Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Data-Driven Competency Assessment Supporting System for Teachers

Taito KANO[a*], Izumi HORIKOSHI[b], Kento KOIKE[b] & Hiroaki OGATA[b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*kano.taito.87a@st.kyoto-u.ac.jp

Abstract: As many countries seek to promote competency-based education, formative assessments are important to capture the learning processes of learners. However, as yet there are no assessments that can fully capture the learning process. Recently, the use of ICT tools for learning has become more general, and learning log data has been accumulated. Using these data, it has become possible to capture learning processes in detail; therefore, data-driven assessment has attracted increasing attention. However, as conventional data-driven competency assessments require experts to map data to competencies, they can only be applied in a defined context. In this study, we proposed an assessment framework that allows teachers to assess their students' competency by freely combining data collected as students used the Learning & Evidence Analytics Framework (LEAF) platform. We created an assessment in a scenario in an assumed educational setting using the proposed framework and examined what kind of assessment would be possible. Then, we created a system for the framework. Finally, interviews were conducted with three teachers regarding the system. The results suggest that the system can achieve context-independent and flexible data-driven assessment, contributing to the continuous improvement of learning and teaching from multiple perspectives in activities that use the system.

Keywords: Competency assessment, data-driven assessment, learning analytics

1. Introduction

Competency-based learning, which emphasizes the development of competencies in K12 education, has been widely implemented (Henri et al., 2017). To develop learners' competencies, it is important to assess their current competencies during the learning process (Guerrero-Roldán & Noguera, 2018). In doing so, it is possible to assess how they have demonstrated and developed their competencies. Competency assessments have also been implemented in K-12 education in Japan (MEXT, 2018). However, most assessment methods focus on products such as tests, reflection sheets, and portfolios; process-based assessments have not yet been developed (Guerrero-Roldán & Noguera, 2018).

Recently, ICT tools for learning have been introduced into the educational field, allowing learners' activities to be recorded in a database as log data that facilitate capturing the learning process in more detail than ever before (Sung et al., 2016; Van der Kleij et al., 2015). Therefore, data-driven competency assessment has attracted considerable attention.

However, in conventional data-driven competency assessments, experts must define the correspondence between data and competencies theoretically (Greene & Azevedo, 2010; Winne & Perry, 2000). Although this approach achieves a robust and highly valid data-driven competency assessment, it can only assess data as defined and cannot respond to diverse contexts.

Therefore, we created a framework that allows teachers themselves to assess competencies by combining data and implementing a data-driven competency assessment system. We believe that this system will also afford each teacher the flexibility to create assessments tailored to their context and improve the quality of observation. We addressed these RQs in this study.

RQ1: How can the data-driven assessment be realized from trace data?

RQ2: How can the data-driven assessment be implemented for teachers' use?

RQ3: How do the teachers expect, use and evaluate the data-driven assessment system?

2. Literature Review

2.1 Data-Driven Competency Assessment

To assess competencies such as critical thinking and self-regulated skills (SRSs), it is necessary to capture the learning process (Guerrero-Roldán & Noguera, 2018). Although many methods have been proposed to capture the learning process, they have certain limitations. For instance, assessments based on observations of daily learning activities are limited to what teachers can capture (Greene, 2015). Even if the teacher asks learners to write a reflection sheet, there is still much noise in the sheet, such as the learners' ability to express themselves and remember (Renninger & Bachrach, 2015). Among the many methods available, assessment based on learning log data has attracted attention as appropriate for assessing the learning process (Siadaty et al., 2016). To assess competencies from the data, the following procedure is necessary.

First, target competency must be defined. Second, the competencies must be subdivided into activity levels. For example, SRSs are divided by Greene and Azevedo (2009) into "macro-level" and "micro-level" activity phases. Third, we must determine the activities in the system that correspond to the activity phase definitions. For instance, if the target competency is SRSs, they are subdivided into "Planning," "Engagement," and "Evaluation & Reflection"; "Engagement" is subdivided into "Working on the Task" and "Applying Strategy Changes"; and "Working on the Task" is described as "to consistently engage with a learning task using tactics and strategies" (Siadaty et al., 2016, p. 192). Based on the description, it can be considered to correspond to an activity such as using a marker function to highlight an important word in "Working on the task". In this case, the system activity must be recorded as data for assessing the activity. By linking activities and data in this way, the assessment value of activities can be quantified from the data, from which the competency can be quantified.

