Having an Einstein in Class. Teaching Maths with Robots is Different for Boys and Girls *

Ivy S. Huang and Johan F. Hoorn

***Abstract*— We conducted an experiment with Chinese schoolchildren being taught symbolic reasoning and solving equations of the form 2y + 3 = 7, using a Hanson Einstein robot together with video instruction. The children were unfamiliar to solving equations like these (baseline 0) and after robot tutoring, they were tested for learning outcomes, while they rated their appreciation of the robot. We found a crossover interaction effect of boys learning more from video and girls more from the robot. The effect was catalytic in that only the combination of factors worked (no main effects). Results are discussed in view of gender-specific technology application.**

I. Introduction

The gap between urban and rural development in China is reflected not only in industrialization but in educational resources as well. In first-tier cities, primary schools have a better student-staff ratio, subject-dedicated teachers, and they are equipped with high-quality technology. In rural areas, with fewer staff and little technology available, primary school teachers have to teach multiple topics. The most advanced teaching technology would be to watch a video instruction. With a shortage of funds, methods are rather simple, which may put rural children at a disadvantage in terms of learning effect and motivation.

Therefore, it would be of great help if in China's rural areas, simple and low-cost social robots could assist teachers in their tasks. If robots do maintenance rehearsal of, for example, the times tables [1], teachers may dedicate their time to creating new content materials or to support pupils that need special attention.

A physically embodied agent such as a robot supposedly receives more attention than an on-screen avatar [2]. Moreover, simple behaviors that are robust and flexible would produce better results than complex behaviors [2]. In general, children have a positive attitude towards robots as a teacher with great interest for the robot's performance [3].

The literature commonly reports beneficial effects of social robots on learning and motivation (e.g., [4][5]), while executing repetitive tasks of content matter already known (e.g., [1]). However, we pushed the envelope by letting our Hanson Einstein machine teach new mathematical contents in the hope that robots may mean more to teachers than just being a minion.

*I. S. Huang is with the School of Design, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: zhengkh0310@gmail.com).

J. F. Hoorn is with the School of Design and Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong and with Vrije Universiteit Amsterdam, The Netherlands (e-mail: csjfhoorn@polyu.edu.hk, t: 852 2766 4509, f: 852 2774 5067).

In view of the results of previous studies, we formulated two hypotheses: Our first hypothesis is that a robot teaching new contents increases the learning outcomes as compared to baseline (H1). While being instructed through video, the addition of a robot will motivate children more than video alone (H2).

To test our assumptions, we devised a lesson that concerned solving mathematical equations and symbolic reasoning for two primary schools in NanXing and BoMei. The childrens' tutor was no one less than Professor Einstein in the form of a simple toy-like social robot.

II. Method

A. Participants, Design, and Measurement

Participants were fourth grade pupils of two rural primary schools (N = 76, NanXing: n = 34, boys: n = 41). For each school, the children were equally divided into an experimental and control group, with an about equal number of boys and girls, and an about equal division of math skills as indicated by their teachers. These children were unfamiliar with robots. While sitting in half a circle, the experimental group received an 8-minute Chinese spoken video instruction supported by a robot. The topic was symbolic reasoning and solving equations. The control group did the same but received video instruction without robot support.

Solving simple equations is taught in the first semester of fifth grade. This way, we were certain that children had a baseline knowledge of zero, as confirmed by their teachers. Measurements were fourth-fold: On the first answering sheet, children indicated their appreciation of the video on a 6-point rating scale, using smiles with Chinese labels (☹ = not happy, ☺ = very happy). The experimental group did the same for the video and separately for the robot. On the next sheet, children translated two pictorial items into symbolic representations (Section B: 'apples in a bag'). Then they solved six equations with one variable unknown (e.g., 2 = 2y ÷ 3).

B. Materials and procedure

Hanson's Professor Einstein is a 15-inch robot that has an offline, online, and mobile mode (Figure 1).[1] During the experiment, the robot was in mobile mode connected to a laptop (Intel-i5) playing the animations. Another laptop (Intel-i7) projected the instructions on a whiteboard next to the robot while playing the sound to a speaker behind the robot. The projection's diagonal was 36.1 inch. Einstein spoke in Putonghua with the voice of an older adult at a speed just

[1] https://professoreinstein.com/

below average for good understanding. Set-up for the group without robot was the same with sound coming from the speaker and the video running on an Intel-i3 laptop.

