Research and Practice in
Technology Enhanced Learning

RESEARCH Free and Open Access

YINSIGHT: Supporting data-informed competency assessment with customizable indicators in a self-regulated learning context

Taito Kano [1] *, Izumi Horikoshi [2], Kento Koike [3] and Hiroaki Ogata [4]

*Correspondence:
kano.taito.87a@kyoto-u.ac.jp
Graduate School of Informatics,
Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
Full list of author information is available at the end of the article

Abstract

Recently, educational assessment has shifted focus toward evaluating not only performance but also learners' attitudes and behaviors toward learning, known as competency assessment. Traditional methods, such as self-report sheets and teacher observations, are limited by bias and reliability. With the rise of ICT tools, learning trace data offer a promising solution for assessing learning processes more reliably. However, existing frameworks for competency assessment based on trace data lack flexibility in real-world applications, prompting the need for customization of the framework according to user needs. To address this gap, this study introduces YINSIGHT, a system that allows users to customize competency assessment indicators according to specific contexts and needs. We outlined the framework for creating these indicators, implemented the YINSIGHT system, and evaluated its effectiveness through semi-structured interviews within a scenario of self-regulated learning. The participants were two English and one math teacher from a high school in Japan. The thematic analysis of interviews revealed that while traditional competency assessments rely heavily on performance-based methods, teachers expressed expectations for YINSIGHT's ability to capture self-directed learning activities, particularly in extensive reading contexts. However, significant concerns emerged regarding system usability problems and compatibility with current practices. Teachers also provided constructive suggestions for gradual implementation and system improvements to address these barriers. This study thus contributes to the continuous improvement of learning and teaching from multiple perspectives on the activities that use the system.

Keywords: competency assessment, data-informed assessment, learning analytics

Introduction

Recently, as active learning, 21st-century skills, and lifelong learning have attracted worldwide attention (Bennett et al., 2010), educational assessments have become increasingly focused not only on what students understand and can use but also on how they approach learning (Gunderman et al., 2003). This assessment of learners' learning attitudes and behaviors, rather than just their academic ability, is called competency assessment (Wong, 2020). When performing a competency assessment, it is important to unravel and assess the learner's learning and thinking processes, both to make a highly valid assessment and guide the learners appropriately (Greene & Azevedo, 2010; Grover et al., 2017). Current assessment methods based on the learning process include self-report sheets by learners (Appleton et al., 2006; Fredricks & McColskey, 2012; Greene, 2015) and observational methods by teachers (Fredricks & McColskey, 2012; Puspita & Suyatno, 2020). However, the problem with each of these methods is that sampling can be biased and the reliability of the assessment basis is low (Appleton et al., 2006; Fredricks & McColskey, 2012; Greene, 2015; Renninger & Bachrach, 2015).

At the same time, ICT tools have become widespread in educational settings, and learners' learning processes are being recorded as trace data (Sung et al., 2016). As learning can be tracked using such data (Faber et al., 2017; Kaliisa & Dolonen, 2023), there is a growing interest in using learning trace data as a highly reliable method for assessing the learning process and its applications to competency assessment (Greene & Azevedo, 2010; Grover et al., 2017; Siadaty et al., 2016; Winne & Perry, 2000; Zhang et al., 2023; Zhidkikh et al., 2023).

Although there are some studies on competency assessment based on trace learning data, they have not yet fully permeated the real-world environment. One reason is the limited applicability of competency assessment frameworks based on trace data proposed by researchers in real-world environments. These frameworks typically involve predefined definitions of assessor purposes, activities, data, and metric calculations (Grover et al., 2017; Zhang et al., 2023; Zhidkikh et al., 2023). However, the definitions and evaluation methods of competencies vary depending on assessor, context, and needs (Vazirani, 2010). Therefore, such inflexible frameworks may be useful in specific scenarios but are challenging to apply in a wider range of learning environments.

This adoption issue has also been widely addressed in the field of learning analytics (Dawson et al., 2019). To address it, learning analytics dashboards (LADs) that allow users to customize indicators according to their needs have been proposed (Muslim, 2022; Pérez-Berenguer et al., 2020). Therefore, in this study, to achieve tailored competency assessment based on data applicable to a wide range of learning environments, we propose a system called YINSIGHT, which allows users to customize indicators for competency assessment according to their own contexts and needs.

Specifically, we designed a framework to customize competency indicators using trace data and conducted scenario analysis using self-regulated skills (SRSs). Based on this framework, we implemented YINSIGHT. Finally, we interviewed three teachers regarding YINSIGHT and evaluated its effectiveness. We believe that this system will afford each user the flexibility to make customized assessments tailored to their context and improve the quality of observations. We address the following research questions (RQs) under the self-regulated learning (SRL) context.

RQ1: How can the customizable data-informed competency assessment be realized?

RQ2: How can the customizable data-informed competency assessment be implemented?

RQ3: How do teachers expect, use, and evaluate data-informed competency assessment systems?

This manuscript is an extension that adds the following points to the authors. First, in the Literature Review section, competency and assessment are considered. Then, customizability in terms of learning analytics is explained as background information for this study. In the Research Foundation and Design section, an explanation of the overall research design is provided. In RQ1 section, we explain the framework of competency indicator creation and conduct a scenario analysis along with user flow. In addition, in RQ2 section, an overview of the system is given. In the Discussion section, the implications, limitations, and future work are discussed.

Literature Review

The subsection titled Competency Assessment introduces the distinction between subject-specific and generic competencies, emphasizing their implications for assessment and the importance of considering individual differences in behaviors and processes. The second subsection examines process-based competency assessment methods, comparing self-reports and teacher observations with data-informed approaches to highlight their respective strengths and limitations. The third subsection explores customizability using LADs, thus demonstrating its potential to enhance assessment flexibility and address existing research gaps. Finally, the fourth subsection refines the study's objectives, ensuring a clear focus for using trace data to provide tailored insights into SRL skills.

Competency Assessment

Generally, competency refers to measurable individual attributes that contribute to performance improvements (Chung & Lo, 2007; Page & Wilson, 1994; Spencer & Spencer, 1993). The background of the increased focus on competency lies in the business world, where personal abilities such as leadership, communication skills, and goal-setting skills are considered important, alongside skills and knowledge, to achieve high performance (McClelland, 1973). Furthermore, competency-based education has gained popularity, as the focus of education has shifted toward cultivating learners who acquire the skills necessary for success in society.

Assessment is essential to develop competencies (Wong, 2020). Proper assessment tailored to learning objectives and activities leads to the efficient development of learners' competencies and achieves high validity in evaluation. First, given that competency is too complex to be treated as a single concept, it is divided into two categories: subject-specific competency (SSC; i.e., the level of knowledge and skills required in each subject and the understanding of course content) and generic competency (GC; i.e., the abilities and knowledge that should be developed independently of specific subjects, such as leadership and communication skills) (González & Wagenaar, 2003). This classification allows for a clearer explanation of their distinct characteristics. This subsection also explains the methods used to assess each type of competency, with a particular focus on how behavioral processes are utilized.

Subject-Specific Competency Assessment

On the one hand, in an SSC assessment, a variety of methods, such as paper-based tests, quizzes, and assignments, are combined for evaluation (Bacquet, 2020; Thippayacharoen et al., 2023), with a primary focus on outcomes. On the other hand, in computer-based testing (CBT), as exemplified by PISA, numerous competency estimation models have been proposed to enhance the accuracy of competency assessment by combining response outcomes with the time taken for each item (e.g., Jiao et al., 2019; Zhang et al., 2023). Additionally, research has been conducted on methods for detecting abnormal behaviors in CBT environments (e.g., Van der Linden & van Krimpen-Stoop, 2003; Zhu et al., 2023).

Therefore, while SSC assessments primarily focus on outcomes, the integration of ICT tools has made it possible to record behavioral processes during assessment tasks, such as the time spent and behavior patterns. Leveraging these data to examine individual differences in processes holds potential for improving the accuracy and trustworthiness of SSC estimations.

Generic Competency Assessment

GC refers to broad competencies such as leadership, communication skills, and self-regulation, which are valuable across all subject areas. GC is a social and personal competency (Le Deist & Winterton, 2005) demonstrated in collaborative processes with others and in personal learning processes. It is generally assessed using broadly standardized methods, such as surveys, and locally developed methods, such as self-reports and teacher observations (Chapman & O'Neill, 2010). On the one hand, the former approach, which investigates learners' characteristics and traits through psychological methods, has high theoretical validity (e.g., Ibrahim et al., 2017; Lechner et al., 2022). However, as these approaches are evaluation methods independent of context, they cannot assess the actual learning processes and performance taking place in the learning environment. On the other hand, the latter approach offers greater flexibility in adapting to organizational policies and contextual factors, as it involves observing, describing, and evaluating the processes within actual learning environments.

Given the increasing prevalence of ICT tools, data on daily learning processes, such as engagement in collaborative tasks or self-regulation, are now available. Consequently, there are studies on competency assessment methods, such as evaluating collaboration skills (Alozie et al., 2020) and self-regulation (Siadaty et al., 2016; Zhidkikh et al., 2023), using these data as a form of locally developed assessment. These assessment methods are attracting attention as more objective and reliable approaches compared to other locally developed assessments, such as observations and self-reports.

Therefore, although SSC and GC have different characteristics, research has demonstrated the significance of considering the processes and behaviors for both.

Competency Assessment Alignment

In the context of competency assessment, ensuring that learning objectives, task content, and evaluation methods are appropriately aligned helps students achieve their learning goals (Koenen et al., 2015). Typically, competency consists of three elements: (1) a description of the competency, (2) a description of activities where the degree of competency manifests, and (3) an explanation of how to determine the degree of competency (Champion et al., 2011; Parry, 1996). Based on these elements, we divided the competency assessment alignment into six processes, as shown in Figure 1: (1) competency modeling, (2) task design, (3) evaluation design in the design phase, (4) task execution and data collection, (5) evaluation execution and data analysis, and (6) assessment during the implementation phase. Each phase is explained below.

(1) Competency Modeling

In the competency modeling phase, assessors define the competencies to be assessed based on their own evaluation perspectives and contexts, considering what states are desirable. This phase aims to set clear learning goals.

(2) Task Design

In this phase, learning activities designed for learners to acquire the targeted competencies are planned.

(3) Evaluation Design

In the evaluation design phase, the methods for assessing activities are designed based on the competencies to be evaluated and the content of the tasks for acquiring them. This includes designing evaluation methods for activities, such as tests and presentations, to define competency proficiency.