However, such assessment processes are mostly conducted by experts and researchers, and their applicable context is limited. Essentially, the axis of competency assessment, interpretation of data, and context of assessment can be changed by the teacher. In this study, we propose a method that allows teachers to flexibly create assessments according to their own contexts.

2.2 Competency Assessment in Japan

Japan's K12 program uniquely fosters three pillars of competencies as shown in Table 1 with their descriptions which are considered based on MEXT (2018). Each pillar consists of multiple competencies that are mainly assessed based on learning products such as tests, presentations, and interview tests.

Table 1. Pillars of Competency and their Descriptions (MEXT, 2018)

Pillar of Competency	Description
Knowledge & Skills	How much knowledge and skills learners have acquired
Thinking, Judgment & Expression	How much acquired knowledge and skills learners can use
Meta-Cognitive & Self-Regulated Skills	How well learners can recognize their state and adjust their learning to acquire knowledge and skills

The recent global digitalization of education has also progressed in Japan. All learners have begun using devices in accordance with the digital education reform known as the GIGA School Concept (MEXT, 2020). There are high expectations regarding the use of learning-log data obtained from ICT tools for competency assessment. However, the use of learning-log data for assessment has not yet been fully realized.

3. Method

3.1 Learning & Evidence Analytics Framework (LEAF)

The LEAF system that our laboratory is developing is an integrated learning platform that integrates LMS (Moodle), e-books (BookRoll), a learning support module, a learning analytics module (LogPalette), and a database (LRS) (Ogata et al., 2018). LEAF can serve as a sensor that records learners' detailed learning behaviors and a database that stores learning logs. The following table presents example BookRoll log data. In this study, a data-driven assessment system was created using the learning-log data obtained from LEAF. However, this assessment method can be applied in other environments as well.

Table 2. Example BookRoll Learning Log Data

Operation_time	Student_id	Course_id	Operation_name
2022-04-05 13:32:01	S_1	C_1	"ADD_MARKER"
2022-04-06 07:14:43	S_2	C_1	"OPEN"
2022-04-06 08:01:02	S_2	C_1	"NEXT"

3.2 Research Questions

・**RQ1: How can the data-driven assessment be realized from trace data?**

We began by creating an assessment framework that allows teachers to combine data and create assessments. As an example application of this framework, we attempted to assess learners' SRSs in a first-year junior high school mathematics course using a set of created indicators. The results indicated the type of assessment possible using the framework.

・**RQ2: How can the data-driven assessment be implemented for teachers' use?**

This section describes the implementation of this assessment. In addition, we describe some of the support functions added to assist teachers in the assessment.

・**RQ3: How do the teachers expect, use and evaluate the data-driven assessment system?**

We conducted interviews with three Japanese teachers who used the implemented system. We compared it with the current competency assessment and asked participants about their expectations and concerns. We then asked them to use the system, and based on observations of their behavior, we discussed its positive aspects and possible improvements.

4. Results and Discussions

4.1 RQ1: Data-Driven Assessment Framework

Here, the proposed assessment framework is introduced and applied to a given scenario.

4.1.1 Indicators and Data Processing Flow

The following explains the process for calculating the assessment value using trace log data, as shown in Figure 1.

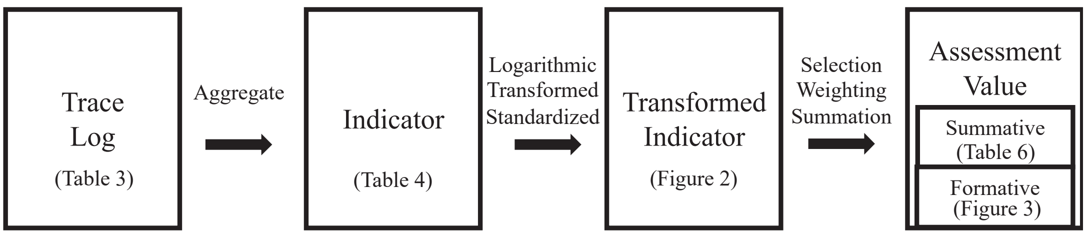

Figure 1. Data Processing Flow

First, the indicators were created by aggregating the trace log data obtained from the LEAF. Second, the indicators were narrowed down by the time range of the activity to be assessed, and summed. Next, the logarithmic transformation of each indicator across learners was perfomed to make the distribution closer to a normal distribution, and standardization was performed to align the distribution means of different indicators. In this way, the influence of outliers can be suppressed, and different indicators can be summed. Finally, by selecting the indicators to be used for assessment and setting weights for each, the values of the selected indicators were weighted and summed to calculate the assessment value. In conventional assessment activities, the scores on several tests or the results of a presentation are weighted and summed to create a rating. Based on these suggestions, ratings were created by adding weights to the indicators. In addition, summative assessments can be created by aggregating the indicators over the entire period, and formative assessments by aggregating the indicators for each week.

