

Renovating Assessment for the Future: Design-Based Implementation Research for a Learning-in-Class Monitoring System Based on the Learning Sciences

Hajime Shirouzu, CoREF, University of Tokyo, shirouzu@coref.u-tokyo.ac.jp
Moegi Saito, CoREF, University of Tokyo, saito moegi@coref.u-tokyo.ac.jp
Shinya Iikubo, CoREF, University of Tokyo, iikubo@coref.u-tokyo.ac.jp
Takahiro Nakayama, CoREF, University of Tokyo, nakayama@coref.u-tokyo.ac.jp
Kimihiro Hori, CoREF, University of Tokyo, k.hori@coref.u-tokyo.ac.jp

Abstract: Renovating assessment requires a shift to continuous formative assessment including lesson improvement which entails a collaborative Plan-Do-Check-Act cycle of teachers. However, it takes too much time to turn the cycle in ordinary life. Thus, we propose two ideas to develop an assessment system: first, a collaboration between humans and AI by utilizing teachers' assessment efforts to advance the knowledge of both, and second, an embedment of the collaboration in the DBIR community. Study 1 introduces the case of a veteran teacher who turned the PDCA cycle twice by finding a problem in his lesson with his colleagues using a dialogue analysis tool and by hitting upon a new plan, which engendered more productive students' dialogues, which referred to the expected keywords. Study 2 demonstrates the effects of the pre-registration of keywords into a dialogue transcription system on the recognition rate. Both results led us to propose a Learning-in-Class Monitoring System.

Keywords: learning assessment, knowledge constructive jigsaw, DBIR

Introduction

Assessment reform requires a radical shift from a “summative assessment which ranks individuals” to a “formative assessment of the learning environment that helps all the students reach the next level of learning” (Scardamalia *et al.*, 2012). Educational policy makers are also aiming for a similar shift in the revision of national curriculum guidelines (OECD, 2017). For example, the newly revised Japanese curriculum guidelines, the Courses of Study, introduced “active learning”, emphasizing its use of formative assessment. However, teachers lack appropriate tools to do this. Therefore, we developed a Learning-in-Class Monitoring System and tested it in educational settings in a series of research, in order to check whether it could serve as a foundation for assessment.

Research indicates that one-off seminars or short courses do not contribute much to teacher learning, and effective professional development involves opportunities for teacher to have actual classroom practice (Voogt, 2010), to reflect upon their practices (Ross & Bruce, 2007), and to participate in a community of practice with peers and experts over an extended period of time for continuous improvement (Penuel *et al.*, 2007). This means that a collaborative Plan-Do-Check-Act cycle holds the most important place for teachers' development. However, for most teachers, too much time and effort is required in completing this cycle. If teachers engage in designing student-centered lessons, it takes much time to collect, transcribe and analyze students' conversations. Thus, we put forward two ideas: first, a reciprocal collaboration between human beings and AI by utilizing teachers' usual efforts of assessment to advance the knowledge of both, and second, embedment of such a collaboration in the design-based implementation research (hereafter DBIR: Penuel *et al.*, 2007). If we can succeed in helping teachers engage in the time-consuming but central step of the PDCA cycle, namely, the “Check” step, teachers will be able to learn from classroom practices as well as raise the quality of their practices more effectively.

Then, how can we help teachers check students' learning and the quality of lessons? We hypothesize that the other steps of Plan, Do, and Act can also serve to raise the quality of the Check activity, since assessment benefits from teachers' explicit plans and re-plans of what kind of learning they want to take place in the classroom. Thus, the PDCA cycle should be turned iteratively in a way where the assessment itself improves its quality. We represent such a cycle in Figure 1a, which has a teachers' collaborative PDCA cycle of lesson improvement in its inner circle. Support systems on the outer circle (Figure 1a: pale green and blue shadow boxes) help teachers turn the cycle in a timely manner. In addition, as teachers use these systems to plan, discuss, conduct, reflect upon, and share the lessons iteratively, the system including AI evolves to become smarter. We call this whole set of systems a Learning-in-Class Monitoring System, specifically, “CoREFs: Co-Renovation of Evaluation for the Future system”.

