Kuromiya, H., Majumdar, R., & Ogata, H. (2020). Fostering Evidence-Based Education with Learning Analytics: Capturing Teaching-Learning Cases from Log Data. *Educational Technology & Society*, *23* (4), 14–29.

Fostering Evidence-Based Education with Learning Analytics: Capturing Teaching-Learning Cases from Log Data

Hiroyuki Kuromiya[1*], Rwitajit Majumdar[2] and Hiroaki Ogata[2]
[1]Graduate School of Informatics, Kyoto University, Japan // [2]Academic Center for Computing and Media Studies, Kyoto University, Japan // khiroyuki1993@gmail.com // dr.rwito@gmail.com // hiroaki.ogata@gmail.com
[*]Corresponding author

(Submitted March 26, 2020; Revised April 23, 2020; Accepted August 27, 2020)

ABSTRACT: Evidence-based education has become more relevant in the current technology-enhanced teaching-learning era. This paper introduces how Educational BIG data has the potential to generate such evidence. As evidence-based education traditionally hooks on the meta-analysis of the literature, so there are existing platforms that support manual input of evidence as structured information. However, such platforms often focus on researchers as end-users and its design is not aligned to the practitioners' workflow. In our work, we propose a technology-mediated process of capturing teaching-learning cases (TLCs) using a learning analytics framework. Each case is primarily a single data point regarding the result of an intervention and multiple such cases would generate an evidence of intervention effectiveness. To capture TLCs in our current context, our system automatically conducts statistical modelling of learning logs captured from Learning Management Systems (LMS) and an e-book reader. Indicators from those learning logs are evaluated by the Linear Mixed Effects model to compute whether an intervention had a positive learning effect. We present two case studies to illustrate our approach of extracting case effectiveness from two different learning contexts – one at a junior-high math class where email messages were sent as intervention and another in a blended learning context in a higher education physics class where an active learning strategy was implemented. Our novelty lies in the proposed automated approach of data aggregation, analysis, and case storing using a Learning Analytics framework for supporting evidence-based practice more accessible for practitioners.

Keywords: Learning analytics, Evidence-based education, Technology-enhanced Evidence-based Education & Learning (TEEL), Learning Evidence Analytics Framework (LEAF), Mixed effects model, Teaching-learning case

1. Introduction

Evidence of good practices in education has been getting prominent around the world (OECD, 2007; European Commission/EACEA/Eurydice, 2017). Governments have developed many applications to store and retrieve evidence in education. For example, What Works Clearinghouse (WWC) by U.S. Department of Education or Evidence Library by UK EEPI Centre were developed to review existing educational researches focusing on the results from a high-quality study. However, since they are often just a database of evidence, users are not supposed to register evidence from their own experience. What users can do is only to search for evidence which is already registered by researchers or educational offices. In some cases, users can register evidence from a form, but users have to input evidence manually, which takes time for users. Currently, practitioners tend to be kept away from the evidence generation process.

To solve these problems, we focus on Learning Analytics approaches. Learning Analytics (LA) is the field which handles educational big data for improving teaching and learning activities. LA has the potential to investigate the impacts of different learning strategies or systems by using students' behavioural logs (Hwang, Spikol, & Li, 2018). In this paper, we propose a LA platform for supporting the evidence generation process for practitioners. Here are the main contributions of the article.

- We propose an automatic case extraction process on a learning analytics platform. Each case is just a description of an intervention, but gathering these cases, they will be evidence of the intervention.
- We show that mixed effects model is useful for case extraction and we develop an interface in our learning dashboard to visualize the results to the users.

In the field of LA, researchers often investigate the evidence of the effectiveness of LA. However, the agenda of how technology can support the process of evidence extraction from big data and building decision support technology is not focused. Hence, we engaged in supporting evidence-based education by using educational big data collected from different learning systems.

2. Related works

2.1. Learning analytics platform to support evidence-based education

To the best of our knowledge, few studies investigate evidence-based practice from a practitioners' perspective. One example is the Learning Analytics Community Exchange (LACE) project's Evidence Hub (Ferguson & Clow, 2016). LACE Evidence Hub followed the evidence-based medicine paradigm to synthesize published LA literature and meta-analyze four propositions about learning analytics - improving learning outcomes, improving learning support and teaching, deployment at scale, and ethics. Evidence Hub collected evidence that will enable the learning analytics community to assess the effectiveness and relative desirability of outcomes resulting from the use of learning analytics tools and techniques. So, the purpose is to collect evidence of the effectiveness of learning analytics practices.

