CoWriting Kazakh: Learning a New Script with a Robot

Anara Sandygulova
Nazarbayev University
Nur-Sultan, Kazakhstan
anara.sandygulova@nu.edu.kz

Wafa Johal
UNSW & EPFL
Sydney, Australia
wafa.johal@unsw.edu.au

Zhanel Zhexenova
Nazarbayev University
Nur-Sultan, Kazakhstan
zhanel.zhexenova@nu.edu.kz

Bolat Tleubayev
Nazarbayev University
Nur-Sultan, Kazakhstan
bolat.tleubayev@nu.edu.kz

Aida Zhanatkyzy
Nazarbayev University
Nur-Sultan, Kazakhstan
aida.zhanatkyzy@nu.edu.kz

Aizada Turarova
Nazarbayev University
Nur-Sultan, Kazakhstan
aizada.turarova@nu.edu.kz

Zhansaule Telisheva
Nazarbayev University
Nur-Sultan, Kazakhstan
zhansaule.telisheva@nu.edu.kz

Anna CohenMiller
Nazarbayev University
Nur-Sultan, Kazakhstan
anna.cohenmiller@nu.edu.kz

Thibault Asselborn
CHILI Lab, EPFL
Lausanne, Switzerland
thibault.asselborn@epfl.ch

Pierre Dillenbourg
CHILI Lab, EPFL
Lausanne, Switzerland
pierre.dillenbourg@epfl.ch

ABSTRACT

In the Republic of Kazakhstan, the transition from Cyrillic to Latin alphabet raises challenges to training an entire population in writing the new script. This paper presents a CoWriting Kazakh system, an extension of the existing CoWriter system, aiming to implement an autonomous social robot that would assist children in transition from the old Cyrillic alphabet to a new Latin alphabet. With the aim to investigate which learning strategy yields better learning gains, we conducted an experiment with 67 children, aged 8-11 years old, who interacted with a robot in a CoWriting Kazakh learning scenario. Participants were asked to teach a humanoid NAO robot how to write Kazakh words using one of the scripts, Latin or Cyrillic. We hypothesized that a scenario in which the child is asked to mentally convert the word to Latin would be more effective than having the robot perform conversion itself. Results show that the CoWriter was successfully applied to this new script-switching task. The findings also suggest interesting gender differences in the preferred method of learning with the robot.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI); **• Social and professional topics → Children**; **• Applied computing → Education**; **• Computer systems organization → Robotics**.

HRI '20, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-6746-2/20/03…$15.00
https://doi.org/10.1145/3319502.3374813

KEYWORDS

Human-Robot Interaction, Child learning, Language learning, Social Robot, NAO, Gender differences

ACM Reference Format:
Anara Sandygulova, Wafa Johal, Zhanel Zhexenova, Bolat Tleubayev, Aida Zhanatkyzy, Aizada Turarova, Zhansaule Telisheva, Anna CohenMiller, Thibault Asselborn, and Pierre Dillenbourg. 2020. CoWriting Kazakh: Learning a New Script with a Robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3319502.3374813

1 INTRODUCTION AND CONTEXT

Kazakhstan has recently adopted a state program for the development and functioning of languages for 2011-2020. This new trilingual education policy is aimed at development among Kazakhs for fluency in three languages: Kazakh, Russian, and English. Additionally, a recent decision on the transfer of Kazakh language from Cyrillic to the Latin alphabet was approved by the Kazakh authorities in October 2017 [1]. While there are clear reasons for these reforms, there are numerous risks facing the transfer, including risks to cause disinterest and lack of motivation learning to write and read the Latin-based Kazakh among children and adults.

1.1 Transliteration and Script Learning

With world-wide globalisation, English-related digraphia is more and more common. Digraphia consists of using an English-related Latin alphabet to represent another language (such as Kazakh) [2]. Biscriptal practices consist of associating two graphemes (graphical symbol of the alphabet) to one phoneme (sound). Rivlina et al. (2016) [2] discuss the global phenomenon of using Latin script in combination with Russian language native script to represent Russian words in writing. According to the results analysis based on web scraping, the authors concluded that digraphic content is in most cases present to attract people's attention to the texts, to increase recognition and memorability, and to play with words. Moreover, digraphia contributes to the erasing of the boundaries between linguistic, national, cultural and domestic aspects of the language.

The study conducted by Al-Azami et al. (2010) [3] evaluates transliteration as a tool for learning to write in Bengali. This technique is adapted in London schools for British Bangladeshi students aged 7-11 years old. Transliteration of Bengali into Roman script consists of converting speech into written form in order for children to communicate with parents and teachers, as well as, learning a new method of being bilingual. For instance, when students do not know the spelling of a particular word, transliteration can help them visualize it, and, therefore, they could understand the meaning of the spoken word and improve their cognitive development. The process of using English sound-symbols and transferring these into a Bengali (Sylheti) system was quickly learnt by children. This research suggests transliteration is a useful technique for teachers to draw children's attention by maximizing their linguistic abilities and to encourage them to express their ideas in more than one script.

