

# Kanji Workbook: A Writing-Based Intelligent Tutoring System for Learning Proper Japanese Kanji Writing Technique with Instructor-Emulated Assessment

Paul Taele, Jung In Koh, Tracy Hammond

Sketch Recognition Lab

Department of Computer Science & Engineering

Texas A&M University

{ptaele, jungin, hammond}@tamu.edu

## Abstract

Kanji script writing is a skill that is often introduced to novice Japanese foreign language students for achieving Japanese writing mastery, but often poses difficulties to students with primarily English fluency due to their its vast differences with written English. Instructors often introduce various pedagogical methods—such as visual structure and written techniques—to assist students in kanji study, but may lack availability providing direct feedback on students' writing outside of class. Current educational applications are also limited due to lacking richer instructor-emulated feedback. We introduce *Kanji Workbook*, a writing-based intelligent tutoring system for students to receive intelligent assessment that emulates human instructor feedback. Our interface not only leverages students' computing devices for allowing them to learn, practice, and review the writing of prompted characters from their course's kanji script lessons, but also provides a diverse set of writing assessment metrics—derived from instructor interviews and classroom observation insights—through intelligent scoring and visual animations. We deployed our interface onto novice- and intermediate-level university courses over an entire academic year, and observed that interface users on average achieved higher course grades than their peers and also reacted positively to our interface's various features.

## Introduction

For students studying Japanese as a foreign language, one of the main components for achieving mastery of the language is fluency in reading and writing the language. Unlike in English and its sole script of the Latin alphabet, the Japanese language heavily utilizes three separate scripts, including the kanji script and its characters that differ greatly from alphabet letters in English. Not only does do kanji script characters have characteristics that challenge language students such as their large number, diverse structure, and tendency to have visual similarity among each other (Taele and Hammond 2009), but is one of several factors that makes the Japanese language one of the most challenging languages for native English users to learn (National Security Agency 2017).

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

In order to address these challenges, language instructors often employ various pedagogical approaches through stroke writing techniques for enabling students to improve their mastery in kanji script characters (Ye 2011), since such techniques are important in Japanese language writing (Takezaki and Godin 2008). However, for conventional classroom settings, instructors may not have the resources to provide direct assistance on assessing students' character writing performance with large classes or outside of class (Taele and Hammond 2009), and existing educational applications for character writing practice remain constrained on the depth of assessment that emulates those of human language instructors.

In this work, we propose *Kanji Workbook*, a writing-based intelligent tutoring for supporting novice Japanese foreign language students in their study and practice of kanji script character writing. Our work not only allows students to employ writing interactions for their study of the characters, but also provides a variety of assessment metrics that provide more granular feedback similarly to human instructors. We designed our interface based on qualitative insights from language instructors, and deployed the interface onto novice university courses in Japanese as a foreign language. From the classroom and interaction studies for evaluating our interface, we observed that students benefited with improved academic performance on their character writing capabilities, and discovered that students found the interface to be intuitive and the learning to be rewarding.

## Related Work

As a writing-based intelligent tutoring system, *Kanji Workbook* relates to other educational applications in terms of factors such as pedagogical goals, interface features, and recognition techniques. These applications include *Mechanix* (Valentine et al. 2012) for civil engineering, *iCanDraw* (Dixon, Prasad, and Hammond 2010) for art drawing, *MathPad2* (LaViola, Jr. and Zeleznik 2004) for mathematical formulas, and *Maestoso* (Taele, Barreto, and Hammond 2015) for music composition. While such applications share similar themes to *Kanji Workbook*, they provide educational assistance for domains that are not

directly related to foreign language writing.

Since *Kanji Workbook* also provides recognition capabilities for kanji script character writing, works on automated recognition of East Asian languages that use characters such as Japanese and Chinese also relate to *Kanji Workbook*. Works from this research area that have focused on recognition techniques for character scripts have continued to exist for the past several decades, and leverage machine learning approaches ranging from neural networks to hidden Markov models (Zhang, Bengio, and Liu 2017). Furthermore, systems that incorporate or advance the approaches from these works have become familiar in commercial tools that are available to the public (Pittman 2007). However, systems that incorporate these approaches focus more on optimizing the classification rates of casual handwriting of fluent writers, as opposed to providing more human-like assessment of novice handwriting of foreign language students.