Analytics4Action Evaluation Framework (Rienties et al., 2016) developed at Open University UK is another example. It provides an evaluation flow of the intervention in the context of distance learning education. They map the six key steps in the evidence-based intervention process; (1) determine key metrics while working together with the key stakeholders, (2) decide on a type of intervention, (3) plan an experimental design, (4) analyze outcomes, (5) store evidence into evidence hub, (6) compare different interventions. Through the cycles of these six steps, one illustrates which types of interventions under which conditions have a positive effect and store the results in to OU Evidence Hub. OU Evidence Hub is a branch of LACE Evidence Hub project. In practice, Rienties, Cross, and Zdrahal (2017) investigated the effectiveness of emails and predictive modelling to learners at-risk through the system. In that paper, they examined how evidence-based research design can be implemented to actual learning analytics intervention.

Table 1 summarizes the features of current learning analytics approaches for evidence-based education. Compared to these two studies, our approach is unique in the following points.

- We take evidence as the accumulation of cases. Each case is not itself evidence, but we can gather cases to evaluate the strength of an evidence.
- Our system includes an automated statistical inference step in case extraction while other studies do not automate statistical inference.

We believe these two points are essential for democratizing evidence-based education with the practitioner by harnessing learning analytics tools and techniques.

Table 1. Feature of current learning analytics approach for evidence-based education

	LACE Evidence Hub (Ferguson & Clow, 2016)	Analytics4Action (Rienties et al., 2016)	Proposed system
Evidence from what?	Literature	Designed Experiment	Case Study (Teaching-Learning Case)
How to support Evidence Generation?	Aggregate existing papers	Suggest the evaluation flow of the intervention	Apply statistical models to the data
Statistical calculation	-	Manual	Automated

2.2. Case extraction models

In our system, we aim to estimate the effectiveness of an intervention by automatic adoption of statistical models. At this stage of development, we consider a variety of models that is able to handle the effectiveness of an intervention. We prepared three candidates for case extraction - classical testing method, time-series model, and mixed effects model. Here, we review each statistical model and examine the suitability to our system.

The first candidate is a group comparison approach. As a representative of it, we consider t-test as a popular pre-post comparison method. Since it is widely used in the learning analytics field, it is easy to interpret for users. However, there are two disadvantages for our system. First is the information loss by aggregation of the data. Since it does not deal with repeated measurement, we should gather the measurement points by person. The second is that t-test has a problem with multiple comparisons where the power of the test decreases if we adopt it many times. As our system allows users to repeat the analysis, it will be a problem in our context.

The second candidate is a time-series model. Here, we consider Interrupted Time Series (ITS) model as a representative of it. Mathematically ITS is the segmented regression model with dummy variables representing the period of the intervention (Bernal, Cummins, & Gasparrini, 2017). In an educational context, Hansen et al. (2014) used ITS model to estimate the impact of a central educational program in German. ITS is a good estimation of the data because learning logs has a time-series structure in nature. However, ITS model compares an intervention period with a baseline, and not necessarily a control period. In ITS, it is difficult to compare two different periods - intervention and control.

The last candidate is a mixed effects model. Mixed effects model is a statistical model that includes a combination of fixed effects and random effects, which represents overall tendency (fixed effects) and individual difference (random effects) for each. It is also used for effectiveness estimation in educational context (Dawson, Jovanovic, Gašević, & Pardo, 2017). For automated extraction of cases from a learning log, such an approach is rational as standard learning logs in any e-learning environment often contain repeated measurement of students. Since it is not a time-series model, it cannot consider the time-series of data. However, instead, mixed effects model can consider students' individual differences. More importantly, it is able to handle two different time periods - intervention and control period and can compare the measurements over those periods.

So far, we discussed the pros and cons of each model. We summarize our discussion in Table 2. According to the table, we concluded that mixed effects model is most appropriate for our purpose. Hence, we adopt a mixed effects model in our current proposed system.