Figure 1. Professor Einstein

The test materials consisted of two items on symbolic representation. They showed a number of apples and bags and asked the children to translate the picture into formal symbols. The correct answer to the example in Figure 2 would be 4 + 2y. The other six questions consisted of simple equations (e.g., 3y = 9). Figure 3 depicts the single-item rating scale for appreciation of video and robot.

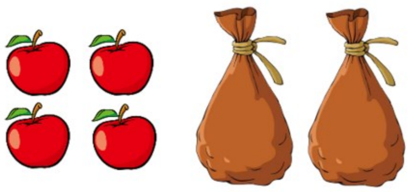

Figure 2. Represent the picture with formal symbols

Figure 3. Appreciation rating scale

Children were gathered in their classroom and were informed that they would receive video instruction. They were not informed about the presence of a robot. Then half the children were brought to the room with the robot while the control group was waiting (NanXing, serial sampling) or the control group was brought to another room for video instruction alone (BoMei, parallel sampling). Experimenters made sure that children in NanXing could not talk to each other in between sessions. For a detailed account of set-up, room layout, and timing, consult [6]. After measurement was completed, children of control and experimental groups were allowed to interact with the robot, eat a snack, and have fun.

III. RESULTS

To test our hypotheses (H1 and H2), we ran a number of tests and variants of tests, the details of which are in [6]. Here, we present the most relevant findings and considerations.

H1 stated that children would learn new content matter from robot instruction compared to baseline, which in our case was zero. We had two dependent measures to test this assumption: the percentage of correctly solved equations (%C) and number of correct symbolic representations (#SR).

For %C, we conducted two one-sample t-tests with test value 0. With the robot, $t_{(36)} = 10.33$, $p < .0001$. Without the robot (video alone), $t_{(37)} = 11.41$, $p < .0001$. Children learned new math in both conditions.

To compare video-with-robot to video-alone instruction, we ran a 2 (School: NanXing-BoMei) by 2 (Robot: yes-no) by 2 (Gender: male-female) between-subjects ANOVA (GLM Univariate) on %C. We used the data of those children who missed no more than one equation. That means that three children were dismissed from analysis because they skipped two or more equations, which left us with N = 73. Figure 4 shows the means and SDs.

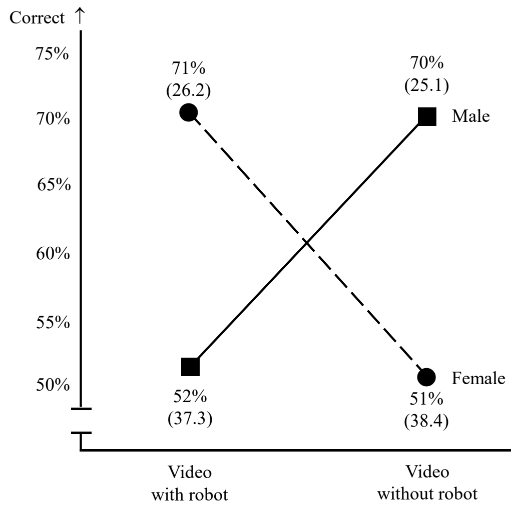

Figure 4. Mean percentage of correctly solved equations and standard deviations (N = 73)

The higher-order interaction of School*Robot*Gender was significant ($F_{(1,65)} = 10.14$, $p = .002$, $\eta_p^2 = .14$). This effect was supported by the significant interaction between Robot and Gender ($F_{(1,65)} = 9.28$, $p = .003$, $\eta_p^2 = .13$). It was not supported, however, by the interaction between School and Robot, which was a trend ($F_{(1,65)} = 3.75$, $p = .057$, $\eta_p^2 = .06$) of which none of the underlying effects were significant, according to unique sums of squares. None of the other interactions and main effects were significant either.

Because the non-theoretical factor School was merely effective in interaction with Robot and Gender but was unsupported by School*Robot, School*Gender, nor by the main effect of School, we focused on the significant interaction between Robot and Gender for further analysis.

To evaluate simple effects (contrasts), we used the unique sums of squares procedure in MANOVA [7]. We preferred this approach because individual t-tests use but half the participants to estimate the error term while significance relies on merely half the *df*. To gain more power, using within-cell variability for all cases led to smaller and more reliable error terms [8].

