Einstein Robot Teaching Math
(Technical Report)

Ivy S. Huang and Johan F. Hoorn

The Hong Kong Polytechnic University

Abstract

We conducted experiments on robotics among a total of 76 students in two rural schools in China. The content of the teaching included symbolistic representation and simple equations. The students were equally divided into a group teaching with robot (VyesR) and a group without robot (VnoR) in advance, with an about equal number of boys and girls, and about equal division of math skills as indicated by their teachers. The only difference between two kinds of groups was whether having robot standing next to the instructional video and performing simple animation. Those students are from fourth grade and have little knowledge about simple equations (baseline 0). Students should watch the 8-minutes instructional video with or without robot, and then fill in the Likert evaluation and test paper. Likert evaluation was on video and robot (if had), and test paper contained two symbolic representations and six simple equations to solve (eg. 3y=9). Afterwards, we found all students achieved over 50% gain from baseline zero in solving equations, in which boys in VnoR gained 18% more than those in VyesR (70% and 52% respectively), while girls in VyesR gained 20% more those in VnoR (71% and 51% respectively). Conclusion was made that students can learn both from the video and robot, but girls benefit more from robot teaching, while boys benefit more from video. Results are discussed that gender difference must be taken into account in robot teaching promotion.

Content

Abbreviation List

VyesR	Video with robot (group with robot)
VnoR	Video without robot (group without robot)
NanXing / N	NanXing primary school
BoMei / B	BoMei primary school
#SR/ Sym	Score of correct symbolic representation
%C / EqC	Percentage of correct equations
App-r	Appreciation for robot
App-v	Appreciation for video
F	Female
M	Male

Experimental Plan

Hypothesis

Based on the previous studies, we formulated two hypotheses: Our first hypothesis is that a robot teaching new contents increases the learning outcomes compared to baseline (H1); 2) While being instructed through video, the addition of a robot will motivate children more than video alone (H2).

Participants

We have two sets of fourth-grade students from NanXing primary school (NanXing) and BoMei Secondary Primary School (BoMei). The experiment was performed independently in two schools separately.

NanXing primary school is in Meitang Town, Puning City and BoMei Secondary primary school in BoMei Town, Lufeng city. Both are countryside-level and defined as third-tier cities, and they are sharing the similar educational equipment and never exposed in high technology teaching environment. They are satisfied the first requirement of participant we want to find, who are in rural countries and know little about the robotics.

Besides, they have the resemble language (whose mother language are Teochew dialect and teaching language are Mandarin) and culture. So it can be confirmed that they do not have language barrier during the experiment. The students from these two schools have the similar family financial condition and most of them are taken care by their grandparents because their parents are working out of the towns. In general, grandparents focus less on their grandchildren' academic performance, therefore students do not have excessive pressure on learning, which makes them less motivated to learn the knowledge in advance. This makes them meet the requirement that the participant should have a zero baseline towards the knowledge in the experiment.

We consulted some experienced teachers about what grade students are suitable for the experiment. We got the feedback that the students below grade two might be too young to focus on the experiment and those higher than four grade might too mature to psychologically accept what the experiment displays. So we picked the grade-fourth children. In table 1, you can find the gender and group distribution of these two schools.

School & RoboYN	NanXing		BoMei	
	VnoR	**VyesR**	**VnoR**	**VyesR**
Female	8	8	11	8
Male	9	9	11	12
Total	17	17	22	20

Table 1: School, gender and group distribution

Materials

We prepared a simple robot from Hanson Robotics called Professor Einstein, an 8-minute Chinese spoken video instruction, test paper and appreciation rating scale.

Hanson's Professor Einstein is a 15-inch robot with impressive and expressive facial features (Figure 1). Its arms move and one of its hands has an articulated finger that can point. The Professor's feet have wheels to have a walk. Einstein has a microphone and a camera that supposedly help it to look at people when they talk to him. The robot has two voice, a kind of narrator that sounds like a young man and gives you instructions, and Einstein's voice that sounds like a wisdom old man and gives you response. The volume of the voice cannot be controlled. The Professor has three modes: an offline, online, and mobile mode. When Einstein's online, everything you say to it gets processed in the cloud, and you should wait a few seconds for the robot to answer you back. Only the questions began with "Who is" or "What is" can successfully get responses from the robot. In offline mode, you can "chat" with the professor and ask it to do things, like smile, stick out his tongue, talk a walk, or go crazy. Mobile mode connected the Professor directly to the app for playing games and taking quizzes through the supplementary application called Stein-O-Matic on the tablet, or to the laptop to play the animation that you pre-set on the Scratch-X. It fits the experiment setting as a very simple robot. During the experiment, the robot pretended as a teacher by playing the corresponding animation and as speaker source.

Figure 1. Professor Einstein

An 8-minute Chinese spoken video instruction teaches symbolic representation and simple equation solving. Symbolic representation is the base of the simple equation. Only knowing the symbol, like letter and shape, can represent an unknown number do they understand the meaning of equation. In the first two and half minutes of the video, the students will learn to use y to represent the unknown quantity of a bag of apple and how to simplify the expression from, for example, $3+y+y$ to $3+2y$ (Figure 2). During the next one minute of the video, students will learn the meaning of equation through the concept of balanced scales (Figure 3), and in the rest time, the video will use three examples ($y+2 =5$, $y-2=3$, $y+y =6$) to demonstrate the equation solving procedure (Figure 4). Solving simple equations is taught in the first semester of fifth grade. Combined with the learning culture we mentioned above, we were certain that the teaching content was totally new to the students. The video was recorded in old man sound pretending it was the voice of the Einstein robot.

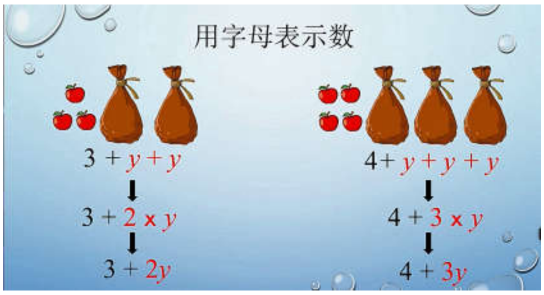


Figure 2. Represent the number with formal symbols

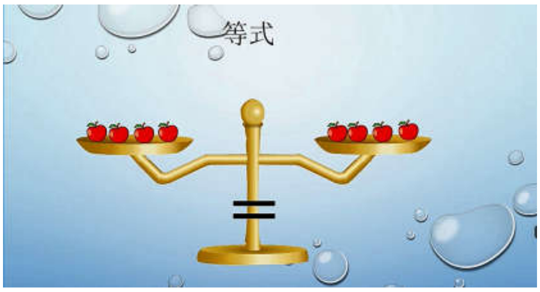


Figure 3. The meaning of equation

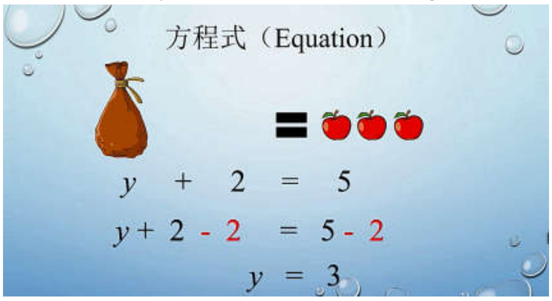


Figure 4. Simple equation solving

Test paper is to test the students' learning outcome, consisting two items on symbolic representation and six items on simple equation. Each symbolic representation item shows a number of apples and bags and children are asked to translate the picture into the formal symbol expression. The correct answers to items of symbolic representation in Figure 5 would be 5+*y* and 4+2*y*. The correct answers to items of simple equation sequentially would be 1, 8, 2, 2, 3 and 4.

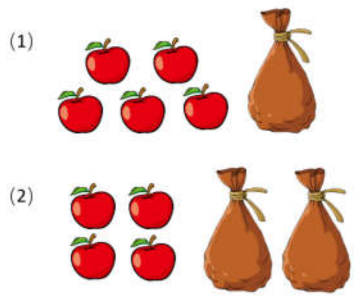


Figure 5. Items of symbolic representation

(1) y + 2 = 3 (2) y – 3 = 5

y = _____ y = ________

(3) 2y + 3 = 7 (4) 4y - 4 = 8

y = _____ y = _____

(3) 3y = 9 (3) 14 = 3y + 2

y = _____ y = _____

Figure 6. Items of simple equation solving

For the appreciation rating scale, there are two versions of the groups with robot and the groups without the robot. The version for the groups with robot consists of two single-item rating scale for appreciation for video and robot. And the version for the groups without the robot only has one single-item rating scale for appreciation for the video. Figure 7 and Figure 8 depict the rating scales.

Figure 7. Appreciation rating scale for group with robot

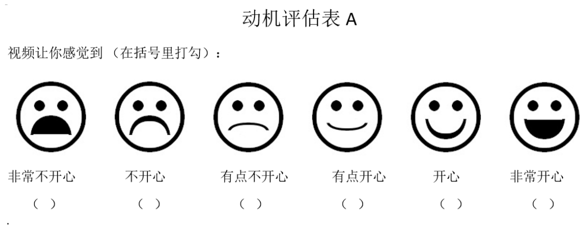


Figure 8. Appreciation rating scale for the group without robot

A piece of appreciation rating scale, a draft paper and a copy of test paper are clipped to a clipboard from top to bottom, and children are not allowed to turn to the next page before they finish the top one. In this case, children would not add the influence of test paper for them to their appreciation for solely lecture experience.

Procedure

NanXing primary school and BoMei primary school provided us a bit different supporting. According to the conditions, grouping, experiment environment setting, and the order of execution were made a corresponding adjustment between these two school.

Here we go to the grouping. Our initial number of each group was less or around 15 children in order to well maintain discipline during the experiment as well as ensure children have clear view of robot and video. In NanXing primary school, there are total 34 fourth grade students took part in our experiment, they were divided into two group and each group had 17 children. However, there were totally 42 children in BoMei primary school. To control the number in each group, we divided them into four groups, two of which are groups with robot and the rest are groups without robots.

We conducted the experiment in NanXing primary school on weekdays so they could not provide us two standardized classrooms, so we made use of the music room as experiment room and multimedia room as buffer room. These two rooms were opposite to each other. The Figure 9 and Figure 10 depicts the environmental map and photo in NanXing primary school.

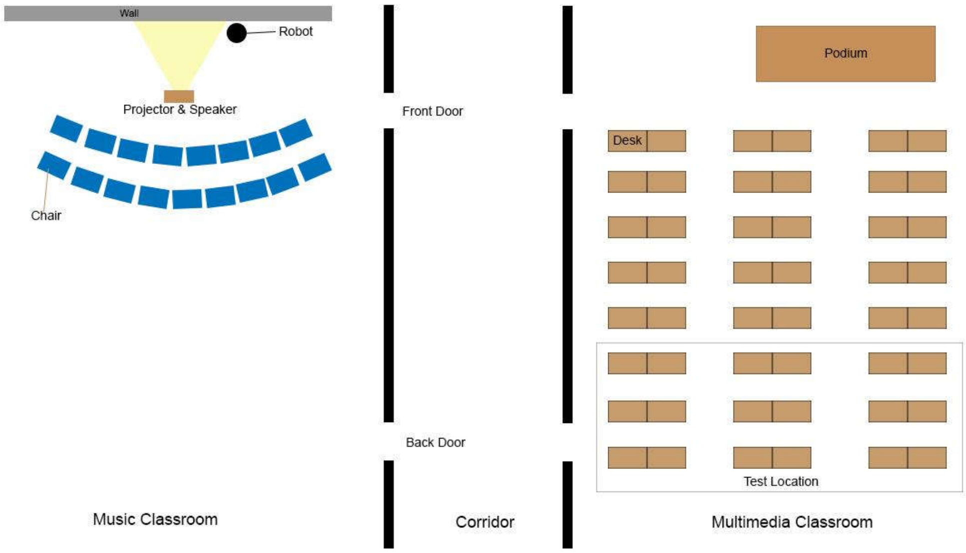


Figure 9. Environmental map in NanXing primary school

Figure 10. Photo of the experiment room in NanXing primary school

The BoMei primary school permitted us to conduct our experiment on Saturday and supplied us three standardized classrooms which we could reset. Figure 11 shows the environmental setting in BoMei primary school. The right classroom was for the group with robot (Room A) and the left one was for the group without the robot (Room B). The only difference between them was without having the table, on which stood the robot, on the left front of the portable screen. Figure 12 is the live photo during the experiment on the group with robot in BoMei primary school.