(4) Task Execution and Data Collection

The task execution and data collection phases involve assigning and implementing the tasks designed during the task design phase to actual learners. During the implementation phase, observation and trace data are collected for evaluation.

(5) Evaluation Execution and Data Analysis

In the evaluation execution and data analysis phase, the assessments designed in the evaluation design phase are implemented.

(6) Assessment

Assessment involves summative and formative assessments based on the results of an evaluation (Garrison & Ehringhaus, 2009). Summative assessment refers to evaluations such as end-of-term grading, whereas formative assessment involves providing motivation and learning support based on the process of tasks, as needed.

Process-based Competency Assessment Method

As highlighted in the previous subsection, the significance of understanding the behavioral process under both SSC and GC has been suggested. This subsection describes three representative assessment methods in competency assessment that focus on behavioral processes: self-report, observation, and data-informed competency assessment.

Self-Report

Self-reporting is one of the most commonly used methods for evaluating learners' behavioral processes and engagement (Appleton et al., 2006; Fredricks & McColskey, 2012). Learners respond to questions about their learning either in an open-ended format or through multiple-choice options. Self-reporting provides valuable insights into learning situations and the inner voice from learners' perspectives, contributing to improving teacher awareness about their learners. However, it has low reliability as an assessment criterion due to the dependency on learners' expressive abilities, meta-cognitive skills (Nederhand et al., 2021; Renninger & Bachrach, 2015; Weil et al., 2013), introduction of biases, and lack of objectivity. Therefore, this method is often used in formative assessments for instructional improvements and individual guidance.

Observation

The observational method, alongside self-reports, is another representative technique for capturing behavioral processes (Fredricks & McColskey, 2012; Puspita & Suyatno, 2020). Observation has the advantage of flexibility in adjusting the focus to specific students or aspects based on context, making it practical for instructional practice and evaluation. It also allows for the observation of learners' learning and thought processes in natural settings. However, the scope of teachers' observations is limited, and interpretation can vary among teachers (Turner & Meyer, 2000), leading to low reliability as an assessment criterion.

Data-informed Competency Assessment

The proliferation of e-learning tools has enabled the recording of learners' learning processes as data (Sung et al., 2016). Such data, known as trace data, are increasingly being utilized in the field of learning analytics, particularly for improving learning and instruction (Ferguson, 2012). Trace data have also garnered attention in the context of competency assessment.

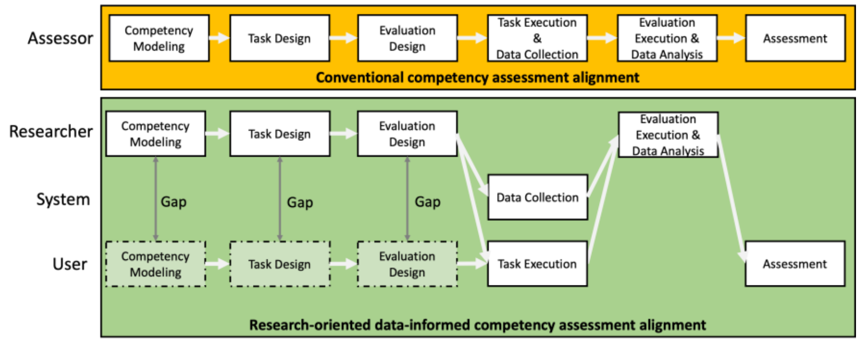
Data-informed competency assessments using trace data have been applied to various competencies. For example, in the 2012 Program for the International Assessment of Adult Competencies (PIAAC) survey, conducted by the Organization for Economic Co-operation and Development (OECD), a problem-solving in technology-rich environments (PSTRE) competency estimation model was proposed, which combined response results with the time spent on each item (Zhang et al., 2023). This model demonstrated higher reliability in estimation compared to a model that used only response results. Furthermore, Grover et al. (2017) proposed a series of analytical methods that combine the data- and hypothesis-driven approaches to discover behavior patterns or anti-patterns related to computational thinking skills in K-12 block-based programming environments. These patterns could potentially be used for both the formative and summative assessments of computational thinking skills. Zhidkikh et al. (2023) proposed and evaluated a method that combines event scales based on trace data with aptitude scales based on self-assessment in the context of SRL, offering comprehensive insights for SRL assessment.

Therefore, data-informed competency assessment provides quantitative and objective insights into behavioral processes, complementing qualitative assessment methods such as self-reports and observation.

Additionally, data-informed competency assessment alignment is achieved through similar phases. One key difference from traditional competency assessment lies in the fact that researchers bear the responsibility for much of the alignment in assessment. These research-oriented frameworks often predefine the assumed assessors, purposes, target activities, data used, and indicators calculation methods by researchers (Grover et al., 2017; Zhang et al., 2023; Zhidkikh et al., 2023). While these less flexible frameworks may be useful in limited, specific scenarios, applying them to a wide range of learning environments is challenging.

Figure 1

Conventional & data-informed competency assessment alignment



Customizability in Learning Analytics

In the previous subsection, we explained data-informed competency assessment in research environments; however, it is rarely conducted in practical environments. However, the use of trace data through LADs has been widely implemented (Matcha et al., 2019). LADs refer to a display that presents various indicators related to learners or the learning process in one or more graphs (Schwendimann et al., 2016).

In the development of LADs, the customizability of visualization, analysis, and indicators is considered essential for meeting user needs (Wise & Vytasek, 2017). For example, the Visualizer developed by Mutlu et al. (2018) allows users to adjust various properties of the dashboard's visualization, such as the types and sizes of graphs and background colors. Muslim et al. (2022) developed a tool that enables users to create indicators by inputting data into a simple form. Within this tool, simplicity and ease of creation are important to ensure that even users who are unfamiliar with the technology can create indicators.

These LADs have the potential to be used for competency assessments. If data-informed competency assessments were to be implemented in a practical environment, it would be necessary to identify trace data related to the tasks, integrate it to create evaluation indicators, and visualize it in relation to competencies. While typical LADs may not meet these requirements, LADs that allow customizability in terms of indicators could potentially address these needs. However, as the LADs that allow the customization of indicators are still limited, especially for competency assessment purposes, such tools were not within the scope of our investigation (Matcha et al., 2019).

Related Works and Objectives

We now summarize the strengths and weaknesses of competency assessment based on the learning process discussed thus far. The traditional process-based assessment methods used in K12 education are flexible in adapting to the learning environment. However, they have mainly relied on unreliable sources such as self-reports and observational data, with little utilization of reliable trace data. However, recent data-informed competency assessment methods have the advantage of providing reliable information about detailed learning processes captured from trace data. However, there is hitherto no tool that supports flexible data-informed competency assessment capable of addressing the diverse learning activities and evaluators' needs in practical environments. The aim of this study is to enable teachers to gain tailored insights into the micro-level learning processes based on trace data when conducting competency assessments. To achieve this, we modeled and implemented a user-centered data-informed competency assessment system, drawing from the LADs with customizability of indicators, and evaluated these implementations. Table 1 shows our research in relation to previous studies.

Table 1

Comparison between our research and conventional process-based assessments

	Self-report, Observation, etc. (Appleton et al., 2006; Fredricks & McColskey, 2012; Greene, 2015)	Conventional Data-informed competency assessment (e.g., Grover et al., 2017; Zhang et al., 2023;)	Customizable Indicator Dashboard (Muslim, 2022)	Our research
Customizability	✓		✓	✓
Data-informed		✓	✓	✓
Competency Assessment	✓	✓		✓

Research Foundation and Design

Here, we describe the Learning and Evidence Analytics Framework (LEAF) system, which serves as the research foundation for the implementation of YINSIGHT. Additionally, an explanation of the overall research design, as well as data collection and analysis, will be provided.

LEAF

The LEAF system developed by our laboratory, shown in Figure 2, is a learning platform that integrates a learning management system (Moodle), electronic teaching materials (BookRoll), a learning support module, a learning analysis module (LogPalettes), and a learning record store (Ogata et al., 2022). LEAF also serves as a sensor that records learners' detailed learning behavior and a database that stores learning logs, improving the experience gained through learning activities and supporting more effective and efficient learning. A significant feature of LEAF is its LMS independence. The log data used in

LEAF adheres to a global standard called the Experience API (xAPI), allowing it to be seamlessly integrated with logs from other learning tools following the xAPI standard (Flanagan & Ogata, 2018). Table 2 shows examples of the log data of BookRoll, which we used for data-informed competency assessment.

Figure 2

LEAF architecture

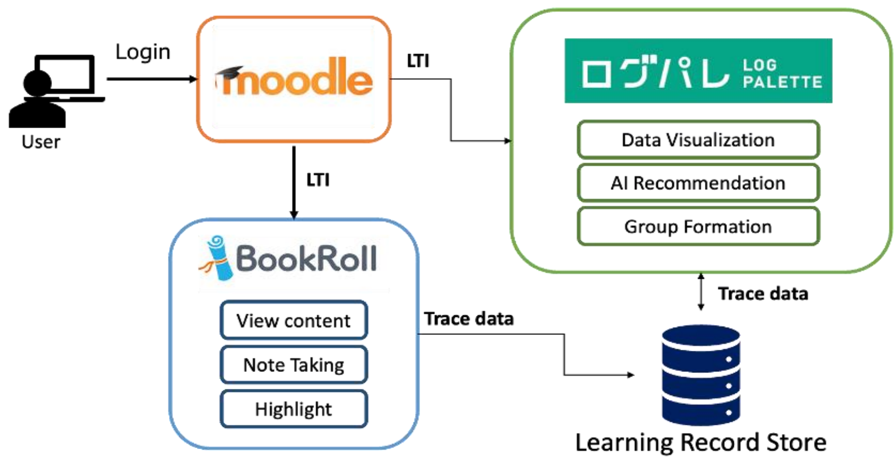


Table 2

Examples of learning log data recorded by BookRoLL

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

Research Design

In this study, we employed a design-based research approach with qualitative evaluation components, structured in three phases to answer the three RQs on model design, system implementation, and evaluation. Design-based research was selected as our methodological framework because it allows for the iterative development and refinement of educational interventions in real-world contexts (Wang & Hannafin, 2005), which aligns with our goal of creating a customizable data-informed competency assessment system that meets users' needs in authentic educational settings. The qualitative evaluation component was chosen because this approach is suitable for investigating phenomena that are not yet well understood or theorized (Robinson, 2014; Palinkas et al., 2015). In this case, the study aims to explore how teachers perceive and use a newly introduced data-informed competency assessment support system (YINSIGHT) for student learning evaluation. The exploratory nature of the study allows for identifying early patterns, user needs, and contextual limitations based on actual teacher practices.