This procedure was expected to enable not only the generation of learner ratings from the data but also continuous monitoring of the changes and growth of learners. Nine indicators were created in this study, as listed in Table 3.

Table 3. List of Indicators

	Indicator Name	Explanation
TS	Time Spent	Time viewed (minutes)
NT	Number of Trans	Number of page transitions made
NJ	Number of Jumps	Number of page jumps made
NYM	Number of Yellow Markers	Number of times yellow marker was drawn
NRM	Number of Red Markers	Number of times red marker was drawn
NB	Number of Bookmarks	Number of times bookmarks were used
NM	Number of Memos	Number of notes
NHM	Number of Hand Writing Memos	Number of handwritten notes
NAQ	Number of Attempts of Quiz	Number of quiz responses

4.1.2 Scenario Analysis

To demonstrate the application of the framework, we conducted an assessment with these nine indicators in a scenario that assumed an actual educational setting as follows:

A teacher teaching a first-grade math course in 2022 will assess his or her students' first semester SRSs. This course will also have a test on June 3. Test scores will be used to assess Knowledge and Skills; therefore, learning activities in the LEAF from April 1 to June 3 will be assessed as a component of SRSs. Figure 2 shows the results of the assessments of the SRSs of the three students based on the indicators shown in Table 3 using the assessment process shown in Figure 1.

Figure 2(a) shows the value of the assessment indicators calculated for the three students during the entire period. Red and blue lines are drawn for indicators in the top and bottom 30% of the values, respectively. These show that student_1 scored high only in the number of red markers, student_2 in four indicators, and student_3 in five indicators. Each student varied in whether indicators were high or low, which may indicate differences in students' approaches. We believe that these figures are important for understanding the characteristics of learning in LEAF enhanced learning activities.

Second, we considered the actual assessment of SRSs based on these nine indicators. Among SRSs, we will focus only on Working on the Task activities for the assessment target because the indicators are considered to be related to Working on the Task, which is described as "to consistently engage with a learning task using tactics and strategies" (Siadaty et al., 2016, p.192).