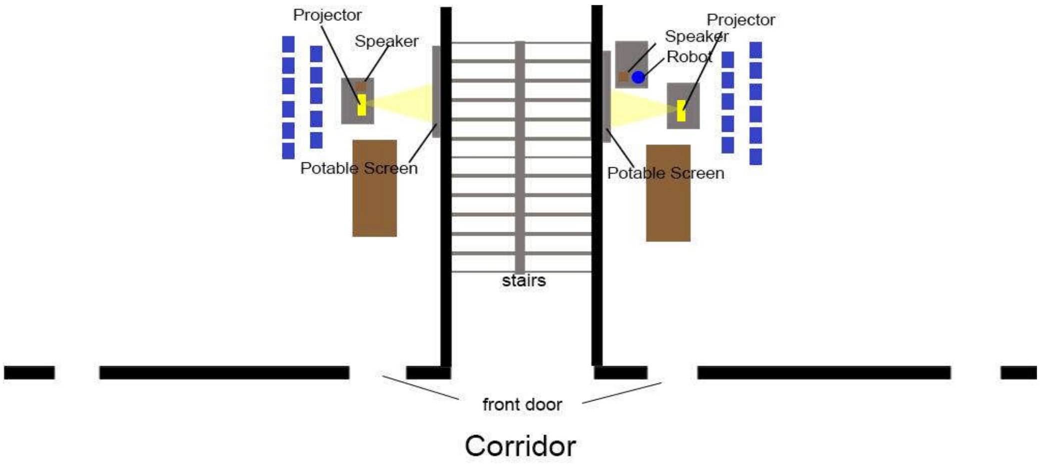


Figure 11. Environmental map in BoMei primary school

Figure 12. Live photo of the experiment in BoMei primary school

Then we designed the order of experiment execution in two schools. NanXing primary school could only give us 45 minutes to perform the experiment, combining that we only had one experiment classroom abovementioned (Figure 9), we made the experiment run half-serially. Table 2 is the experiment implementation timetable in NanXing primary school.

26th March, 2018		
Time	VnoR	VyesR
14:30	Students gathered in their own classroom. They were told (1) firstly they would see an 8-minute video; (2) after video instruction they need to do the appreciation rating scale and test paper	
14:35	Remained in their classroom	Brought to the music classroom

14:40		Had the video instruction with robot
14:48	Began to setting off to the music classroom	Brought to the multimedia classroom
14:50	Had the video instruction with robot	Do the test paper and the appreciation rating scales
14:55		Collecting test papers and the appreciation rating scales
15:00	Did the test paper and appreciation rating scales	Waited in the multimedia classroom
15:05	Were collected test papers and the appreciation rating scales	
15:10	All groups and instructors gather in the music classroom, interacting with the robot	

Table 2. Implementation timetable in NanXing primary school

As we mentioned before, there were four groups in BoMei primary school and two experiment rooms. We divided the groups into two sections, each of which has a group with robot and a group without robot. The two sections performed the experiment serially and two groups in a same section run parallel. Table 3 depicts the implementation timetable in BoMei primary school.

<table>
<tr><td colspan="5">5th May, 2018</td></tr>
<tr><td rowspan="2">Timeslot</td><td colspan="2">Section 1</td><td colspan="2">Section 2</td></tr>
<tr><td>VyesR</td><td>VnoR</td><td>VyesR</td><td>VnoR</td></tr>
<tr><td>14:25 –14:35</td><td colspan="4">Students gathered in their own classroom. They were told (1) firstly they would see an 8-minute video; (2) after video instruction they need to do the appreciation rating scale and test paper</td></tr>
<tr><td>14:40 – 14:50</td><td rowspan="2">● Watched video
● Filled out the appreciation rating scales
● Did the test paper
(in Room A)</td><td rowspan="2">● Watched video
● Filled out the appreciation rating scales
● Did the test paper
(in Room B)</td><td colspan="2">Waited in their own classroom</td></tr>
<tr><td>14:50 – 14:55</td><td colspan="2">Lined up to go to the experimental rooms</td></tr>
<tr><td>14:55 – 15:00</td><td colspan="2">Lined up to go back to the classroom</td><td colspan="2">Prepared for the experiment</td></tr>
<tr><td>15:00 – 15:15</td><td colspan="2">Waited in their own classroom</td><td>● Watched video
● Filled out the appreciation rating scales
● Did the test papers
(in Room A)</td><td>● Watched video
● Filled out the appreciation rating scales
● Did the test papers
(in Room B)</td></tr>
<tr><td>15:15 – 15:45</td><td colspan="4">Interaction with the robot in the classroom</td></tr>
</table>

Table 3. Implementation timetable in BoMei primary school

To summarize, none of the students were informed about the presence of the robot, and they were informed before the experiment that they would watch the video instruction and later did an appreciation rating scale. Then the group would take turns to the corresponding experiment room to participate the experiment. They firstly watched the video instruction, and then the research assistants would distribute the clipboard with paper materials on to them (refer to ***Materials***) to fill out the test paper and appreciation rating scale. After finished all the task, the paper materials would be collected by the research assistant and scored.

During the whole experiment, children were not allowed to chat with classmates or research assistants to avoid peer influence and assistant preference influence on the experiment result. Though the two groups in NanXing primary school (Table 2) would see another group at classroom exchange around 14:48, we make the previous group move from the front door and the latter group came in through the back door without any chatting. It is the same in BoMei primary school where there was a crossed time when two sections of children exchange their position. To avoid the information of robot was leaked, we required the group with robot in the first section to keep it as secret what happened in the classroom and promised that they would get a reward if they made it. In fact, after we interviewed the groups in the second section afterward, we found they did not know the presence of the robot. So we could say, the groups are all independent and at the same beginning status.

Equipment

In the groups with robot, they had the robot pretending as a teacher by playing the simple animation, like turning its head and blinking its eyes and being speaker source. To successfully play the animation, the Einstein robot should be set on mobile mode and remotely controlled by the experiment assistant through the computer. But before the experiment in NanXing primary school, the robot was found could not work well in mobile mode, so we turned it into online mode. It could maintain power on in eight minutes and turned its head and blinked its eyes since there was sound (from the speaker) around it. The imperfection was that it would say something (like encourage asking it questions) spontaneously and played pointing animation during speaking. We placed some transparent tape on its built-in speaker trying to reduce the irrelevant sound, but the sound still could be heard. While in the BoMei primary school, it could work well in the mobile mode so we set it that mode. In Table 4, it could be found the status and information of the equipment we used during the experiment.

School	NanXing primary school		BoMei primary school	
Group	Group with robot (VyesR)	Group without robot (VnoR)	Groups with robot (VyesR)	Groups without robot (VnoR)
Laptop (video playing)	Intel-i5	Intel-i5	Intel-i7	Intel-i3
Laptop (robot controlling)	/	/	Intel-i5	/
Robot	Online mode	/	Mobile mode	/
Projector	36.1 (diagonal)	36.1 (diagonal)	36.1 (diagonal)	36.1 (diagonal)
Speaker	√	√	√	√

Table 4. Equipment information

Measurement

The test paper was used to measure whether they got the information the video instruction taught them rather than judge whether their answer was correct. So except the standardized answer above mentioned (***Materials***), the answer of symbolic representation items with the extra information and the answer which did not be simplified will be judged as getting score. For example, a child's answer for the first item (whose standardized answer should be 5+y) in symbolic representation was *5+y=10 y=5* and we could find the *=10 y=5* was the extra information. We still marked it correct because the student really learns something from the video though she or he did not master it well. Taking the time (only 8 minutes) of lecture instructing a new mathematical knowledge as consideration, the unfamiliar was understandable. So the child giving *14-2÷3* as the answer of, for example, the sixth item in simple equation (14 = 3y+2) should be judged having learned something from the video. The children learn from the video that *y* should be got through simultaneously subtracting 2 and dividing by 3 for both sides of the equation though he or she forgot or did not know how to use brackets to declare the calculation priority. And bracket was not stressed in the video instruction. For the appreciation rating scale, the items from left to right represents 1 to 6 points.

Data Analysis

Overview

The data was collected from the children' test paper and appreciation rating scale (Figure 13). The column with title *School* has two options, which are NanXing and BoMei. NanXing represents the NanXing primary school and BoMei represents BoMei primary school. It used to differentiate the school the children are from. The column *RoboYN* is to recognize which group the children are in. They are from either VyesR (group with robot) or VnoR (group without robot). The data from *Gender* column are afterward added according to their names and confirmation by their head teachers. The columns with the main title *Test paper* included the raw data in their test papers and scores of symbolic representation items and simple equation items. The columns with subtitle in the format of *Ans number1 (number2)* recorded the initial answer the children gave. *number 1* is the index of test paper, representing two question types – 1 is symbolic representation and 2 is simple equation. The second number in brackets marked the sub-index in their own question group.

According to the rating method explained in ***Measurement*** in the last section, we marked the cell as red which was seen as the subject had learned from the video instruction. In the column with subtitle *#SR*, we add up the number of previous two red cell. It would be 0, 1 and 2. Among the children, four of them skipped both items of symbolic representation, so we marked their total scores as N/D. The column with subtitle *%C* calculated the percentage of correctly solved equations. For those who solve more than one to six, we took that as hundred percent. So those people, for example, who solve five and missed one will still be considered full score. There were seven children skipped one slot and two children skipped two and three separately. Only a child who did not fill out one of the equation would not be included in the analysis of the percentage correct and marked as N/D. Table 5 depicts the distribution of missing slot in the test paper. In the last second column *MissingYN*, we marked whether there was the missing slot in the simple equation part. Column *CountYN* will be explained later.

School & RoboYN	**NanXing**				**BoMei**			
	VnoR		**VyesR**		**VnoR**		**VyesR**	
Number of Missing slot	#SR	%C	#SR	%C	#SR	%C	#SR	%C
0	15	15	16	13	22	21	19	17
1	0	1	0	3	0	1	0	2
2	2	0	1	1	0	0	1	0
3		0		0		0		1
6		1		0		0		0

Table 5. Distribution of missing slot

For the appreciation for robot and video, we marked them in the column with subtitle *App-r* and *App-v* respectively. The groups without the robot (VnoR) would skipped the rating for the robot, and you could see the blank cell in the corresponding location. There was one child in the group without robot (VnoR) did not fill out the evaluation for the video and we took it as N/D.