Overall, the literature about digraphia and introductions of a new script mainly investigate social impacts without presenting an educational perspective on how to address the introduction of the new script into school curriculum. However, several methods could be used to present the new alphabet to students. Gonzalez et al. [4] investigated two methods of tracing versus copying to learn a handwritten pattern. The authors found the two tasks showed varying benefits depending on short or long-term learning measures: immediate retention was found to be better when tracing, while long-term performances showed no difference between the two methods. In this paper, we propose to explore the effect of these two methods for the learning of the new Latin-based Kazakh alphabet.

1.2 Social Robots for Learning

A substantial increase in social robots in various areas of application highlights the importance of human-robot interaction research, especially in the application of education [5]. Recent years have seen the increase of investigations using social robots for reading [6–8] and language learning [9, 10].

Since 2014, the CoWriter project has explored how robotic technologies can help children with the training of handwriting via an original paradigm known as, Learning by Teaching (LbT) [11–13]. In this approach, children act as the teachers who help the robot to learn handwriting, thus the children practice their handwriting without noticing, staying committed to the success of the robot via the Protége effect. Additionally, previous research has shown the motivational aspect of LbT with a robot for handwriting [12].

A recent survey by [14] reports the potential benefit of LbT for: task commitment, motivation, and self-confidence and mental states attribution. Finally, Lubold et al. [15] proposed a series of design suggestions to adapt dialog strategies, finding that individual differences affected the LbT effect.

We believe the CoWriter activity could boost children's self-esteem and motivation to learn the Latin-based Kazakh alphabet and its handwriting. This paper presents the CoWriting Kazakh project that aims to enhance the new language planning in Kazakhstan to address challenges of training and motivating children to learn and use a new alphabet.

In particular, we propose to explore the interaction scenario derived from the existing CoWriter to extend it to the new context of learning the new Latin-based Kazakh script. Prior work by Kim et al. (2019) [16] presented a study with 48 children that interacted with a robot in a script learning scenario. However, the study had a limitation in the way the pre-test was performed, which made its findings inconclusive. The experiment presented in this paper was designed as a direct reproduction of the study presented in [16], where the only difference between the two studies was in the way the pre- and post-tests were performed. The pre-test of the [16] study was performed on a Wacom tablet for children to write a Cyrillic letter and its Latin version one by one for 23 out of 42 letters. When children made mistakes in the pre-test, they were shown the correct answers by the researcher. It provided a means of learning these letters from the pre-test. We believed this approach led to a potential confounding factor, which caused the insignificant differences between the two developed conditions: Cyrillic-to-Latin vs Latin-to-Latin. It is worth mentioning in our previous study [16], there were interesting gender differences found: boys' speed was slower and they chose to complete less words than girls did, but they learned more words in comparison to girls. Due to a major limitation in the way the pre-test was conducted, we replicated the study with the same system design, interaction scenario, and the two experimental conditions, but with a different execution of the pre- and post-tests (the current study opted for paper-based tests), which included all 42 letters without showing the correct answers in cases of mistakes or hesitations, as well as availability of video recorded interactions for facial expression analysis.

2 COWRITING KAZAKH SYSTEM

Since the project is motivated by the recent decision of Kazakh authorities to transition from Cyrillic to Latin script, students are trained for the new script in a learning scenario with a social robot. In contrast to the original CoWriter's LbT paradigm, where robot's handwriting improved gradually via several demonstrations by the child, the CoWriting Kazakh system does not have a handwriting improvement component. In the presented system, the robot and a child engage in co-operative learning where the robot learns from the child the new vocabulary in Kazakh, while the child learns from the robot the spelling in a new script. Thus, they take turns in writing words in Kazakh (see Figure 1). There are thirteen English words asked by the robot in total.

2.1 Robot Role

In the scenario, the NAO robot plays a role of a peer. The robot is introduced to a child as a native English speaker, of approximately his or her age, who wants to learn Kazakh. The robot asks the child for help, especially, to demonstrate how to write Kazakh words using the new Latin alphabet because that is convenient for the robot to read. In a controlled condition, the robot does not ask to

Figure 1: Screenshot of a Wacom tablet demonstrating Cyrillic-to-Latin condition. The top part of the screen shows the robot's writing space and the bottom one, the child's writing space.

write explicitly in Latin script, so the child writes words in their preferred script, which is Cyrillic script, as that is what they are accustomed to.

The child is told that the robot does not understand either Kazakh or Russian languages, so the children have to listen intently to understand what the robot says. Because of an absence of child voices in commercially available Kazakh text-to-speech engine, it was important to compose simple robot speech utterances for the children to understand. We developed these utterances and then verified it was fitting the level of English understanding of pupils with the help of their English teacher.

2.2 Dialogue

Interaction with the child consists of several stages during which the robot sustains the interaction: greets the child in the beginning, provides instructions in the form of the questions, and says goodbye at the end.