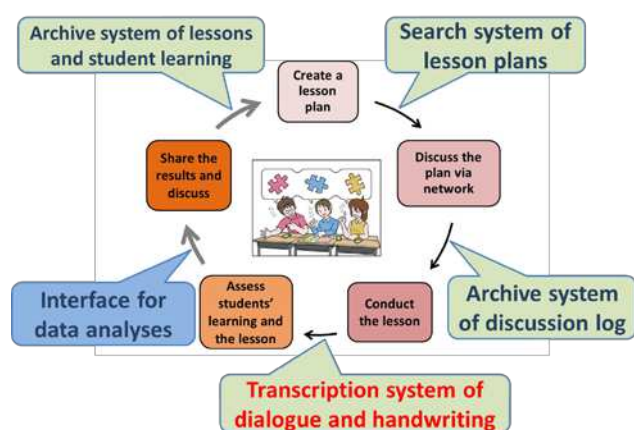


Figure 1a. PDCA cycle and Learning-in-Class Monitoring System, “CoREFs”.

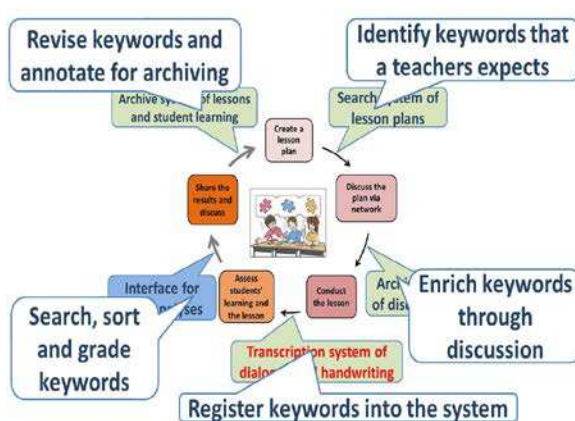


Figure 1b. How the cycle and system works in the case of keyword detection.

For example, Figure 1b shows how such collaboration between humans and AI contributes to their evolution. When a teacher creates a lesson plan, s/he anticipates keywords that the students will refer to or write in the lesson according to her or his own goal of the day. Several teachers can simulate this lesson plan from multiple perspectives and build on a variety of keywords. On conducting the lesson, however, the teachers often find that students’ expressions are more diverse than expected. In assessing student learning, the teachers put keywords into the analytic tool, learn how students use, share and improve their keywords, and sort and grade the keywords according to his or her newly-created criterion. The teachers feed the results back into the archive system with their reflections. When another teacher tries to conduct a similar lesson, s/he will be able to refer to an enriched database of students’ actual written and dialogue data. If s/he is able to register appropriate keywords into the transcription system, the speech recognition rate will become higher, which is the most difficult part of dialogue analysis (red letter in the pale blue box of Figure 1a). Such kind of speech-to-text translation can promote in-depth analyses of collaborative learning by researchers, high-quality lesson improvements by teachers, and context-driven approaches to semantic analyses of learning by engineers, contributing to renovating the foundation of assessment.

But how? The scheme described above might sound plausible, but how can we create a community wherein teachers “not only share ideas but generate and refine new ideas through the dynamics of networked social interaction” (Scardamalia *et al.*, 2017)? It is difficult for diverse teachers to collaborate on one lesson without a shared vision of learning that they expect to take place in children. The shared task holds the key for *constructive interaction* (Miyake, 2008), in which every participant deepens her or his understanding through role exchanges of task-doing and monitoring. Here the DBIR can provide a promising support for diverse teachers to gather, learn from each other and form a community which shares a vision and/or instructional framework.

So, this research has introduced constraints of an instructional and assessment framework of the lesson as a simple but strong core of constructive interactions among all (Miyake, 2013). Specifically, this study builds on a Japanese learning sciences project by the Consortium for Renovating Education of the Future (hereafter CoREF), which utilizes a concrete lesson framework, “Knowledge Constructive Jigsaw” method (hereafter KCJ: depicted in the center of Figure 1a). The project has seen the teaching of approximately 2,000 lessons every year by teachers of all subjects across all grades from 1st to 12th, supported by 30 regional boards of education. The participating teachers, experienced as well as novice teachers, work together to design, practice and reflect lessons across subjects, schools and districts, either face-to-face or via the internet. Using this set of a common method and learning community around it, teachers can obtain “streamlined” data, reflect upon it collaboratively, and become “annotators” by using the system that we propose. Yet, we do not want to propose that the KCJ is the only way to do this, but wish to provide a set of constraints spanning across multiple levels in which this kind of cycle can be turned in any way.