School	RoboYN	Gender	Ans1_1	Ans1_2	SymR	Ans2_1	Ans2_2	Ans2_3	Ans2_4	Ans2_5	Ans2_6	EqC	AppR	AppV	MissingYN	CountYN
School	RoboYN	Gender	Test paper										Motivation score sheet		MissingYN	CountYN
			Ans 1 (1)	Ans 1 (2)	#SR	Ans 2 (1)	Ans2 (2)	Ans 2 (3)	Ans 2 (4)	Ans2 (5)	Ans 2 (6)	%C	App-r	App-v		
NanXing	VnoR	Female	5=y	4=y	0									3	Y	N
NanXing	VnoR	Female	5=y	4=y+y	0	2+3	5-3	3+7	8-4	3	14+3	0.00%		5	N	Y
NanXing	VnoR	Female	y=5	y2=4	0	2	3	2	4	3	3	33.33%		5	N	Y
NanXing	VnoR	Male	5+y=8	4+2y=12	2	1	8	2	4	3	4	83.33%		3	N	Y
NanXing	VnoR	Male				1	8	2	3	3	4	100.00%		6	N	Y
NanXing	VnoR	Male	8	10	0	1	8	4	4	6	1y	33.33%		6	N	Y
NanXing	VnoR	Male	5+y=8	4+2y=10	2	1	8	2	1	3	4	83.33%		6	N	Y
NanXing	VnoR	Male	y+5 =8	2y+4 =12	2	1	8	2	12	3	4	83.33%		6	N	Y
NanXing	VnoR	Female	5=y	4=y+y	0	2+3	5-3	7+3	8-4	9-3	2+3	0.00%		5	N	Y
NanXing	VnoR	Female	5=y	4=y+y	0	2+3	3-5	3+7	8-4	9-3	y+2	0.00%		5	N	Y
NanXing	VnoR	Female	y=5	2y=4	0	5	8	10	3		5	40.00%		5	Y	Y
NanXing	VnoR	Male				1	8	2	4	3	4	83.33%		4	N	Y
NanXing	VnoR	Male	5+y=5y	4+2y =42y	2	1	8	2	3	3	4	100.00%		4	N	Y
NanXing	VnoR	Male	5+y	4+2y	2	1	8	2	3	3	4	100.00%		6	N	Y
NanXing	VnoR	Male	5+y	4+2y	2	1	8	2	3	3	4	100.00%		6	N	Y
NanXing	VnoR	Female	y=5	2y=4	0	1	8	2	1	3	3	66.67%		3	N	Y
NanXing	VnoR	Female	y=5	y=2	0	1	8	2	1	3	4	83.33%		3	N	Y
NanXing	VyesR	Male	y=5	2y=4	0	1	8	2	3	3	4	100.00%	5	4	N	Y
NanXing	VyesR	Male	5=2y	4=2y	0	1	8	2	3	3		100.00%	4	4	Y	Y
NanXing	VyesR	Male	5=1y	4=2y	0	1	8	2	3	3	4	100.00%	3	4	N	Y
NanXing	VyesR	Male	5 y	4 2y	0	1	8	2	3	3	4	100.00%	5	5	N	Y
NanXing	VyesR	Female	5y	5y	0	1	8	2	3	6	9	66.67%	6	4	N	Y
NanXing	VyesR	Female	5+y	4+y+y	2	1	8	2	3	3		100.00%	5	4	Y	Y
NanXing	VyesR	Female	5+y	4+y+y	2	1	8	2	3	3		100.00%	6	5	Y	Y
NanXing	VyesR	Male				5	2	10	4	9	5	0.00%	6	6	N	Y
NanXing	VyesR	Male	5+y	4+y+x	1	5	2	10	4	9	2	0.00%	6	6	N	Y
NanXing	VyesR	Male	5+y	4+y+y	2	5	2	10	4	9	5	0.00%	6	6	N	Y
NanXing	VyesR	Female	5+y	4+y+y	2	1	8	2	3	3	4	100.00%	6	6	N	Y
NanXing	VyesR	Male	y=5	y=2	0	1	2	2	4	3	9	50.00%	6	4	N	Y
NanXing	VyesR	Female	y=5	2y=4	0			5y	4y	6y	5y	0.00%	4	4	Y	N
NanXing	VyesR	Female	5+y	4+y+y	2	1	8	2	1	3	4	83.33%	6	5	N	Y
NanXing	VyesR	Female	5+y	4+y+y	2	1	8	2	1	3	4	83.33%	5	6	N	Y
NanXing	VyesR	Female	5+2y	x 4=2y	0	1	8	2	3	3		83.33%	6	5	N	Y
NanXing	VyesR	Female	5=5	y+y=8	0	1	8	2	3	3	4	100.00%	6	2	N	Y
BoMei	VnoR	Female	5+y=10 y=5	2y =4 2y÷2 = 4÷2=2 4×2=8	1	1	8	2	12	3	4	83.33%		5	N	Y
BoMei	VnoR	Female	5+y=10 y=5	4+2y=8 y=2	2	1	8	2	12	3	4	83.33%		5	N	Y
BoMei	VnoR	Female	5+y=10 y=5	4+2y=8 y=2	2	1	8	4	12	3	12	50.00%		4	N	Y
BoMei	VnoR	Female	5+y=10 y=5	4+y=8 y=2	1	1	8	4	12	3	4	66.67%		4	N	Y
BoMei	VnoR	Female	5+y =10 y=5		1	1	8	2	12	3	4	83.33%		6	N	Y
BoMei	VnoR	Female	5+y=8	4+y+y =10	2	3	8	6	12	9	14	16.67%		4	N	Y
BoMei	VnoR	Male	5+y=8	4+y+y=10	2	1	8	4	12	9	12	33.33%		6	N	Y
BoMei	VnoR	Male	5+y	4+y×2	2	1	8	4	12	9	12	33.33%		6	N	Y
BoMei	VnoR	Male	5+y=8	4+y×2	2	3	8	4	12	3	4	50.00%		4	N	Y
BoMei	VnoR	Female	5+y×1	4+y×2	2	3	2	2	4	9	9	16.67%			N	Y
BoMei	VnoR	Male	5+y =8	4+y×2	2	1	2	2	4	3	4	66.67%		4	N	Y
BoMei	VnoR	Female	y=5 5+5=10	y=4 4+4+4=12	0	3-2	3+5	7-3+2	4+8+4	9+3	14-2+3	100.00%		5	N	Y
BoMei	VnoR	Male	y+5=5x	4y+2y=2	1	3-2	3+5	7-3+2	4+8+4	9+3	14-2+3	100.00%		5	N	Y
BoMei	VnoR	Male	y=5 5+y=10	y=4 4+2y=12	1	3-2	3+5	7-3+2	4+	9+3	14-2+3	83.33%		4	N	Y
BoMei	VnoR	Male	5+y=10 5=y	y=4 4+y+y	2	1	3+2	7-3+2	4+	9+3	14-2+3	66.67%		4	N	Y
BoMei	VnoR	Male	5=y	4=1y	0	1	8	7-2+3	12	3	4	50.00%		4	N	Y
BoMei	VnoR	Male	5+y	4+y	1	1	8	2	12	3	12	66.67%		6	N	Y
BoMei	VnoR	Male	5+y	4+y+y	2	1	8	2	8	6	12	50.00%		6	N	Y
BoMei	VnoR	Male	5+y	4+y+y	2	5	8	5	8	3	6	33.33%		6	N	Y
BoMei	VnoR	Female	5+y	4 y+y 2y	1	2	2	3y	16	9y		0%		6	Y	Y
BoMei	VnoR	Female	y=5	y=2	0	1	8	2	3	3	4	100.00%		6	N	Y
BoMei	VnoR	Female	y=5	y=2	0	1	8	2	3	3	4	100.00%		6	N	Y
BoMei	VyesR	Male	1y×5=	y+y=2y 2y×4=	0	1	8	2				100.00%	5	5	Y	N
BoMei	VyesR	Male	5+y =8	4+y+y=4+2y	2	3	5	6	8	9	14	0%	6	5	N	Y
BoMei	VyesR	Female	5+y=8	4+y+y=4+2y	2	3	8	6	12	3	9	33.33%	5	3	N	Y
BoMei	VyesR	Female	5+y=10	4+2×y=14	2	1	8	4	4	3	6	50.00%	6	6	N	Y
BoMei	VyesR	Female	5+y=10	4+2y=12	2	1	8	2	1	3	4	83.33%	6	6	N	Y
BoMei	VyesR	Female	5+y=1-	4+2y=12	2	1	2	4	4	3	6	33.33%	6	6	N	Y
BoMei	VyesR	Male	5+y=6	4+y+y+2y=12	1	5	8	4	12	3	4	50.00%	6	6	N	Y
BoMei	VyesR	Male	5+y=6	4+2y=12	2	5	8	4	12	3	12	33.33%	6	4	N	Y
BoMei	VyesR	Male	5+y=2y	y+y=2y 4×2y=	1	1	8	4	3	3	4	83.33%	5	5	N	Y
BoMei	VyesR	Male				5	2	2y		3	4	40.00%	6	5	Y	Y
BoMei	VyesR	Male	5+3=8	4+6=10	0	1	8	2	3	3	4	100.00%	6	6	N	Y
BoMei	VyesR	Female	5+y=8	4+2y=10	2	1	8	2	12	3		80.00%	6	6	Y	Y
BoMei	VyesR	Male	5+y=8	4+2y=10	2	1	8	3	2	8	6	33.33%	6	6	N	Y
BoMei	VyesR	Male	y×5=10	4×y×y=12	0	1	8	4	12	9	12	33.33%	6	5	N	Y
BoMei	VyesR	Female	5+y=8	4+(y×2)=10	2	1	8	4	12	3	6	33.33%	6	6	N	Y
BoMei	VyesR	Male	5+y=10	4+y+y=12	2	1	8	4	3	3	4	83.33%	4	5	N	Y
BoMei	VyesR	Male	5+y=7	4+y+y=7	2	1	2	4	2	3	12	33.33%	5	4	N	Y
BoMei	VyesR	Male	5+2=7	y+4=3	0	11	8	2	4	6	5	33.33%	3	1	N	Y
BoMei	VyesR	Female	5+y=8	4+2y=10	2	1	2	2	4	3	12	50.00%	6	5	N	Y
BoMei	VyesR	Female	5+y=8	4+y+y=10	2	1	8	5	12	3	12	50.00%	6	6	N	Y

Figure 13. EinsteinRobotInXinanBoMeiTeachingMath_forSAS.xlsx

Analysis process

Table 5 shows the means and standard deviation for correctly solved equation, correctly represented symbol expression and appreciation for media. For those who missed more than one equation were treated as outliner, which means that we performed the analyses once with outliers included and once without to see whether that made a difference. In the cells ④ ⑤ ⑥ ⑯ ⑰ ⑱, the words in black are the results of analysis with outliers and words in red are those without outliers. We found there were no much difference between, so we would include the outliers in further analysis.

<table>
<tr><th colspan="11">Within Subjects</th></tr>
<tr><th colspan="3"></th><th colspan="2">#SR</th><th colspan="2">%C</th><th colspan="2">App-r</th><th colspan="2">App-v</th></tr>
<tr><td rowspan="4">B</td><td rowspan="2">Video with Robot VyesR</td><td>NanXing</td><td>①
$M = 0.81$
($S = 0.99$)</td><td rowspan="2">③
$M = 1.17$
($S = 0.95$)</td><td>④
$M = 69\%$
($S = 0.42\%$)
$M = 73\%$
($S = 0.39\%$)</td><td rowspan="2">⑥
$M = 60\%$
($S = 0.35\%$)
$M = 60\%$
($S = 0.34\%$)</td><td>⑦
$M = 5.35$
($S = 0.93$)</td><td rowspan="2">⑨
$M = 5.46$
($S = 0.87$)</td><td>⑩
$M = 4.71$
($S = 1.10$)</td><td rowspan="2">⑫
$M = 4.89$
($S = 1.20$)</td></tr>
<tr><td>BoMei</td><td>②
$M = 1.47$
($S = 0.84$)</td><td>⑤
$M = 52\%$
($S = 0.27$)
$M = 49\%$
($S = 0.25\%$)</td><td>⑧
$M = 5.55$
($S = 0.83$)</td><td>⑪
$M = 5.05$
($S = 1.27$)</td></tr>
<tr><td rowspan="2">Video Without Robot VnoR</td><td>NanXing</td><td>⑬
$M = 0.80$
($S = 1.01$)</td><td rowspan="2">⑮
$M = 1.11$
($S = 0.91$)</td><td>⑯
$M = 62\%$
($S = 0.38$)
$M = 62\%$
($S = 0.38\%$)</td><td rowspan="2">⑱
$M = 61\%$
($S = 0.33\%$)
$M = 61\%$
($S = 0.33\%$)</td><td rowspan="2" colspan="2"></td><td>⑲
$M = 4.76$
($S = 1.20$)</td><td rowspan="2">㉑
$M = 4.92$
($S = 1.05$)</td></tr>
<tr><td>BoMei</td><td>⑭
$M = 1.32$
($S = 0.78$)</td><td>⑰
$M = 61\%$
($S = 0.30\%$)
$M = 61\%$
($S = 0.30\%$)</td><td>⑳
$M = 5.05$
($S = 0.92$)</td></tr>
</table>

Table 6. Means and Standard Deviation for correctly solved equation and appreciation for media

The analysis is divided into 4 parts. The first part is analysis on #SR, followed by %C and then App-r and App-v. A regression analysis between %C and App-r as well as App-v is given in the last section. Observations with respect to different independent variables were brought into appropriate analysis models for hypothesis testing, with the model explained, followed by discussions of the results obtained from SAS.

1. Score of correct symbolic representation (#SR)

The observations can be grouped into School, Gender and Robot (robot-assisted video teaching). It is possible to investigate whether the distribution of #SR is affected by the any of the abovementioned independent variables. The objective here is to evaluate how likely the observation difference grouped by category ***G*** (***G*** = School, Gender, Robot) comes by chance.