Writing-based systems that both provide the intelligence as educational applications and focus on character scripts such as the Japanese language's kanji script have seen exploration since the 1990s (Ma, Huo, and Shi 2009), and many of these works focus on assessing the visual correctness of character writing. Other systems have considered the technical production of the character's strokes (An and Li 2011; Taelle and Hammond 2009; 2010; Chu, Taelle, and Hammond 2018), but limitations still linger. Specifically, the depth of assessment on students' character writing tend to focus only on a limited set of popular pedagogical approaches such as stroke order and stroke direction, while other types of assessments—such as for visual structure and writing precision—are little considered. As a result, *Kanji Workbook* aims to provide an writing-based intelligent tutoring system with richer assessment of students' character writing performance that better emulate those provided in human language instructors.

### Qualitative Study

Prior to developing our interface, we first conducted a qualitative study—consisting of instructor interviews and classroom observations—to discover real-world classroom best practices for character writing instruction. These practices would eventually be qualitatively analyzed into a list of insights that would lead to the design of the user interface and assessment system. In the study, we interviewed four instructors and observed two classrooms, all from different university-level foreign language programs that offered Chinese character writing instruction.

The materials (i.e., transcripts and notes) produced from the semi-structured interviews and classroom observations were qualitatively analyzed into encoded data units (Glaser 1965), where we define our units as concepts related to character writing instruction and also emphasized, repeated, or deemed important from the materials. From our initial selection of 269 data units, we applied a constant comparative method (Glaser 1965) that organized these data units into initially 54 data codes, then seven categories, and finally three emergent themes. Two of the emergent

themes—*in-class instruction* and *learning activities & resources*—led to insights that defined the interaction features of our user interface. The remaining emergent theme—*class challenges*—yielded valuable insights on a variety of students' common character writing mistakes that instructors considered important to assess for correctness. This particular insight led to the development of ten assessment metrics that was used in our interface and is further elaborated in the following section.

### Assessment System

The assessment system provides the automated recognition of students' kanji script character writing performance through our proposed collection of assessment metrics. Prior works have incorporated a small subset of the interface's assessment metrics for students' character writing such as stroke order and direction (Taelle and Hammond 2009; 2010). One of the novel contributions in our work was further expanding these existing assessments to a total of ten metrics, based on the insights on our earlier qualitative study that confirmed and expanded upon the few assessment metrics employed of those past works. Furthermore, these ten assessment metrics measured the students' character writing performance relative to model templates from an expert instructors' collection of character writing data. The following goes into greater details of the AI techniques that were employed in our assessment system.

### Character Writing Data Model Templates

In order to assess the students' character writing performance, their writing data was compared to model templates from an expert's character writing data. Specifically, the data was collected from one of the Japanese foreign language instructors who helped consult with the development of the interface. The expert was tasked with providing character writing data—consisting of the 448 characters that were used in the interface—as model templates for our assessment system. The expert was instructed to write the characters as models for comparison with students' character writing data through a specialized data collection application. The data format of the model character writing was stored in JSON format, and consisted of the following relevant components similarly utilized in prior sketch recognition systems such as LADDER (Hammond and Davis 2005):

- **Point.** The writing canvas samples points from the user's input contact on the interface's writing canvas. Each point collected is stored as an object with an x- and y-coordinate value as spatial data, and with a timestamp value as temporal data.
- **Stroke.** For each collection of points that was sampled from the user's start of writing (e.g., touching the screen) to end of writing (e.g., lifting off the screen), the interface groups them together as a stroke.
- **Sketch.** For each collection of strokes that was collected from an empty writing canvas to submission for assessment, the interface groups them together as a sketch.

- **Metadata.** Additional information is stored in the data as metadata for supplemental interface purposes. These metadata attributes include the character’s identification label for retrieval purposes, and its original canvas dimensions for scaling to the dimensions of the interface’s writing canvas.

Once the expert’s character writing data was collected as model templates for our assessment system, this collection was then pre-processed through transformation algorithms. This pre-processing step was performed to achieve faster assessment by normalizing the model template into a consistent size, so that only the students’ character writing data—which can be input from devices of multiple screen sizes (e.g., smartphones, tablets)—during the assessment process. For the normalization steps of the expert data, we leveraged the following transformation algorithms that were similar utilized in stroke-based template-matching recognizers (Wobbrock, Wilson, and Li 2007):

- **Resampling.** This transformation rearranges the template’s stroke points into 64 equally-spaced points.
- **Scaling.** This transformation resizes the template’s bounding box to a size of 250 pixels.
- **Translating.** This transformation shifts the template’s bounding box to the origin in Cartesian system-based coordinates.