In sum, we wish to propose the DBIR (CoREF project) for the Learning-in-Class Monitoring System (CoREFs) as a promising candidate for renovating assessment of the future in a series of research, among which this paper reports its initial trial. Hereafter, we first illustrate the KCJ method. Then, we introduce the case of a veteran teacher who turned the PDCA cycle twice using the KCJ and a dialogue analysis tool. This study examines the effect of double loops of the PDCA cycle on raising the quality of the “C” step, as well as giving hints on what functionalities are needed in the Learning-in-Class Monitoring System. In Study 2, we examine if the pre-

registration of keywords into an automatic transcription system of student's dialogues would raise the speech recognition rate. This study examines the effect of "P" step on raising the quality of data for "C" step in the cycle.

Framework of the Knowledge Constructive Jigsaw (KCJ)

The KCJ consists of five learning activities: (1) writing an answer to the day's given problem based on a rule of thumb, (2) an expert-group activity which allows each individual student to accumulate some pieces of knowledge relevant in solving the problem, (3) a jigsaw-type activity where students from different expert groups get together to exchange and integrate the accumulated pieces of relevant knowledge and form an answer, (4) a cross talk activity where the students exchange their ideas for solutions, involving the entire class, and (5) writing down an individual answer again to the same problem and newer questions. Compared with the original Jigsaw method (Aronson, 1978), this method emphasizes the role of a shared "problem" for knowledge construction. Although strongly scripted, this method is dynamically adaptable to any learning contents and situations in the sense that each teacher can decide the "problem (jigsaw task)" and "learning materials (for the expert activity)." The essential flow of activity allows constructive interaction among task-doers and monitors to take place naturally and repeatedly: the design requires each student to become a task-doer in the jigsaw group, and provides each student with the chance to become a monitor who infers what the other students say and why they say that, in order to integrate the ideas of others with their own.

From perspectives of data analysis and engineering, this method has two features. First, since it lets students write down their answers to the same question twice, that is, at the beginning of the lesson and its end, the change or improvement between two answers give hints to infer what kind of interaction including dialogue takes place during two time points. Second, since students externalize their ideas or prior knowledge at step (1) above, intake information from reading and discussion on the material as well as verbalize their ideas at step (2), exchange and integrate ideas at step (3), present their group's ideas and listen to the other groups' ideas at step (4), and finally externalize their ideas and next question at step (5), we can follow the information source and flow that crystalizes into each student's final understanding of the day.

Study 1

Study 1 demonstrates the effect of a keyword-detecting tool of students' dialogues for a teacher's "reflection on action" (Schön, 1987) when properly embedded in a collaborative PDCA cycle of lesson improvement. In the knowledge society, teachers should improve their lessons continuously for students' better knowledge construction. As "reflective practitioners" (Schön, 1987), teachers should not only be able to implement researchers' advice for improvement but also be able to find problems in their lessons, propose plausible solutions and implement them by themselves. Students' dialogues are precious resources for such reflection as analyzing students' learning processes, evaluating the lesson and planning the next one. In the field of Learning Analytics, it becomes important to best make information on students' learning processes accessible and useful to teachers, or to design an analytic tool from a "teacher-centered" perspective (Erkens *et al.*, 2016; Matuk, Cocco & Linn, 2016). Although there have been many conversation analysis tools in the field of CSCL, such as KBDeX (Oshima, Oshima & Matsuzawa, 2012), PolyCAFe (Trausan-Mau, 2013), and Tatiana (Dyke, Lund & Girardot, 2009), most of them are best used by researchers who have enough time to dig through the process data of different situations to yield generalizable analytic methods and findings. Instead of applying such methods to the automatization of analysis and recommending analytic results to teachers, our focus is on 1) raising the quality of the teacher's own "process of analyses" and 2) providing teachers with a PDCA cycle to "use" their analytic results for lesson improvements, since we assume that such experiences will have the effect of raising the quality of teachers' reflection and selection of keywords.

