PRIT - Personally Relevant Instruction and Teaching

Nanjie Rao
Zeyuan Jing
Huan Kuang
Yunjie Tang
Zhang Dong

ABSTRACT

Having teachers use information personally relevant to their students within a lesson would create a more personalized lesson that could resonate with the students. This project seeks to support the teacher to teach a lesson in a manner that resonates with those learning by making use of personal experiences of the learners and displaying them in an aggregated manner for the teacher during a lesson. In this study, we specifically focused on one independent variable, system help or not. We hypothesized that using PRIT - an automatic speech-based personal information retrieval system can help the teacher to include more personal information of students into a lesson, easier and better, compared to a normal notecard for retrieving personal information. Our results showed that the PRIT section did get better usability, the lower cognitive load on the teacher and more frequent personalized teaching behaviors compared to the notecard section. Our work paves the way for needed future research on the use of teaching assistant system that can integrate into teachers' school lives.

ACM Reference Format:
Nanjie Rao, Zeyuan Jing, Huan Kuang, Yunjie Tang, and Zhang Dong. 2020. PRIT - Personally Relevant Instruction and Teaching. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Content in lessons is not personalized enough causing students to develop misconnections between concepts. It has been stated in previous research that implementing relevant information into a lesson can cause positive effects with learning [2, 9]. Having teachers use information personally relevant to their students within a lesson would create a more personalized lesson that could resonate with the students. This project seeks to support the teacher to teach a lesson in a manner that resonates with those learning by making use of personal experiences of the learners and displaying them in an aggregated manner for the teacher during a lesson.

We hypothesize that using PRIT - an automatic speech-based personal information retrieval system can help teacher to include more personal information of students into a lesson, easier and better, compared to a normal notecard for retrieving personal information. Our independent variable is system help or not, where in both levels, personally relevant information is used to make the teaching personalized. Our dependent variables (DV) are: (1) teacher's perception (perceived usability, importance and competence), (2) cognitive overload of the teacher, and (3) frequency of personalized teaching behaviors.

We designed a within-subjects study where each participants went through both the PRIT section and the notecard section in a 60-minute study session. Results from the paired-samples t-tests showed that there were statistically significant differences between the PRIT section and the notecard section on the scores from Post Study System Usability Questionnaire (PSSUQ) [8], NASA Task Load Index (TLX) [6], and frequency of personalized teaching behaviors from recorded videos of teaching sessions. There was no significant difference in Intrinsic Motivation Inventory (IMI) [10].

2 BACKGROUND

It has been found in many research that including personal info will result in better study motivation as well as results for students, as seen by Lipka et al.'s study [9] focusing on Yupik culture incorporated into a math lesson which resulted in improved learning. Bouillion and Gomez [2] investigated whether connecting science to social context can support science teaching in urban settings. They used real-world problems and school-community partnerships as contextual scaffolds to bridge community and school-based knowledge and found that students learned more key concepts and were more apt to discuss what they learned with their family members. So in our study, we encouraged our participants to put the personal info they have learnt through the materials we collected and provided beforehand into their lesson content, whether it's assisted with our PRIT system or done with the old-style notecard.

Meanwhile, teaching is already a multi-task behavior by nature. Teachers' work is increasing in complexity and intensity because of societal changes, reformed and increased work tasks, the changed moral and normative character of teacher work, and teachers' experiences of doing more than one thing at the same time [3].

As we were trying to design a novel system to help teachers integrate students' personal experiences into their lessons without overwhelming the teachers on top of their intense and complex multi-task teaching work, Cognitive load theory (CLT) provides guidelines for designing instructional materials and systems. The basic premise of CLT is that learners have a working memory with very limited capacity when dealing with new information [11]. When handling new information, working memory is severely limited in both capacity and duration; that is, working memory can only hold about seven (plus or minus two) items, or chunks of information, at a time. Additionally, when processing information (i.e., organizing, contrasting, and comparing), rather than just storing it, humans are probably only able to manage two or three items of information simultaneously, depending on the type of processing