> NAO: -Hello. I am a robot. My name is Mimi. [Waves its hand]
> Child: -...
> NAO: -I study Kazakh language. Can you help me?
> Child: -...
> NAO: -How do you say "Hello" in Kazakh?
> Child: -Sálem
> NAO: - How do you write it? [In Latin-to-Latin case: Please write it using Latin letters so that I can read it.]
> Child: -[Writes on a tablet the word in one of the scripts]
> NAO: -Let me try to write it too [gesticulates]. This is a correct writing using Latin letters.
>
> ... repeated for another 12 words
>
> NAO: - You are a great teacher. Thank you very much! Goodbye! [waves]

2.3 Software and Hardware components

The Wacom Cintiq Pro tablet is a graphics tablet which can serve as the second monitor. Its pen has 8,192 levels of pressure sensitivity and tilt recognition. This allows it to acquire the trajectory of handwriting, including pressure and tilt at every point.

A humanoid robot NAO is a 58-cm programmable robot developed by Aldebaran Robotics. We extended the original CoWriter project[1] using ROS and NaoQi API in order to design a new software handling the two scripts and the new learning scenario. The child and the robot's writing occur on the same tablet on a blank writing space with three icons, "Eraser", "Done", and "Finish" (see Figure 1).

3 EXPERIMENT

The methodology of the experiment was designed and then verified in the prior work [16]. This section reports on the method, participants and recruitment process, hypotheses and conditions, procedure, and measurements used in the experiment.

3.1 Method

The experiment was conducted at the primary school in the capital of Kazakhstan. It involved one meeting with a robot for each child participant. All participants were assigned to a condition in a between-subject design, with a conversion type as a between-subject variable.

Each child interacted with a robot for approximately 20 minutes. Half of the children interacted with the Latin-to-Latin robot condition, while the other half of the children interacted with the Cyrillic-to-Latin condition. Counterbalancing was also applied in terms of gender and year group, so that each condition had a balanced number of boys and girls. Assignment to each of the robot conditions was otherwise random for any particular child.

3.2 Recruitment

This research was approved by the ethics committee of Nazarbayev University. Informed consent was obtained in writing from all children and their parents. Supporting information included an assent form for children and an informed consent form for parents or guardians. Children received a brief explanation of the purpose of the study and the procedures involved in data collection. Assent and consent forms were distributed to children in their classrooms in the presence of their teachers. Children were asked to show the forms to their parents at home and submit them to their teachers, who then collected the forms for us during the days that followed.

3.3 Participants

There were 67 children (32 females) aged 8-11 years old. The children came from diverse socio-economic backgrounds and all were native or fluent Kazakh speakers. At the time of the experiment, children in the third grade had spent approximately 30 months writing in Cyrillic and fourth-graders had spent approximately 42 months (3 years and 6 months) writing in Cyrillic. The children practiced handwriting for 6 hours per week, which started from simple shapes, and moved to Cyrillic letters after approximately

[1]https://github.com/chili-epfl/cowriter

6 weeks in grade 1. All children had 2 hours of English per week where they also started writing English letters from grade 1 (i.e., 30 months of handwriting in English for the third graders). However, the children had not been introduced to a Latin-based Kazakh alphabet (extended Latin alphabet with six additional letters) and its associated handwriting. Thus, in contrast to the Cyrillic script, the learning time for the new script was the same across all grade levels, i.e. it was equal to 0.

3.4 Hypotheses and Conditions

Using the CoWriting Kazakh system described in Section 2, we investigated whether it is more effective for the child to perform conversion mentally and observe correctly written Latin spelling by the robot. Our specific hypotheses to address this are as follows:

- H1: The CoWriting Kazakh learning scenario will result in significant improvement in the number of learned letters, which will serve as a check that the provided intervention results in learning a new script.
- H2: It is more effective for the child to perform conversion of Kazakh words mentally and then observe correctly written Latin spelling by the robot.

In order to address these hypotheses, we distinguish two conditions that are different in who performs the conversion:

- Latin-to-Latin: the child hears the word to be written and has to write it directly in Latin. Then the robot writes the word in Latin as a corrective feedback. In this condition, the child needs to perform mentally the script conversion.
- Cyrillic-to-Latin: The child hears the word, and she writes it in Cyrillic. Then the robot performs the script conversion by writing the same word using the Latin-based Kazakh alphabet.

During the interaction, we did not help children in writing and did not correct their mistakes. We would only help them in case they did not understand or hear the robot.

3.5 Procedure

The procedure of the experiment consisted of the following stages: a survey, a pre-test, a learning activity, an interview, and a post-test. The whole procedure for one child lasted approximately 30-40 minutes.

Each child was called out of the class and walked with the first researcher for approximately two-three minutes to a separate room. While walking with the child, the first researcher started with an icebreaker warm-up talk to relax and engage the youngster. "My name is Aida and what is your name?", "Have you even seen a robot before?", "When I was in school, I liked Mathematics and what is your favourite subject? Upon entering the room with the robot, children were invited to take a seat at the table with questionnaires and answer a few questions about their age, gender, and mood prior to the interaction with the robot. Then, children were asked to take a seat next to second researcher to complete a pre-test to find out if children know Latin-based Kazakh alphabet. After the questionnaire and tests were filled in, children were invited to change tables and take a seat facing the robot. After the interaction with the robot, children were interviewed by the first researcher who conducted a structured interview about their perception of the robot. Finally, children were given a post-test similar to the pre-test to evaluate their knowledge of Latin-based Kazakh, again. In the end, the first researcher brought the child back to the class and called for the next participant.