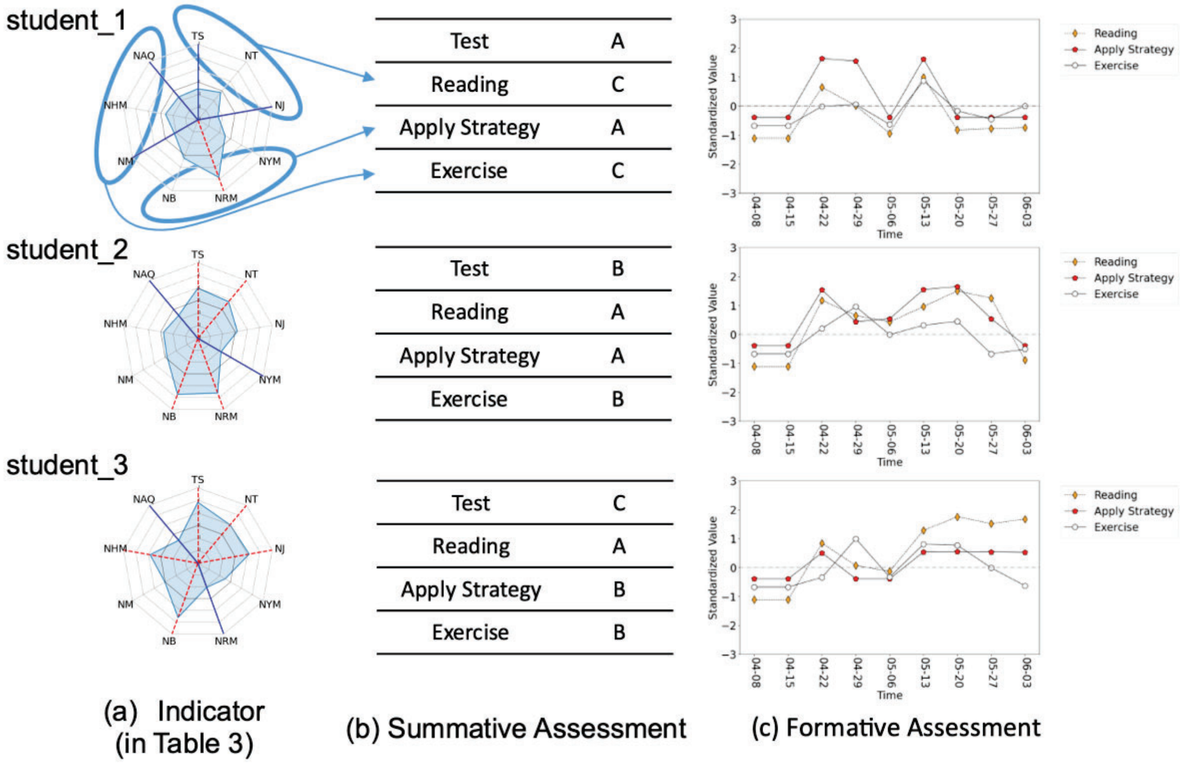

Figure 2. The Transformed Indicators and the Assessment Values

The activities performed in Working on the Task were further classified into three categories (Table 4), and each indicator was mapped to a category based on its characteristics.

Table 4. Correspondences between Categories and Indicators

Category	Indicator
Reading	Viewing time, number of page transitions, number of page jumps
Applying Strategy	Number of yellow markers, red markers, bookmarks
Exercise	Number of notes, handwritten notes, quiz responses

The calculated assessment value of each activity was assigned A for the top 30%, B for the top 30–70%, and C for the remaining activities to be graded. The grades for each activity from April 1 to June 3 are listed in Figure 2(b). Test grades are also provided for reference purposes.

The results of this summative assessment provide insights into students' learning activities that were not found in test scores alone. For example, *student_1*, who also scored high on the test, did not perform many Reading or Exercise activities but was highly involved in the Apply Strategy activity, suggesting that he or she was trying to learn creatively. *Student_3*, who has high scores in each activity but a low score on the test, is considered to have some problem with the way he or she performs the activity. These differences in the learning activities of each student afford a perspective for assessment that cannot be seen from grades alone.

Thus far, we have learned how to incorporate the data obtained from LEAF into a conventional summative assessment. Such summative assessments provide suggestions on how to conduct learning activities but do not capture the learning process. Therefore, we calculated the indicators for each week and performed weighted additions to calculate the assessment values for the time series. Figure 2(c) shows the transitions in the assessment values of the three activities for all students.

This formative assessment makes it possible to see how long the activities had continued and when they were decreasing or increasing. For example, student_2 improved on Reading and Apply Strategy activities from the first week of May to the third week of May, but then dropped from the fourth week of May. Student_3 improved on Reading and Apply Strategy activities as well from the first week of May, and continued to do so until the test. Using a time-series graph allows us to capture student activities' real-time changes and also allows us to observe multiple activities simultaneously. We believe that this will allow for a more multifaceted and continuous formative assessment in LEAF enhanced classes.

In other cases, when a teacher makes a specific intervention for students in a LEAF enhanced classroom activity, it is possible to evaluate their intervention based on how the student activities have changed. We believe that this will help activate the cycle of classroom practice and improvement.

To summarize the assessment process, summative and formative assessments were created by selecting indicators from a given set of indicators according to their purpose. Thus, the framework allowed us to flexibly tailor the assessment to each context. Furthermore, this could serve not only as a grading instrument but also to improve learning and teaching in the system.

However, such freedom in assessment may reduce its validity. Similar issues have been raised in previous studies (Azevedo, 2015; Guerrero-Roldán & Noguera, 2018). When assessing competencies based on data, it is necessary to establish a correspondence between the indicators and competencies. This mapping is mostly performed by experts and researchers who are familiar with both data and competencies, as we did here, and is considered difficult for teachers, whose comprehension of competencies and indicators varies widely. However, teachers should be able to perform the mapping to create assessments flexibly according to their own context. That is, it is necessary to consider how teachers can learn to define competencies and make appropriate correspondences. Accordingly, in this study we incorporated a function to store and share the correspondences made by teachers between indicators and competencies in the system.