Simple-effects analysis in the significant interaction between Robot and Gender, then, showed (Figure 4) that with a robot, boys (M_{male} = 52%, *SD* = 37.3%) performed significantly poorer than girls whereas girls benefitted from the robot (M_{female} = 71%, *SD* = 26.2%) ($F_{(1,69)} = 2.89$ (or $t = -1.70$), $p = .094$) when tested one-sided against $\alpha = .10$. If tested two-sided, this would be a trend. Without a robot, boys (M_{male} = 70%, *SD* = 25.1%) correctly solved significantly more equations than girls (M_{female} = 51%, *SD* = 38.4%) ($F_{(1,69)} = 3.13$ (or $t = 1.77$), $p = .081$) when tested one-sided but this should be regarded a trend when tested two-sided.

We tested H1 also with the number of correct symbolic representations (#SR). Because #SR was an ordinal measure (0, 1, 2 items correctly represented), we calculated X^2, which for the condition with vs without the robot remained insignificant ($p > .40$). For the two Schools, however, $X^2_{(2)} = 13.1$, $p = .0014$, indicating that BoMei (11.1% #SR = 0, 12.5% #SR = 1, 33.3% #SR = 2) was significantly better than NanXing (25% #SR = 0, 1.4% #SR = 1, 16.7% #SR = 2). For Gender, no significant distributions occurred ($p > .50$).

H2 expected that children learning math with a robot would be more motivated than children with video alone. Therefore, all children scored their Appreciation for the video (App-v, n = 75 (1 missing)) and those in the experimental group for the robot as well (App-r, n = 37).

We ran a 2 (School: NanXing-BoMei) by 2 (Robot: yes-no) by 2 (Gender: male-female) between-subjects ANOVA on App-v. No effects were significant except for the interaction between School and Gender ($F_{(1,74)} = 4.23$, $p = .044$). This result depended on girls of NanXing having more appreciation of the video than the BoMei girls ($t_{(32)} = 2.21$, $p = .03$). For the robot group, we ran a 2 (School: NanXing-BoMei) by 2 (Gender: male-female) between-subjects ANOVA on App-r but none of the effects were significant. Then we performed multiple linear regression on %C with App-v and App-r as the predictors but for the whole model, $F < 1$.

IV. Conclusion

Based on the literature, H1 predicted that children would learn from robot instruction and as we daringly added, also when it concerns new mathematical content. And indeed they did: One-sample t-test with zero knowledge as the baseline confirmed that children could positively solve equations based on robot instruction, confirming H1. However, they also could do that based on video instruction alone. This does not disconfirm H1 but urges to reconsider its exclusiveness.

Regarding H1, we also found a significant crossover interaction effect that particularly girls benefitted from the robot whereas boys benefitted from the video without the robot. The related main effects remained absent, indicating that it is the combination of girls-with-robots and boys-with-videos that renders positive results. This is called a catalytic effect. Thus, the girls, not the boys, supported H1.

With respect to symbolic representation, one school performed significantly better than the other did, which is theoretically inconsequential.

H2 predicted that children would be more motivated to work with the robot than with a mere video but this was not confirmed by our (way of) measurement. The only significant effect was the interaction between School and Gender on appreciation for the video, which again is theoretically of little interest.

V. Discussion

From our results, robots have no exclusive rights to be more motivational or able to teach than plain video instruction does – if teachers do not differentiate according to gender. Both robots and videos provide learning gains. After all, Figure 4 shows that irrespective of medium, children on average could solve 50% of the novel mathematical equations correctly as confirmed by one-sample t. The upshot is, however, that a surplus in learning gain can be obtained if the robots are teaching the girls and videos the boys. If gender is not taken into consideration, the smaller learning effects of robots for boys cancel out the higher effects for girls and with videos vice versa.

That girls learned more from robots than boys is in line with [9], who found that boys barely improved in learning prime numbers as compared to girls. However, that study had no video-alone condition, as in our study, in which boys did improve significantly. Overall, boys find computers, as in our video-alone condition, significantly more positive than girls [10]. They perceived computers as being more "enjoyable," "special," "important," "friendly," and "cheaper" than girls [10]. Yet, our appreciation measure did not reveal any significant differences between robot and video-alone conditions in interaction with gender (which may be a flaw of the measure).

With [11] we wonder whether girls view robots differently than boys. In that study, conceptual differences were that males tended to think of a robot as more human-like, whereas females saw the robot as more machine-like, exhibiting less socially desirable responses to the robot. Perhaps the boys were more distracted by the robot 'as a human' than the girls were.

In general, gender typicality of HRI tasks substantially influences HRI as well as humans' perceptions and acceptance of a robot [12]. In [13], for instance, participants tended to rate the robot of the opposite sex as more credible, trustworthy, and engaging. Perhaps this explained the performance of the girls as related to our robot, which represented a male scientist. Contrariwise, impression formation of a same-gender robot was found more positive when genders matched, exerting more psychological closeness to the same-gender robot [14].