### Generalized Assessment Recognition Process

In order for our interface to perform instructor-emulated assessments on the students’ character writing input, the assessment system follows a series of automated processing and recognition steps. The following list details a high-level summary description of each step for all the assessment metrics:

1. **Retrieve Model Templates.** Once the user starts the interface, perform a one-time retrieval of the model templates from the server.
2. **Record Input Data.** After the user writes each character and submits it for assessment, record the user’s current character writing data into local data structures.
3. **Normalize Input Data.** Run the same transformation algorithms that were used on the model templates on the user’s input data.
4. **Assess Input Data.** Perform each of the ten assessment metrics on the input data.
5. **Calculate Input Data’s Score.** Based on each assessment metric’s calculated results, calculate the input data’s performance score.

More specialized steps are required for the ten assessment metrics. Due to these metrics sharing similarity in both assessment concept and recognition process with other metrics, they have been grouped together into the following types: structure, technique, and precision. The following sections provide further details about the three assessment types.

### Structure Assessments

The first set of assessment metrics focuses on the correctness of students’ visual structure in their character writing. These assessments are important to instructors for better informing them on whether a students’ character writing is visually correct. Instructors conventionally assess students on their visual structure by reviewing the students’ input strokes on several factors such as whether they exist and are visually similar enough. However, instructors are challenged in assessing students’ visual structure such as determining lacking and extraneous strokes and whether a stroke was unambiguously written together or not. Our assessment system accommodates instructors by providing metrics for the following assessments (Figure 1): *stroke match*, *stroke valid*, and *stroke exist*.

An important step for assessing the visual structure correctness of students’ character writing is first finding the existing corresponding pairs between the students’ input strokes to the expert’s model strokes. This step is achieved by leveraging a strategy that is commonly employed by stroke-based template matching algorithms for classifying different classes of pen and touch gestures. Specifically, we leverage the Hausdorff distance metric for calculating the distance error between the user inputs and the model templates’ corresponding points from relevant template matching algorithms such as the \$P\$ gesture classifier (Vatavu, Anthony, and Wobbrock 2012). However, instead of automatically classifying user’s input gestures to labels of most likely corresponding model templates with the smallest distance error, we adapt this strategy to automatically classify students’ strokes to indices of most likely corresponding expert’s model strokes.

- **Stroke Match.** This metric assesses how many correct matches exist between the user’s input strokes and the expert’s model strokes.
- **Stroke Valid.** This metric assesses how many model strokes correctly match to the user’s corresponding input strokes, and also assists instructors in assessing whether a student has missing correct strokes.
- **Stroke Exist.** This metric assesses how many input strokes correctly match to the expert’s corresponding model strokes, and also assists instructors in assessing whether a student has extraneous strokes.



Figure 1: Structure assessment visualizations for: stroke match (left), stroke valid (middle), and stroke exist (right).

### Technique Assessments

The second set of assessment metrics focuses on the correctness of students’ written technique in their character

writing. These assessments are important to instructors for better informing them on whether a student's character writing is technically correct. Instructors conventionally assess students on their visual structure by physically observing their character writing in class. However, instructors are challenged in assessing students' visual structure due to time constraints of reviewing each student's character writing or less optimal solutions of requiring students to physically label their strokes in homework assignments. Our assessment system accommodates instructors by providing metrics for the following assessments (Figure 2): *stroke order* and *stroke direction*.

An important step for assessing the written technique correctness of students' character writing is first finding the existing correspondences of strokes and stroke endpoints between the student's input strokes to the expert's model strokes. This step helps determine whether the student wrote their strokes temporally correct in both order and direction. However, finding the correspondences of the endpoints between the user's input strokes and the expert's model strokes is non-trivial for assessing stroke direction, since there may be cases where both endpoints of a user's stroke may have similar distances to an endpoint of the corresponding expert stroke (e.g., small strokes, long strokes with close endpoints). As a result, we leverage the Euclidean distance metric for calculating the distance error between the user's input stroke and the corresponding expert's model stroke from relevant template matching algorithms such as the \$1 gesture classifier (Vatavu, Anthony, and Wobbrock 2012). This strategy differs from similar prior systems that also assess stroke direction (Taele and Hammond 2009) since we calculate the entire distance error to more accurately classify the correct corresponding endpoints. From this strategy, we summarize the AI techniques employed by our technique assessment metrics:

- **Stroke Order.** This metric assesses how many of the user's input strokes are written in temporal correct order.
- **Stroke Direction.** This metric assesses how many input of the user's input strokes are written in temporal correct direction.