3.5.1 Survey. A short survey was administered by the first researcher who recorded the child's name, age, gender, class, and general mood on a 5-point Likert scale.

3.5.2 Pre-test. The next phase was the pre-test, where each child was presented with a table of 42 Cyrillic letters of Kazakh Cyrillic alphabet with a task to convert each letter to a corresponding Latin version of a new Kazakh alphabet. The pre-test was needed in order to determine the child's level of knowledge of Latin script.

3.5.3 Activity with the robot. Once the child was done with the pre-test, the researcher invited the child to take a seat in front of the robot. The researcher then launched the CoWriting Kazakh system detailed in Section 3 (Figure 3). The scenario was structured as an interactive lesson with a standing NAO robot.

There was a Wacom tablet on the table, in between the robot and the child. As the robot asked for a word translation, the child had to write the word in Kazakh, in either Cyrillic or Latin script. The robot would then gesticulate in the air with its arm as it was "writing" on the tablet, while the strokes of the letters appeared with the motion of the robot simultaneously. The robot's font was created using children's handwriting data collected in the previous work [12]. The robot always wrote during its turn using correct spelling in Latin. After the robot was done writing a word, it would then ask for another word's translation. The game would be stopped either by the child or after thirteen words were tried. The words were chosen to be simple words for the children's level of English, which were first verified with their English teacher. All thirteen words contained all 33 letters of the new Latin alphabet at least once. The robot then thanked the child for being a great teacher. Apart from small differences in the robot's speech detailed in Section 3.2, the interaction flow and robot's non-verbal behaviours were otherwise exactly the same between robot conditions.

The interaction was developed using both a robot's text-to-speech in English and also a face recognition engine. Throughout the interaction, the robot performed a series of alive mode animations with the use of arm gestures and head movements. In addition, the robot expressed non-verbal social cues, such as acknowledging the child's presence with eye contact and deictic gestures.

3.5.4 Interview. After the interaction, the child was asked to take a seat with the first researcher who then conducted a structured interview which consisted of several questions:

(1) What is your mood? (5-point Likert scale)
(2) What can you compare the robot to? (Options were: a toy, a computer, a human, or a pet)
(3) Who can you compare the robot to? (Options were: a friend, a parent, a sibling, a classmate, a teacher, a stranger)
(4) Funometer scale [17] was explained to a child with an example how the meter worked: the weather is very cold in the winter (at the bottom of the meter) while it is very hot in the summer (at the top of the meter). Where would you rate

Figure 2: Interaction with the robot

today's weather? Then, the next example was demonstrating a fun measurement: "imagine that it is your birthday and you get many presents, you have a lot of fun (lets place your feeling at the top of the meter), or the opposite when you feel bored while waiting, for example for a bus (your fun metric is at the bottom of the meter). Now lets place a small robot (a physical paper NAO robot) where you would rate how boring or fun it was to interact with the robot?" (Children rating was recorded on a scale from 0 to 100).

(5) Sorting task: children were explained with an example how the researcher believed five items were the most/least interesting. For that there were five small paper items: a book, a tablet, a NAO robot, a computer, and a teacher. The sorted position of the NAO robot was recorded on a 5-point Likert scale).
(6) Similarly, children were asked to sort the five items according to what/who is the least/most effective for learning? (The sorted position of the paper robot was recorded on a 5-point Likert scale).
(7) They were also asked to sort a book, a tablet, a robot, a computer, and a teacher according to what/who they like the least/most? (The sorted position of the paper robot was recorded on a 5-point Likert scale).
(8) Finally, children were asked to sort the five items according to what/who is the easiest way to learn with/from? (The sorted position of the paper robot was recorded on a 5-point Likert scale).

The questions' aim were to determine the extent to which children liked the interaction. Moreover, we asked for their mood after the interaction, to see whether their mood worsened or improved. Finally, we gave them a series of sorting tasks where children were able compare their feeling of the robot in comparison to common learning approaches. We utilized various techniques to deliver questions to be understandable as possible. For example, the Funometer scale and a picture of a NAO were printed on a paper for children to physically drag the printed robot and position it on a scale. It proved to be more appropriate than having pictorial 5-point Likert scales, since the majority of children placed the robot near 70-90 percent instead of 100.

3.5.5 Post-test. The last phase of the procedure was the post-test. During the post-test, children were presented with a similar table of 42 Cyrillic letters as in the pre-test. The task was to write Kazakh letters in their Latin version again. The post-test was needed in order to determine the difference between the number of correct letters in both tests to calculate the number of learned letters. After they completed the test, children received a book for participation.

3.6 Measurements

Data were collected from both self-reported questionnaires and a camera that recorded the interactions.