Method

Dialogue analysis tool

We developed two assessment tools utilizing observation chances in the KCJ format that enable "visualization" of the students' learning processes. The first tool is "a comparison of pre- and post-class comprehension," which simply asks the same question twice in the steps (1) and (5) above. Thanks to this, children are able to compare their own answers, and confirm whether they have seen progress. Teachers can also compare the answers with their expectations and ascertain to what extent children have deepened their understanding and how diverse their expressions are. The second tool is "multilateral dialogue analysis," which aims to auto-transcribe the students' conversations in all of the groups during the class and provide transcripts electronically searchable by keywords. Study 1 focuses on the latter tool, even though we here used a wizard of oz method (humans performing the speech transcription).

This tool supports an analysis of transcription after the lesson. It has two simple functions: one function is highlighting utterances that include keywords of the lesson, and the other is changing the scope of analysis. The tool has one window for analyzing one group dialogue carefully (Figure 2a), which represents each student in each column and each utterance in each cell. If a user enters a keyword in a colored area, then the cell including the keyword turns into a colored cell. The other window is for comparing dialogues of all groups (Figure 2b), which shows each student in each column and each utterance in just one line. Thus the user cannot see the contents of the utterances, but is able to grasp an overall image of the distribution of keywords. The user can change these two views by clicking on any part. By using these functions, teachers can both analyze whether and how each student discusses the topic in the expert or jigsaw group using keywords related to the lesson contents or social processes, and examine to what extent a found pattern is universal among the groups. According to Soller *et al.* (2005), our tool is classified as a *mirroring system*, which just displays states and changes of students' interaction.

Delving into one group

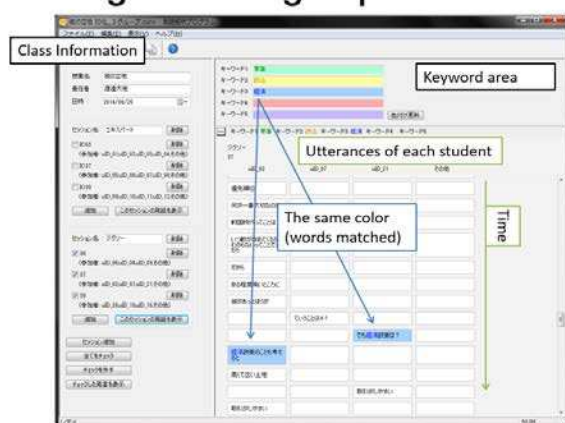


Figure 2a. The dialogue analysis tool: Close-up window.

Comparing all groups

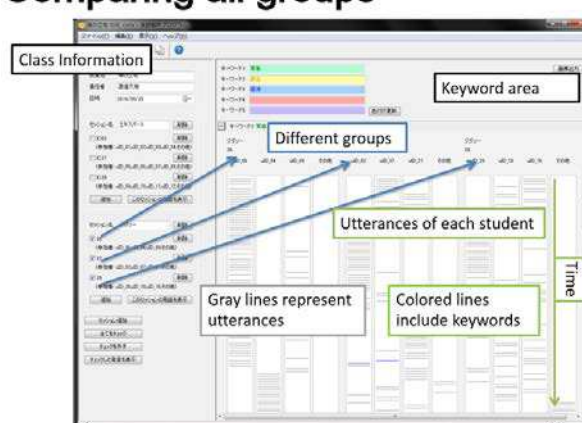


Figure 2b. Bird's-eye window.

Research setting and Teacher K

As shown in Table 1, this study focused on Teacher K's lesson improvement process spanning over the three-year period from the year 2014 to 2016, using the above dialogue analysis tool in a seminar described below in the year 2015. Since this research was conducted using a case study method, we will be giving its details together in the Results section. Underlined phrases and words in Table 1 are explained later.

The teacher (hereafter "Teacher K") who was the subject of this study is a junior-high school male teacher. He has been working as a science teacher for 28 years and has taught at seven junior-high schools in several cities and towns of a rural prefecture over his teaching career. He participated in the CoREF project from its initial stage, the year 2008, having developed more than thirty KCJ lesson materials. He has not only attended CoREF's teacher training workshops more than 30 times, but has also served on five occasions as a lecturer in these workshops and symposiums. He also participated in a seminar called "KCJ and Lesson Study Masters" in the year 2014, which saw 40 teachers taking part from all around Japan, providing them with more advanced contents of learning sciences, enabling the teachers to connect them with their own KCJ experiences, and also to discuss their learning sciences. As a master teacher, he not only opened his classes up to other teachers, but also traveled to other districts in order to conduct KCJ lessons in other teachers' classes.