First, in the Model Design section (RQ1), we designed a flow for calculating indicators from trace data based on the user's input and created a model in which the user plays a central role in implementing data-informed competency assessment. As an example of applying this model, we demonstrated how to evaluate students' SRSs based on trace data collected when Japanese second-year junior high school students used BookRoll. The reason for focusing on SRSs is its strong relevance to "agentic learning disposition," one of the three competencies that the Japanese government emphasizes in education (MEXT, 2017). Additionally, there are many theoretical and practical examples of measuring SRSs, making it highly reliable and valid (Greene & Azevedo, 2010; Siadaty et al., 2016; Winne & Perry, 2000); it is also a competency that can be mapped to the trace data obtained in this study.

In System Implementation section (RQ2), we implemented a system called YINSIGHT based on the designed indicator calculation flow and the user-centered competency assessment alignment model. This system allows the user to customize indicators and link them to competencies.

Finally, in Evaluation section (RQ3), we evaluated the implemented system in two stages. First, we asked two English teachers in Japan to review the system and provide feedback on two points: (1) their previous methods for conducting competency assessments and (2) their expectations and concerns about using YINSIGHT for competency assessment. We organized their responses and compared the expectations and concerns regarding the use of YINSIGHT for competency assessment with traditional methods. Second, we had one math teacher review the system similarly and, following the analysis in Model Design section, asked the teacher to design indicators for evaluating SRSs. From the indicators designed by the teacher and their feedback, we gained insights into areas for improving the system.

Through these three steps, we demonstrated how user-centered data-informed competency assessment could be realized, its effectiveness, and the challenges that should be addressed in future research.

RQ1: How can the customizable data-informed competency assessment be realized?

Methods

We initiated the process by developing an indicator customization framework based on trace data. The indicator customization framework was developed from the perspectives of computational processes and user flow. Regarding the computational processes, we designed a computational process flow based on the input variables provided by the users to enable them to customize the indicators. Subsequently, we designed a user flow for a customizable data-informed competency assessment. In the conventional data-informed competency assessment alignment model, most of the work is done by researchers; however, we designed a user-centered data-informed competency assessment alignment that allows users to work almost entirely by themselves. Finally, we conducted a scenario analysis of competency assessment for SRSs to elucidate an assessment that can be implemented using the proposed indicator customization framework. Here, we used trace data related to engagement from an electronic learning material collected between April 1 and June 8, 2023, from a class of second-year junior high-school students in Japan.

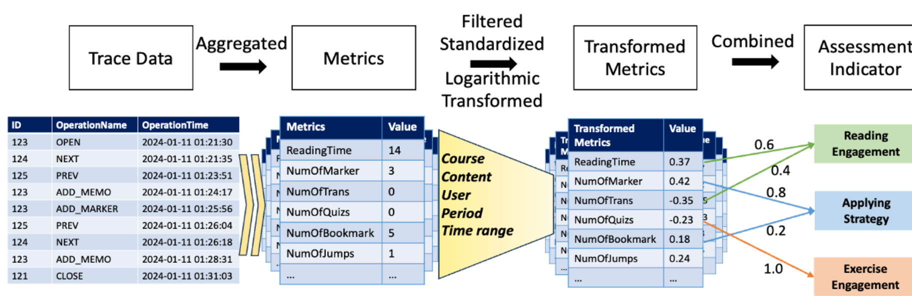
Results

Indicators and data processing flow

The following explains the process of calculating the indicator using the trace data shown in Figure 3.

Figure 3

Computational processing flow



First, the metrics were created by aggregating the trace log data obtained from LEAF. This phase is fully automated, with no involvement from the user. While there are LADs like those described by Pérez-Berenguer et al. (2020), in which the user manually creates indicators from trace data, such tools require users to create indicators using domain-specific language. However, handling domain-specific language is not easy for general teachers. To make it easier to create indicators, we provided predefined metrics, as suggested by Muslim et al. (2022) in their LD design. Additionally, as mentioned in the subsection titled LEAF, the trace data used in this study from LEAF follows the xAPI format. Therefore, any external tools that adhere to the xAPI format can convert the metrics in the same way. This approach responds to the call for increased interoperability in LADs through the use of the xAPI format or Caliper (Schwendimann et al., 2016).

Second, the metrics are filtered using the time range input by the users. Next, the logarithmic transformation of each indicator across learners to make the distribution closer to a normal distribution and standardization to align the distribution means of the different metrics were performed. Therefore, the influence of outliers can be suppressed, and the different metrics can be summed.

Finally, by selecting the metrics to be used to create an indicator and setting weights for each, the values of the selected metrics were weighted and summed to calculate the indicator. In conventional assessment activities, scores on several tests or the results of a presentation are weighted and summed to create a rating. Based on these suggestions, ratings are created by adding weights to the metrics.

In this manner, we designed an indicator processing flow that allows users to intervene by performing indicator calculations based on input variables, incorporating filtering and weighting. This calculation method incorporates the simple indicator creation approach proposed by Muslim et al. (2022), while enhancing the flexibility of indicator creation through operations such as time filtering and combinations. Nine metrics were created in this study, as listed in Table 3.

Table 3

List of metrics

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

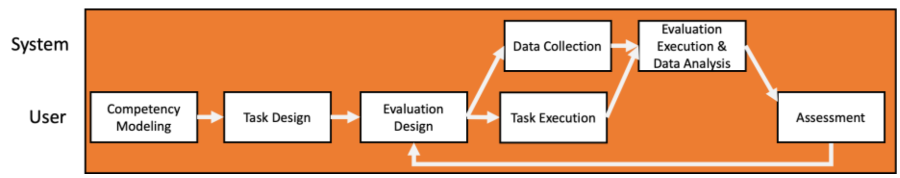
User-centered Assessment Flow

To achieve a data-informed approach applicable to diverse learning activities, most phases of the data-informed competency assessment were designed to be executable by users, as shown in Figure 4, leading to the practical implementation of a data-informed competency assessment aligned with the real learning environment.

Furthermore, ensuring that the assessment results align with the user's perspective is crucial for conducting competency assessments. Therefore, in this model, we anticipate an iterative process of evaluation creation, in which users adjust their creation methods if the results do not align with their expectations after viewing them, thus iterating the creation process. Specifically, if indicator values are skewed or the results differ from user expectations, adjustments of indicator weights, redesigning activities, or redefining targeted competencies are possible. Through this iterative process, it is expected that the indicators will become optimal for the users' intuition and the actual situation.

Figure 4

User-centered data-informed competency assessment alignment



Scenario Analysis

To demonstrate the application of the alignment model, we conducted an assessment of SRSs using these nine indicators in a scenario that assumed an actual educational setting as follows.

(1) Competency Modeling

As the initial step, we decided to assess SRSs as the target of the competency assessment. Based on theory, micro and macro-level processes in the SRL are structured as shown in Table 4 (Siadaty et al., 2016). We defined the ability to carry out these processes as SRSs. Furthermore, we believe that it is desirable for learners to demonstrate a mindset of making efforts and persevering in their daily learning activities. This aligns with the "working on the task" micro-level process within the macro-level process of "engagement" in SRL. However, such efforts are difficult to evaluate objectively and fairly for all students through self-reports or direct observations by teachers. Therefore, we decided to use trace data to quantify the degree of students' engagement in "working on the task" and use this as reference data when evaluating the SRSs.

(2) Task Design

Regarding "working on the task," we decided to perform three tasks, both inside and outside the classroom: reading, exercise, and applying strategy activities. For the reading activities, students are asked to read textbooks and understand theorems and problem-solving approaches. In exercise activities, students work on tasks using their acquired mathematical knowledge to master the subject matter. Finally, apply strategy activities involve the use of organizational and marker strategies carried out in conjunction with reading and exercise activities. These activities, both inside and outside the classroom, may be conducted according to the teacher's instructions or can be initiated by the learners themselves. Additionally, for the collection of trace data, all activities were conducted on an e-book platform.

Table 4

Micro- and macro-level processes in the SRL model and their descriptions (Siadaty et al., 2016)

Macro-level SRL process	Micro-level SRL process	Description
Planning	Task Analysis	To get familiar with the learning context and the definition and requirements of (learning) task at hand
	Goal Setting	To explicitly set, define, or update learning goals
	Making Personal Plans	To create plans and select strategies for achieving a set learning goal
Engagement	Working on the Task	To consistently engage with a learning task using tactics and strategies
	Applying Strategy Changes	To revise learning strategies, or apply a change in tactics
Evaluation and Reflection	Evaluation	To evaluate one's learning process and compare one's work with the goal
	Applying Strategy Changes	To reflect on individual learning and share learning experiences

(3) Indicator Design

Based on the characteristics of the metrics in Table 3 and the content of the tasks, we mapped tasks to their respective metrics, as shown in Table 5. For this study, we calculated three indicators, Reading, Apply Strategy, and Exercise, by summing the corresponding metrics using equal weights; in other words, the value of each indicator is the average of the metrics.

Table 5

Mapping between tasks and metrics

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

(4) Task Execution and Data Collection

From April 1 to June 3, we conducted activity practice based on task design, and the activities were recorded as trace data.

(5) Indicator Calculation and Data Analysis

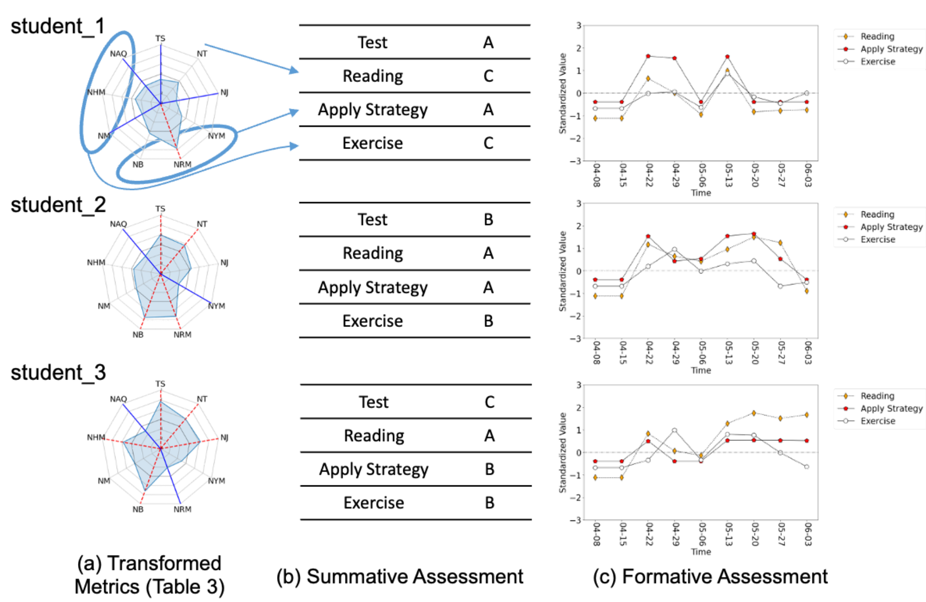
Following indicator design, the indicators were calculated from the trace data. This process was conducted entirely on the system, and the calculated results were provided.

