.-H., Boticki, I., Sun, J., & Looi, C.-K. (2011). Improving the scaffolds of a mobile-assisted Chinese character forming game via a design-based research cycle. *Computers in Human Behavior*, *27*(5), 1783-1793.

Yin, W., He, S., & Weekes, B. S. (2005). Acquired dyslexia and dysgraphia in Chinese. *Behavioural Neurology*, *16*, 159-167.

Zhang, X. Y., Wu, Y. C., Yin, F., & L., L. C. (2019). Deep learning based handwritten Chinese character and text recognition. In H. K., H. A., W. Q. F., & Z. R. (Eds.), *Deep Learning: Fundamentals, Theory and Applications* (pp. 57-88).

Zou, J. Y., Zhang, J. L., & Wang, L. (2019). Handwritten Chinese Character Recognition by Convolutional Neural Network and Similarity Ranking. *arXiv preprint*. https://arxiv.org/abs/1908.11550