- **Change in Mood: pre- and post-mood.** Children were asked to rate their mood before and after the interaction with the NAO robot.
- **Funometer.** Child was asked to rate how much they liked interacting with the robot from 0 to 100 on a Funometer scale [17].
- **Automatic Emotion Analysis** There was a camera placed in front of the child capturing facial expressions for real-time emotion analysis. We employed Sophisticated High-speed Object Recognition Engine (SHORE) [18] software, which gives the intensity values of the following emotional states: happiness, sadness, surprise and anger. According to Alonso et al. (2013) [19], SHORE has 100% success rate for recognizing happiness. Thus, we only accounted for the expressed happiness of the participants. The intensity score of the expressed happiness was recorded every second to calculate the average score of the expressed happiness for each participant [0-100]. In addition to SHORE's data, all interactions were also manually video coded to validate SHORE's recognition accuracy.
- **Robot Type.** The question was about child's comparison of the NAO robot via a forced-choice question: *a toy, a computer, a human, or a pet.* This question was also used by Belpaeme et al. (2012) [20].
- **Robot Role.** Children were asked to compare whether the robot is similar to one of a forced-choice options: *a friend, a parent, a sibling, a classmate, a teacher, or a stranger.* This question was also used in [20].
- **Sorting Robot.** Children were asked to physically sort five small pictures (a book, a tablet, a NAO robot, a computer, and a teacher) in ascending order according to their 1) Effectiveness to teach, 2) Easy to learn from, 3) Interesting, and 4) Enjoyable. We then noted the order (as a 5-point Likert scale) they placed the robot in.
- **Number of learned letters.** Number of learned letters was calculated after the post-test as the difference between known letters in the post-test and the pre-test (e.g. 18 correct letters in the post-test and 10 correct letters in the pre-test equals to 8 learned letters).

4 RESULTS

There were 67 children (32 females) aged 8-10 years old. Exactly half of the participants interacted with a robot in a Latin-to-Latin condition. A series of Kolmogorov-Smirnov (K-S) and Shapiro-Wilk tests was conducted on all dependent variables overall and within

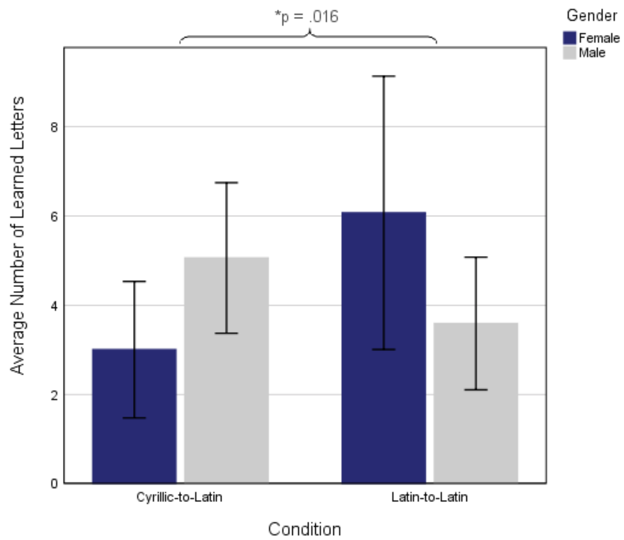

Figure 3: Average number of learned letters for boys and girls, split by the robot conditions. Error bars show 95% Confidence Interval, * indicates significance at the 0.05 level.

groups (i.e. gender and robot condition) to check the assumption of normality. Since some scores were significantly non-normal, non-parametric tests were used for the statistical data analysis presented in some of the next sections.

4.1 Learned Letters

In general, children improved their knowledge of Latin-based Kazakh alphabet during the experiment. The average number of new learned letters was 4.35 (SD = 3.7, Max = 18, Min = 0).

A two-way ANOVA was conducted examining the effect of gender and robot condition on a number of learned letters. There was a statistically significant interaction between the effects of gender and robot condition, $F(1, 64) = 6.17$, $p = .016$. Boys learned more in Cyrillic-to-Latin condition (5.06 ± 3.28 vs 3.59 ± 2.89) while girls learned more in Latin-to-Latin condition (3.00 ± 2.87 vs 6.07 ± 5.31). A separate one-way ANOVA was conducted for girls: $F(1, 29) = 4.017$, $p = 0.05$. Thus, the learning strategy of performing mental conversion themselves was more effective for girls, in contrast to boys who learned more when the robot performed the conversion for them (5.06 ± 3.29 vs 3.29 ± 2.89 in Cyrillic-to-Latin and Latin-to-Latin respectively), though not significant (F = 1, 33) = 1.916, $p = 0.17$). Figure 3 presents the results.

A series of one-way ANOVA tests was conducted to determine if there was any difference in learning. It revealed a non-significant difference in the number of learned letters between different robot conditions, which rejects our hypothesis that Latin-to-Latin condition is more effective in the current learning scenario. Children did not have significant differences neither in learning gains, nor in pre-tests or post-tests. Boys learned 4.32 ± 3.14 while girls learned 4.39 ± 4.33 letters. Girls scored slightly better in a pre-test (14.25 ± 5.98 vs 12.69 ± 6.22) and in a post-test (18.26 ± 5.8 vs 16.85 ± 6.05), though not significant.

4.2 Funometer and Mood Change

Children rated how fun it was to interact with the robot on a scale from 0 to 100. An average rating for all children was 90 ± 15. Children rated Latin-to-Latin robot condition as 93 ± 10 while Cyrillic-to-Latin condition as 87 ± 18. The difference was not significant.