Results

Teacher K taught a science lesson to 8th graders using the KCJ method on a unit of the mechanics of human body movements in the year 2014, and reflected on the lesson directly after teaching it, especially on the gap between what he had asked the students to do and what the students understood to be "today's task". By analyzing the students' dialogues in a seminar about a year later, however, he became aware of the difficulties students faced in extracting and integrating the information from the expert materials as well as the need to describe the task more clearly. After a further year, when given the chance to try teaching the unit again, he revised his lesson plan, which ultimately led to deeper understanding on the part of the students. Table 1 summarizes the whole process.

Table 1: Teacher K's lesson improvement

Phase	Event (Planning, implementing, reflecting and re-planning of the lesson)
1	<p>Planning of the first version of the lesson (30th June – 2nd July, 2014; on CoREF’s mailing list)</p> <p>Teacher K first intended to make students more <u>conscious of their body movements in their club activities or in their everyday lives</u>. Therefore, he planned to show a video of a person hitting a tennis ball, a baseball, or a volleyball, and to have the students explain the body movements of any player, handed out <u>three expert materials</u> of the “<u>nervous system</u>,” “<u>skeletal structure</u>” and “<u>mechanism of muscles</u>” and formed jigsaw groups by student club (tennis, baseball, or volleyball) to allow the students to apply what they had learned in their activities. To this plan, one researcher expressed <u>the concern</u> that the task seemed too difficult for students in spite of the attractiveness of “<u>solving personal problems</u>”. Teacher K, however, did not follow this advice and took the risk of implementing his own initial idea.</p>
2	<p>Implementing the lesson and reviewing it in a meeting (4th July, 2014)</p> <p>The lesson went over the allotted time since students had difficulty in integrating the information in the jigsaw groups and presenting it in the crosstalk. Directly after the lesson, Teacher K said in the meeting, “The children did not seem to be <u>conscious of their own movements</u>, so I should have used the video or whatever to enable them to think deeply about the inside workings of their bodies,” which implied he had found some problems with his lesson but had just resorted to another <i>know-how</i> of teaching.</p>
3	<p>Reviewing the lesson in a CoREF reflection sheet (July 2014)</p> <p>In the CoREF reflection sheet, the teacher is requested to select any three students and compare their pre and post class answers. Teacher K reflected upon these answers and the impression of the class, writing “Although even <u>in bending one’s arm</u>, “<i>the biceps brachii muscle contracts while the triceps brachii muscle relaxes</i>,” the students’ descriptions were too obscure to reach this level. This kind of goal demanded too much from the students because I myself did not have a clear image of what I had wanted from them. In redesigning the lesson, I would present the same movement to everybody regardless of the club the student belongs to,” suggesting he reached the idea of making the task clearer.</p>
4	<p>Reviewing students’ dialogues of the lesson in the CoREF seminar (1st August, 2015)</p> <p>About a year later from the lesson, a three-hour workshop in the “KCJ and Lesson Study Masters” seminar was held, utilizing <i>hand-transcribed</i> dialogues from Teacher K’s lesson as described above. Fifty-three teachers including Teacher K used the dialogue analysis tool collaboratively to dig through one to three jigsaw groups of students. Before using the tool, Teacher K had planned to model students’ learning from “understanding of the task” and “externalizing one’s own initial thoughts” through “reacting to others’ thoughts (keywords of “<u>I see</u>”, “<u>Why?</u>”, “<u>Uh?</u>”, “<u>Aha?</u>”)” to “restructuring one’s own thoughts”. During actual use, a group including Teacher K entered a keyword “<u>In a word</u>,” but found that 8th graders did not use such a formal expression, and came to focus more on content words like “<u>ball</u>.” After hearing other groups’ findings, Teacher K wrote down not only communicative words like “<u>For example</u>” or “<u>Why?</u>” but also content words integrating three expert materials like “<u>command</u>”, “<u>transmission</u>”, “<u>reaction</u>” as promising keywords. Teacher K’s change implies that even though he had assumed the content was not so difficult for his students, he found it was not the case. Actually, it was difficult for the students to discuss the materials by using and connecting the content keywords as he had expected.</p>
5	<p>Planning of the second version of the lesson and its implementation (18th October, 2016)</p> <p>When Teacher K was given the chance to teach the same unit and open up that class, he decided to try the KCJ lesson again. Although he changed <u>the expert materials</u> only slightly, he drastically changed the aim of lesson and the main task. He dropped the aim of “explaining everyday body movement” from the expected outcomes, and clarified the task by asking “<u>What are you doing to catch a falling ruler?</u>” and making it a common target for students to explain. He also clarified the expected outcomes such as “Students should explain “<i>I get stimulus that the ruler starts to move from the eyes. The stimulus is transmitted from the optic nerve through the sensory nerve and the spinal cord to the brain. Then the brain commands the muscles inside the thumb and forefinger to contract, the command of which is transmitted from the brain through the spinal cord and the motor nerve to the muscles inside the fingers, and the fingers successfully catch the ruler</i>””. By implementing this plan, almost all the jigsaw groups of students reached the expected outcomes and furthermore they tried to connect their learning outcomes to the following lessons.</p>
6	<p>Reviewing the lesson in a CoREF reflection sheet (October 2016)</p> <p>Teacher K wrote in the reflection sheet, “I felt that most of the students clearly understood <u>the mechanics of body movement</u>, due to a change in the task which dealt with a simple behavior. I realized that even though I changed the expert materials a little, the reaction of the students drastically changed with revision of the task and targeted material, at which I was surprised.”</p>