(6) Assessment

Figure 5 shows the results of the indicator calculations. Figure 5(a) presents the values of the metrics for three students throughout the study period. The red and blue lines represent the metrics at the top and bottom 30% of the values, respectively. It is evident that student_1 scored high only in terms of the number of red markers, student_2 in four metrics, and student_3 in five metrics. Each student exhibited variations in high and low metrics, suggesting differences in their approaches. The interpretative results of the indicators considering activity content are presented in Figures 5(b) and (c).

Figure 5

Transformed metrics, summative, and formative assessment indicators



Figures 5(b) and (c) demonstrate the overall average and temporal changes of indicators calculated based on the metrics and give us some insights for summative and formative assessments, respectively, of SRSs.

As shown in Figure 5(b), the assessment values derived from each activity were categorized as A for the top 30%, B for the 30–70% range, and C for the remaining activities. Test grades were also included for reference purposes.

The results of the average summative indicator assessment provided insights into students' learning activities that were not found in the 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. Despite not being reflected in the results, it can be inferred that student_3, who achieved high scores in each activity but scored low on the test, persevered and engaged in persistent learning. Such outcomes provide information on learning attitudes and methods that are not apparent through tests alone, thereby contributing to the triangulation of competency assessments. Furthermore, because data acquisition and assessment calculations are conducted mechanically, the results are highly reliable, and the process requires minimal effort, while presenting additional advantages. Therefore, in addition to traditional evaluations, it has the potential to serve as a practical method that complements qualitative assessments such as self-reports and observations by providing objective insights into learners' SRSs.

Figure 5(c) shows the changes in the indicators of the three activities for all students. Time series graphs make it possible to observe how long the activities continued and when they decreased or increased. For example, student_2 improved the Reading and Apply Strategy activities from the first week of May to the third week of May, but then decreased from the fourth week of May. Student_3 improved their Reading and Apply Strategy activities from the first week of May and continued to do so until the test.

Variations in these indicator values provide insights into learning slumps and sustained efforts in SRL, even when observed individually. These insights are likely to enhance teachers' understanding of learners' SRSs and contribute to improving the quality of feedback for the improvement of SRSs. Furthermore, the description of reliable learning processes contributes to making teachers' interpretations of learners' SRSs based on activity observations and qualitative assessments more reliable.

RQ2: How can the customizable data-informed competency assessment be implemented?

Methods

Here, we describe the implementation of the customizable data-informed competency assessment support system YINSIGHT based on the created indicator customization framework. YINSIGHT comprises three main modules: a customization module for generating indicators, a visualization module for displaying indicators, and a sharing module created by other teachers to share indicators. Whereas the customization and visualization modules are essential for realizing customizable data-informed competency assessment, the sharing module serves as a supportive component for customization. We considered the requirements that each module should meet and proceeded with the implementation.

Results

Customization Module

To enable users to customize indicators, it is necessary to provide simple and straightforward input items (Muslim et al., 2016). Additionally, context-related metadata are crucial for interpreting the indicators customized by users. Therefore, in the customization module illustrated in Figure 6, similar to the system proposed by Muslim et al. (2022), simple questions are prepared for each input item, allowing users to input their responses. Furthermore, in addition to the items for converting metrics into indicators, input items related to competency modeling and task design were provided. For example, as input items for converting metrics into indicators, users can select aggregation time periods and weights for each metric. For context-related input items, they can select the names of the competencies to be evaluated and types of activities. The information input in this manner is used for indicator calculations on visualization dashboards and for recommending related indicators based on the input context information. Therefore, the input items for user-customized indicators enable customization competency assessment indicators, and the input items for interpreting indicators support the users in customizing indicators, as expected.

Sharing Module

In participatory design systems, such as customization modules where data use is user-dependent, users' lack of expertise is often problematic (Dilmore et al., 2013; Holt et al., 2015). Sharing prior knowledge of data use with users is effective in addressing this issue (Dollinger et al., 2019). This can complement teachers' expertise. Furthermore, showing examples of data used by other users improves their reliability. Therefore, in this study, after inputting the context information into the customization module, information in a similar context was shared by pressing the search button in the sharing module (Figure 6).

s-1. The three most recently created assessment indicators

s-2. The percentage of selected indicators

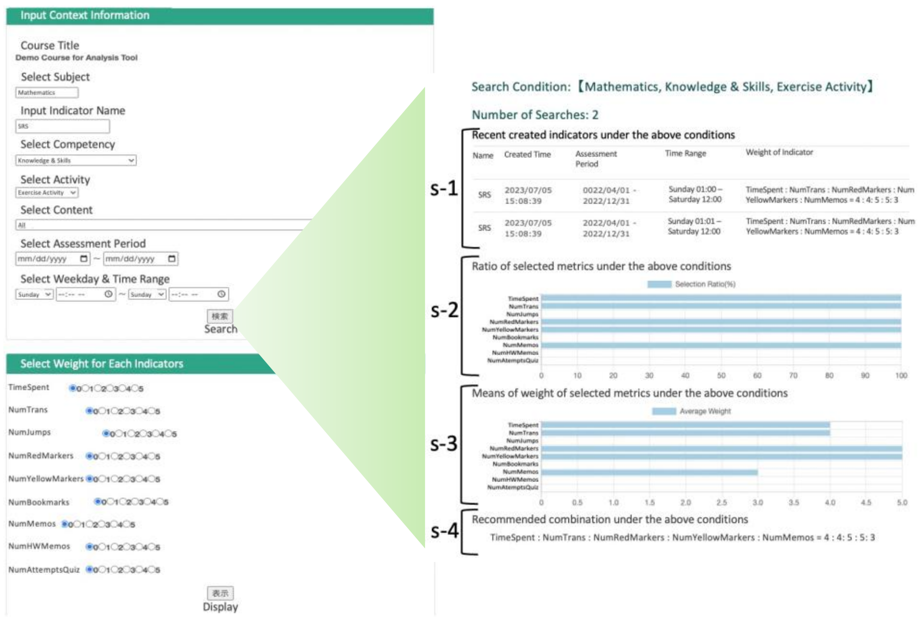
s-3. A weighted average of indicators when selected

s-4. Estimated weights predicted by the system

With this information, we attempted to bridge the gap between the nature of the data and teachers' understanding.

Figure 6

Customization and sharing modules



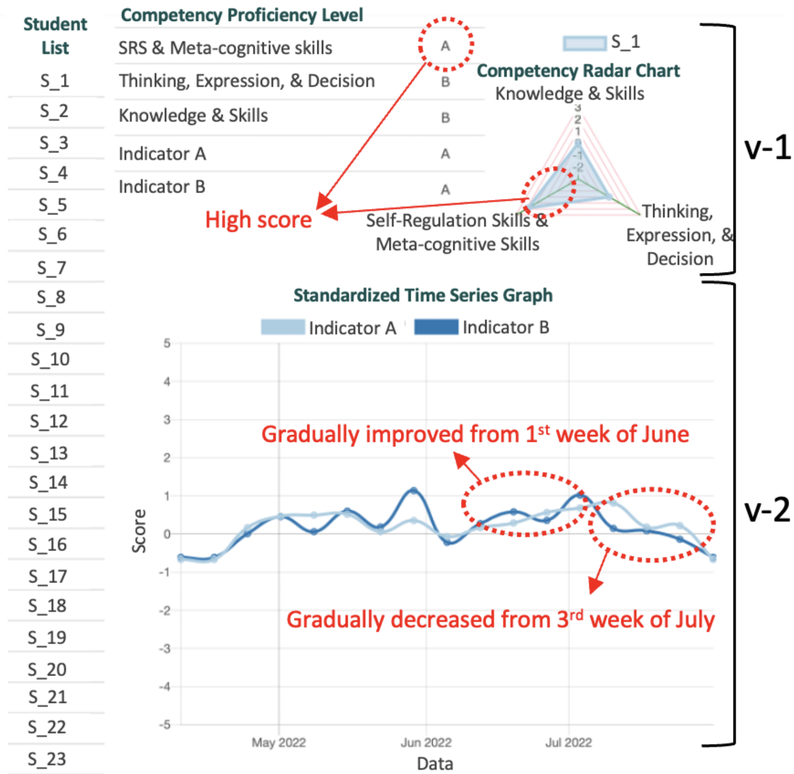
Visualization Module

The purpose of using the customized indicators for assessment includes the two aspects mentioned in subsubsection titled Scenario Analysis: giving insights for summative assessment and formative assessment. Therefore, the visualization module depicted in Figure 7 provides visualization for both summative assessment (v-1) and formative assessment (v-2).

Based on the results of the scenario analysis, we considered the requirements of each section. For v-1, there were two requirements: (i) multiple assessment values should be visible together, and (ii) the average of the indicators over the entire period should be available. To achieve (i), we visualized multiple assessment values on the same graph and table so that all the assessment values could be grasped simultaneously. To fulfill (ii), the average values over the entire assessment period were depicted in a radar chart and classified into the three levels of ABC, allowing the average characteristics of learning over the period to be captured. For example, as shown in Figure 7, it is easy to understand whether the values of each competency are high or low.