Table 2. Feature of statistical models for case extraction

	Group Comparison Models	Time-series models	Mixed effects model
Example	*t*-test	Interrupted Time-Series	Mixed Effects Model
Advantage	Easy to interpret	Able to capture the time series nature of data	Able to consider students' individual difference
Disadvantage	Information loss by aggregation	Difficult to compare two periods	Cannot handle the time-series behavior in data

3. Our solution: Case extraction system on learning analytics platform

3.1. LA system overview

We had proposed an evidence extraction system integrated into the learning analytics platform as a solution to extract and store evidence from practice (Majumdar, Akçapınar, Akçapınar, Flanagan, & Ogata, 2019; see Figure 1). Teachers can use LEAF in the context of face-to-face learning in higher and secondary education. It has a learning management system *moodle* (see https://moodle.org/), e-book reader *BookRoll* (Ogata et al., 2015), and the dashboard called LAViEW. BookRoll is a learning tracker as well as e-book reader facilitating many learning analytics studies related to the seamless learning environment (Mouri, Uosaki & Ogata, 2018). In our context, teachers conduct an intervention and control method, using these learning tools in their class. We collect data from the Learning Record Store (LRS) for each period – intervention and control. Students' learning log stored in LRS is represented as xAPI format. The accumulated learning log will be an indicator to measure the effect of the intervention in our system.

In LEAF, we have a module which captures teaching-learning activities as a case to generate evidence from it. We call it Teaching-Learning Case (TLC), which is a description of an intervention, which is captured from a single teaching-learning scenario. In LEAF, TLC has a specific format - Context, Problem, Indicator, Intervention, Control, and Results) structure as described in Table 3. This format is modified based on PICO structure, which is commonly used in medicine for making research questions and literature review (Jacobs, 2008). Population, Intervention, Control, and Outcome of Interest corresponds to Context, Solution, Baseline, and Indicator respectively. In addition to PICO format, we added "Problem" and "Results" fields in order to arrange and gather records to make evidence. Although there are other formats of evidence - SPICE framework (Booth, 2006) and PECO format (Morgan, Whaley, Thayer, & Schünemann, 2018), our structure of TLC shares many properties with other evidence formats.

TLC itself is not strong evidence of the intervention because it does not have any strict experimental designs like Randomized-Control Study (RCT). However, by gathering many Teaching-Learning Cases (TLCs), we can generate evidence in the future.

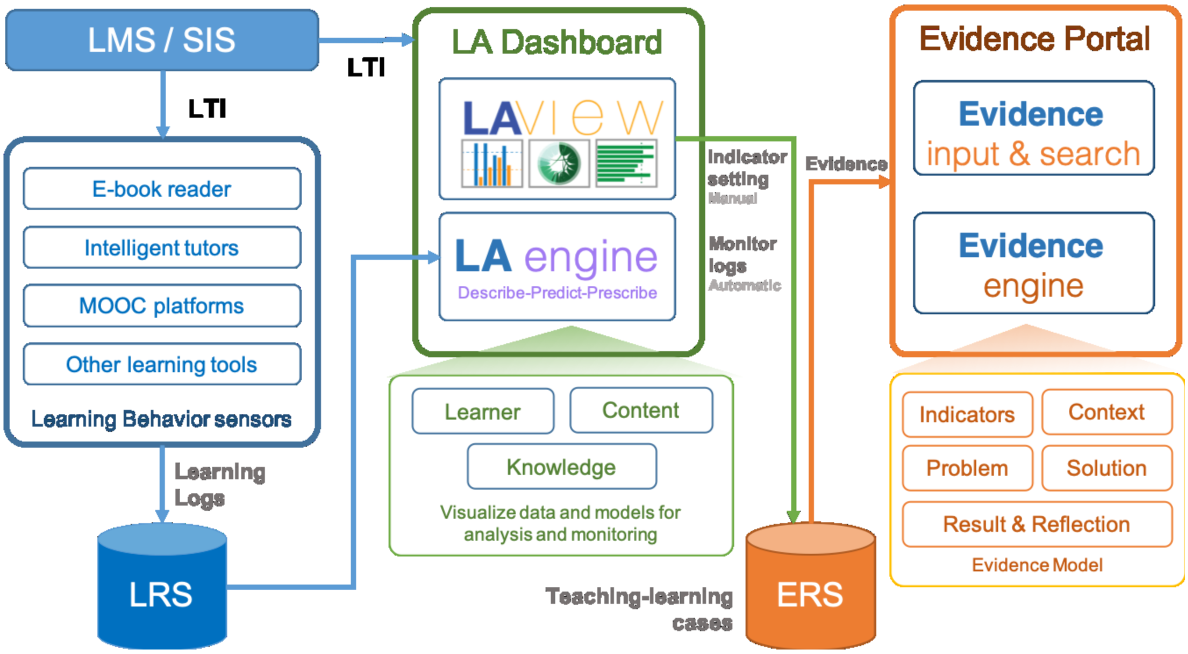

Figure 1. Evidence extraction system on learning analytics platform (Majumdar et al., 2019)