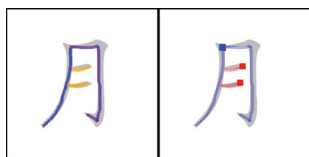


Figure 2: Technique assessment visualizations for: stroke order (left) and stroke direction (right).

### Precision Assessments

The third set of assessment metrics focuses on the correctness of students' precision performance in their character writing. These assessments are important to instructors for better informing them on whether a student is seamlessly writing the characters. Instructors conventionally

assess students on their precision performance through direct observation, but becomes a challenge for consistent assessment physical presence for larger class sizes. Our assessment system accommodates instructors by providing metrics for the following assessments (Figure 3): *stroke edit*, *stroke speed*, *stroke length*, *stroke closeness*, and *symbol speed*.

An important step for assessing the precision performance of students' character writing is determining the appropriate thresholds of the different assessment metrics similar to the assessment feedback of an instructor. To achieve this type of assessment, we employed a similar strategy found in sketch recognition systems such as (Hammond and Davis 2005) and its follow-up works that assigns pre-defined thresholds for feature-based classification of sketches. We adapt this strategy of our own pre-defined thresholds that were empirically derived from the expert model templates to develop features that classify the assessment level of students' character writing. From this strategy, we summarize the AI techniques employed by our technique assessment metrics:

- **Stroke Edit.** This metric assesses how many times the student edits their input strokes during their character writing session.
- **Stroke Length and Closeness.** These metrics assess the stroke length and spatial distances, respectively, between each matched pair of input and model strokes.
- **Stroke and Symbol Speed.** These metrics assess the differences in relative writing time duration between the input and models' strokes and characters, respectively.

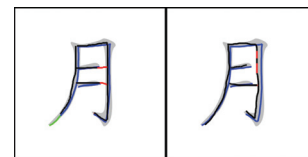


Figure 3: Precision assessment visualizations for: stroke length (left) and stroke closeness (right).

### User Interface

For *Kanji Workbook*, we implemented a user interface that takes inspirational design cues from systems for other domains (Dixon, Prasad, and Hammond 2010; LaViola, Jr. and Zeleznik 2004; Valentine et al. 2012), and also adapted similar visual and interaction cues to other related systems (Taele and Hammond 2009; Sloniger et al. 2018). We also incorporated novel components and features in our user interface that is elaborated in the following.

### Lesson Characters

The kanji script characters that were adapted in the interface were derived from the lesson chapters in the Genki textbook series (Banno, Ikeda, and Ohno 2011; Banno 2011) (Figure 4). This textbook was utilized since



Set 1									
日	本	一	二	三	四	十			
Set 2									
月	火	水	木	金	土	日	五	六	七
八	九	四	千	万	四	十			
Set 3									
曜	月	火	水	木	金	土	日	神	
Set 4									
上	下	右	左	半	間	日	田	今	大
学	生	母	親	父	中	見	行	私	印
人	先								
Set 5									
手	好	上	下	天	気	今	日	友	母
父	一	時	間	半	毎	先	生	男	女
朝	月	菜	食	子					
Set 6									
家	月	中	国	元	気	小	学	校	高
会	入	日	本	語	保	明	天	生	飲
年									

Figure 4: Subset of characters in *Kanji Workbook*.

it was the required textbook for the language courses that deployed the interface (see *Deployment Results*). The interface corresponds to the characters used in the chapters from the Genki textbooks for a total of 23 selectable lessons, which consisted of 448 characters combined.



Figure 5: Practice mode view.