Perhaps we should look better into our statistics and interpretation of results. One could object that perhaps we tested too leniently. After all, the p-values of the t-tests that sustained the crossover interaction were in the range between .05 and .10, which for hard-nosed frequentists is beyond the cut-off point. Yet, testing against .10 or .05 depends on one's perspective. If one follows the line that from the beginning, the prediction was in favor of the robot, then the test is one-sided: $\alpha = .10$. However, because this merely counted for girls, which was not predicted, one could argue that a two-sided test is required for the boys: α = .05. One could counter that we should apply Bonferroni correction, rendering effects insignificant again. On the contrary, intra-individual variability of data obtained from young children is notoriously high [15] so that one may argue that a 10% rejection area is strict enough already.

All statistical intricacies and possible explanations aside, there is a strong overall trend for girls to benefit from support by a simple robot whereas for boys, video alone works better. That may be hopeful news for Chinese teachers in rural areas with little time and little budget on their hands to increase their pupils' performance and who would want to relax some of the stress they are in. For them, we added an extended abstract of this paper in English and Chinese in the technical report [6]. For us, extra data sampling is required to make our promising results come out just a little better.

ACKNOWLEDGMENT

We would like to express our gratitude to the principals, teachers, and students of NanXing elementary school and BoMei second elementary school for their participation and helpful suggestions. We also thank Lis Chan and Johnny Ho for assisting us during the experiments.

REFERENCES

[1] E. A. Konijn and J. F. Hoorn, "Humanoid Robot Tutors Times Tables: Does Robot's Social Behavior Match Pupils' Educational Ability?," *Proceedings IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2017 Workshop Mutual Shaping of Human-Robot Interaction*, 2017.

[2] P. Baxter, T. Belpaeme, L. Canamero, P. Cosi, Y. Demiris, V. Enescu, and A. Hiolle, "Long-term human-robot interaction with young users," *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop)*.

[3] Z. J. You, C. Y. Shen, C. W. Chang, B. J. Liu, and G. D. Chen, "A Robot as a Teaching Assistant in an English Class," *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT06)*.

[4] E. Rintjema, R. van den Berghe Berghe, A. Kessels, J. de Wit, and P. Vogt, "A Robot Teaching Young Children a Second Language," *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI 18*, pp. 219–220, 2018.

[5] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains," *CogSci*, 2012.

[6] I. S. Huang, J. F. Hoorn, "Having an Einstein in Class," *https://ivyhuangshiming.github.io/lab/EinsteinRobotTechingMathTechRep/.* Hong Kong, The Polytechnic University of Hong Kong, 2018.

[7] G. Levine, M. C. Page, S. L. Braver, and D. P. MacKinnon, *Levines guide to SPSS for analysis of variance*. Hillsdale, NJ: Erlbaum, Lawrence, Associates, 2003.

[8] J. T. Newsom, "USP 634 Data Analysis I," *http://web.pdx.edu/~newsomj/da1/ho_simple effects.pdf*..

[9] E. A. Konijn and J. F. Hoorn, "Humanoid Robot Tutors Times Tables: Does Robot's Social Behavior Match Pupils' Educational Ability?," *Proceedings IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2017 Workshop Mutual Shaping of Human-Robot Interaction*, 2017.

[10] T. Levin and C. Gordon, "Effect of Gender and Computer Experience on Attitudes toward Computers," *Journal of Educational Computing Research*, vol. 5, no. 1, pp. 69-88, 1989.

[11] P. Schermerhorn, M. Scheutz, and C. R. Crowell, "Robot social presence and gender: Do females view robots differently than males? ," *Proceedings of the 3rd international conference on Human robot interaction - HRI 08*, 2008.

[12] D. Kuchenbrandt, M. Häring, J. Eichberg, and F. Eyssel, "Keep an Eye on the Task! How Gender Typicality of Tasks Influence Human–Robot Interactions," *Social Robotics Lecture Notes in Computer Science*, pp. 448–457, 2012.

[13] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive Robotics: The influence of robot gender on human behavior," *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[14] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, "If you sound like me, you must be more human," *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI 12*, 2012.

[15] H. Meindertsma, M. van Dijk, H. Steenbeek, and P. van Geert, "Stability and Variability in Young Childrens Understanding of Floating and Sinking During One Single-Task Session," *Mind, Brain, and Education*, vol. 8, no. 3, pp. 149–158, 2014.