Conference'17, July 2017, Washington, DC, USA
2020. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

required [7]. Finally, new information held in working memory, if not rehearsed, is lost within about 15 to 30 seconds [1]. We don't want to overload our participants while they were teaching. Teachers have to practice new skills over and over until they become automatic or during points where cognitive load is too heavy they will fall back on unconscious automatic processes that could backfire (e.g. stereotypical comments, prejudices, disciplinary actions) [5]. And since they never used our PRIT system before our studies, using a new system for the first time would more likely to be frustrating and anxious. That's why we adopt a design similar to a teleprompter, to reduce their cognitive load while providing simple automatic display of info memo.

We used paired-samples t-tests for all our measures. Student's t-test is a test comparing means, while Wilcoxon's tests the ordering of the data. The assumptions of t-test are met even for our small sample, it has greater statistical power than Wilcoxon's test [4].

3 EXPERIMENT

Our research investigated the impact of using PRIT or not for retrieving students' personal information to support personalized teaching. In a personalized teaching scenario, we addressed the following research questions:

Are there significant differences between teacher's use of PRIT as opposed to notecard in terms of:

RQ1: teacher's perception, operationalized as perceived usability, importance and competence?;

RQ2: cognitive overload of the teacher?; and

RQ3: frequency of personalized teaching behaviors?

To answer these research questions, we designed a within-subjects study with *system help or not* as the single independent variable with two levels. There were two conditions – 'PRIT' condition and 'notecard' condition. For the PRIT condition, we used an automatic speech-based personal information retrieval system to help teacher retrieve personal information of students. For the notecard condition, we prepared a Word document that contains the same information as it would be in the PRIT system to serve as teachers' notecard while they were teaching. Examples of participant's screen when doing PRIT condition on topic of Christmas, and notecard condition on topic of sleep are shown in the Figure 1.

Our dependent variables are teacher's perception (perceived usability, importance and competence), cognitive overload of the teacher, and frequency of personalized teaching behaviors. They were measured accordingly, by Post Study System Usability Questionnaire (PSSUQ) [8], Intrinsic Motivation Inventory (IMI) [10], NASA Task Load Index (TLX) [6], and normalized count of personalized teaching behaviors from recorded videos of teaching sessions.

Each participants went through both PRIT and notecard in a 60-minute study session. The order of the two conditions and two topics was counterbalanced across participants under a design of Latin Squares.

In our study, during the first 10-15 minutes, each participant will be given a detailed explanation and instruction about this study. They will have 5 minutes to prepare their teaching of the first topic. 10 minutes to teach the first topic, with or without the assistance of PRIT system depending on their assigned order. And then 5-10

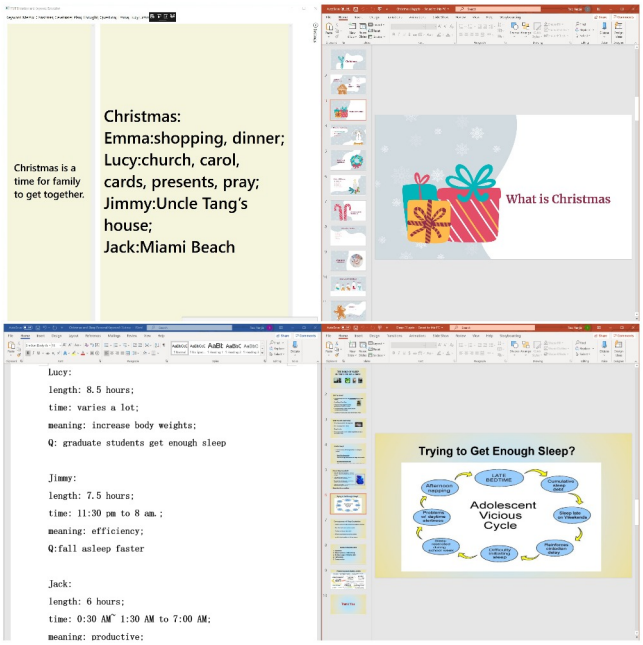

Figure 1: Examples of participant's screen for 2 conditions

minutes to answer the survey on Qualtrics. And then, 5 minutes to prepare another topic, 10 minutes to teach the second topic without or with the PRIT system. And then 5-10 minutes to answer the survey again but based on their experience of their 2nd condition.