If ***G*** is independent to #SR, then the distribution of #SR is the same regardless of the groups of ***G***, or else a statistically significant difference would be revealed. There are only 3 possible values of #SR: 0, 1 and 2, so while #SR is numeric, it is treated as categorical.

Chi-square test of independence is used, with the assumptions of simple random sample, sufficiently large sample size, adequate cell count (10 or greater) and independent observations. The null hypothesis is that the distribution of #SR in the two groups of each ***G*** (NanXing and BoMei for ***G*** = school, Female and Male for ***G*** = gender, and VyesR and VnoR for ***G*** = Robot) is statistically independent, i.e. they follow the same distribution. The alternative hypothesis is that the #SR distributions in ***G*** are different. The significance level α is set to 0.10. Only observations with all questions answered are taken into analysis, so one VyesR male pupil and 2 VnoR male from the 34 observations in NanXing, as well as 1 VyesR male from the 42 pupils in BoMei were taken out from the following analysis. The SAS code for the chi-square tests on #SR with each ***G*** is shown as follows:

```
proc freq data=WORK.Rawdata_exNullSymR order=formatted;
   tables SymR*School/chisq;
   title "a.Chi square test of School (B-N) on SymR";
run;

proc freq data=WORK.Rawdata_exNullSymR order=formatted;
   tables SymR*Gender/chisq;
   title "b.Chi square test of gender (b-g) on SymR";
run;

proc freq data=WORK.Rawdata_exNullSymR order=formatted;
   tables SymR*RoboYN/chisq;
   title "c.Chi square test of RoboYN (y-n) on SymR";
run;
```

Table 7 summarizes the test results. The sample size of the groups of each ***G*** is shown in the row *Total* and the distributions of #SR with different ***G's*** can be seen in the columns *Total*. It can be seen that the null hypotheses of the tests concerning Gender and Robot are not rejected, meaning that gender and robot instruction do not affect the #SR performance. Statistical significance is found on observations grouped by School, leading to a rejection of null hypothesis, inclining to the belief that the pupils in the two schools contributed to different #SR distributions of performance. When looking into the raw data, VyesR and VnoR pupils are half-half in NanXing for each of the #SR = 0 and 2 which dominante the #SR distribution, while in BoMei, 11 VnoR and 13 VyesR observations constitute the mode of the distribution (#SR = 2), which is also close to half-half contribution. Therefore #SR performance in BoMei appears to have a systematic upshift compared to that in Nanxing. This may be attributed to the disturbance of irrelevant sound produced by the robot during the experiment in Nanxing, as described in the Equipment section.

From Table 6, it is found that #SR covers virtually the entire possible range with 1 standard deviation (①-③, ⑬-⑮). There is some difference between schools for VyesR and VnoR groups, but the grand mean shows little difference. The large S.D. results in difficulty in drawing concrete conclusion for #SR with and without robot instruction. It is suggested the need of increasing the number of questions (range of #SR) or the sample size to reduce the relative error.

	School			Gender			Robot		
#SR	NanXing	BoMei	Total	Female	Male	Total	VnoR	VyesR	Total
0	18	8	26	15	11	26	13	13	26
1	1	9	10	4	6	10	7	3	10
2	12	24	36	17	19	36	17	19	36
Total	31	41	72	36	36	72	37	35	72
DF	2			2			2		
χ^2 value	13.1102			1.1265			1.6568		
Prob	0.0014			0.5694			0.4367		

Table 7 Summary of chi-square test of independence on #SR.

2. Percentage of correct equations (%C)

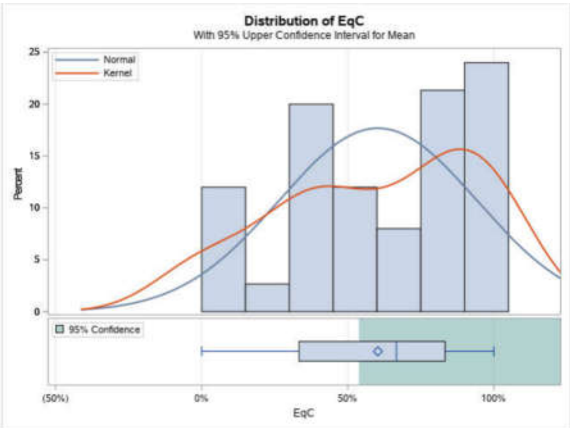
2.1 One sample *t*-test

A good beginning for the %C correlation is a one sample *t*-test to check if there is any difference in performance of teaching with and without robot. The objective in this section is to determine if the means of %C of the targeted sampling groups (all pupils (union set of VyesR and VnoR, or VyesR ∪ VnoR), VyesR, VnoR) statistically significantly differs from the baseline (zero). If the ability of solving equations of the pupils got improved in the experiment, the means of %C of the targeted population groups will be statistically greater than zero.

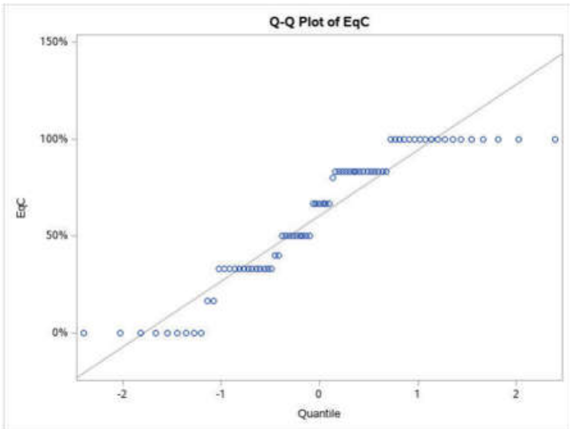
In one sample *t*-test, simple random sample, independent observations and normally distributed sampling population (with sufficiently large samples) are assumed. The null hypothesis is that the mean of %C of the group ***G*** (***G*** = (VyesR ∪ VnoR, VyesR, VnoR) equals 0 without prior knowledge of population variance. The corresponding alternative hypothesis is that the mean of %C of the group ***G*** is greater than 0, indicating an one-tail test. The significance level α is set to 0.10. One female VnoR observation from NanXing without completing any one question were taken out in the following hypothesis testing. The SAS code for the one sample *t*-test on %C with each ***G*** is shown as follows:

```
proc sql;
create table WORK.Rawdata_exNullEqC as
    select * from WORK.Rawdata where EqC is not null ;
run;
```

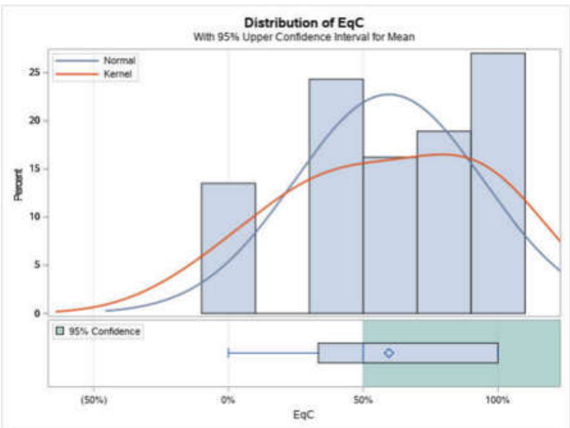
```
proc ttest h0=0 sides=u alpha=0.05 data=WORK.RawData_exNullEqC;
var EqC;
title "(4) One Sample test on %C - a. all";
run;

proc sql;
create table WORK.RawData_exNullEqC_RoboY as
    select * from WORK.Rawdata where EqC is not null && RoboYN ="VyesR";
run;

proc ttest h0=0 sides=u alpha=0.05 data=WORK.RawData_exNullEqC_RoboY;
var EqC;
title "(4)One Sample test on %C - b. RoboY";
run;

proc sql;
create table WORK.RawData_exNullEqC_RoboN as
    select * from WORK.Rawdata where EqC is not null && RoboYN ="VnoR";
run;

proc ttest h0=0 sides=u alpha=0.05 data=WORK.RawData_exNullEqC_RoboN;
var EqC;
title "(4)One Sample test on %C - c. RoboN";
run;
```



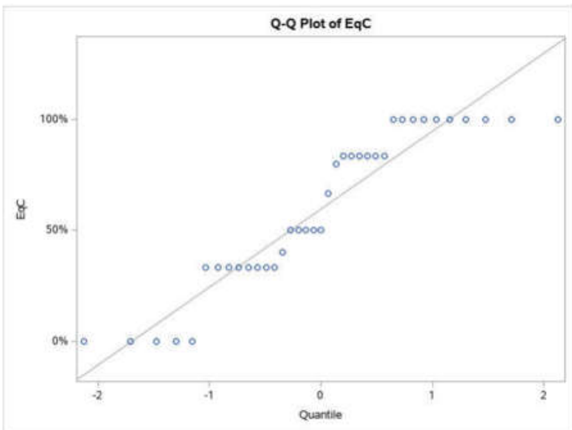
(a)



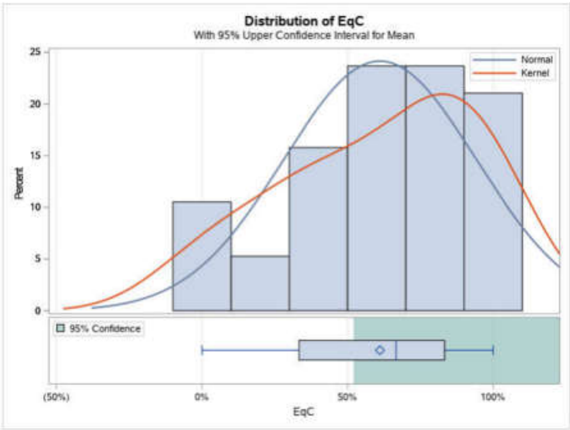
(b)



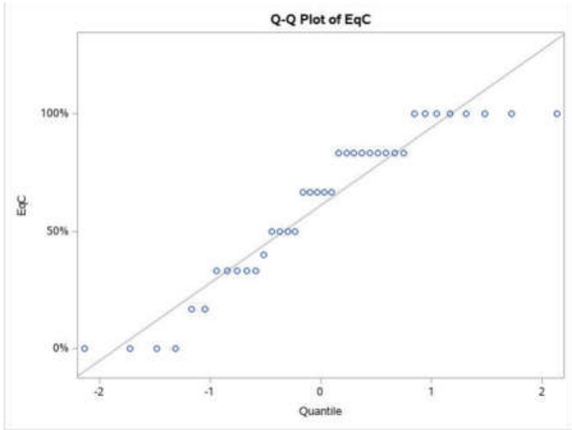
(c)



(d)



(e)



(f)

Figure 14 Distributions of %C of VyesR ∪ VnoR (a)-(b), VyesR (c)-(d) and VnoR (e)-(f). The left column shows the histograms of the distribution, with the fitted curves of normal distribution (blue) and kernel density estimation (KDE) (orange). The right column shows the Q-Q plot using standard score as the quartile axis.

	Mean (%)	S.D. (%)	# of samples	DF	*t* value	Pr > t
VyesR ∪ VnoR	60.36	33.84	75	74	15.45	<0.0001
VyesR	59.55	35.08	37	36	10.33	<0.0001
VnoR	61.14	33.04	38	37	11.41	<0.0001

Table 8 Summary of one sample *t*-test on %C.

Table 8 shows a summary of the one-sample *t*-test. All three groups shows statistical significance on the zero-mean %C, suggesting the adoption of the alternative hypothesis, that all three groups of pupils acquired better ability of solving equations in the experiment. The means of the three groups shows little difference, so a deeper understanding in the distribution itself may be needed, as shown in Figure 14. While normal sample distribution is assumed, multi-peak behavior is observed in these distributions (7%, 36% and 93% in VyesR ∪ VnoR, 40% and 100% in VyesR, and 0% and 70% in VnoR). The effect is readily observable with the kernel density estimation (KDE). Such phenomenon can also be read from the Q-Q plot. Observations in the region right to the straight line indicates excess frequency compared to the theoretically predicted normal distribution, and vice versa for those left to it. Therefore multi-peak behavior may be indicated by left-right-left pattern across the EqC axis. In VyesR ∪ VnoR, such regions include quartile ranges <-2 to -1, -1 to 0 and 0.15 to >2, similar to VyesR (-1 to 0 and 0 to >2) and VnoR (<-2 to -1 and 0 to >2). It can be seen that the observations in the Q-Q plots appear step-wise increment because of the limited possibilities of the values of %C with only 6 questions in the test paper, making it look only fairly fit to normal distribution. Therefore, more detailed analysis should be carried out to draw more meaningful test results.