Table 3. The Structure of Teaching-Learning Case in LEAF

Factor	Description	Example
Context course name: subject: grade: class size:	Information regarding the context of evidence. All the information is automatically retrieved from LMS.	course name: Linear Algebra I, subject: Math, grade: B1, class size: 120
Problem	The problems to be addressed in the classroom	Low engagement to homework materials
Indicator	Measurable indicators that a user wants to look at.	Reading time on the materials
Intervention title: dates	The details of the intervention conducted by a teacher. Users need to specify the date of the intervention	title: Send email to students dates: “2019-05-01,” “2019-05-08,” ...
Control title: dates	The details of the control method that a teacher wants to compare with the intervention. The format is the same as the intervention.	title: In-class reminder dates: “2019-04-05,” “2019-04-12,” ...
Results	Description of analysis results by a user. In proposed system, the analysis results are given by the system.	Email intervention increased the reading time by 5.5 min (p = .01)

In this paper, we propose a case extraction module on LEAF. We aim to capture Teaching-Learning Cases from teaching-learning logs stored in Learning Record Store. The characteristics of our system are three listed below.

- Users can estimate the effectiveness of their intervention by just clicking the tool integrated in the learning analytics platform.
- The system provides a database integrated to learning analytics dashboards where users can store and share the effectiveness of their intervention with actual learning context.
- Users can use our system in either online learning environments or face-to-face classroom teaching.

Next, we describe a workflow of capturing TLC in actual teaching-learning context. The statistical aspect of capturing TLC is explained in the last section of this part.

3.2. Workflow of capturing TLC with LEAF

Consider a teacher who teaches their classroom with our learning tools on LEAF. When the teacher conducts a specific intervention, and wants to know the effectiveness of it, our TLC extraction module allows users to register evidence by three steps using the power of educational big data stored in LRS.

The first step is to fill the analysis setting panel (Figure 2A). In this step, users specify the information needed to analyze learning data to conduct the statistical analysis. The course information is automatically retrieved by LTI information, so users do not have to choose. The “Intervention” and “Control” fields are composed by two input fields - name and dates. Users are supposed to write the content of their intervention in name and specify when they conducted the intervention by date picker dialog. Intervention period can be either before or after the control period because we do not consider the order of intervention and control. For source and indicator fields, users can choose the best metric available that fits to measure their problem. Currently, two sources are available - moodle and BookRoll. Considering Moodle as a data source, there are five indicators that can be choose - number of access to moodle contents, moodle pages, moodle quizzes, moodle forum, and external resources. Considering BookRoll as the source, five indicators are available - reading time, number of markers, number of memos, number of bookmarks and number of access to the contents. Once users fill the five fields, push the “Analyze” button to conduct statistical analysis based on the learning log stored in LRS.

The second step is to interpret the results from the system (see Figure 2B and Figure 2C). The overview panel (Figure 2B) represents the inferential statistics estimated by mixed effects model. Overview panel consists of two pieces of information - the effectiveness of the intervention and the strength (reliability) of the case. The effectiveness is retrieved from the coefficients of mixed effects model and the strength of the evidence is determined based on the *p*-value of the coefficient. We refer to the previous studies about the thresholds for p-value interpretation (Raiola & Di Tore, 2012) to categorize p-values into sentences. The strength of the evidence is determined by the following criteria (Table 4). In the example in Figure 2, users can see that the intervention increased the reading time on e-book readers by 11.42 minutes and the results are highly reliable because the strength of the case is “Very Strong.” The data panel (Figure 2C) represents descriptive statistics. It compares average reading time over intervention dates and control dates.