The two topics that the participants will teach are the Christmas holiday, and sleep. The 2 topics are chosen in the manners that: (1) It's familiar for every participant, yet everyone would have their own experiences and opinions, also there is uncommon knowledge that can be taught through the topic. (2) Whichever the topic appears in what condition or what order, there will not be a learning effect in terms of carry-on knowledge from previous topic. In teaching each topic, the participants are expected to teach it to four students. For simulated classroom teaching set up, every participant will always be the teacher and 4 of our researchers will be the students. Simulated students are trained so that for each different teacher, they always have a similar behavior pattern (be a professional researcher/actor). At the end of each condition, questionnaires through Qualtrics online service will be given to the participants (the simulated teachers), not the students.

4 DATA COLLECTION

A screenshot of Qualtrics survey service we used to collect questionnaire responses is shown in the Figure 2.

Aside from the Qualtrics survey and the questionnaires that were contained inside it, we used quite a bit of other apparatuses as well. In fact, due to coronavirus pandemic outbreak, we changed the technology we used inside the PRIT system to support teaching due to converting the study to remote. Previously we utilized Microsoft Kinect and its built-in microphone array to detech where the voice is coming from, and based our teleprompter-alike display on the direction of the teacher's voice. How our setup was in a physical

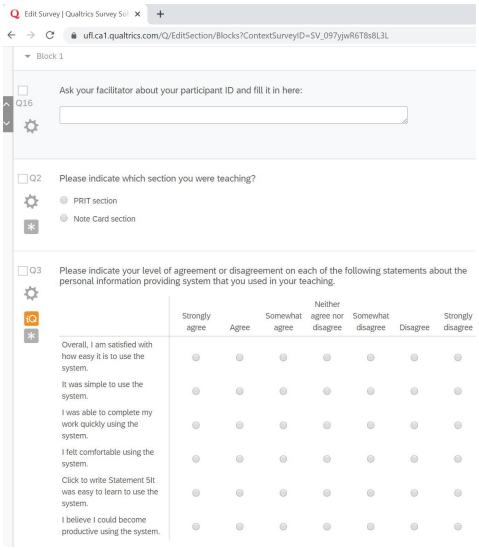

Figure 2: A screenshot of Qualtrics survey service we used

study room is shown in the Figure 3. That requires physical presence

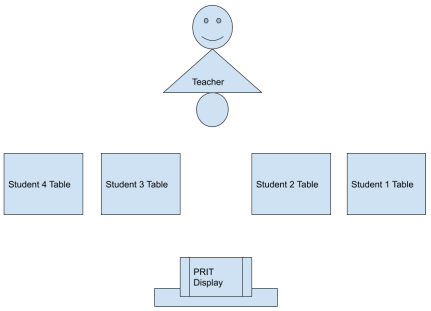

Figure 3: How our setup was in a physical study room

and can be understood as a proximity-based approach to decide which personal information should be displayed to the teacher. As per the guidelines of Conducting Approved Human Subject Research during the COVID-19 Pandemic, for studies that have low to no direct health benefit, only activities with no face to face interactions with study subjects may continue.

Migrating the technologies we integrated for it to be adapted to use in Zoom meeting service enabled us to carry on our human subject studies virtually. How our setup was in a Virtual Zoom Study is shown in the Figure 4. We used Microsoft Azure API and cabled internet to ensure the speech recognition (STT) is stable behind keyword extraction. We specifically tested our PRIT system to pick out the names and keywords that have the highest recognition success rates before all studies. Despite the participants' teaching style and skill, our PRIT system was working fine in every study.

We had a total of 8 study participants (5 male, 3 female). All participants were university students recruited through CISE SONA system, and compensated with course credit.