It is worth noting that, when comparing the peak positions of the 3 distributions, it can be induced that the peak at around 36% in VyesR ∪ VnoR probably comes from VyesR as VnoR contains no peak at that position. However, since both VyesR and VnoR contains multiple peaks, it is of meaningful motivation to identify the sources of contribution of these peaks in the distributions, which is to be discussed in section 2.3.

2.2 2 × 2 × 2 ANOVA of School (B–N) × Gender (M-F) × Robot (VyesR–VnoR) on %C

While it is not surprising to see both video-with-robot and video-alone instructions give rise to surplus in equation solving ability, comparison between them is of primary interest. The objective in this section is to check the statistical equivalence of pupils' performance with explanatory variables, School – BoMei (B) vs NanXing (N), Gender – male (M) and female (F) and Robot – VyesR and VnoR using a 3-way (between subjects) ANOVA (GLM Univariable) on %C. There are 8 groups in total. Independent observations, normal distribution of residues and homoscedasticity are assumed in ANOVA. Prior knowledge of normal distribution of sample is not needed. The null hypothesis is that all the 8 groups have the same population mean of %C, i.e sampling mean difference comes by chance. The corresponding alternative hypothesis is that some specific groups of observations have a different mean of %C. The significance level α is set to 0.10 (one-tail test). Observations of missing no more than one equation was used in the analysis, and thus one female VnoR pupil from NanXing, and two VyesR pupils, one male from NanXing and one female from BoMei, were taken out from the analysis, resulting in a total of 73 observations. Type III sum of squares (SS) is used for the nature of unbalanced data (see Table 9) and neglection of factor order in the hypothesis testing. It is also sensitive for any interaction effect. The SAS code for the 2 × 2 × 2 ANOVA on %C is shown as follows:

```
proc sql;
create table WORK.Rawdata_exNullEqCnomissingMore as
    select * from WORK.Rawdata where EqC is not null && CountYN="Y";
run;

proc glm data = WORK.Rawdata_exNullEqCnomissingMore ;
    class School Gender RoboYN;
    model EqC = School|Gender|RoboYN / SS3;
    title "(2)  2 × 2 × 2 Anova of School (B - N) × Robot (y - n) × Gender (b - g) on %C ";
    Means Gender|RoboYN/ALPHA=0.1  LSD cldiff;
run;
```

Gender	Robot	Mean %	S.D.	Sample size
F	VnoR	51.30	38.40	18
	VyesR	70.62	25.36	16

M	VnoR	70.00	25.13	20
	VyesR	51.23	38.13	19

Table 9 Means and standard deviations (S.Ds.) of %C in different groupings of gender and robot.

Source	DF	Sum of squares	Mean square	*F* value	Pr > *F*
Model	7	2.078	2.078	3.29	0.0047
School	1	0.2000	0.2000	0.88	0.1416
Gender	1	0.0106	0.0106	0.65	0.7325
School × Gender	1	0.1004	0.1004	1.42	0.2955
Robot	1	0.005089	0.005089	0.00	0.8131
School × Robot	1	0.2852	0.2852	1.47	0.0803
Gender × Robot	1	0.7946	0.7946	5.51	0.0042
School × Gender × Robot	1	0.8703	0.8703	7.67	0.0028

Table 10 Summary of 2 × 2 × 2 ANOVA on %C.

A summary of the 2 × 2 × 2 ANOVA on %C is shown in Table 10. The overall test (Model) is significant, implying that the 8 groups of samples have some different means of %C. The difference of the effect of each explanatory variable is also given. The test results of Gender × Robot and School × Gender × Robot suggest rejection of null hypothesis, so these two groups should have a different %C mean to the others. The means and standard deviations of %C with respect to robot assistance and gender is shown in Figure 15. A cross interaction is observed, so whether or not the pupils gained better equation solving ability depends on gender: for females it is better to have robot instruction, but for males without the robot. The contributions from sole robot and sole gender, as well as School × Gender and School × Robot are not statistically significant, it may be deduced that neither gender nor robot-instruction is regarded as the dominant main effect. This is supported by the least significant difference (LSD) *t*-test summarized in Table 11. The mean difference between gender only and between robot-instruction only are both tested to be statistically insignificant under a significance level of 0.1 (or 90% confidence interval). Therefore, further analysis about the interaction between Robot and Gender is carried out in the next section.

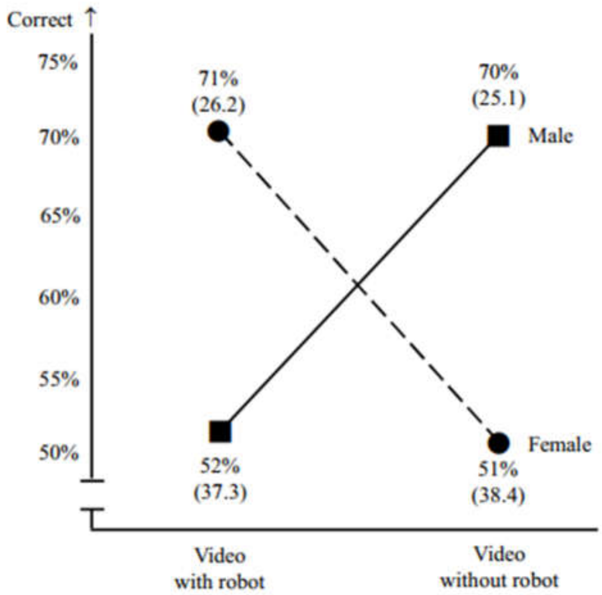


Figure 15 Mean percentage and standard deviations (in brackets) of correctly solved equations %C.

Comparisons significant at the 0.1 level are indicated by ***.				
Gender Comparison	Difference Between Means	90% Confidence Limits		
Male - Female	0.004625	-0.113052	0.122303	
Female - Male	-0.004625	-0.122303	0.113052	

Comparisons significant at the 0.1 level are indicated by ***.				
RoboYN Comparison	Difference Between Means	90% Confidence Limits		
VnoR - VyesR	0.01045	-0.10705	0.12795	
VyesR - VnoR	-0.01045	-0.12795	0.10705	

Table 11 Summary of least significant difference (LSD) *t*-test for %C.

2.3 Independent sample *t*-test

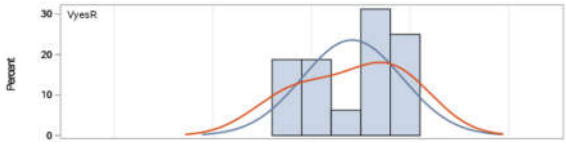
To understand in more detail how %C depends on the combination of explanatory variables Gender and Robot, it is introduced the objective here to determine if the difference in means of equation solving ability %C between 2 targeted population groups G_1 and G_2 (G_i = female performance with and without robot (F_{VyesR} & F_{VnoR}), male performance with and without robot (M_{VyesR} & M_{VnoR}), performance of female and male with robot ($VyesR_F$ & $VyesR_M$), performance of female and male without robot ($VnoR_F$ & $VnoR_M$)) is statistically significant. If there is a statistically significant difference between the means of %C between certain two groups, then the two groups have different performance (two-tail test), or one group may be considered better performance than the other (one-tail test). With the same assumptions and sampling conditions of the *t*-test, the null hypothesis is defined as equal means of %C in G_1

and G_2, and the corresponding alternative hypothesis is the mean of %C in the G_1 and G_2 are different (two-tail test), or that in G_1 is greater than that in G_2 (one-tail test). As there are multiple hypotheses to be tested with the same observations, Bonferroni correction is used, making the the significance level α divided by the number of hypothesis to be tested, i.e. $0.10 \div 4 = 0.025$. The SAS code for the independent sample *t*-test on %C with each set of G_i's is shown as follows:

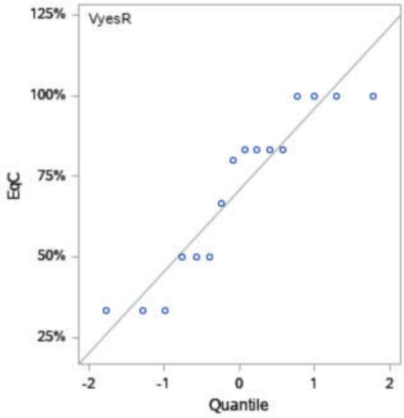
```
proc sql;
create table WORK.Rawdata_exNullEqC_Female as
    select * from WORK.Rawdata where EqC is not null && Gender = "Female"&& CountYN="Y";
run;

proc TTEST data = WORK.Rawdata_exNullEqC_Female;
    class RoboYN;
    var EqC;
    title "(3)  Independent samples t-test on %C - a.Female: Ry vs Rn"
Run;

proc sql;
create table WORK.Rawdata_exNullEqC_Male as
    select * from WORK.Rawdata where EqC is not null && Gender = "Male"  && CountYN="Y";
run;

proc TTEST data = WORK.Rawdata_exNullEqC_Male;
    class RoboYN;
    var EqC;
    title "(3)  Independent samples t-test on %C - b.Male: Ry vs Rn";
Run;

proc sql;
create table WORK.Rawdata_exNullEqC_RoboY as
    select * from WORK.Rawdata where EqC is not null && RoboYN = "VyesR" && CountYN="Y";
run;

proc TTEST data = WORK.Rawdata_exNullEqC_RoboY;
    class Gender;
    var EqC;
    title "(3)  Independent samples t-test on %C - c.RoboY: Female vs Male";
Run;

proc sql;
create table WORK.Rawdata_exNullEqC_RoboN as
    select * from WORK.Rawdata where EqC is not null && RoboYN = "VnoR" && CountYN="Y";
run;

proc TTEST data = WORK.Rawdata_exNullEqC_RoboN;
    class Gender;
    var EqC;
    title "(3)  Independent samples t-test on %C - d.RoboN: Female vs Male";
Run;
```

G_1	G_2	Test Type	DF	*t* value	Pr > |*t*|
F_{VyesR}	F_{VnoR}	one-tail	29.68	-1.75	0.0906
M_{VyesR}	M_{VnoR}	one-tail	30.93	1.81	0.0808
$VyesR_F$	$VyesR_M$	two-tail	34.731	1.80	0.0823
$VnoR_F$	$VnoR_M$	two-tail	28.802	-1.76	0.0898

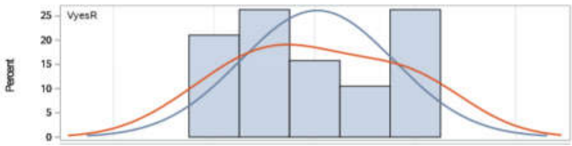
Table 12 Summary of independent sample *t*-test on %C.