Discussion for the system requirement

Teacher K first stuck to his own hypothesis of the lesson. Thanks to the PDCA cycle supported by the KCJ method and its learning community embedded in the DBIR, Teacher K had been exposed to another person's hypothesis (researcher's concern) about the lesson and found "some" problems in his lesson. Yet, with the help of tool-assisted dialogue analysis in the community, he first realized the hidden problem of the lesson and came up with a newer plan. The dialogue analytics embedded in the collaborative PDCA cycle promoted teachers' own action research. It also implies that even a veteran teacher like K cannot imagine all of the dialogues of the students, and so can benefit from an activity of dialogue analysis.

Then, how are we able to reduce this whole process spanning more than two years in order to help teachers gain more timely feedback, especially when we aim to support a large number of teachers in our community simultaneously? Reviewing Table 1 from the perspective of keyword detection, enrichment and revision, we are able to consider the requirements of the Learning-in-Class Monitoring System. Single underlined phrases are the corpus from which keywords can be detected, and double underlined words or phrases are candidate keywords by themselves.

When a teacher tries to plan a KCJ lesson, she or he is supposed to make not only expert materials but also a "lesson design sheet" which includes a lesson objective, main task, and gist of expert materials. For example, Teacher K created the lesson plan and materials, which can provide the corpus for keywords (single underlined in phase 1 of Table 1). A "search system of lesson plans" (Figure 1a) should automatically conduct a morphological analysis to propose "hot words" in the texts when the plan and materials are uploaded. In addition, if the teacher sets the keywords by her- or himself (double underlined in phase 1 in Table 1 like "nervous system"), the system should be able to register those words. When the teacher consults with other teachers in the KCJ community, she or he posts the lesson design sheet and materials on the bulletin board or mailing list system. Others made comments on them such as suggestion, advice or mere concern (single underlined in phase 1 of Table 1). An "archive system of the discussion log" (Figure 1a) should automatically accumulate logs and learn to detect keywords from the discussions. A "transcription system of dialogue and handwriting" (Figure 1a) should have components of registering keywords, recognizing speech or handwritten letters, representing the recognized results and allowing correction of them. The system also should be seamlessly connected with an "interface for data analyses" (Figure 1b) like the dialogue analysis tool described above (Figure 2), that is, the transcribed dialogue is automatically imported from the transcription system into the analysis tool and represented there with its voice information. This integrated system should keep track of a teacher's search of keywords by registering words or phrases that the teacher looks for with their frequency (see keywords double underlined in phase 4 in Table 1). An "archive system of lessons and student learning" (Figure 1a) should include not only the lesson plan and materials but also all the transcribed data of each child of pre- and post-class answers to the same question and in-between dialogues (expert, jigsaw, cross talk activities), with teacher's reflection upon it (single underlined in phases 2, 3, and 6 in Table 1) and all the log data of search or dialogue analysis above. Then, when the teacher or another teacher tries a similar lesson, all those data can be used in the search system of lesson plans, upon which teachers can register newer keywords and corpora (single underlined in phase 5 in Table 1) and use them for the transcription system.

Study 2

We are now developing a whole set of systems and have reported here only a part of it: an automatic transcription system of student's dialogues.