The difference between boys' and girls' ratings was also not significant: boys rated interaction with the robot as 92 ± 16 while girls rated as 87 ± 14.

Children's mood difference was also not significant between robot type and gender groups.

4.3 Happiness Expression Analysis

The data of only 50 children was available for facial expression analysis. The average expressed happiness score did not deviate significantly from normal ($p > 0.05$). A series of one-way ANOVA tests was conducted revealing a non-significant difference between robot type groups, however girls' expressed happiness (10.8 ± 7.2) was significantly higher than boys' expressed happiness (6.03 ± 3.9), $F(1, 49) = 9.002$, $p = .004$.

4.4 Robot Perception

When asked to position the robot according to its effectiveness to teach, easiness to learn from, being interesting and enjoyable in comparison with a book, a tablet, a computer, and a teacher, children did not have a significant difference in their ratings for most items. However, there was a significant difference ($F(1, 64) = 9.4$, $p = .003$) between robot conditions when asked how easy it was for children to learn from the robot: the rating of 4.06 ± 1.17 was in the Cyrillic-to-Latin case while only of 3.13 ± 1.28 in the Latin-to-Latin case.

There were no significant differences between gender groups: girls rated the robot as 2.93 ± 1.14 and 4.00 ± 1.4, while boys rated the robot as 3.28 ± 1.4 and 4.12 ± 0.9, in Latin-to-Latin and Cyrillic-to-Latin cases, respectively.

A series of chi-square tests of independence were conducted to examine the effect of independent variables on children's answers for Robot Type and Robot Role. We did not find any statistical significant results between groups for these measurements.

5 DISCUSSION AND LIMITATIONS

All participants were from the same school and we can not generalize to confidently say the same result will be valid in other schools in Kazakhstan.

From the analysis of the results, we can conclude that H1 is supported, confirming the intervention with the system had a positive effect on the children's performance in the pre- and post-tests at a highly significant level ($p < 0.001$). Since the learning gains of children were not affected by the robot strategy, we can conclude that H2 was not supported. However, boys and girls had different learning gains in two robot conditions: learners who were girls performed significantly better in a post-test after they attempted the conversion themselves and then watched their mistakes corrected by the robot.

5.1 Gender Effect

It is worth mentioning that children's performance was significantly different in two robot conditions for boys and girls. Latin-to-Latin condition was a more effective approach for learning a new script for girls - when they made an attempt at converting words themselves and then saw a correct spelling by the robot - it resulted in significantly more learned letters than in Cyrillic-to-Latin condition. On the contrary, boys learned more letters when they observed the correct spelling produced by the robot and it was more effective for them to learn it, as they could see their Cyrillic and robot's Latin spellings.

The gender difference could be explained by referring to the literature on gender studies ,which often show that girls outperform boys in languages. For example, according to the Program for International Student Assessment (PISA), a worldwide assessment conducted by the Organisation of Economic Co-operation and Development (OECD), girls consistently show high achievement in reading which "measures the capacity to understand, use and reflect on written texts in order to achieve goals, develop knowledge and potential, and participate in society" [21]. Drawing from PISA data, Breda and Napp (2019) indicate that girls are more advanced than boys at reading [22]. This finding is consistent with other large-scale studies indicating girls' language development matures quicker and more robustly than boys' language [23]. As the type of source text (e.g., humor, adventure) appear to appeal to girls and boys differently [24], future studies could examine word choice in child-robot interactions and their effect on gender.

5.2 Task Difficulty

Children rated Latin-to-Latin condition as significantly harder than Cyrillic-to-Latin case, which raises a trade-off debate: whether it is worth making children work harder and perform mental conversion themselves when a learning gain for all children is not significantly different from the easier condition.

Moreover, a behavioural analysis demonstrated girls smiled significantly more in comparison to boys even though self-reported Funometer-based ratings of boys were slightly better than girls' ratings. Unfortunately, there were some missing video data, which is a limitation to reliably interpret the behavioral data. In addition, as stated in Ros et al. [25], behavioural measures, such as facial expressions have their advantages, but are also subject to individual differences between children (expressive vs. non-expressive interaction styles). Lastly, these differences could also have a gendered component effecting the output.

This experiment has raised a number of questions and further work is needed in several directions to adequately determine the best scenarios and strategies for learning a new script. For example, a choice of the vocabulary should be investigated further as it might be effective to enable children to use their knowledge of foreign language vocabulary to advance foreign script learning. Contrary to this strategy, it might be more effective to use unknown/non-existing words in languages children know, in order to avoid confusions with prior knowledge. We believe it is important to leverage various strategies to find the most effective cognitive learning scenario, as the robot is situated in the physical world, interactions with a social robot can be multimodal (verbal, visual, and tactile) and be adapted according to all perceptual modalities, including events on the tablet, its stylus data and child's feedback.