Figure 7

Visualization module

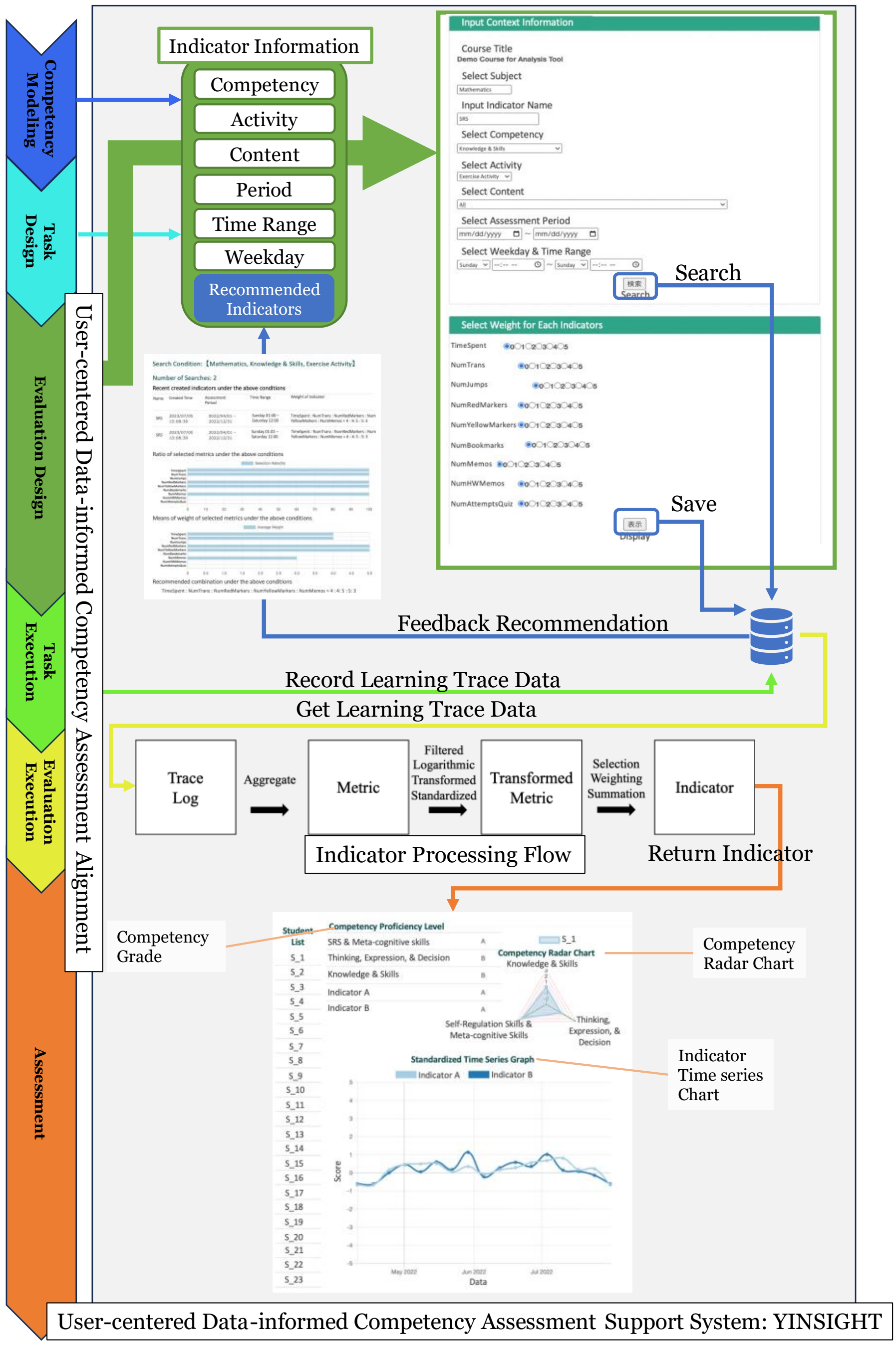


In v-2, three requirements were identified: (iii) multiple assessment values should be visible together, (iv) the process of changing assessment values should be visible, and (v) it should be possible to see when a change occurs. To address (iii), multiple assessment values were visualized on the same graph as in v-1. In (iv), the assessment values were calculated according to a time series to illustrate the process of the change in assessment values. Finally, to satisfy condition (v), the assessment values were calculated weekly, enabling users to identify the week in which the change occurred. For example, as shown in Figure 7, when each competency improves, learning progresses smoothly, whereas a gradual decline may suggest a need for intervention in learning. Moreover, if an increase is observed since June, it could be interpreted that the learning activities or guidance provided at the school during that month contributed to learners' motivation. Therefore, the module performs a comprehensive and continuous assessment based on multiple assessment values. This is expected to contribute to the observation of the characteristics and changes in learning activities.

System User Flow

Figure 8

User flow of YINSIGHT



Here, we explain the flow of system usage, aligning the three modules described thus far—the creation, sharing, and visualization modules—with the designed user-centered data-informed competency assessment alignment. Figure 8 presents an overview of the YINSIGHT interface and usage flow. Initially, in the creation module, the evaluation indicators are modeled based on the competencies to be evaluated and the design of tasks to enhance those competencies. However, it is conceivable that some users may not know which indicators to create or how to create them. In such cases, pressing the search button in the creation module recommends similar cases from the accumulated creation examples in the database, thereby providing a basis for inexperienced users.

Next, task execution based on task design was performed, and learning activities were recorded as data. The collected log data were transformed into indicators following the indicator creation flow based on the user creation method. The transformed indicators were visualized on the visualization dashboard as radar charts and time-series graphs, representing competency assessments, and were utilized for formative and summative assessments.

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

Method

Interview Process

We conducted semi-structured interviews with two English participating teachers and one math teacher from a Japanese high school using the LEAF system. The three participating teachers were selected through purposive sampling (Palinkas et al., 2015), based on their actual usage of the LEAF system in their teaching practices. The two English teachers included one who was responsible for student evaluation at the school and played a key role in coordinating the interview, and another who regularly used LEAF in daily instruction. At the time of the study, no other English teachers at the school were regular users of the LEAF system. The mathematics teacher was selected because the indicator design scenario used in the evaluation specifically focused on a mathematics lesson. Moreover, this teacher was the only one at the school who had actively used LEAF in their practice. Owing to these constraints, the small number of participants reflects the real-world limitations of system adoption rather than an arbitrary decision. In this school, the LEAF system is used only in English and mathematics, which limited our potential participant pool. While we acknowledge that the small sample size limits the generalizability of our findings, this exploratory evaluation aligns with the early-stage assessment typical in design-based research.

First, the participants had never used the system directly; therefore, we demonstrated how to use YINSIGHT, showing actual visualizations derived from trace data collected from their students' use of BookRoll to ground the discussion in concrete examples. However, this demonstration was primarily aimed at helping teachers understand the system's capabilities rather than conducting a comprehensive evaluation of how these indicators might impact their assessment practices.

Subsequently, the two English teachers and the mathematics teacher were asked about their expectations and concerns regarding YINSIGHT for competency assessment. After that, we asked the mathematics teacher to use the system and observe its usage to obtain some reflections from the teachers. This was a specific task, and we asked him to select metrics for assessing student agency in exercise activities. In this case, we asked the participant to select metrics for two patterns: with and without the use of the function for sharing assessment cases. Under both patterns, the metrics were selected from a state in which nothing was selected. During this phase, we observed differences between the scenarios with and without the sharing module to determine the actual utility of the sharing module in indicator creation.

All interviews were conducted for approximately 60 minutes each and were audio-recorded and transcribed verbatim for analysis. The transcribed interviews generated a total of 238 utterances (126 teacher utterances and 112 interviewer utterances), of which 116 teacher utterances related to impressions and perceptions of the system were selected for coding.

Data Analysis Process

We employed a hybrid thematic analysis approach (Roberts et al., 2019; Fereday & Muir-Cochrane, 2006) to analyze the interview data, combining both deductive and inductive coding methodologies. Initially, the first author designed a preliminary codebook deductively based on a close reading of the raw data and the RQs, following Boyatzis's (1998) guidelines for codebook development. Subsequently, three researchers (including the first author and two co-authors) independently coded the 116 selected utterances using this preliminary codebook, with each utterance receiving a single code. Utterances unrelated to the codebook themes were assigned "Other" codes. This process was iterative, involving multiple rounds of coding. After each round, the coders discussed and refined the codes, adding new codes that emerged from the data, merging similar existing codes, and restructuring themes until a consensus was reached.

For the final analysis, only codes that achieved consensus from at least two of the three coders were retained from the initial 39 codes, ensuring reliability in our coding framework. Finally, inter-coder reliability for the final coding was assessed using Fleiss' Kappa, achieving a value of 0.510. This systematic process resulted in 31 final codes across 5 main themes: Current Assessment (8 codes), Device Usage (6 codes), ICT Environment and Tool Access (3 codes), Learning Support Tools (8 codes), and Competency Assessment with YINSIGHT (7 codes), combining deductive codebook creation with inductive refinement through multiple coders and reliability checks to ensure methodological rigor in our qualitative analysis.

Results

Expectations and concerns

To understand teachers' perceptions of YINSIGHT in context, we first analyzed current assessment practices at the participating school based on our coded data. In Japan, teachers conduct three-perspective assessment covering agency, thinking-expression-judgment skills, and knowledge-skills. The participating teachers' assessment practices were primarily performance-based, focusing on oral examinations and tests ('Performance Tests/Oral Examinations', 2 utterances). For agency assessment, teachers collect e-portfolios ('E-portfolio', 2 utterances), audio data ('Audio Data', 1 utterance), and surveys ('Surveys', 1 utterance) to infer students' behind-the-scenes efforts from observable outcomes ('Agency Assessment', 1 utterance). Thinking-expression-judgment assessment relies on application problems in regular tests and performance tests ('Thinking-Expression-Judgment Assessment', 3 utterances), while knowledge-skills assessment uses textbook-based problems ('Knowledge-Skills Assessment', 1 utterance). Additional assessment activities include Moodle's quiz functions for make-up tests ('Moodle Make-up Tests', 1 utterance).

Each teacher was asked about his expectations and concerns about the system compared with these conventional competency assessment methods. Tables 6 through 9 present the systematic thematic analysis results organized by main themes, showing sub-themes, codes, descriptions, and frequency counts for each category.

Device usage for education

Teachers identified significant barriers to educational device adoption, with multiple challenges spanning learner and practical factors. Student resistance to digital tools emerged as the most prominent concern, with teachers noting that "if paper is available, they think it is definitely better to do it that way." Additional barriers included students' reluctance to engage in self-directed activities, curriculum constraints specific to their preparatory school context, and teacher workload pressures. Teachers mentioned the need for enforcing device use and expanding device-based activities to promote device use.

Table 6

Device usage for education

Type	Factor/ Perspective	Code	Description	Frequency
Challenge	Learner	Student Resistance	Student resistance to using PCs and digital tools	4
		Self-directed Activity Barriers	Learners' reluctance to engage in self-directed activities results in less tool access	2
	Practical	Curriculum Constraints/Char-acteristics	School curriculum constraints/characteristics result in less tool access	3
		Teacher Workload	Teachers' busy schedules preventing daily tool usage in practice	2
Suggestion	Practical	Mandatory Usage	Need for mandatory system usage to promote tool access	2
		Expanding Usage Contexts	Need to expand digital tool usage contexts to promote tool access	1

ICT Environment and Tool Access

Infrastructure limitations present substantial obstacles to system access. Network and server limitations cause significant disruptions, as noted: "When 40 students open tablets simultaneously, about five experience connection failures." System accessibility issues create additional barriers to tool utilization. Teachers mentioned the need to improve accessibility, such as by making the content available as an app or by making it accessible with just one click.