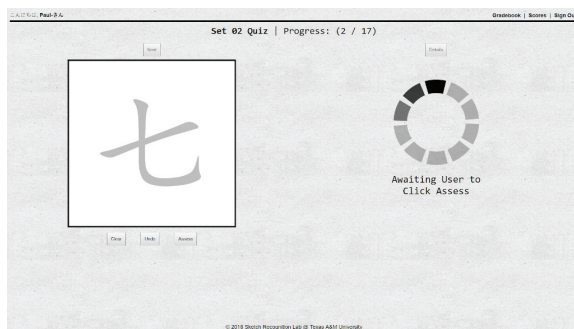


Figure 6: Quiz mode view.

## Writing Area Space

In some views of the interface, there is an area for users to provide their kanji script character writing on the left half portion of the interface (e.g., Figure 5). This writing space area is devoted to two smaller areas worth highlighting.

**Writing Canvas** The first smaller area consists of the writing canvas, which is an interaction space for users to

write the characters using stylus, touch, or mouse. The writing canvas also displays a semi-transparent visualization of the current character in the lesson for the user to trace over, and also displays animations and feedback strokes that overlay the user's own character writing.

**Interaction Buttons** The second smaller area consists of interaction buttons that surround the writing canvas. Conventional buttons include navigation (i.e., *Back*, *Next*), editing (i.e., *Clear*, *Undo*), and submit (i.e., *Assess*). Two addition buttons display informative animations: *Demo*, for showing the full character writing animation; and *Steps*, for a similar stroke-by-stroke animation.

## Character Information Space

While the user is practicing character writing, the right half portion of the interface displays supplemental information about the character (Figure 5).

- **Character.** A text visualization of the character.
- **Pronunciations/Translations.** The Japanese pronunciations and English translations of the character.
- **Vocabulary.** The vocabulary words from the Genki textbooks that use the character, and their corresponding Japanese pronunciations and English translations. The highlighted vocabulary words indicate words that are tested in the textbooks' supplemental paper workbooks.

## Assessment Area Space

Once a user submits a character for assessment, the right half portion of the interface displays assessment information on the user's character writing performance (Figure 7). More specifically, this space shows each of the assessment metrics and their corresponding animation buttons and scoring scores. Details of these metrics are described in the *Assessment System* section.

- **Animation Buttons** These buttons displays an animation that relates to their corresponding assessment metric.
- **Scoring Stars** This area shows a visualization of three stars for providing a three-point score of the user's character writing performance, where the number of displayed stars indicate the level of correctness for that particular assessment metric.
- **Color Key** This visualization is located in the writing area space, and is dynamically shown for certain animations to inform users of the purpose of the different color strokes that are displayed in the animation.

## Lesson Modes

Two modes are available in the interface for users to accommodate their lesson studies, in order to emulate their existing character writing practice and quiz activities .

- **Practice Mode.** The first lesson mode is the practice mode, which allows users to pursue character writing study in a workbook writing-like activity (Figure 5). In this mode, users are able to progress forwards and backwards for characters multiple times in the lesson, and

are also able to view supplementary character information during their writing.

- **Quiz Mode.** The second lesson mode is the quiz mode, which allows users to pursue character writing study in a quiz taking-like activity (Figure 6). In this mode, users can only progress forward for characters once in the lesson, are not shown supplementary character information during their writing. Furthermore, they receive a comprehensive assessment of their lesson-wide character writing performance at the end of the lesson (Figure 8).

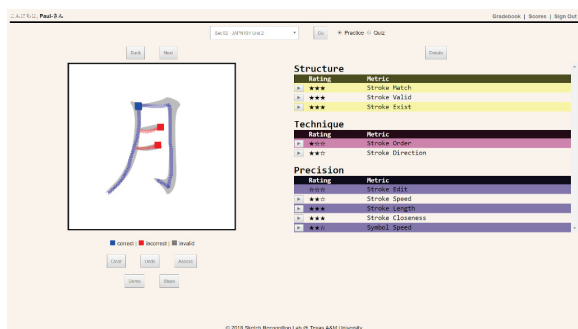


Figure 7: Character results view.

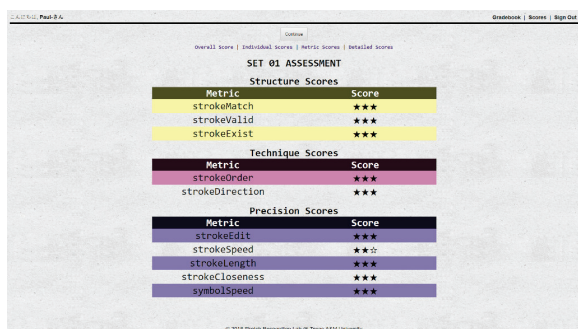


Figure 8: Quiz results view.