4.2 RQ2: Implementation of the Proposed Assessment Method

Figure 3 shows the three main modules included in the system. The main components of the system are the Setting Module, which sets the context of the assessment and selects indicators; the Sharing Module, which shares information to support the selection of indicators in the Setting Module; and the Visualization Module, which visualizes the calculated assessment values.

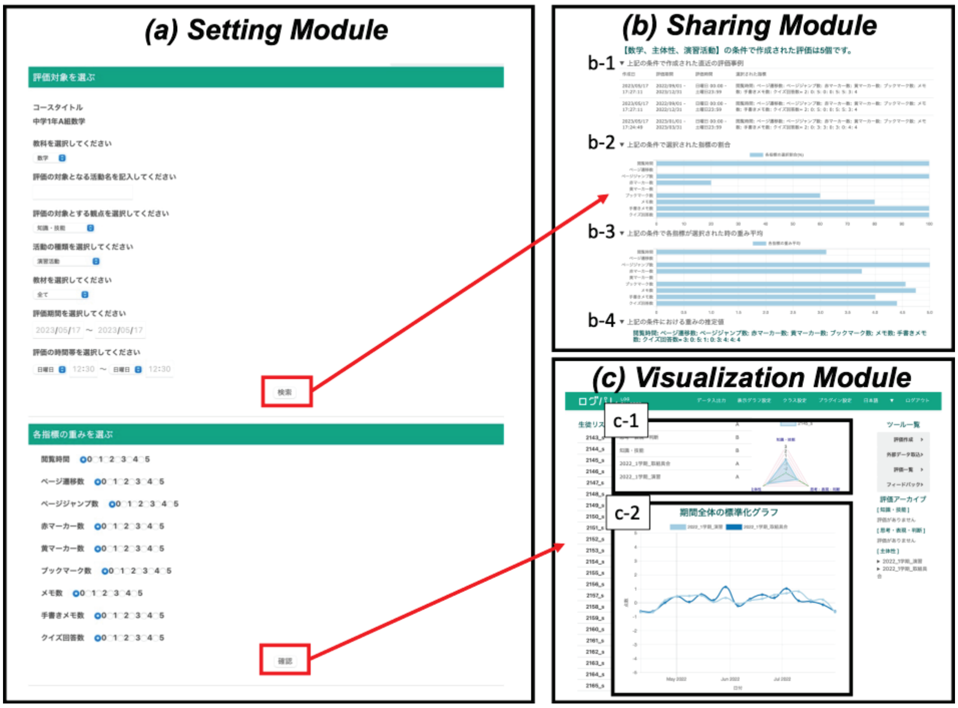

Figure 3. Main Modules of the System

4.2.1 Setting Module and Sharing Module

The Setting Module is shown in Figure 3(a). It is divided into two sections, one for inputting the context of the assessment and the other for selecting indicators. In the context input section, the name of the activity to be assessed, competency to be assessed, type of activity, and assessment period are input. In the indicator selection section, a combination of indicators can be set to calculate the assessment values by setting a weight of 1 or more for the indicators to be used, and a weight of 0 for the others.

In participatory design systems, such as the Setting Module, where data use is teacher-dependent, teachers' lack of expertise is often problematic (Dilmore et al., 2013; Holt et al., 2015). Sharing prior knowledge of data use with teachers is said to be effective in addressing this issue (Dollinger et al., 2019). This can complement teachers' expertise. Furthermore, showing examples of data use by other teachers improves the reliability of these examples. Therefore, in this study, after inputting the context information into the Setting Module, the following information in a similar context was shared by pressing the search button in the Sharing Module (Figure 3(b)).

- b-1. The three most recent assessment cases
- b-2. The percentage of selected indicators
- b-3. A weighted average of indicators when selected
- b-4. Estimated weights predicted by the system

With this information, we attempted to bridge the gap between the nature of the indicators and the teachers' understanding of them. As the assessments created by the project continue to accumulate as case studies, the generalizability of shared information will increase. We believe that this will lead to improvements in the validity of the correspondence between the selected indicators and competencies.

4.2.2 Visualization Module

The Visualization Module, shown in Figure 3(c), provides the visualization of teacher-generated assessments. It is divided into the following two sections

- c-1. Visualization of summative assessments
- c-2. Visualization of formative assessments

Based on the results obtained in 4-1-2, we considered the requirements for each section. c-1 has two requirements: (i) multiple assessment values should be visible together, and (ii) the average of the indicators over the entire period should be available. First, to achieve (i), we visualized multiple assessment values on the same graph and table so that all assessment values could be grasped at once. Second, to achieve (ii), the average values over the entire assessment period were described in a radar chart and in three levels of ABC, so that the average characteristics of learning over the period could be captured. Achievement of these two requirements is expected to lead to a comprehensive understanding of the characteristics of learning activities.