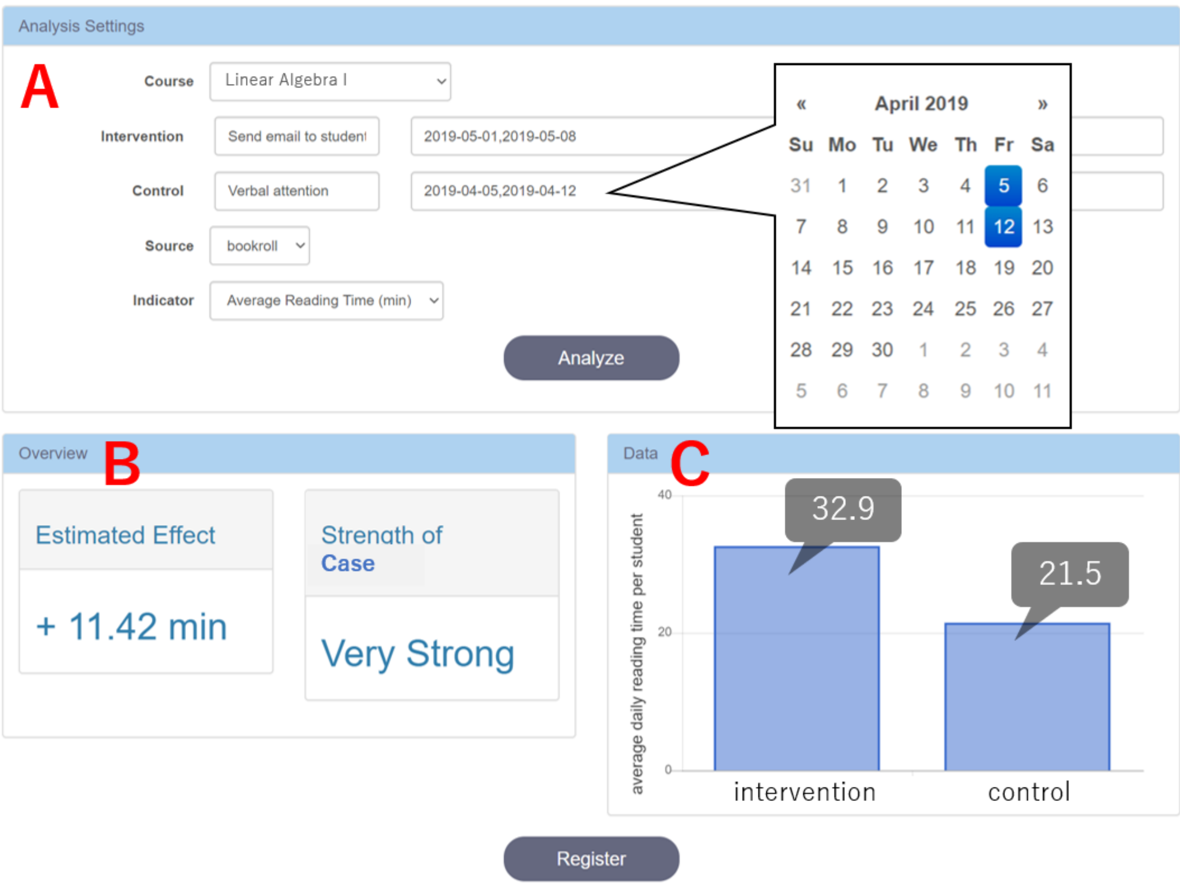

Figure 2. Case extraction page in evidence portal

Once the user interprets the results, users are encouraged to save the results of their intervention in the Evidence Record Store (Figure 3). It has seven input fields – Institution, Course name, Problem, Intervention, Control, Indicator, and Results. It seems to be tough to fill all the fields by manual. However, the fields users have to fill is just one - Problem. Other fields are automatically filled by the system based on the results in the Case Extraction Page. In this example, the population would save the course name "Example Course" from Moodle, Intervention and Control are filled as "Send email to students" and "Teacher-Centered Approach" for each. Indicator is filled as "average reading time on BookRoll." Results section are written as "estimated effect: 11.42 (p = .00)" because the results from the statistical model are retrieved from the system as well. Users can store additional information such as "subject" or "intervention details" by clicking "Details" to expand the form.

Table 4. Conventional Thresholds for *p*-value interpretation (Raiola & Di Tore, 2012)

p-value	Interpretation	Strength of case
$p > .1$	Absence of evidence against the null hypothesis	Null
$.05 \leq p < .1$	Low evidence against the null hypothesis	Low
$.01 \leq p < .05$	Moderate evidence against the null hypothesis	Moderate
$.001 \leq p < .01$	Strong evidence against the null hypothesis	Strong
$p < .001$	Very strong evidence against the null hypothesis	Very Strong