Table 7

ICT environment and tool access

Type	Factor/ Perspective	Code	Description	Frequency
Challenge	Environment	Network/Server Constraints	School network/server constraints result in less tool access	3
	Tool	Low System Accessibility	Low accessibility of digital tools results in less tool access	3
Suggestion	Tool	Improving Accessibility	Need to improve accessibility to promote tool access	1

Learning Support Tools

Teachers expressed positive expectations for learning support tools, particularly appreciating digital material convenience and AI recommendation features. One teacher emphasized that "having access to so many libraries is a tremendous benefit." However, significant challenges emerged regarding system functionality, with system usability problems being the most frequently cited concern. Additional issues included model performance limitations, lack of model explainability, and insufficient intervention effectiveness. Teachers proposed various improvements, including usability enhancements and model performance improvements.

Table 8

Learning support tools

Type	Factor/ Perspective	Code	Description	Frequency
Expectation	Tool	Digital Material Convenience	Digital material convenience as a benefit of tool usage	4
		AI Recommendation Convenience	AI recommendation convenience as a benefit of tool usage	2
Challenge	Tool	Low System Usability	Tools not being used continuously due to low system usability	6
		Low Model Performance	Tools not being used continuously due to low performance of embedded models	2
		Low Model Explainability	Tools not being used continuously due to low explainability of embedded models	1
		Low System Intervention Effectiveness	Tools not being used continuously due to low effectiveness of system intervention	2
Suggestion	Tool	Improving System Usability	Need to improve system usability for continuous tool usage	2
		Improving Model Performance	Need to improve embedded model performance for continuous tool usage	2

Competency Assessment with YINSIGHT

Teachers expressed specific expectations and concerns regarding YINSIGHT for competency assessment. They recognized the potential value of learning situation monitoring, appreciating the system's ability to track students' self-directed learning activities that are typically invisible in traditional assessment methods. However, significant challenges emerged, including concerns about inappropriate data collection due to students not following instructions, difficulty assessing non-uniform learning activities, and low compatibility with current educational practices. One teacher noted, "When it comes to assessment, looking at it as scores, the story changes. Some students keep no records but do lots at cram school and perform well on mock exams." Despite these challenges, teachers provided constructive suggestions for implementation, with preparing appropriate practice forms and environments being the most frequently mentioned recommendation, alongside appropriate usage method guidance and experimental data usage approaches.

Table 9

Competency assessment with YINSIGHT

Type	**Factor/ Perspective**	**Code**	**Description**	**Frequency**
Expectation	Practical	Digital Material Convenience	Digital material convenience as a benefit of tool usage	4
Challenge	Learner	Inappropriate Data Collection	Difficulty in using YINSIGHT for assessment due to inappropriate data collection	3
	Practical	Difficulty Assessing Non-uniform Learning Activities	Difficulty in using YINSIGHT for assessment due to non-uniform student engagement	2
		Low Compatibility with Current Practices	Difficulty of using YINSIGHT for assessment due to low compatibility with current assessment and learning practices	2
Suggestion	Practical	Preparing Appropriate Practice Forms/Environments	Need to prepare appropriate practice forms/environments to promote YINSIGHT usage	9
		Appropriate Usage Method Guidance	Need to guide learners on appropriate tool usage methods when implementing YINSIGHT assessment	1
		Experimental Data Usage	Need to use data experimentally when implementing YINSIGHT assessment	1

These findings, systematically organized across Tables 6 through 9, reveal a complex landscape where teachers appreciate the potential benefits of data-informed assessment tools while recognizing substantial implementation barriers across technical, pedagogical, and institutional dimensions.

Experience and Observation

Table 10 shows the comments of the teacher when using the sharing function, and Figure 9 shows the results of the metric selection for each case. These differences provide insights into the effects of these features on the teacher. First, in statement (u3), while he considered the quiz important to assess SRSs, he gave less weight to the number of quiz attempts than to the other metrics without the sharing function. However, when using the sharing function, they assigned the maximum weight to the metric. Furthermore, in statement (u4), while he said that the selection was not significantly affected, the actual selection changed between the condition with the sharing function and the other one. These findings suggest that the information about others' choices shared using the sharing function provided him with some awareness of the metrics that were actually important to him.

However, as stated in statement (u1), the name of the indicator or the purpose of creating the indicator was not indicated. This implies that, to refer to shared information, it is important that the purpose of the information matches one's own purpose. In the future, it will be necessary to improve the sharing function and share the purpose itself so that teachers can refer to the shared information.

Finally, concerning statements (u2) and (u3), some focused on the differences in subjects and emphasized the importance of the quiz activities that they often use. This suggests that teachers may choose metrics based on their contexts and teaching styles. Therefore, we examined whether these factors actually influenced teachers’ choice of metrics and investigated the possible influences of other factors.

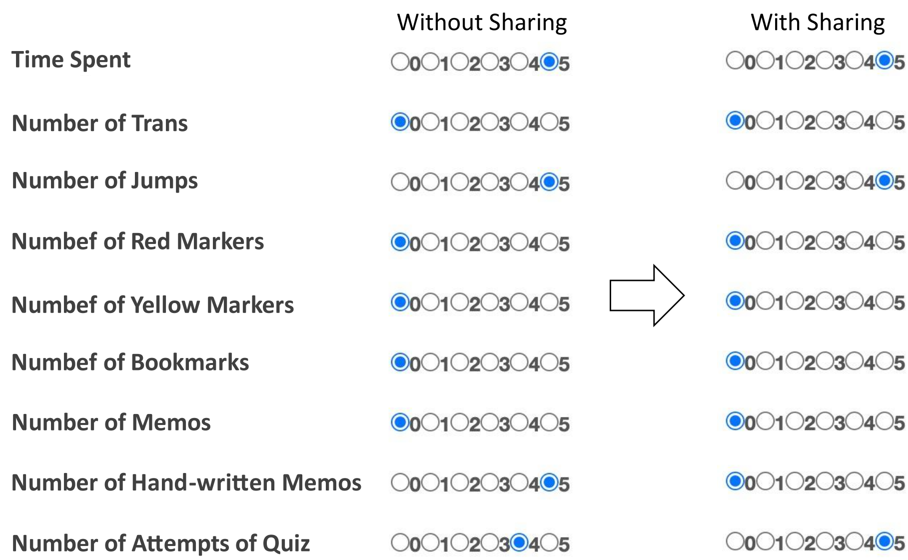
Table 10

Utterances by a teacher when using the sharing function

Summary	Utterance
Improvement points for recommended indicators	(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.
Insights regarding the indicators	(u2) I guess different subjects have different choices.
Insights into important indicators	(u3) Just a quiz would be fine.
Awareness of the effect of sharing	(u4) Even if the others’ selection of metrics is shared, the selection would not be affected very much.

Figure 9

Two patterns of metric selection by the teacher



General Discussion

Answers to the RQs

RQ1: How can the customizable data-informed competency assessment be realized?

In this study, we designed a framework for a data-informed competency assessment. When designing this framework, we considered two perspectives: calculation flow and user flow. The main features of the calculation flow are twofold: the ability to calculate indicators based on user-input variables and the creation of time-series representations for formative assessment and average representations for summative assessment, depending on the aggregation scope. These two features enable users to flexibly customize competency indicators. Next, the main features of user flow are twofold: designing a competency assessment alignment along with a traditional assessment alignment, and assuming iterations during the evaluation design. These two features lower the barriers for users in conducting competency assessments based on data and anticipate autonomous evaluation improvements. Furthermore, a scenario analysis demonstrates the application of this indicator customization framework to the SRL context and its utility. Therefore, designing such an indicator customization framework would enable flexible assessments based on user context. Such an assessment, applicable to various contexts, enhances the granularity of traditional assessments and broadens coverage. Consequently, triangulation with traditional assessments has progressed, contributing to increased validity.

RQ2: How can the customizable data-informed competency assessment be implemented?

In this study, we designed and implemented YINSIGHT, comprising three modules, to demonstrate the implementation of a data-informed competency assessment support system. The system comprises customization, sharing, and visualization modules.

In the customization module, we created a simple and easy-to-use interface for non-technical users to create indicators. In the sharing module, we enabled sharing indicator creation examples by other users as a basis for users unfamiliar with data-informed competency assessment. Finally, in the visualization module, we designed a dashboard where customized indicators can be viewed to give insights for formative and summative assessments based on data. With these three elements, a system supporting user-centered data-informed competency assessment can be implemented.

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

Our thematic analysis revealed complex teacher perspectives on data-informed competency assessment through YINSIGHT. Teachers expressed strong expectations for YINSIGHT's potential to capture self-directed learning activities that traditional assessment methods struggle to evaluate. This aligns with the benefits of using trace data noted in previous studies (Sung et al., 2016; Van der Kleij et al., 2015), which enables precise tracking of what, when, and how much learners have done.

However, significant challenges emerged across multiple dimensions. Implementation barriers included system usability problems, infrastructure limitations, and device adoption challenges. More fundamentally, teachers revealed a conceptual mismatch between their current assessment philosophy and data-informed approaches. Teachers currently evaluate student agency by inferring "behind-the-scenes efforts" from observable outcomes, as one teacher explained: "By looking at the scores from these tests, if a student has achieved a certain score, we allocate points for autonomy based on how much they have been doing behind the scenes". This outcome-focused approach deliberately avoids bias against specific learning methods.

In contrast, teachers expressed concerns about YINSIGHT's limitation to digital learning activities, noting that "the methods and locations for self-directed learning vary from person to person" and that "forcing students to use e-books instead of what they are doing with paper does not offer any benefits to the students". Despite these challenges, teachers provided constructive suggestions for implementation, indicating potential for gradual, context-appropriate deployment.

Finally, regarding the sharing feature, a scaffolding effect on users during indicator creation was observed. However, there are suggestions for improvements in the content of the shared information. Therefore, further improvements were required. In addition, it is essential to consider the utilization of accumulated methods for indicator creation to clarify what constitutes a valid indicator in each context. This requires revealing the similarity and discriminability of indicators from various perspectives, such as the activities, competencies, situations, and evaluation purposes targeted by each indicator, and systematizing indicators suitable for each context. This enhances the effectiveness of indicator sharing and supports more efficient and effective indicator creation.