Method

System

The system consisted of unidirectional microphones attached to each student, and automatic transcription tools with a cloud engine. This engine accepts pre-registration of keywords by the user up to 100 words. Using this system, we iterated the high-school Japanese lessons four times.

Participants and procedure

Fifteen students from high schools and/or prep school participated in the lesson, totaling up to sixty in the four iterations. The lesson let students deepen their understanding of a short story, a Japanese novel "*Kaban* (The Bag)" by Kōbō Abe, through the KCJ method. Students first read the story to ponder the main message of the story, and then were divided into three expert groups, one of which considered how the main character feels and changes his thinking as the story develops, another of which reflected how another character does so, and the third of which thought about the meanings of two expressions found in the novel. Three members from the expert groups formed

one jigsaw group to think again about the message of the story, reported their thoughts to the class and wrote down their answers individually at the end.

Through the first three iterations, we were able to get the accuracy rate of automatically transcribed data compared with the whole truth transcription, that is, 100% minus WER (the word error rate), of about 23% to 36% depending on the group at step (3) of the jigsaw activity. The reason why the rate differed greatly by group is that the acoustic and language models were not adjusted to the particular group, where group members simultaneously talk, take over each other's utterances, and sometimes suddenly burst into laughter.

Thus, we focused on transcribing the crosstalk at step (4), where a representative of each group presented their final answer in turn and the other groups quietly listened to that. Yet, they used many colloquial expressions instead of formal ones found in the material (the novel), and thus an extra "dictionary" was needed produced by their own conversations and writings. During the three iterations, we acquired the conversation data of a total of 100,000 Japanese characters (roughly corresponding to thirty to fifty thousand English words), which were classified into about 2,000 "different" (Japanese) words. We, the researchers, chose 49 hot words from them which also matched the objective of the lesson, registered the words into the system beforehand, and applied the system to the crosstalk at step (4) of the fourth (final) trial of the lesson.

Results

Table 2 represents the result. The "recognition rate" represents the accuracy of the words transcribed by the system, while the "recognition accuracy" means the accuracy of the transcription compared with the whole truth transcription. Pre-registration of keywords resulted in an accuracy gain of about 10%, since the recognition of the registered words (often hard-to-recognize words) contributed to that gain. Still, the rates are still too low to use them online by teachers. We need more words to register and need to know how to choose words to raise the rates.

Table 2: The accuracy rate of the transcription system with or without pre-registration of keywords

	Recognition rate	Recognition accuracy	Recognition rate of registered words
Third trial of the lesson (Without pre-registration)	66.7% (SD: 16.9%)	58.2% (SD: 20.2%)	---
Fourth trial of the lesson (With pre-registration)	74.0% (SD: 11.9%)	64.4% (SD: 5.8%)	63.5% (varied from 0 to 100%)

General discussion

We examined the effect of turning the cycle twice on raising the quality of the "C" step in the PDCA cycle in Study 1, and the effect of pre-registration of the keywords, that is, specification of "Plan" on raising the quality of data for "C" in Study 2. Combining the results of Study 1 and Study 2, we found promising results on the possibility of recruiting teachers as "annotators" of the PDCA cycle. First, although even a master teacher tends to reflect upon their lesson subjectively without records of learning processes, s/he can reflect upon it precisely and concretely with just a simple analytic tool. Second, the keywords that the teacher inputs can change as his or her understanding of the lesson topic and the students develops like the keywords Teacher K considered in Study 1 (italics in Table 1). Third, when the teacher deepens their understanding of the topic and how the students learn, s/he gets clearer in what kind of keywords s/he wants and gets better in re-designing the lesson, which results in gaining a higher rate of recognition of students' keywords, while the students feel freely to talk with their colloquial expressions in a more focused lesson. Such a reciprocal collaboration between systems and humans can contribute to enriching the iterative assessment cycle as a foundation for renovating assessment. In other words, an intertwined cycle of teaching and assessment can set an arena for mutual growth of humans and AI.

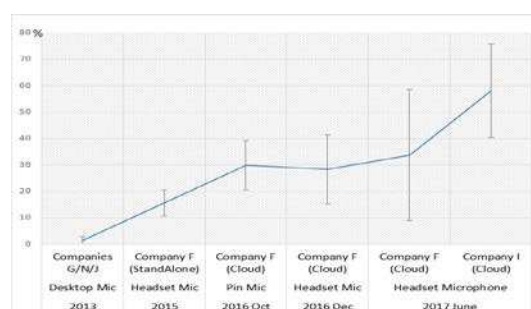


Figure 3. Recognition accuracy of transcription of jigsaw dialogues.