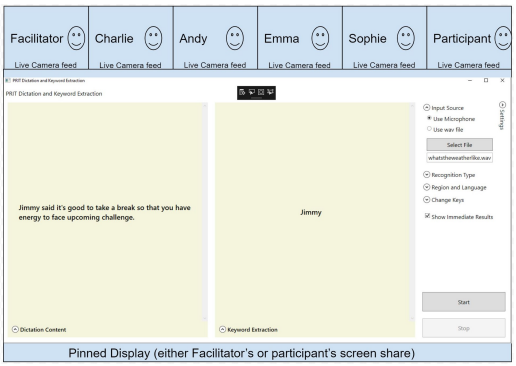

Figure 4: How our setup was in a Virtual Zoom Study

Part of the IMI questionnaire items (the 7 items about perceived choice) we had inside Qualtrics were excluded from subsequent analysis, participants had complained about these 7 items didn't make much sense to our specific study. And after serious consideration, we decided that the 7 items about perceived choice were indeed not really related to what we want to measure and also didn't fit our study. So we didn't do data analysis on the 7 items from all our 8 participants. However, all the other items in the Qualtrics survey were fully analysed.

5 DATA ANALYSIS

We have 3 major questionnaires contained inside the Qualtrics survey that the participants need to fill out after each condition: (1) Post Study System Usability Questionnaire (PSSUQ) [8], to measure teacher's perceived usability about the system, which is our DV1; (2) Intrinsic Motivation Inventory (IMI) [10], to measure teacher's perceived importance, choice and competence, which is also our DV1; and (3) NASA Task Load Index (TLX) [6], to measure the cognitive overload of the teacher, which is our DV2.

There were some pre-processing work needed before the statistical tests. All these questionnaire items were using 7-point Likert scales. We exported our collected survey data from Qualtrics into Excel, and substitute the choice texts with number 1 7, e.g., replace "Strongly agree" to 1, and "Strongly disagree" to 7. Some survey items are inverted in the way it stated the rating, for example, in the NASA TLX questionnaire, "I were successful in accomplishing what I were asked to do." All the other items in TLX were in the same direction as "It was mentally demanding to use the system." For these inverted questionnaire items, we reorganized the rating score by replacing it using "eight minus original score".

We used paired-samples t-tests for all our measures. Student's t-test is a test comparing means, while Wilcoxon's tests the ordering of the data. The assumptions of t-test are met even for our small sample, it has greater statistical power than Wilcoxon's test [4].

Results, as shown in the Table 5 and Figure 6, showed that there were statistically significant differences between the PRIT section and the notecard section on the scores from Post Study System Usability Questionnaire (PSSUQ), NASA Task Load Index (TLX), and frequency of personalized teaching behaviors from recorded videos of teaching sessions. There was no significant difference in Intrinsic Motivation Inventory (IMI).

	PRIT_Mean	NC_Mean	PRIT_SD	NC_SD	p-Value
PSSUQ	2.91	3.23	1.21	1.57	0.05
IMI-Effort	4.00	4.56	1.72	1.58	0.07
IMI-Comp	3.72	3.91	1.62	1.59	0.31
TLX	4.56	3.83	1.60	1.75	0.01
BehaviorNum	4.00	2.38	1.31	1.16	0.00

Figure 5: A summary table of our results

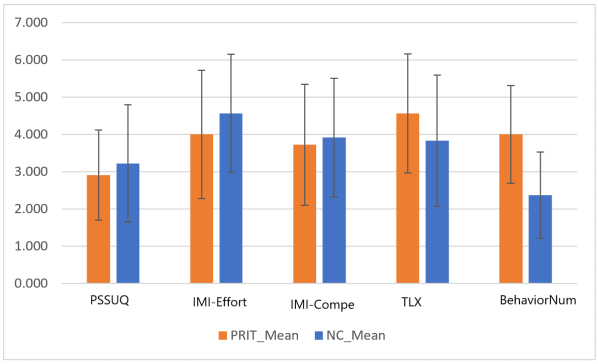

Figure 6: A bar chart of our results

Result from our subsequent test for TLX is shown in Figure 7.