5.3 Handwriting Recognition

At first, and in order to evaluate the children's handwriting performances, we implemented handwriting recognition for the Cyrillic script. We achieved 98% of accuracy on the Cyrillic-MNIST dataset on a validation set utilising state-of-the-art algorithms i.e. 784-500-500-2000 network described in Hinton and Salakhutdinov [26] and CNN similar to Le-Net-5 [27] with custom parameters. However, when deployed, the recognition on children's handwriting data, the recognition accuracy was only 38%. Similar to reports that explicitly tested state-of-the-art speech recognition engines [28] on children's speech, age and gender recognition approaches on children's faces [29], the OCR also does not perform as well with children's handwriting.

Quality of handwriting can only be assessed when taking into account the age and gender of children [30]. Collecting a dataset representative of children's Cyrillic handwriting will, in the future, help us to achieve satisfactory evaluation of Cyrillic handwriting in real-time.

5.4 Social Robots and Learning Scenario

We can not argue that the CoWriting Kazakh system was effective at helping children learn new letters due to a humanoid robot being present. The focus of this experiment was to test that the setup as a whole was effective enough in teaching in a single session. Future work will involve conducting an acceptability and usability study to determine the value of the robot as an engaging peer learner. Nevertheless, the whole system and the implemented scenario (a social robot + handwriting on a tablet) managed to engage children as the average number of completed words was 12.63 (out of 13 words in total). In addition, we believe that a background story of the robot was crucial for the motivational aspect of having a child committed to the interaction, which can not be achieved without the robot.

The importance of comparing our system with a classical approach of teaching is evident at the current stage. We plan to conduct a long-term study comparing different conditions: pen and paper, tablet, and tablet + robot. This long-term study will evaluate and compare children's learning outcomes and long-term engagement with the system.

Our ultimate goal is to build an adaptive system using differentiated learning strategies applicable across learning settings and individuals, for learning the new script. This study revealed gender differences in learning preferences and effectiveness of the strategies (Cyrillic-to-Latin vs Latin-to-Latin). We aim to conduct a follow-up study to validate these findings and adapt the learning scenario to the gender of the learner. We expect higher learning outcomes in the case of Latin-to-Latin used for learners who are girls and Cyrillic-to-Latin used for learners who are boys. Having said that, we believe other parameters can be adjusted to optimize the learning experience to each learner. Future studies will aim to determine what other parameters (apart from gender), such as diversity markers and learning-styles, can be used to personalize the learning and the interaction with the robot.

6 CONCLUSION

The CoWriter system was designed to help children with dysgraphia by using a Learning by Teaching approach, children would be engaged to practice their handwriting with a social robot. In this paper, we proposed to extend this system for a new context: learning the new script of the Kazakh language. We proposed two approaches for children learning: Latin-to-Latin and Cyrillic-to-Latin. Our results showed that the CoWriter System was successfully applicable in this new context, as all children achieved positive learning gains in terms of number of letters learned. Besides, children were engaged and appreciated learning with the robot. Our future work will focus on improving the automatic assessment of handwriting quality and test the scenario in a long-term interaction.

ACKNOWLEDGMENTS

We would like to thank the children and their parents for participating in this research. The authors would also like to express great appreciation to the members of staff of the participating primary school for their time and help with the experiment.

This work was supported by the Nazarbayev University Collaborative Research Program grant, "CoWriting Kazakh: Learning a New Script with a Robot" (Award number is 091019CRP2107).

This research was also partly funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) as parts of the CIS Region Seeding Grants (Project SFG 1038), and the NCCR Robotics (SNSF).

REFERENCES

[1] E. Altynsarina. Kazakhstan adopts new version of Latin-based Kazakh alphabet. NATION. https://astanatimes.com/2018/02/kazakhstan-adopts-new-version-of-latin-based-kazakh-alphabet/, accessed at September 2019.

[2] A. Rivlina. 2016. Global English-related Digraphia and Roman-Cyrillic Biscriptal Practices. *Procedia - Social And Behavioral Sciences*, 236, 207-212.

[3] S. Al-Azami, C. Kenner, M. Ruby, and E. Gregory. 2010. Transliteration as a bridge to learning for bilingual children. *International Journal of Bilingual Education And Bilingualism*, 13(6), 683-700.

[4] C. Gonzalez, J. Anderson, P. Culmer, M. R. Burke, M. Mon-Williams, and R. M. Wilkie. 2011. Is tracing or copying better when learning to reproduce a pattern?. *Experimental Brain Research*, 208(3), 459âĂŞ465. doi: 10.1007/s00221-010-2482-1.

[5] O. Mubin, C.J. Stevens, S. Shahid, A. Al Mahmud, and J.-J. Dong. 2013. A Review of the Applicability of Robots in Education. *Technology For Education And Learning*. doi:10.2316/journal.209.2013.1.209-0015.

[6] E. Yadollahi, W. Johal, A. Paiva, and P. Dillenbourg. 2018. When deictic gestures in a robot can harm child-robot collaboration. In Proceedings of the 17th ACM Conference on *Interaction Design and Children* (IDC '18), New York, NY, USA, 195-206. doi: https://doi.org/10.1145/3202185.3202743

[7] G. Gordon and C. Breazeal. 2015. Bayesian Active Learning-Based Robot Tutor for Children's Word-Reading Skills. In Proceedings of the 29th *AAAI Conference on Artificial Intelligence*.