Implications

The findings from our teacher interviews provide critical insights for the broader learning analytics community regarding the complex relationship between technological capabilities and educational practices. Three key implications emerge for future research and development in data-informed competency assessment systems.

First, successful implementation requires a hierarchical approach to system development that prioritizes foundational elements before advanced analytics features. The prevalence of usability concerns and infrastructure limitations suggests that sophisticated competency assessment capabilities are meaningless without robust, user-friendly underlying systems. Future learning analytics development should therefore follow a staged approach: establishing reliable, intuitive basic functionality before introducing complex analytical features.

Second, our findings reveal a fundamental philosophical divide between outcome-focused and process-focused assessment approaches that extends beyond technical implementation challenges. The participating teachers' deliberate choice to infer student efforts from learning outcomes rather than directly observing learning processes represents a pedagogical strategy designed to maintain assessment equity across diverse learning contexts. This outcome-focused philosophy contrasts sharply with the process-focused nature of trace data analytics, suggesting that data-informed systems should be positioned as complementary rather than replacement tools for traditional assessment methods.

Third, the scope limitation of data-informed assessment to digital learning activities presents significant equity concerns that require careful consideration in system design. Teachers' emphasis on accommodating diverse learning methods highlights the risk that data-informed assessment could inadvertently create systematic bias against students who prefer non-digital approaches or lack consistent technology access (Ramdani et al., 2022). This suggests that future research should focus on developing hybrid assessment frameworks that combine the precision of trace data analytics with the inclusivity of traditional evaluation methods.

These implications point toward a need for learning analytics systems that prioritize adaptability and inclusivity over technological sophistication, ensuring that the benefits of detailed process analytics enhance rather than constrain diverse educational practices.

Limitations and Future Work

Evaluation

One significant limitation of this study is the small sample size of three teachers from a single high school. While this sample was appropriate for the exploratory phase of our design-based research approach and provided valuable insights, it limits the generalizability of our findings. Future work should include larger-scale evaluations with more diverse participant groups across different educational contexts.

Additionally, the inter-coder reliability for our thematic analysis achieved a Fleiss' Kappa value of 0.510. While this demonstrates some consistency in our coding process, there is room for improvement in future studies through further codebook refinement and enhanced communication between coders to achieve higher reliability. Nevertheless, our in-depth interviews with teachers who were actual users of the LEAF system provided rich qualitative data that informed the initial design and potential application of YINSIGHT.

Another important limitation is that our current evaluation focused primarily on teachers' initial impressions and expectations of the system, rather than a thorough assessment of how the trace data-based indicators would actually inform their pedagogical decisions and assessment practices. Future research should include longitudinal studies of teachers actively using the system in their daily practice to evaluate how the data-informed indicators influence assessment activities and educational outcomes. This would address the current disconnect between the system's technical capabilities and its practical educational impact.

Scalability

In this study, the application of the user-centered competency assessment alignment model was limited to SRSs, focusing on the "working on the task" aspect of content learning. Therefore, further research is necessary to explore the potential for expanding this approach to other competencies. There are three directions in which future investigations can proceed: data sources, data granularity, and metric calculation methods.

First, regarding data sources, this study utilized data solely from the digital learning material BookRoll, which restricted the competency assessment to content learning. As such, it is important to explore how assessments might differ when using trace data from other systems, such as CBT and Learning Management Systems, or data recorded by teachers and students. Understanding the evaluation outcomes from these different data sources will be a key area for future research.

Second, concerning data granularity, the metrics in this study were based on aggregated trace data collected in one-hour intervals. However, when using trace data from systems such as CBTs or online surveys to estimate competencies, it may be necessary to work at a more granular level, such as the item level. Therefore, it is essential to investigate the level of data granularity required, depending on the competency being assessed and the data being utilized.

Finally, regarding indicator calculation methods, this study employed standardization, logarithmic transformation, and weighted summation to compute the indicators. However, depending on the evaluation objective, data used, and the competency being assessed, other calculation methods may be more appropriate. Therefore, it is necessary to expand the available calculation methods to align with user needs.

By expanding and investigating the implementation of data sources, data granularity, and metric calculation methods, future research can explore the potential for expanding the proposed user-centered data-informed competency assessment and YINSIGHT to other competencies.

Reliability

During the interviews, it was pointed out that "learners often do not adhere to the activity design." In such cases, the reliability of the assessment indicators may decrease. For instance, dishonest behaviors in the assessment indicators, such as meaningless page navigation or indiscriminate button pressing, may result in extremely high values. One solution to maintain data reliability is to detect and eliminate such behaviors. A method proposed for this purpose is a data-driven approach that detects and excludes abnormal data from the trace data (Alexandron et al., 2017). Implementing such methods is crucial for enhancing the reliability of the assessment indicators.

Validity

To enhance the validity of the assessment indicators, YINSIGHT allows users to customize indicators according to their context. However, for teachers unfamiliar with data interpretation and manipulation, customizing indicators can be challenging. If customization is not performed appropriately, it can lead to a decrease in assessment validity. To address this concern, this study proposes and implements support through the accumulation and sharing of assessment cases among teachers. This is expected to provide a basis for teachers with limited experience in data-informed competency assessments. Furthermore, as the accumulation and sharing of cases progresses, the generalization of assessment methods is expected to occur, revealing more valid assessment methods. However, this support method may face a cold-start problem when there are few accumulated cases to share, and appropriate examples for sharing are lacking. In cases with limited examples, methods for enhancing the validity of teacher-created indicators can be considered by drawing on previous research.

First, the triangulation of evaluation grounds involves assessing competency from three perspectives—data-based assessment, observation-based assessment by teachers, and self-assessment by learners—which is said to increase the validity of the results (Fan et al., 2020; Winne, 2020). Second, sharing methods for creating indicators through a hypothesis-driven approach could be considered (Greene & Azevedo, 2010, Siadaty et al., 2016, Winne & Perry, 2000). If there are no examples created by teachers, referencing experts' theoretically constructed creation methods can enhance construct validity (Messick, 1987). However, this method may suffer from low applicability, as previously mentioned. Finally, support for indicator creation methods using a data-driven approach can be considered (Grover et al., 2017). This involves analyzing other assessment indicators related to the targeted competency and creating indicators from data showing a relationship. However, the validity enhanced through this method is criterion-related. Therefore, the original assessment indicators should have construct validity (Messick, 1987). Introducing this method also requires aligning the procedure such that teachers can conduct data analysis. Combining these approaches is expected to further improve the validity of assessment indicators.

Conclusions

To achieve a user-centered data-informed competency assessment, we created an indicator customization framework and implemented a system called YINSIGHT based on this framework. This system comprises three modules that support data-informed competency assessment. The results of the scenario analysis suggest that the proposed framework has the potential to provide insights into learners' SRL processes in both the formative and summative assessments of SRSs. Interviews with teachers revealed that the system seemed to allow independent learning and detailed efforts that could not have been captured before. Furthermore, when teachers were asked to use the system, the Sharing Module helped them gain awareness regarding indicator selection and that their indicator selection had changed.

Future challenges include examining methods for assessment that consider the diversity of learning and equality in evaluation, as well as investigating what data sources, data granularity, and calculation methods are required when applying the proposed framework and system to competencies other than SRSs. We hope that our proposed framework and system will contribute to increasing the practical use of trace data for competency assessment and provide insights into micro learning to continuously improve learning and teaching using activities using the system from multiple perspectives.

Abbreviations

LAD(s): learning analytics dashboard(s); SRS(s): Self-regulated skill(s); RQ(s): Research question(s); SRL: Self-regulated learning; SSC: Subject-specific competency; GC: Generic competency; CBT: Computer-based testing; PIAAC: Program for the International Assessment of Adult Competencies; OECD: Organization for Economic Co-operation and Development; PSTRE: Problem-solving in technology-rich environments; xAPI: experience API.

Author's contributions

Kano designed the overall research, analyzed data, implemented system, and drafted the manuscript. Horikoshi provided oversight for the entire research project, including supervision of system design, manuscript, and evaluation methods. Koike contributed to the design of proposed frameworks and models and offered advice on manuscripts. Ogata served as the Principal Investigator and supervised the LEAF system, including the system developed in this study. He also offered guidance on research design and system architecture.

Author's information

Taito Kano is a doctor student at the Graduate School of Informatics, Kyoto University, Japan. His research interests include Learning Analytics, especially data-informed competency assessment.

Izumi Horikoshi is a senior researcher at Uchidayoko Institute for Education Research, Uchida Yoko Co., Ltd., Japan. Her research interests include Learning Analytics and classroom visualization for formative assessment and reflection. She is a member of APSCE and SoLAR.

Kento Koike is an Assistant Professor at Faculty of Engineering, Tokyo University of Science. He received a Ph.D. degree in Engineering from Graduate School of Engineering, Tokyo Polytechnic University. His research interests include: Intelligent Tutoring Systems, Cognitive Science and Knowledge Engineering.

Hiroaki Ogata is a professor at the Academic Center for Computing and Media Studies and the Graduate School of Informatics, Kyoto University, Japan. His research interests include Learning Analytics, educational data science, CSCL, CSCW, and CALL.

Funding

This work is partially supported by the following research grants: JSPS KAKENHI JP23H00505, CSTI SIP Program Grant Number JPJ012347, and JST BOOST, Grant Number JPPMJBS2407.

Availability of data and materials

Not applicable.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Kyoto University, Japan

[2] Uchidayoko Institute for Education Research, Uchida Yoko Co., Ltd., Japan

[3] Tokyo University of Science, Japan

[4] Kyoto University, Japan

Received: 29 February 2024 Accepted: 4 September 2025
Published online: 3 March 2026

References

Alexandron, G., Ruipérez-Valiente, J. A., Chen, Z., Muñoz-Merino, P. J., & Pritchard, D. E. (2017). Copying@Scale: Using harvesting accounts for collecting correct answers in a MOOC. *Computers & Education, 108*, 96–114. https://doi.org/10.1016/j.compedu.2017.01.015

Alozie, N. M., Dhamija, S., McBride, E., & Tamrakar, A. (2020). Automated collaboration assessment using behavioral analytics. In M. Gresalfi & I. S. Horn (Eds.), *Proceedings of the International Society of the Learning Sciences* (pp. 1071–1078). International Society of the Learning Sciences. https://repository.isls.org/handle/1/6296

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Bennett, K. B., Davis, C., & Weddel, K. S. (2010). *Learner centered instruction: An independent study course for adult education and family literacy teachers*. Northern Colorado Professional Development Center (NCPDC).