## Evaluation

### Classroom Study

Following the interface’s completed development, the interface was deployed onto 1<sup>st</sup>- and 2<sup>nd</sup>-year Japanese foreign language courses at Texas A&M University’s Department of International Studies. The study was conducted in collaboration with two instructors teaching these 1<sup>st</sup>- and 2<sup>nd</sup>-year courses, respectively, where enrolled students were observed over a time period spanning nine consecutive months (i.e., during the Fall 2018 and Spring 2019 academic semesters). A total of 94 students—29 females—from these courses took part in the classroom study. These participating students were partitioned into case and control groups, where the case group were tasked with using the interface in their kanji script character writing

study for their enrolled semester, and the control group was tasked with using conventional study habits with paper-based materials and optional online resources, (Table 1).

Table 1: Total number of classroom study participants.

LANGUAGE YEAR	CONTROL	CASE	TOTAL
1st Year only	34	16	50
2nd Year only	16	28	44
Total	50	44	94

During the study, students from both case and control groups were introduced to kanji script characters for study, as normally scheduled by their instructor in the curriculum. Over the span of their enrolled semesters, students were assigned five in-class tests with strong emphasis on vocabulary and sentence writing from the latest introduced characters. Students in the control group were instructed to utilize paper workbook assignments and any optional educational resources for learning the characters. Those in the case group were assigned to log into the interface and encouraged to practice their character writing on the interface’s practice mode sessions for learning the characters, in addition to their required paper workbook assignments.

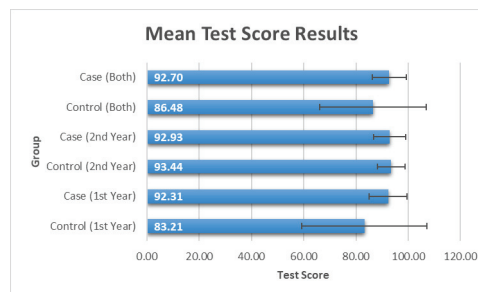


Figure 9: Mean test score results.

At the end of the deployment period, the pre-adjusted averaged test scores—on a 100-point scale from the five semester tests—were shared by the collaborating instructors of the 1<sup>st</sup>- and 2<sup>nd</sup>-year courses (Figure 9). From these scores:

- the 1<sup>st</sup>-year case group scored higher on average than the 1<sup>st</sup>-year control group by 9.10 points
- the 2<sup>nd</sup>-year case group differed on average from the 2<sup>nd</sup>-year control group by 0.51 points
- the entire case group scored higher on average than the entire control group by 6.22 points

One noticeable observation was that the 1<sup>st</sup>-year control group had a large standard deviation, due to the wide range of scores between low- and high-performing students in the 1<sup>st</sup>-year control group. In order to determine whether this observation affected whether the entire case group’s higher average scores were statistically significant, a t-test was conducted between the scores of the entire

case group and the entire control group. From the two-tailed t-test, the calculated p-value was  $p=.046<.05$ , which demonstrated that the test scores from the entire case group was still considered statistically significant (Table 2) despite the observation. When the two-tailed t-test was similarly performed for the test scores between the test scores of the 1<sup>st</sup>-year control group and the 1<sup>st</sup>-year case group, the calculated p-value of  $p=.048<.05$  also demonstrated statistical significance for the 1<sup>st</sup>-year case group's higher average scores compared to the 1<sup>st</sup>-year control group.

Table 2: Test score results statistics for test scores between the entire case group and the entire control group.

Statistics	Control	Case
Mean	86.48	92.70
Variance	416.7	42.26
Observations	50	44
P(T<=t) two-tail	0.046	

### Interaction Study

An interaction study was conducted that asked students to provide Likert-scale and freeform responses—with constructive feedback on the interface's features of buttons, ratings, and animations—in terms of intuitiveness and usefulness. For the Likert-scale responses, most students agreed or strongly agreed on the intuitiveness and usefulness of these three features. For the freeform, students shared positive feedback on the interface for providing a self-improving and rewarding experience while practicing their character writing, and also shared that the interface overall was intuitive and that the feedback and animations were useful. Of the three interface features, the ratings feature—which consisted of the three-star scoring system—received relatively lower Likert-scale agreement scores in terms of intuitiveness. That is, less students provided *Strongly Agree* responses for the ratings' intuitiveness, though the majority of students provided *Strongly Agree* responses for all the ratings in terms of usefulness. The freeform responses on the ratings' lower agreement scores for intuitiveness revealed that students either generally felt that the ratings were overly strict, or that ratings for specific assessment metrics (i.e., stroke speed, order, closeness, and length) were less intuitive.