t-Test: Paired Two Sample for Means

	PRIT	*NC*
Mean	4.5625	3.833333
Variance	2.549202	3.078014
Observatic	48	48
Pearson C	0.193689	
Hypothesi	0	
df	47	
t Stat	2.370383	
P(T<=t) on	0.010962	

Figure 7: A sample result from our subsequent test for TLX

6 CONCLUSION

As seen in the Table 5 and Figure 6, almost all aspects met our expectation except for IMI. For perceived usability (PSSUQ, which belongs to our DV1), higher the mean, harder to use the system, so PRIT was perceived as having better system usability than notecard. For cognitive overload of the teacher (TLX, which belongs to our DV2), higher the mean, lower the cognitive load, so PRIT has lower cognitive load than notecard, thus easier for the teacher. For frequency of personalized teaching behaviors (normalized count of the behaviors from recorded videos of teaching sessions, which belongs to our DV3), Higher the mean, more frequent personalized teaching behavior, so PRIT prompted the teachers to lead personalized teaching almost twice as often as notecard. These findings matched the basic premise of CLT, which is learners have a working memory with very limited capacity when dealing with new information [11], as PRIT system retrieved the relevant information when needed automatically, and this saved the effort of the teacher searching through the notecard. As it's more convenient and easier for the teacher to use PRIT than notecard, they tend to engage in more personalized teaching.

For perceived effort and competence (IMI-Effort and IMI-Compe, which belongs to our DV1), there was no significant difference between PRIT and notecard. This finding might seem surprising at first, but it's also understandable because of 2 reasons, the participants never used PRIT system before and most of them don't have teaching experience. Although we adopted a design similar to a teleprompter, but most likely they also never used a teleprompter before, because the participants were all from the CISE department. The source of the participants is also why most of them don't have teaching experience. Usually not many people can do challenging things on their first trial. Also, almost every participants felt that their preparations were in a rush, the 5-minute preparation time was not enough for the 10-minute teaching.

Issues mentioned above actually showed some threats to the internal and external validity of our study. For external validity, participants of our study differ substantially from the population of teachers. For internal validity, the 5-minute preparation time could have been longer for the participants to have better organization and planning on their teaching. We could resolve these 2 issues in the future by recruiting real teachers and increasing the study time to 1.5 hour. Another limitation is that our sample size is small. And you can see the impact of this from our results' standard deviations. Again, the coronavirus outbreak disrupted our plan in timeline, we actually developed 2 sets of system and study protocol to respond. For more convincing results, bigger sample size is needed.

REFERENCES

[1] R Anthony Jr et al. 2008. Cognitive load theory and the role of learner experience: An abbreviated review for educational practitioners. *Aace Journal* 16, 4 (2008), 425–439.

[2] Lisa M Bouillion and Louis M Gomez. 2001. Connecting school and community with science learning: Real world problems and school–community partnerships as contextual scaffolds. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 38, 8 (2001), 878–898.

[3] Göran Brante. 2009. Multitasking and synchronous work: Complexities in teacher work. *Teaching and Teacher Education* 25, 3 (2009), 430–436.

[4] YH Chan. 2003. Biostatistics 102: quantitative data–parametric & non-parametric tests. *Blood Press* 140, 24.08 (2003), 79.

[5] David F Feldon. 2007. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist* 42, 3 (2007), 123–137.

[6] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[7] Paul A Kirschner, John Sweller, and Richard E Clark. 2006. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist* 41, 2 (2006), 75–86.

[8] James R Lewis. 1992. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 36. SAGE Publications Sage CA: Los Angeles, CA, 1259–1260.

[9] Jerry Lipka and Esther Illutsik. 1999. Transforming the Culture of Schooling: Teacher Education in Southwest Alaska. (1999).

[10] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.

[11] Fred Paas, Alexander Renkl, and John Sweller. 2003. Cognitive load theory and instructional design: Recent developments. *Educational psychologist* 38, 1 (2003), 1–4.