Bacquet, J. N. (2020). Implications of summative and formative assessment in Japan: A review of the current literature. *International Journal of Education and Literacy Studies, 8*(2), 28–35.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Sage.

Chapman, E., & O'Neill, M. (2010). Defining and assessing generic competencies in Australian universities: Ongoing challenges. *Education Research and Perspectives, 37*(1), 105–123. https://search.informit.org/doi/10.3316/ielapa.086552707509215

Chung, R. G., & Lo, C. L. (2007). The development of teamwork competence questionnaire: Using students of business administration department as an example. *International Journal of Technology and Engineering Education, 55*–57.

Dawson, S., Joksimović, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (LAK ’19) (pp. 446–455). https://doi.org/10.1145/3303772.3303784

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. https://doi.org/10.1111/cts.12030

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. https://doi.org/10.18608/jla.2019.62.2

Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*, 83–96. https://doi.org/10.1016/j.compedu.2016.12.001

Fan, Y., Van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., … Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition & Learning, 17*(3), 949–987. https://doi.org/10.1007/s11409-022-09291-1

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80–92. https://doi.org/10.1177/160940690600500107

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal, 10*(4), 469-484. https://doi.org/10.34105/j.kmel.2018.10.029

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of Research on Student Engagement* (pp. 763–782). Springer. https://doi.org/10.1007/978-1-4614-2018-7_37

Garrison, C., & Ehringhaus, M. (2009). *Formative and summative assessment in the classroom*. Westerville, OH: AMLE. https://www.amle.org/BrowsebyTopic/WhatsNew/WNDet/TabId/270/ArtMID/888/ArticleID/286/Formative-and-Summative-Assessments-in-the-Classroom.aspx

González, J., & Wagenaar, R. (Eds.). (2003). *Tuning educational structures in Europe: Universities’ contribution to the Bologna Process* [Final Report – Pilot Project Phase 1]. Universidad de Deusto Press.

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist, 50*(1), 14–30. https://doi.org/10.1080/00461520.2014.989230

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. https://doi.org/10.1080/00461520.2010.515935

Grover, S., Basu, S., Bienkowski, M., Eagle, M., Diana, N., & Stamper, J. (2017). A framework for using hypothesis-driven approaches to support data-driven learning analytics in measuring computational thinking in block-based programming environments. *ACM Transactions on Computing Education (TOCE), 17*(3), 1–25. https://doi.org/10.1145/3105910

Gunderman, R. B., Williamson, K. B., Frank, M., Heitkamp, D. E., & Kipfer, H. D. (2003). Learner-centered education. *Radiology, 227*(1), 15–17. https://doi.org/10.1148/radiol.2271021124

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), e0137446. https://doi.org/10.1371/journal.pone.0137446

Ibrahim, M. Y., Noh, C. H. C., & Abas, M. M. (2017). Confirmatory factor analysis of interpersonal skills of youth’s scale. *International Journal of Academic Research in Business and Social Sciences, 7*(7), 752–762. http://dx.doi.org/10.6007/IJARBSS/v7-i7/3136

Jiao, H., Liao, D., Zhan, P. (2019). Utilizing process data for cognitive diagnosis. In: von Davier, M., Lee, YS. (Eds.), *Handbook of Diagnostic Classification Models: Models and model extensions, applications, software packages* (pp. 421–436). Springer. https://doi.org/10.1007/978-3-030-05584-4_20

Kaliisa, R., & Dolonen, J. A. (2023). CADA: A teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. *Technology, Knowledge and Learning, 28*(3), 937–958. https://doi.org/10.1007/s10758-022-09598-7

Koenen, A., Dochy, F., & Berghmans, I. (2015). A phenomenographic analysis of the implementation of competence-based education in higher education. *Teaching and Teacher Education, 50*, 1–12. https://doi.org/10.1016/j.tate.2015.04.001

Lechner, C., Knopf, T., Napolitano, C., Rammstedt, B., Roberts, B., Soto, C. J., & Spengler, M. (2022). The behavioral, emotional, and social skills inventory (BESSI): Psychometric properties of a German-language adaptation, temporal stabilities of the skills, and associations with personality and intelligence. *Journal of Intelligence, 10*(1), 1–27. https://doi.org/10.3390/jintelligence10030063

Le Deist, F. D., & Winterton, J. (2005). What is competence? *Human Resource Development International, 8*(1), 27–46. https://doi.org/10.1080/1367886042000338227

Matcha, W., Gašević, D., & Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802

McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *The American Psychologist, 28*(1), 1–14. https://doi.org/10.1037/h0034092

Messick, S. (1987). Validity. *ETS Research Report Series, 1987*(2), i–208. https://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2017). *Junior high school course of study*. https://www.mext.go.jp/content/20201008-mxt_kyoiku02-000005242_1.pdf

Muslim, A., Chatti, M. A., Mahapatra, T., & Schroeder, U. (2016). A rule-based indicator definition tool for personalized learning analytics. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (pp. 264–273). ACM. https://doi.org/10.1145/2883851.2883921

Muslim, A., Chatti, M. A., & Schroeder, U. (2022). Supporting indicator personalization and platform extensibility in open learning analytics. *Technology, Knowledge and Learning, 27*(2), 429–448. https://doi.org/10.1007/s10758-021-09543-0

Mutlu, B., Simic, I., Cicchinelli, A., Sabol, V., & Veas, E. (2018). Towards a learning dashboard for community visualization. In *Proceedings of the 1st Workshop on Analytics for Everyday Learning co-located with the 13th European Conference on Technology Enhanced Learning (EC-TEL 2018)*, Leeds, UK. CEUR Workshop Proceedings.

Nederhand, M. L., Tabbers, H. K., De Bruin, A. B., & Rikers, R. M. (2021). Metacognitive awareness as measured by second-order judgements among university and secondary school students. *Metacognition and Learning, 16*(1), 1–14. https://doi.org/10.1007/s11409-020-09228-6

Ogata, H., Majumdar, R., Yang, S. J. H., & Warriem, J. M. (2022). LEAF (Learning & Evidence Analytics Framework): Research and practice in international collaboration. *Information and Technology in Education and Learning, 19*(41). https://doi.org/10.12937/itel.2.1.Inv.p001

Page, C., & Wilson, M. (1994). *Management competencies in New Zealand: On the inside looking in*. Ministry of Commerce.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Parry, S. B. (1996). The quest for competencies: Competency studies can help you make HR decisions, but the results are only as good as the study. *Training*, *33*(7), 48–56.

Pérez-Berenguer, D., Kessler, M., & García-Molina, J. (2020). A customizable and incremental processing approach for learning analytics. *IEEE Access*, *8*, 36350–36362. https://doi.org/10.1109/ACCESS.2020.2975384

Puspita, R. R., & Suyatno, S. (2020). Teacher observation assessment for primary education. *Eternal (English Teaching Journal)*, *11*(2), 395–355. https://doi.org/10.26877/eternal.v11i2.7565

Ramdani, Z., Amri, A., Hadiana, D., Warsihna, J., Anas, Z., & Susanti, S. (2022). Students diversity and the implementation of adaptive learning and assessment. In *Proceedings of the Interdisciplinary Conference of Psychology, Health, and Social Science (ICPHS 2021)* (pp. 157–161). Atlantis Press. https://doi.org/10.2991/assehr.k.220203.025

Renninger, K. A., & Bachrach, J. E. (2015). Studying triggers for interest and engagement using observational methods. *Educational Psychologist*, *50*(1), 58–69. https://doi.org/10.1080/00461520.2014.999920

Roberts, K., Dowell, A., & Nie, J. B. (2019). Attempting rigour and replicability in thematic analysis of qualitative research data: A case study of codebook development. *BMC Medical Research Methodology*, *19*(1), Article 66. https://doi.org/10.1186/s12874-019-0707-y

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology*, *11*(1), 25–41. https://doi.org/10.1080/14780887.2013.801543

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. John Wiley & Sons.

Siadaty, M., Gašević, D., & Hatala, M. (2016). Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics*, *3*(1), 183–214. https://doi.org/10.18608/jla.2016.31.11

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. https://doi.org/10.1016/j.compedu.2015.11.008

Thippayacharoen, T., Hoofd, C., Pala, N., Sameephet, B., & Satthamnuwong, B. (2023). Assessing language or content? A systematic review of assessment in English medium instruction classrooms in different contexts. *LEARN Journal: Language Education and Acquisition Research Network*, *16*(2), 548–571. retrieved from https://so04.tci-thaijo.org/index.php/LEARN/article/view/266971

Turner, J. C., & Meyer, D. K. (2000). Studying and understanding the instructional contexts of classrooms: Using our past to forge our future. *Educational Psychologist*, *35*(2), 69–85. https://doi.org/10.1207/S15326985EP3502_2

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. H. M. (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. https://doi.org/10.3102/0034654314564881

Van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, *68*(2), 251–265. https://doi.org/10.1007/BF02294800

Vazirani, N. (2010). Competencies and competency model—A brief overview of its development and application. *SIES Journal of Management*, *7*(1), 121–131.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23. https://doi.org/10.1007/BF02504682

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271. https://doi.org/10.1016/j.concog.2013.01.004

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Academic Press. https://doi.org/10.1016/B978-012109890-2/50045-7

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior*, *112*, Article 106457. https://doi.org/10.1016/j.chb.2020.106457

Wise, A. F., & Vytasek, J. M. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (1st ed., pp. 151–160). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.013

Wong, S. C. (2020). Competency definitions, development and assessment: A brief review. *International Journal of Academic Research in Progressive Education and Development*, *9*(3), 95–114. http://dx.doi.org/10.6007/IJARPED/v9-i3/8223

Zhang, S., Wang, Z., Qi, J., Liu, J., & Ying, Z. (2023). Accurate assessment via process data. *Psychometrika*, *88*(1), 76–97. https://doi.org/10.1007/s11336-022-09880-8

Zhidkikh, D., Saarela, M., & Kärkkäinen, T. (2023). Measuring self-regulated learning in a junior high school mathematics classroom: Combining aptitude and event measures in digital learning materials. *Journal of Computer Assisted Learning*, *39*(6), 1834–1851. https://doi.org/10.1111/jcal.12842

Zhu, H., Jiao, H., Gao, W., & Meng, X. (2023). Bayesian change-point analysis approach to detecting aberrant test-taking behavior using response times. *Journal of Educational and Behavioral Statistics*, *48*(4), 490–520. https://doi.org/10.3102/10769986231151961

Publisher's Note

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.