### Recognized Improper Characters

An additional observation from the deployment was the types of students' improperly-written characters that were recognized by the interface. Nine different types were noticeably observed as having consistency among students or were uniquely not expected to be written (Figure 10).

### Discussion and Future Work

The results of the grading outcomes in the classroom study and the constructive feedback of students from the interaction study revealed various insights that were worth

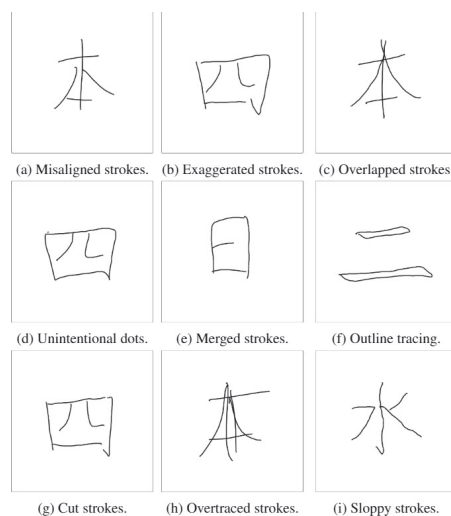


Figure 10: Examples of recognized user's improperly-written characters.

considering for improving and continuing the development efforts of *Kanji Workbook*. The first potential next step is in regards to the visual aesthetics and the intelligent features of the interface. Although the interaction study yielded positive feedback on the rewarding experience and intuitive interactions of the interface, students also expressed strong concerns regarding the weak visual aesthetics of the interface and occasional issues with the correctness of the animations or the discrepancies of the assessment scoring. We suspect that these issues may stem from potential technical issues with the assessment system and the need to further tweak the values for dictating the assessment of students' character writing performance.

The second potential next step is in regards to efforts of employing previous assessment classification techniques for the interface that yielded little successes. Specifically, one technique that was attempted was template matching for classifying the type of characters that students had written. However, this technique was initially unsuccessful due to visual factors in how students wrote their characters that were too noisy for our implemented template-matching algorithms. We are interested in re-visiting this approach for exploring additional assessment metrics that may be of benefit to students and of interest to instructors.

The third potential next step is in regards to further enhancing and specializing our evaluation, in order to discover additional challenges that 1<sup>st</sup>- and 2<sup>nd</sup>-year students face in their character writing study. Specifically, we would like to refine our classroom study's evaluation materials that incorporates more character writing-focused test materials, which we anticipate will be invaluable for more deeply revealing potential barriers—or lack of—that 1<sup>st</sup>- and 2<sup>nd</sup>-year students encounter. Insights from this revised evaluation will allow us to, for example, more confidently target assessments that cause issues to 1<sup>st</sup>-year students and better develop assessments that better challenge 2<sup>nd</sup>-year students.

Overall, the successes of the interface's deployment have not only led to its continued availability to the instructors of the participating courses for subsequent semesters, but have also opened up plans to adapt this interface for characters used in Chinese foreign language classes and for character writing in more advanced Japanese foreign language classes. We believe that *Kanji Workbook* has strong potential for becoming a richer resource supplement to foreign language courses that employ character writing.

## Conclusion

In this paper, we describe our work with *Kanji Workbook*, an educational application for novice students to practice their writing in and receive intelligent assessment on their kanji script character writing. Our work leverages various recognition techniques for assessing students' writing performance through intelligent scoring and visual animations, so that they may be better informed on their proficiency of Japanese language writing proficiency. From our evaluations, we demonstrated that not only did students overall benefited from the interface during course enrollment, but also responded positively to the interface's various features.

## Acknowledgements

We give our huge thanks to instructors George Adams and Yuki Waugh at Texas A&M University for their valuable assistance in providing consultation on our interface and access to their students. We also thank our former undergraduate students Dakota Sloniger, Rupen Sakariya, Carlo De Guzman, Jacob Mathews, and Erwin Susanto for their development work in an initial prototype of our interface.