[8] J. E. Michaelis and B. Mutlu. 2018. Reading socially: Transforming the in-home reading experience with a learning-companion robot. *Science Robotics*, 3(21). doi: 10.1126/scirobotics.aat5999.

[9] K. Balkibekov, S. Meiirbekov, N. Tazhigaliyeva, and A. Sandygulova. 2016. Should robots win or lose? Robot's losing playing strategy positively affects child learning. In Proceedings of the 25th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 706-711. IEEE.

[10] T. Belpaeme, P. Vogt, R. van den Berghe, K. Bergmann, T. GÃűksun, M. de Haas et. al. 2017. Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 10(3), 325-341.

[11] D. Hood, S. Lemaignan, and P. Dillenbourg. 2015. The CoWriter Project: Teaching a Robot how to Write. In Proceedings of the 10th ACM/IEEE *International Conference on Human-Robot Interaction (HRI)*, Portland, USA, March 3-5. doi: 10.1145/2701973.2702091

[12] A. Jacq, S. Lemaignan, F. Garcia, P. Dillenbourg, and A. Paiva. 2016. Building Successful Long Child-Robot Interactions in a Learning Context. In Proceedings of the 11th ACM/IEEE *International Conference on Human-Robot Interaction* (HRI), Christchurch, New Zealand, March 07-10, 239-246.

[13] S. Lemaignan, A.D. Jacq, D. Hood, F. Garcia, A. Paiva et al. 2016. Learning by Teaching a Robot: The Case of Handwriting. *Robotics and Automation Magazine*. doi: 10.1109/Mra.2016.2546700.

[14] F. Jamet, O. Masson, B. Jacquet, J.-L. Stilgenbauer, and J. Baratgin2018. Learning by Teaching with Humanoid Robot: A New Powerful Experimental Tool to Improve Children's Learning Ability. *Journal of Robotics.* [Online]. Available: https://www.hindawi.com/journals/jr/2018/4578762/.

[15] N. Lubold, E. Walker, H. Pon-Barry, Y. Flores, and A. Ogan. 2018. Using iterative design to create efficacy-building social experiences with a teachable robot. In Proceedings of *International Conference of the Learning Sciences*, ICLS, 2, 737-744.

[16] A. Kim, M. Omarova, A. Zhaksylyk, T. Asselborn, W. Johal, P. Dillenbourg, and A. Sandygulova. 2019. CoWriting Kazakh: Transitioning to a New Latin Script Using Social Robots. In Proceedings of the 28th IEEE *International Conference on Robot and Human Interactive Communication* (RO-MAN). IEEE. [in press].

[17] P. Markopoulos, J.C. Read, S. MacFarlane, and J. Hoysniemi. 2008. Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers. *Interactive Technologies*. Elsevier Science.

[18] A. Ernst, T. Ruf, and C. Kueblbeck. 2009. A modular framework to detect and analyze faces for audience measurement systems. In *2nd Workshop on Pervasive Advertising at Informatik*, 75–87.

[19] F. Alonso-Martín, M. Malfaz, J. Sequeira, J. F. Gorostiza, and M. A. Salichs. 2013. A multimodal emotion detection system during human–robot interaction. *Sensors*, 13(11), 15549–15581.

[20] T. Belpaeme, P. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayová, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerincx, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Cañamero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert. 2013. Multi-modal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2), 33–53..

[21] OECD. 2019. Reading performance (PISA) (indicator). doi: 10.1787/79913c69-en (Accessed on 31 September 2019)

[22] T. Breda, and C. Napp. 2019. Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. In Proceedings of the *National Academy of Sciences*, 116(31), 15435-15440.

[23] M. Eriksson, P. Marschik, T. Tulviste, M. Almgren, M. PÃľrez Pereira, S. Wehberg et.al. 2011. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30(2), 326-343.

[24] T. Merisuo-Storm. 2006. Girls and Boys Like to Read and Write Different Texts. *Scandinavian Journal of Educational Research*, 50(2), 111-125.

[25] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi. 2011. Child-robot interaction in the wild: Advice to the aspiring experimenter. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, 335–342, New York, NY, USA. ACM.

[26] G.E. Hinton and R.R. Salakhutdinov. 2006. Reducing the Dimensionality of Data with Neural Networks," *Science*, 313(5786), 504-507.

[27] Y. Lecun, L. Bottou, Y. Bengio, & P. Haffner, 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278-2324.

[28] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulos et.al. 2017. Child speech recognition in human-robot interaction: evaluations and recommendations. In Proceedings of the 2017 ACM/IEEE *International Conference on Human-Robot Interaction*, 82-90. ACM.

[29] A. Sandygulova, M. Dragone, and G.M. O'Hare. 2014. Real-time adaptive child-robot interaction: Age and gender determination of children based on 3d body metrics. In Proceedings of the 23rd IEEE *International Symposium on Robot and Human Interactive Communication*, 826-831. IEEE.

[30] T. Asselborn, T. Gargot, ÅĄ. KidziÅĎski, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg.2018. Automated Human-Level Diagnosis Of Dysgraphia Using A Consumer Tablet. *Npj Digital Medicine*, 1(1). doi:10.1038/s41746-018-0049-x.