VyesR
Percent
30
20
10
0

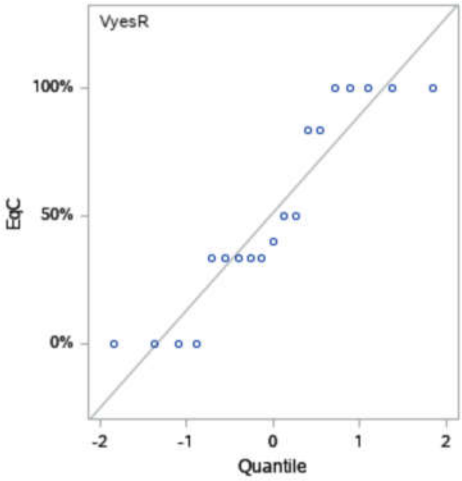
(a)


VyesR
125%
100%
75%
50%
25%
EqC
-2
-1
0
1
2
Quantile

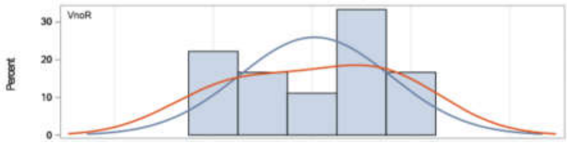
(b)


VyesR
Percent
25
20
15
10
5
0

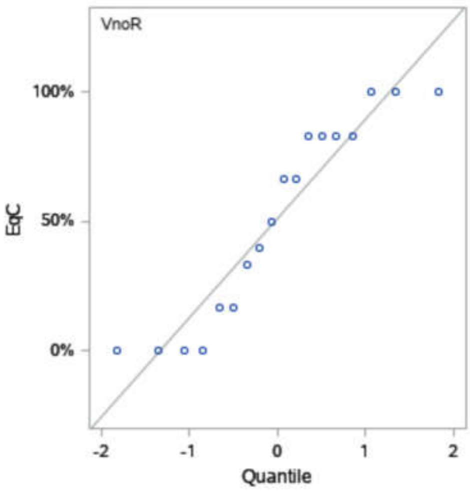
(c)


VyesR
100%
50%
0%
EqC
-2
-1
0
1
2
Quantile

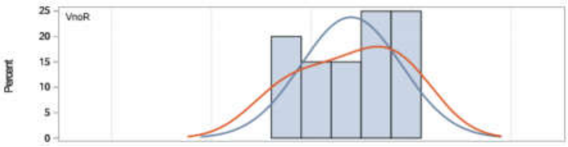
(d)


VnoR
Percent
30
20
10
0

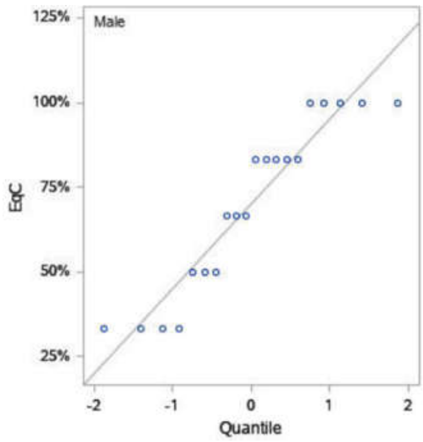
(e)


VnoR
100%
50%
0%
EqC
-2
-1
0
1
2
Quantile

(f)



(g)



(h)

Figure 16 Distributions of %C of female with robot (F_{VyesR}/$VyesR_F$) (a)-(b), male with robot (M_{VyesR}/$VyesR_M$) (c)-(d), female without robot (F_{VnoR}/$VnoR_F$) (e)-(f) and male without robot (M_{VnoR}/$VnoR_M$) (g)-(h). The left column shows the histograms of the distribution, with the fitted curves of normal distribution (blue) and kernel density estimation (KDE) (orange). The right column shows the Q-Q plot using standard score as the quartile axis.

A summary of the statistics and the hypothesis test results is given in Table 12. Without prior assumption of the equivalent variance of the two groups of samples, Satterthwaite approximation is used for the *t*-test. The mean values of %C reveal the following: Female pupils had a better equation solving ability with robot than without robot, but the opposite for male pupils. Under robot teaching, female pupils performed better than male pupils, but the opposite phenomenon was observed under teaching without robot. All of them, however, conform in this test with acceptance of the null hypothesis, meaning that the ability of solving equations is the same for female with and without robot, male with and without robot, female and male with robot, and female and male without robot.

The distributions of the groups are brought into a closer look to evaluate the quality of the test. Distribution of female pupils taught without robot shows double-peak behavior (0% and 75%), and a less significant yet considerable similar behavior for the male (36% and 86%). For distributions of robot teaching, a slightly negative skewed distribution is seen for the female, also interpreted in the Q-Q plot, and double peaks are found in the male (25% and 100%). The double-peak behavior can be read from the Q-Q plots (-2 to 0.25 and 0.25 to 2 for VnoR of females, -2 to 0 and 0 to 2 for VnoR of males, and -2 to -0.5 and 0.25 to 2 for VyesR males). Therefore, the validity of the hypothesis testing results obtained here may be questionable.

When comparing Figures 14(b), 15(a) and 15(c), it can be deduced that the lower %C peak (at 40%) in the VyesR distribution is contributed by the male pupils and the greater %C peak (at 100%) is contributed by both female and male pupils. For the VnoR distribution (Figure 14(c)), when compared to Figures 15(e) and (g), the broad peak in VnoR may come from the lower %C peak of the male whose %C value (36%) is greater than that in VyesR (25%) and the higher %C peaks of the male (86%) and the female (75%). It is noted that the distribution of VnoR male is particularly narrower, implying the lowest %C peak of VnoR (0%) belongs to the female pupils. The difference in distributions of female VyesR and VnoR may suggest that some female pupils may specifically require the robot in the teaching for enhanced learning, and the variance of learning outcome is reduced in male pupils under teaching without robot. Teaching without robot could reduce performance difference between the male and part of the female, but a portion of robot-relying females may get suffered.

3. Appreciation for robot (App-r) and Appreciation for video (App-v)

To examine whether the pupils learning equation solving would be more motivated than those with video alone, two response variables, appreciation for robot (App-r) and Appreciation for video (App-v), both ranged from 0 to 6, are introduced. They are compared against the explanatory variables, trying to recognize if any of them has a dominant effect. The means and standard deviations of App-r and App-v of male and female pupils in different schools are shown in Table 13. Tests for App-r is conducted, followed by those for App-v, and finally a multiple linear regression is done to correlate App-r and App-v to %C.

School	Gender	App-r			App-v			MeanApp*	
		N	Mean	S.D.	N	Mean	S.D.	Mean	S.D.
BoMei	M	12	5.333	0.9847	23	4.870	1.180	5.022	1.027
	F	8	5.875	0.3536	18	5.278	0.9583	5.361	0.8368
NanXing	M	8	5.125	1.126	17	5.059	1.088	5.118	1.068
	F	9	5.556	0.7265	17	4.412	1.121	4.676	0.9510
Total	M								
	F								

Table 13 Means and standard deviations (S.Ds.) of Appreciation for robot (App-r) and Appreciation for video (App-v) in different groupings of gender and school. (* The sample size of MeanApp is the same as that of App-r)

3.1 2 × 2 ANOVA of School (B–N) × Gender (M–F) on App-r

The cross interaction of %C between gender and robot instruction naturally gives rise to the question of whether boys and girls are motivated differently by the robot. The objective in this section is to check the statistical equivalence of pupils' appreciation for robot with the explanatory variables, School – BoMei (B) vs NanXing (N), and Gender – male (M) and female (F) using a 2-way (between subjects) ANOVA (GLM Univariable) on App-r. There are 4 groups in total. Assumptions for ANOVA have been discussed in Section 2.2. The null hypothesis is that all the 4 groups have the same population mean of App-r, i.e sampling mean difference comes by chance. The corresponding alternative hypothesis is that some specific groups of observations have a different mean of App-r. The significance level α is set to 0.10 (one-tail test). Observations without completing the survey question was excluded, and all observations are taken for the analysis. resulting in a total of 37 observations. Type III sum of squares (SS) is used for the nature of unbalanced data (see Table 13) and neglection of factor order in the hypothesis testing. The SAS code for 2 × 2 ANOVA on App-r is shown as follows:

```
proc Anova data=WORK.RawData_exNullEqC_RoboY;
    class School Gender;
    Model AppR = School|Gender;
    Means School|Gender/LSD CLDIFF;
    title "(5)  2 × 2 Anova of School (B - N) × Gender (b - g) on AppR";
Run;
```

Source	**DF**	**Sum of squares**	**Mean square**	***F* value**	**Pr > *F***
Model	3	2.550	0.8501	1.14	0.3478
School	1	0.3568	0.3568	0.48	0.4942
Gender	1	1.910	1.910	2.56	0.1193
School × Gender	1	0.2837	0.2837	0.38	0.5419

Table 14 Summary of 2 × 2 ANOVA on App-r.

A summary of the 2 × 2 ANOVA on App-r is shown in Table 14. The overall (Model) as well as all groups test results are insignificant, implying same means of App-r in all the 4 groups. The LSD *t*-test shows the same results, summarized in Table 11. The mean difference between gender only and between schools only are both tested to be statistically insignificant under a significance level of 0.05 (or 95% confidence interval), so none of Gender and School results in difference in appreciation for robot. While there is no significant difference between the schools and between the genders, the standard deviations of boys are greater than those of girls, as shown in Table 13.

Comparisons significant at the 0.1 level are indicated by ***.				
School Comparison	Difference Between Means	90% Confidence Limits		
BoMei - NanXing	0.1971	-0.2853	0.6795	
NanXing - BoMei	-0.1971	-0.6795	0.2853	

Comparisons significant at the 0.1 level are indicated by ***.				
Gender Comparison	Difference Between Means	90% Confidence Limits		
Female - Male	0.4559	-0.0265	0.9383	
Male - Female	-0.4559	-0.9383	0.0265	

Table 15 Summary of least significant difference (LSD) *t*-test for App-r.

3.2 2 × 2 × 2 ANOVA of School (B–N) × Gender (M–F) × Robot (VyesR–VnoR) on App-v

While the cross interaction of %C between gender and robot instruction may or may not suggest different learning motivations by the robot of boys and girls, the other appreciation factor, namely, appreciation for video, should also be investigated, as it is not yet known whether the use of robot would affect the effect on learning brought by video. The objective in this section is to check the statistical equivalence of pupils' appreciation for video with the explanatory variables, School – BoMei (B) vs NanXing (N), Gender – male (M) and female (F) and Robot – VyesR and VnoR using a 3-way (between subjects) ANOVA (GLM Univariable) on App-v. There are 8 groups in total. Assumptions for ANOVA have been discussed in Section 2.2. The null hypothesis is that all the 8 groups have the same population mean of App-v, i.e sampling mean difference comes by chance. The corresponding alternative hypothesis is that some specific groups of observations have a different mean of App-v. The significance level α is set to 0.10 (one-tail test). Observations without completing the survey question was excluded, and so one female observation from BoMei are taken for the analysis. resulting in a total of 75 observations. Type III sum of squares (SS) is used for the nature of unbalanced data (see Table 13) and neglection of factor order in the hypothesis testing. The SAS code for 2 × 2 × 2 ANOVA on App-v is shown as follows:

```
proc sql;
create table WORK.Rawdata_exNullAppV as
    select * from WORK.Rawdata where AppV is not null;
run;

proc glm data = WORK.Rawdata_exNullAppV;
    class School Gender RoboYN;
    model AppV = School|Gender|RoboYN / SS3;
    Means School|Gender|RoboYN/LSD CLDIFF;
    title "(6) 2 × 2 × 2 Anova of School (B - N) × Robot (y - n) × Gender (b - g) on AppV";
run;
```

Source	DF	Sum of squares	Mean square	*F* value	Pr > *F*
Model	7	9.044	9.044	1.04	0.4126
School	1	2.403	2.403	1.93	0.1790
Gender	1	0.2238	0.2238	0.18	0.6727
School × Gender	1	5.263	5.263	4.23	0.0435
Robot	1	0.01347	0.01347	0.01	0.9274
School × Robot	1	0.04215	0.04215	0.03	0.8545
Gender × Robot	1	1.9476	1.9476	1.57	0.2251
School × Gender × Robot	1	0.00000885	0.00000885	0.00	0.9979

Table 16 Summary of 2 × 2 × 2 ANOVA on App-v.

A summary of the 2 × 2 × 2 ANOVA on App-v is shown in Table 16. The overall (Model) as well as all groups except School × Gender test results are insignificant, implying same means of App-v in all these groups. App-v of female pupils depends on school, with a higher value of NanXing than of BoMei, as shown in Table 13. The appreciation for video is unlikely to be affected by robot instruction, given the statistical insignificance of the test in the Robot grouping together with other interaction grouping involving it. (School × Gender does not have grouping effect due to Robot).

Table 17 shows the LSD *t*-test results for recognition of dominant factor of contribution to App-v.. The mean difference between gender only and between schools only are both tested to be statistically insignificant under a significance level of 0.05 (or 95% confidence interval), so none of Gender and School results in difference in appreciation for video.

Comparisons significant at the 0.1 level are indicated by ***.

School Comparison	Difference Between Means	90% Confidence Limits		
BoMei - NanXing	0.3135	-0.1179	0.7449	
NanXing - BoMei	-0.3135	-0.7449	0.1179	

Comparisons significant at the 0.1 level are indicated by ***.