## References

- An, W., and Li, C. 2011. Automatic matching of character strokes for computer-aided chinese handwriting education. In *Proc. ICEEE '11*, 283–288. IEEE.
- Banno, E.; Ikeda, Y.; and Ohno, Y. 2011. *GENKI I: An Integrated Course in Elementary Japanese I*. GENKI. Tokyo, Japan: Japan Times/Tsai Fong Books.
- Banno, E. 2011. *GENKI: An Integrated Course in Elementary Japanese II*. GENKI. Tokyo, Japan: Japan Times/Tsai Fong Books.
- Chu, T.; Taele, P.; and Hammond, T. 2018. Supporting chinese character educational interfaces with richer assessment feedback through sketch recognition. In *Graphics Interface 2018 (GI 2018)*. Toronto, Canada: CHCCS/SCDHM.
- Dixon, D.; Prasad, M.; and Hammond, T. 2010. icandraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces. In *Proc. CHI '10*, 897–906. ACM.
- Glaser, B. G. 1965. The constant comparative method of qualitative analysis. *Social Problems* 12(4):436–445.
- Hammond, T., and Davis, R. 2005. Ladder, a sketching language for user interface developers. *Computers & Graphics* 29:518–532.
- LaViola, Jr., J. J., and Zeleznik, R. C. 2004. Mathpad2: A system for the creation and exploration of mathematical sketches. In *ACM SIGGRAPH 2004 Papers*, 432–440. ACM.
- Ma, L.; Huo, Q.; and Shi, Y. 2009. A study of feature design for online handwritten chinese character recognition based on continuous-density hidden markov models. In *Proc. ICDAR '09*, 526–530.
- National Security Agency. 2017. Foreign language learning: A comparative analysis of relative difficulty. <https://bit.ly/35mMrAH>.
- Pittman, J. A. 2007. Handwriting recognition: Tablet pc text input. *Computer* 40(9):49–54.
- Sloniger, D.; Sakariya, R.; Guzman, C. D.; Mathews, J.; Susanto, E.; Koh, J. I.; Taele, P.; and Hammond, T. 2018. Tensai: A sketch-based educational game for assessing proper writing of japanese writing scripts. In *Proceedings of Graphics Interface 2018*. Canadian Information Processing Society.
- Taele, P., and Hammond, T. 2009. Hashigo: A next-generation sketch interactive system for japanese kanji. In *Twenty-First Innovative Applications Artificial Intelligence Conference*, 153–158. AAAI.
- Taele, P., and Hammond, T. 2010. Lamps: A sketch recognition-based teaching tool for mandarin phonetic symbols i. *Journal of Visual Languages & Computing* 21(2):109–120.
- Taele, P.; Barreto, L.; and Hammond, T. 2015. Maestoso: An intelligent educational sketching tool for learning music theory. In *Proceedings of the Twenty-Seventh Annual Conference on Innovative Applications of Artificial Intelligence*, 3999–4005. AAAI.
- Takezaki, K., and Godin, B. 2008. *An Introduction to Japanese Kanji Calligraphy*. Tuttle Publishing.
- Valentine, S.; Vides, F.; Lucchese, G.; Turner, D.; hoe Kim, H.; Li, W.; Linsey, J.; and Hammond, T. 2012. Mechanix: A sketch-based tutoring system for statics courses. In *Proceedings of the Twenty-Fourth Innovative Applications of Artificial Intelligence Conference, IAAI 2012*, 2253–2260. AAAI.
- Vatavu, R.-D.; Anthony, L.; and Wobbrock, J. O. 2012. Gestures as point clouds: A \$p recognizer for user interface prototypes. In *Proc. ICMI 2012*, 273–280. New York, NY, USA: ACM.
- Wobbrock, J. O.; Wilson, A. D.; and Li, Y. 2007. Gestures without libraries, toolkits or training: A \$1 recognizer for user interface prototypes. In *Proc. UIST 2007*, 159–168. ACM.
- Ye, L. 2011. *Teaching and Learning Chinese as a Foreign Language in the United States: To Delay or Not to Delay the Character Introduction*. Ph.D. Dissertation, Georgia State University.
- Zhang, X.-Y.; Bengio, Y.; and Liu, C.-L. 2017. Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark. *Pattern Recognition* 61:348–360.