An immediate challenge for us is to raise the quality of the transcription system, as a core of the Learning-in-Class Monitoring System, or CoREFs, especially recognizing the jigsaw dialogues, where several groups simultaneously engage in discussions in the same room. Figure 3 shows an improvement of recognition accuracy of transcription of jigsaw dialogues over five years, with each year showing an average and SD of 3 to 19 groups of various lessons by various teachers. Although we cannot disclose the names of the recognition engines, in the year 2013 the recording environment was of low quality, which was improved in the year 2015 through the use of headset microphones. Cloud engines (used in Study 2) increased the accuracy of the year 2016 by 15% compared with the year 2015. In the year 2017 the engine increased accuracy by 25%, through accepting limitless pre-registration of corpora and dictionaries, which takes advantage of the features of the KCJ described in this paper and opens the arena for mutual growth of humans and AI.

References

- Aronson, E. (1978). *The Jigsaw Classroom*. Beverly Hills, SAGE Publications.
- Dyke, G., Lund, K. & Girardot, J.-J. (2009). Tatiana: An environment to support the CSCL analysis process. In *Proceedings of CSCL'09, Volume 1*, 58-67.
- Erkens, G., van Leeuwen, A., Janssen, J. & Brekelmans, M. (2016). Learning analytics to support teachers: Regulating teaching practices through analytics in CSCL. In C. K. Looi, J. L. Polman, U. Cress & P. Reimann (Eds.). *Transforming Learning, Empowering Learners: ICLS 2016, Volume 1*, 28-29.
- Matuk, C., Cocco, F. & Linn, M. (2016). A teacher-centered approach to designing a real-time display of classroom activity. In C. K. Looi, J. L. Polman, U. Cress & P. Reimann (Eds.). *Transforming Learning, Empowering Learners: ICLS 2016, Volume 2*, 1122-1123.
- Miyake, N. (2008). Conceptual change through collaboration. In S. Vosniadou, (Ed.). *International Handbook of Research on Conceptual Change*, 453-478. New York: Routledge.
- Miyake, N. (2013a). Case report 5: Knowledge construction with technology in Japanese classrooms (CoREF). In P. Kampylis, N. Law, Y. Punie (Eds.), *ICT-enabled innovation for learning in Europe and Asia*, 78-90, European Commission, Joint Research Centre.
- OECD (2017) *Educaiton 2030*. <http://www.oecd.org/edu/school/education-2030.htm> (retrieved 30th, Nov. 2017)
- Oshima, J., Oshima, R. & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60 (5), 903-921.
- Penuel, W. R., Fishman, B. J., Yamaguchi, R. & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921-958.
- Ross, J. & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of randomized field trial. *The Journal of Educational Research*, 101(1), 50-60.
- Scardamalia, M., et al. (2017). Toward a multi-Level knowledge building innovation network. In B. K. Smith, M. Borge, E. Mercier & K. Y. Lim (Eds.). In *Proceedings of CSCL2017, Volume 2*, 703-710.
- Scardamalia, M., Bransford, J., Kozma, R. & Quellmalz, E. (2012). New assessments and environments for knowledge building. In P. Griffin, B. McGaw & E. Care (Eds.). *Assessment and Teaching of 21st Century Skills*. Dordrecht: Springer, 231-300.
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.
- Soller, A., Martinez, A., Jermann, P. & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education (IJAIED)*, 15, 261-290.
- Trausan-Matu, S. (2013). Collaborative and differential utterances, pivotal moments, and polyphony. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplov & N. Law (Eds.). *Productive Multivocality in the Analysis of Group Interactions*, Springer, 123-139.
- Voogt, J. (2010). Teacher factors associated with innovative curriculum goals and pedagogical practices: Differences between extensive and non-extensive ICT-using science teachers. *Journal of Computer Assisted Learning*, 26(6), 453-464.

Acknowledgments

This work was supported by JSPS KAKENHI (No.17H06107), the CKF project, Saitama Prefectural Board of Education, Akiota Town Board of Education, and Iizuka City Board of Education.