Gender Comparison	Difference Between Means	90% Confidence Limits		
Male - Female	0.09286	-0.33760	0.52332	
Female - Male	-0.09286	-0.52332	0.33760	

Table 17 Summary of least significant difference (LSD) *t*-test for App-v.

3.3 Independent sample *t*-test on App-v

Given the significant interaction between School and Gender described in Section 3.2, the next investigation is to understand in more detail how App-v depends on the combination of explanatory variables School and Gender. It is introduced the objective here to determine if the difference in means of appreciation for video App-v between 2 targeted population groups G_1 and G_2 (G_i = female and male performances in NanXing (N_F & N_M), female and male performances in BoMei (B_F & B_M), female performance in NanXing and BoMei (F_N & F_B) male performance in NanXing and BoMei (M_N & M_B)) is statistically significant. If there is a statistically significant difference between the means of App-v between certain two groups, then the two groups have different performance. With the same assumptions and sampling conditions of the *t*-test, the null hypothesis is defined as equal means of App-v in G_1 and G_2, and the corresponding alternative hypothesis is the mean of App-v in the G_1 and G_2 are different. As there are multiple hypotheses to be tested with the same observations, Bonferroni correction is used, making the the significance level α divided by the number of hypothesis to be tested, i.e. $0.10 \div 4 = 0.025$ (two-tail test). The SAS code for the one sample *t*-test on %C with each set of G_i's is shown as follows:

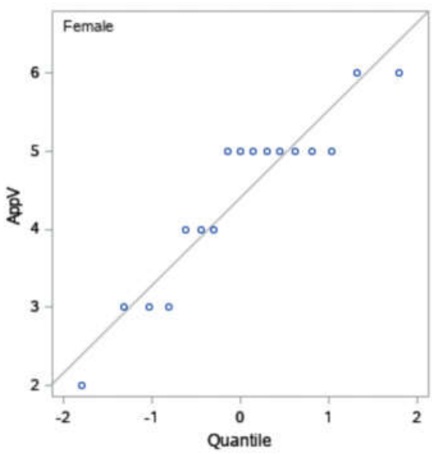
```
proc sql;
create table WORK.Rawdata_exNullAppV_NanXing as
    select * from WORK.Rawdata where AppV is not null && School = "NanXing";
run;

proc TTEST data = WORK.Rawdata_exNullAppV_NanXing;
    class Gender;
    var AppV;
    title "(7) Independent samples t-test on App-v - a. School N: Female vs Male";
Run;

proc sql;
create table WORK.Rawdata_exNullAppV_BoMei as
    select * from WORK.Rawdata where AppV is not null && School = "BoMei";
run;

proc TTEST data = WORK.Rawdata_exNullAppV_BoMei ;
    class Gender;
    var AppV;
    title "(7) Independent samples t-test on App-v - b. School B: Female vs Male";
Run;

proc sql;
create table WORK.Rawdata_exNullAppV_Female as
    select * from WORK.Rawdata where EqC is not null && Gender = "Female";
run;

proc TTEST data = WORK.Rawdata_exNullAppV_Female;
    class School;
    var AppV;
    title "(7) Independent samples t-test on App-v - c. Female: School N vs School B";
Run;
```

```
proc sql;
create table WORK.Rawdata_exNullAppV_Male as
    select * from WORK.Rawdata where EqC is not null && Gender = "Male";
run;

proc TTEST data = WORK.Rawdata_exNullAppV_Male;
    class School;
    var AppV;
    title "(7) Independent samples t-test on App-v - d. Male: School N vs School B";
Run;
```

G_1	G_2	Test Type	DF	*t* value	Pr > |*t*|
N_F	N_M	one-tail	31.97	-1.71	0.0974
B_F	B_M	one-tail	38.93	1.22	0.2289
F_N	F_B	one-tail	30.07	2.19	0.0363
M_N	M_B	two-tail	36.07	-0.52	0.6030

Table 18 Summary of independent sample *t*-test on App-v.

(a)



(b)

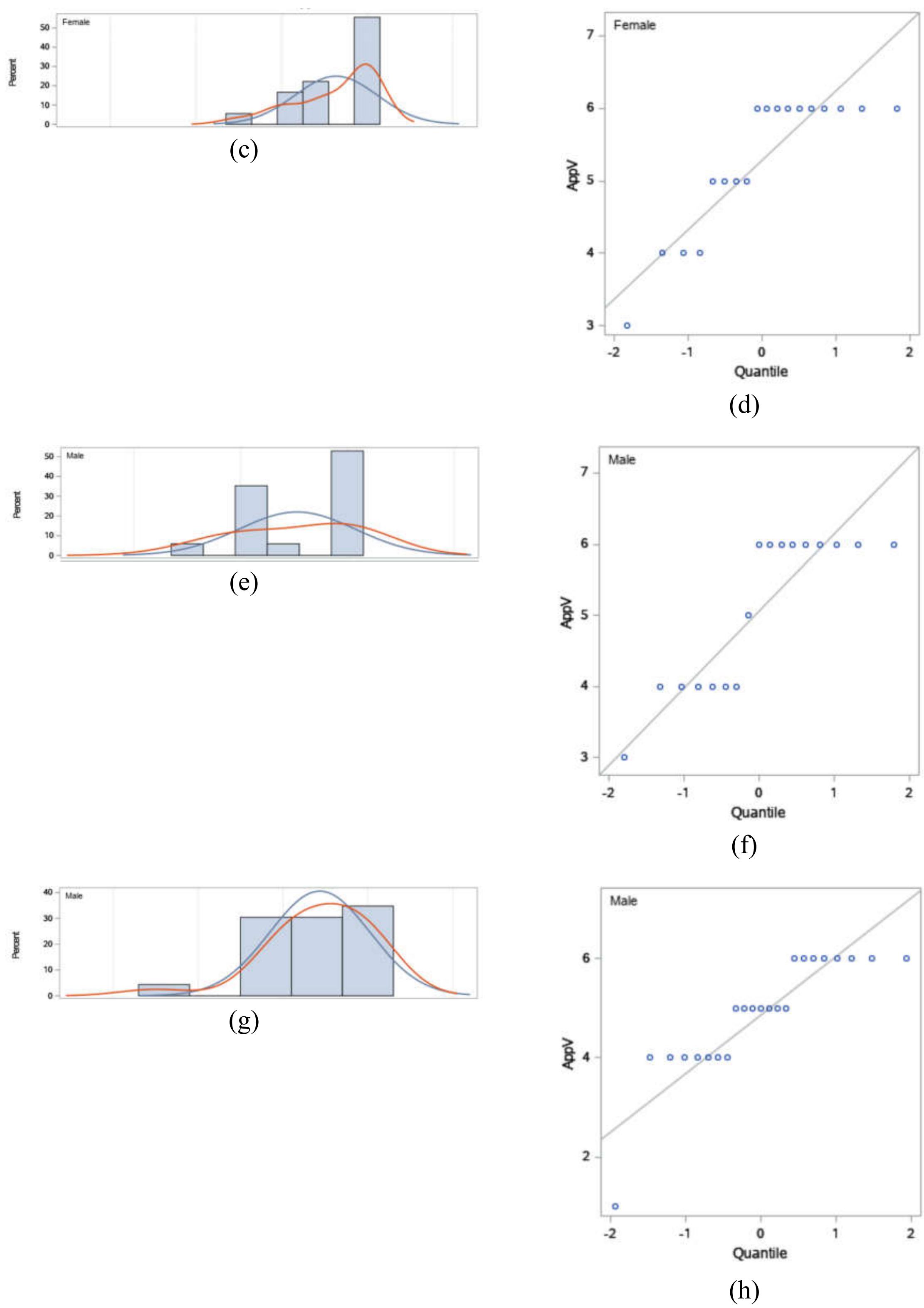


Figure 17 Distributions of App-v of female pupils in NanXing (a)-(b) and in BoMei (c)-(d), and male pupils in NanXing (e)-(f) and in BoMei (g)-(h). The left column shows the histograms of the distribution, with the fitted curves of normal distribution (blue) and kernel density estimation (KDE) (orange). The right column shows the Q-Q plot using standard score as the quartile axis.

A summary of the statistics and the hypothesis test results is given in Table 18. Without prior assumption of the equivalent variance of the two groups of samples, Satterthwaite approximation is used for the *t*-test. The mean values of %C reveal the following: Female pupils had a better equation solving ability with robot than without robot, but the opposite for male pupils. Under robot teaching, female pupils performed better than male pupils, but the opposite phenomenon was observed under teaching without robot. All of them, however, conform in this test with acceptance of the null hypothesis, meaning that the ability of solving equations is the same for female with and without robot, male with and without robot, female and male with robot, and female and male without robot.

The distributions of the groups are brought into a closer look to evaluate the quality of the test. Distribution of female pupils taught without robot shows double-peak behavior (0% and 75%), and a less significant yet considerable similar behavior for the male (36% and 86%). For distributions of robot teaching, a slightly negative skewed distribution is seen for the female, also interpreted in the Q-Q plot, and double peaks are found in the male (25% and 100%). The double-peak behavior can be read from the Q-Q plots (-2 to 0.25 and 0.25 to 2 for VnoR of females, -2 to 0 and 0 to 2 for VnoR of males, and -2 to -0.5 and 0.25 to 2 for VyesR males). Therefore, the validity of the hypothesis testing results obtained here may be questionable.

3.4 2 × 2 × 2 ANOVA of School (B–N) × Gender (M–F) × Robot (VyesR–VnoR) on MeanApp

The overall level of appreciation (MeanApp) during learning can also be tested. It is defined as the average of equal weight of App-r and App-v, MeanApp = (App-r + App-v) ÷ 2, for robot-instructed teaching, and MeanApp = App-v for video-only teaching. Independence of quantities involved in averaging is implicitly assumed. Without any prior knowledge, a crude assumption of equal weight is taken. The objective in this section is to check the statistical equivalence of pupils' appreciation for the combined effect of robot (if any) and video with the explanatory variables, School – BoMei (B) vs NanXing (N), Gender – male (M) and female (F) and Robot – VyesR and VnoR using a 3-way (between subjects) ANOVA (GLM Univariable) on MeanApp. There are 8 groups in total. Assumptions for ANOVA have been discussed in Section 2.2. The null hypothesis is that all the 8 groups have the same population mean of MeanApp, i.e sampling mean difference comes by chance. The corresponding alternative hypothesis is that some specific groups of observations have a different mean of MeanApp. The significance level α is set to 0.10 (one-tail test). Observations without completing the survey question was excluded, and so one female observation from BoMei are taken for the analysis. resulting in a total of 75 observations. Type III sum of squares (SS) is used for the nature of unbalanced data (see Table 13) and neglection of factor order in the hypothesis testing. The SAS code for 2 × 2 × 2 ANOVA on MeanApp is shown as follows:

```
proc sql;
create table WORK.Rawdata_MeanApp as
    select School, Gender, RoboYN, case
        when AppR > -1 then (AppR + AppV)/2
        else AppV
        end as AppRV
        from WORK.Rawdata where AppV is not null && Gender is not null;
run;

proc glm data = WORK.Rawdata_MeanApp;
    class School Gender RoboYN;
    model AppRV = School|Gender|RoboYN / SS3 ;
    Means School|Gender|RoboYN/LSD CLDIFF;
    title "(8)  2 × 2 × 2 Anova of School (B - N) × Robot (y - n) on MeanApp";
run;
```

Source	DF	Sum of squares	Mean square	*F* value	Pr > *F*
Model	7	8.711	1.244	1.32	0.2567
School	1	1.943	1.943	2.05	0.1564
Gender	1	0.03349	0.0349	0.04	0.8513
School × Gender	1	3.172	3.172	3.35	0.0715
Robot	1	1.687	1.687	1.78	0.1862
School × Robot	1	0.002410	0.002410	0.00	0.9599
Gender × Robot	1	2.841	2.841	3.00	0.0877
School × Gender × Robot	1	0.2665	0.2665	0.28	0.5973

Table 19 Summary of 2 × 2 × 2 ANOVA on MeanApp.

A summary of the 2 × 2 × 2 ANOVA on MeanApp is shown in Table 19. School × Gender and Gender × Robot are tested to be statistical significant, implying MeanApp interactions between school and gender as well a gender and robot. Other groups and the overall (Model) test results are insignificant, implying same means of MeanApp in all these groups.