In c-2, the three requirements were (iii) that multiple assessment values be visible together, (iv) that the process of changing assessment values be visible, and (v) that it be possible to see when the change occurred. To achieve (iii), multiple assessment values were visualized on the same graph as in c-1. To achieve (iv), the assessment values were calculated according to a time series so that the process of change in the assessment values could be seen. Finally, to achieve (v), the assessment values were calculated weekly to show in which week the change occurred. Fulfillment of these three requirements is expected to lead to continuous assessments and measurement of the effectiveness of instruction.

In summary, this module has the functionality to perform a comprehensive and continuous assessment based on multiple assessment values. It is expected to contribute to the observation of the characteristics and changes in learning activities and the effects of their own interventions.

4.3 RQ3: Interview about the System

To investigate their impressions of the developed data-based assessment system, we interviewed three teachers (one math, two English) from high schools that have implemented the LEAF system.

4.3.1 Expectations and Concerns

Each teacher was asked about his expectations and concerns about the system compared to traditional competency assessment. As expected, the English teachers responded that it was possible to assess independent reading activities. The mathematics teacher responded that there were activities that were not fully captured in the field and that the system could be used to assess such activities. He also said that it would be good to be able to assess indicators that can only be qualified by data (e.g., average solving time). One concern raised by teachers in both subject areas was that not all students' activities were conducted in the LEAF system, and even when they used it, they were not always performing the activities they were supposed to.

This indicates that teachers expect two things from data-driven assessment: that it will allow them to understand what is beyond their reach, and to quantify students' activities with objective measures. We also found that they were concerned that the activities using the system would not be fully accomplished. Therefore, data from more sensors should be converted into indicators to increase their diversity and promote the activities supported by LEAF.

4.3.2 Experience and Observation

Next, we asked a mathematics teacher to use the system and observe its usage to elicit reflections of teachers. This was conducted shortly after a demonstration of the system. Thus, although he had received a brief explanation of the system's functions and purpose, he had never directly used the system. In the same context as in Section 4.1.2, we asked him to select indicators for assessing students' SRSs in exercise activities in two patterns: with and without the use of the function for sharing assessment cases. In both patterns, he selected indicators from a state in which nothing is selected. Figure 4 shows the results of the indicator selection for each case. Indicators not selected for either pattern are not shown.

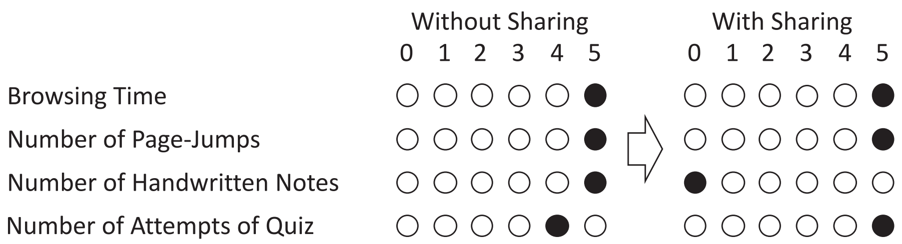

Figure 4. Two Patterns of Indicator Selection by the Teacher

Also, while he silently selected indicators from start to finish when he did not use the sharing function, he commented on it as follows.

(u1) "It would be nice to see the title (author's note: name of the indicator), I would like to know the purpose for which the indicator was created."

(u2) "I guess different subjects have different choices."

(u3) "Just a quiz would be fine."

(u4) "Even if the others' selections of indicators are shared, the selection would not be affected very much."

These differences provide some insight into the effect the feature had on him. First, in statement u3, while he considered quizzes important to assess SRSs, he gave less weight to the number of quiz attempts than the other indicators without the sharing function. However, when using the sharing function, he gave the maximum weight to this indicator. Furthermore, in u4, while he said that the selection was not much affected, the actual selection changed between the condition with sharing function and that without. These findings suggest that the information about others' choices shared by the sharing function provided him with some awareness of which indicators were actually important to him.

Contrarily, a weak point is that, as stated in u1, the name or purpose of the indicator was not indicated. This implies that to refer to the shared information, it is important that its purpose matches one's own purpose. In the future, it is necessary to improve the sharing function and share the purpose itself so that teachers can refer to the shared information.

Finally, concerning u2 and u3, some statements focused on the differences in subjects and emphasized the importance of quiz activities that they often use. This suggests that teachers may choose indicators based on their own context and teaching style. Therefore, we will examine whether these factors actually influence teachers' choice of indicators and also investigate the possible influence of other factors in future research.

5. Conclusion and Future Work

To achieve data-driven competency assessment by teachers, we created a framework that allows teachers to freely select indicators and implemented a system that shares information about other teachers' assessments. The system differs from conventional data-driven competency assessment systems in that teachers select their indicators by themselves. While this has the advantage of allowing teachers to tailor their assessments to their context, it also decreases the validity of the assessment. However, the accumulation of assessment cases is expected to contribute to addressing this issue. In addition, the interviews with teachers revealed that the system seemed to allow the capture of independent learning and detailed efforts that could not have been captured before. However, it became clear that there were problems outside the system, such as a lack of sufficient activities using the system. Furthermore, when teachers were asked to use the system, it was found that the Sharing Module helped teachers gain awareness regarding indicator selection and that their indicator selection changed. In the future, we would like to increase the effectiveness of the shared information by clarifying what teachers consider important in selecting indicators.

In summary, the proposed system and functionalities are expected to achieve a context-independent and flexible data-driven assessment. Furthermore, it will be possible to continuously improve learning and teaching with activities using the system from multiple perspectives. However, in the specific assessment cases presented in this study, only the logs obtained from LEAF were used for assessment, and there was a limitation that the data obtained from off-system activities and other tools were not included in the assessment target. However, xAPI format-compliant data are available in the system. Therefore, if other tools are also xAPI compliant, cross-data source assessment will be possible in the system. Furthermore, if more sensors record off-system activities as data, the system can be extended to off-system activities. In this way, we hope to make it possible to conduct highly valid assessments in a variety of contexts.

Acknowledgments

This study is supported by NEDO JPNP20006, JSPS KAKENHI 22K20246, and 23H00505.

References

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational psychologist, 50*(1), 84-94.

Dilmore, T. C., Moore, D. W., & Bjork, Z. (2013). Developing a competency-based educational structure within clinical and translational science. *Clinical and Translational Science, 6*(2), 98–102.

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. M. (2019). Working together in learning analytics towards the co-creation of value. *Journal of Learning Analytics*, *6*(2), 10-26.

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist*, *50*(1), 14-30.
Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary educational psychology*, *34*(1), 18-29.
Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4), 203–209.
Guerrero-Roldán, A. E., & Noguera, I. (2018). A model for aligning assessment with competences and learning activities in online courses. *The Internet and Higher Education*, *38*, 36-46.
Henri, M., Johnson, M. D., & Nepal, B. (2017). A review of competency-based learning: Tools, assessments, and recommendations. *Journal of engineering education*, *106*(4), 607-638.
Holt, E. A., Young, C., Keetch, J., Larsen, S., & Mollner, B. (2015). The greatest learning return on your pedagogical investment: Alignment, assessment or in-class instruction? *PLoS One, 10*(9).
MEXT. (2018, June 5). *Human Resource Development for Society 5.0 ~Changes to Society, Changes to Learning~.* https://www.mext.go.jp/component/a_menu/other/detail/__icsFiles/afieldfile/2019/09/27/1419637_15.pdf, (ref 2023-8-1).
MEXT. (2020, July 16). *The image of the transformation of learning brought by "1 device for 1 student with a high-speed network".* https://www.mext.go.jp/en/content/20200716-mxt_kokusai-000005414_04.pdf, (ref 2023-8-10).
Ogata, H., Majumdar, R., Akcapinar, G., Hasnine, M. N., & Flanagan, B. (2018). *Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning*. In 26th International Conference on Computers in Education Workshop Proceedings (pp. 493-496). Asia-Pacific Society for Computers in Education (APSCE).
Renninger, K. A., & Bachrach, J. E. (2015). Studying triggers for interest and engagement using observational methods. *Educational Psychologist*, *50*(1), 58-69.
Siadaty, M., Gasevic, D., & Hatala, M. (2016). Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics*, *3*(1), 183-214.
Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, *94*, 252-275.
Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of educational research*, *85*(4), 475-511.
Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In *Handbook of self-regulation* (pp. 531-566). Academic Press.