The LSD *t*-test shows the same results, summarized in Table 20. The mean difference between gender only and between schools only are both tested to be statistically insignificant under a significance level of 0.05 (or 95% confidence interval), so none of Gender and School results in difference in the overall appreciation.

Comparisons significant at the 0.1 level are indicated by ***.				
Gender Comparison	Difference Between Means	90% Confidence Limits		
Male - Female	0.03393	-0.34153	0.40939	
Female - Male	-0.03393	-0.40939	0.34153	

Comparisons significant at the 0.1 level are indicated by ***.				
RoboYN Comparison	Difference Between Means	90% Confidence Limits		
VyesR - VnoR	0.2546	-0.1200	0.6293	
VnoR - VyesR	-0.2546	-0.6293	0.1200	

Table 20 Summary of least significant difference (LSD) *t*-test for MeanApp.

4. Correlation between appreciation (App-r and App-v) and equation solving ability (%C)

This section explores the correlation between appreciation for robot and to video and the equation solving ability. The cross interaction between gender and %C as suggested in Section 2.2. Therefore, we explore the possibility of between appreciation and equation solving ability in male, female, and mixed genders with and without robot instructions. The correlation between App-r and App-v for pupils under robot instruction is also explored.

4.1 Linear regression of App-r and App-v on %C under robot instruction

A multiple linear regression is carried out, looking for a possible relationship of

$$\%\mathrm{C} = k_{\mathrm{v}} \times \text{App-v} + k_{\mathrm{r}} \times \text{App-r} + \mathrm{I/C} \tag{1}$$

where k_v and k_v are constants and I/C is the *y*-intercept, with App-v and App-r as predictors. Normally distributed residuals, no correlation between the independent variables App-r and App-v, and homoscedasticity are assumed in the regression. Observations with robot instruction without missing all questions in the test paper are included. The SAS code for multiple linear regression of App-r and App-v on %C is shown as follows:

```
proc sql;
create table WORK.rawdata_RobotY_CountY_Female as
    select * from WORK.Rawdata where RoboYN = "VyesR" && Gender="Female" && EqC is not null&& CountYN="Y" && AppV is Not null;
run;

proc reg data = WORK.rawdata_RobotY_CountY_Female;
    model EqC = AppV AppR;
    title "(9)  a1.Multiple linear regression of AppR and AppV on %C_Female_robotY ";
run;

proc reg data = WORK.rawdata_RobotY_CountY_Female;
    model EqC = AppR;
    title "(9)  a2.Multiple linear regression of AppR on %C_Female_robotY ";
run;


proc sql;
create table WORK.rawdata_RobotY_CountY_Male as
    select * from WORK.Rawdata where RoboYN = "VyesR" && Gender="Male" && EqC is not null&& CountYN="Y" && AppV is Not null;
run;


proc reg data = WORK.rawdata_RobotY_CountY_Male;
    model EqC = AppV AppR;
    title "(9)  b1.Multiple linear regression of AppR and AppV on %C_Male_robotY ";
run;

proc reg data = WORK.rawdata_RobotY_CountY_Male;
    model EqC = AppR;
    title "(9)  b2.Multiple linear regression of AppR on %C_Male_robotY ";
run;
```

```
proc sql;
create table WORK.rawdata_CountY_RobotY as
    select * from WORK.Rawdata where RoboYN = "VyesR" && EqC is not null && CountYN="Y" && AppV is Not null;
run;

proc reg data = WORK.rawdata_CountY_RobotY;
    model EqC = AppV AppR;
    title "(9)  c1.Multiple linear regression of AppR and AppV on %C_robotY ";
run;

proc reg data = WORK.rawdata_CountY_RobotY;
    model EqC = AppV;
    title "(9)  c2.Multiple linear regression of AppV on %C_robotN ";
run;
```

Gender	*N*	Variable	Parameter estimate	Standard error	*t* value	Pr > |*t*|	R^2	Adj. R^2	*F* value	Pr > *F*
M	19	I/C	1.44 (1.49)	0.41 (0.41)	3.54 (3.66)	0.0028 (0.002)	0.31 (0.26)	0.22 (0.22)	3.61 (5.98)	0.051 (0.026)
		App-v	0.097	0.090	1.09	0.293				
		App-r	-0.27 (-0.186)	0.11 (0.076)	-2.52 (-2.44)	0.0226 (0.057)				
F	16	I/C	-0.81 (-0.82)	1.00 (0.98)	0.82 (0.84)	0.429 (0.416)	0.038 (0.001)	-0.11 (-0.070)	0.26 (0.01)	0.779 (0.909)
		App-v	-0.041	0.058	-0.71	0.4931				
		App-r	0.017 (0.019)	0.18 (0.17)	0.10 (-0.12)	0.925 (0.909)				
Mixed	35	I/C	1.22 (0.80)	0.38 (0.24)	3.23 (3.31)	0.0028 (0.0023)	0.080 (0.022)	-0.02 (-0.0082)	1.40 (0.72)	0.2623 (0.4010)
		App-v	0.0038	0.056	0.07	0.464				
		App-r	-0.12 (-0.041)	0.081 (0.048)	-1.43 (-0.85)	0.1624 (0.4010)				

Table 21 Summary of multiple linear regression of App-r and App-v on %C under robot instruction. Numbers in brackets are regression results with App-r as the sole predictor, i.e. App-v is taken out from Eq. (1).

The results of the analysis is shown in Table 21. For mixed gender, the coefficients of both App-v and App-r as predictors are close to zero, with insignificant *p* values of the *t*-tests on the predictors with respect to the 0.10 significance level (two-tail test). The R^2 and adjusted R^2 values are also close to zero, indicating uncorrelated %C relation to App-r and App-v. *F* test of the model suggests a rejection of the proposed Eq. (1) as a description of the correlation (with significance level 0.10). For male only and female only regressions, significant *t*-test results are obtained for App-r, with a positive and a negative coefficients for female and male pupils respectively. This suggests that the equation solving ability of female pupils increases with their appreciation for robot, but that of males pupils decreases with it. This supports the cross interaction between gender and robot on %C. An almost zero but statistically insignificant coefficient of App-v is obtained for female pupils, and a slightly positive yet statistically insignificant coefficient of App-v for the male. This suggests a possibility of App-v-uncorrelated %C for female pupils and a little positive correlation of %C with App-v for the male. The F test results suggest rejection of Eq. (1) for female %C modeling, possibly due to the near-zero R^2 and adjusted R^2 value, but acceptance for the male. If predictor App-v is removed from Eq. (1), keeping App-r as the sole predictor, then all the 3 groups gives significant t-test and F-test results. This may allow a preliminary proposal of the universality of the behavior of both genders towards the correlation between equation solving ability and appreciation for robot, positive for the female and negative for the male, uncorrelated to appreciation for video.

4.2 Linear regression of App-v on %C without robot instruction

This section explores the correlation between appreciation for robot and to video and the equation solving ability. A linear regression is carried out, looking for a possible relationship of

$$\%C = k_v \times \text{App-v} + \text{I/C} \tag{2}$$

with App-v as the predictor. Assumptions of the multiple regression have been discussed in Section 4.1. Observations with robot instruction without missing all questions in the test paper are included. The SAS code for multiple linear regression of App-v on %C is shown as follows:

```
proc sql;
create table WORK.rawdata_RobotN_CountY_Female as
    select * from WORK.Rawdata where RoboYN = "VnoR" && Gender="Female" && EqC is not null && CountYN="Y" && AppV is Not null;
run;

proc reg data = WORK.rawdata_RobotN_CountY_Female;
    model EqC = AppV;
    title "(9)  d.Multiple linear regression of AppV on %C_Female_robotN ";
run;



proc sql;
create table WORK.rawdata_RobotN_CountY_Male as
    select * from WORK.Rawdata where RoboYN = "VnoR" && Gender="Male" && EqC is not null && CountYN="Y" && AppV is Not null;
run;

proc reg data = WORK.rawdata_RobotN_CountY_Male;
    model EqC = AppV;
    title "(9)  e.Multiple linear regression of AppV on %C_Male_robotN ";
run;

proc sql;
create table WORK.rawdata_CountY_RobotN as
    select * from WORK.Rawdata where RoboYN = "VnoR" && EqC is not null && CountYN="Y" && AppV is Not null;
run;

proc reg data = WORK.rawdata_CountY_RobotN;
    model EqC = AppV;
    title "(9)  f.Multiple linear regression of AppV on %C_robotN ";
run;
```

Gender	N	Variable	Parameter	Standard	t value	Pr > $|t|$	R^2	Adj. R^2	F value	Pr > F

			estimate	error						
M	17	I/C	0.9106	0.2834	3.21	0.0048	0.0310	-0.0229	0.58	0.4580
		App-v	-0.04128	0.05444	-0.76	0.4580				
F	20	I/C	0.5000	0.5142	0.97	0.3462	0.0003	-0.0664	0.00	0.9482
		App-v	0.006910	0.1047	0.07	0.9482				
Mixed	37	I/C	0.6741	0.2760	2.44	0.0198	0.0010	-0.0275	0.04	0.8526
		App-v	-0.0102	0.05441	-0.19	0.8526				

Table 22 Summary of multiple linear regression of App-v on %C without robot instruction.

The results of the analysis is shown in Table 22. For all gender groups, the coefficients of both App-v as predictors are all close to zero, with insignificant *p* values of the *t*-tests on the predictors with respect to the 0.10 significance level (two-tail test). The R^2 and adjusted R^2 values are also close to zero, indicating uncorrelated %C relation to App-v. *F* test of the model suggests a rejection of the proposed Eq. (2) as a description of the correlation (with significance level 0.10). In the case of robot-instructed and video-alone teaching, App-v contributes little to %C. One reason for this may be that appreciation is relevant to motivation, while motivation not mainly or directly lead to well instant knowledge acquirement, which means appreciation cannot guarantee good performance.

4.3 Linear regression of App-v on App-r under robot instruction

In the experiment, the robot was put next to the video, and so pupils were interacting with both media simultaneously. Therefore it is meaningful to explore the correlation between appreciation for robot and for video. A linear regression is carried out, looking for a possible relationship of

$$\text{App-v} = k_r \times \text{App-r} + \text{I/C} \qquad (3)$$

with App-v as the predictor. Assumptions of the multiple regression have been discussed in Section 4.1. Observations with robot instruction without missing all questions in the test paper are included. The SAS code for multiple linear regression of App-v on %C is shown as follows:

```
proc reg data=WORK.rawdata_RobotY_Female;
    model AppV = AppR;
    title "(9)  f2.Multiple linear regression of AppV (female) on AppR(female) ";
run;
proc reg data=WORK.rawdata_RobotY_Male;
    model AppV = AppR;
    title "(9)  g2.Multiple linear regression of AppV (male) on AppR(male)  ";
run;
proc reg data=WORK.rawdata_RobotY;
    model AppV = AppR;
    title "(9)  h2.Multiple linear regression of AppV (all) on AppR(all)  ";
run;
```

Gender	N	Variable	Parameter estimate	Standard error	t value	Pr > |t|	R^2	Adj. R^2	F value	Pr > F
M	20	I/C	0.5469	1.070	0.51	0.6153	0.4765	0.4475	16.39	0.0008
		App-r	0.8101	0.2001	4.05	0.0008				
F	17	I/C	0.8723	2.893	0.30	0.7671	0.1206	0.0619	2.06	0.1721
		App-r	0.7234	0.5045	1.43	0.1721				
Mixed	37	I/C	0.7078	1.069	0.66	0.5122	0.310	0.2900	15.70	0.0003
		App-r	0.7664	0.1934	3.96	0.0003				

Table 23 Summary of multiple linear regression of App-v on App-r under robot instruction.

The results of the analysis is shown in Table 22. While female pupils show positive yet insignificant correlation between App-v and App-r, male and mixed genders shows positive and significant correlation between App-v and App-r (two-tail test of 0.10 significance level), meaning that appreciation for robot increases with that for video. F test of the model suggests accepting Eq. (3) as a description of the correlation (with significance level 0.10) for male only and mixed gender observations, but a rejection for female pupils, possibly due to the near-zero R^2 and adjusted R^2 values for the fit.