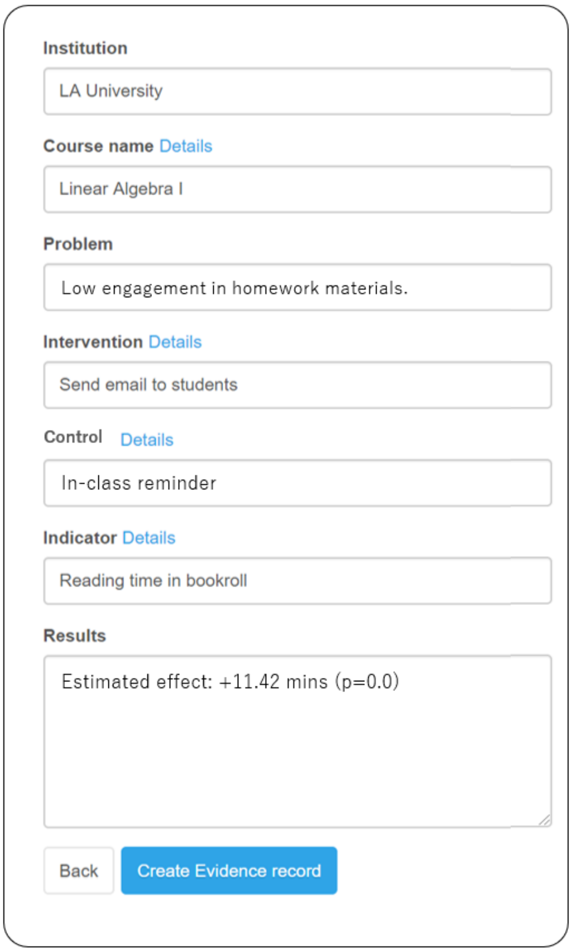

Figure 3. Evidence registration form

Users can see all evidence registered in the Teaching-Learning Cases Table (see Figure 4). It lists every Teaching-Learning Cases with course name, intervention, control, indicator, and results. Users can filter or sort the TLC by subject, grade, or class size. The user will find the cases from similar contexts that match problems that the user has. Users can edit their registered TLC if they get additional results regarding the intervention.

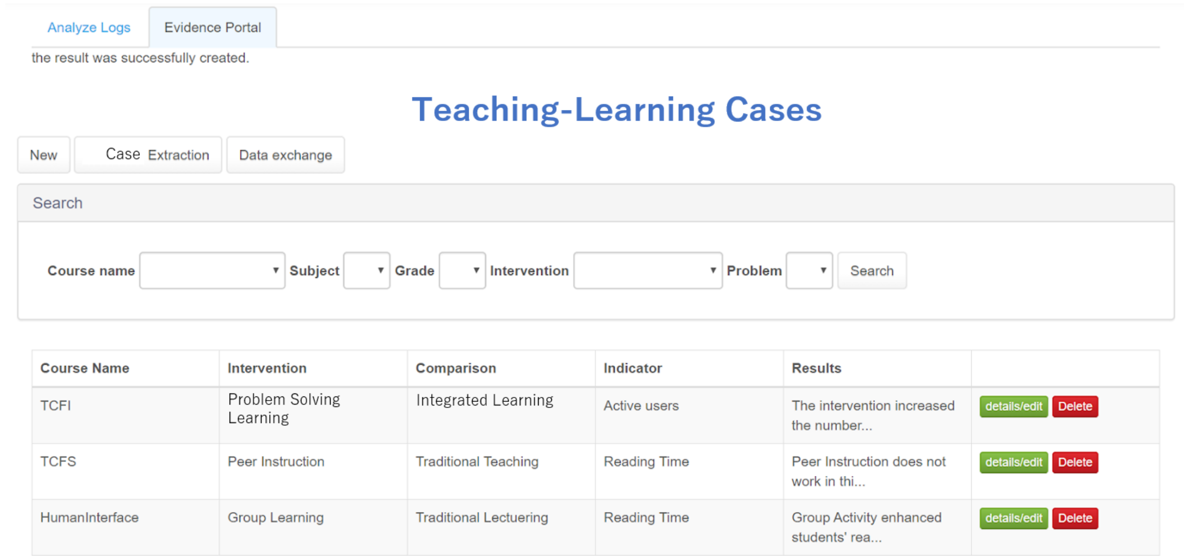

Figure 4. TLC record table

3.4. Evidence extraction process

So far, we have described users' perspective of case extraction. Here we introduce the technical part of our evidence extraction system. It has a four-step process as described in Figure 5. Since we have already introduced result visualization, and case storing step in the previous section, here we describe the details of the first two steps - data collection and modeling and testing step.

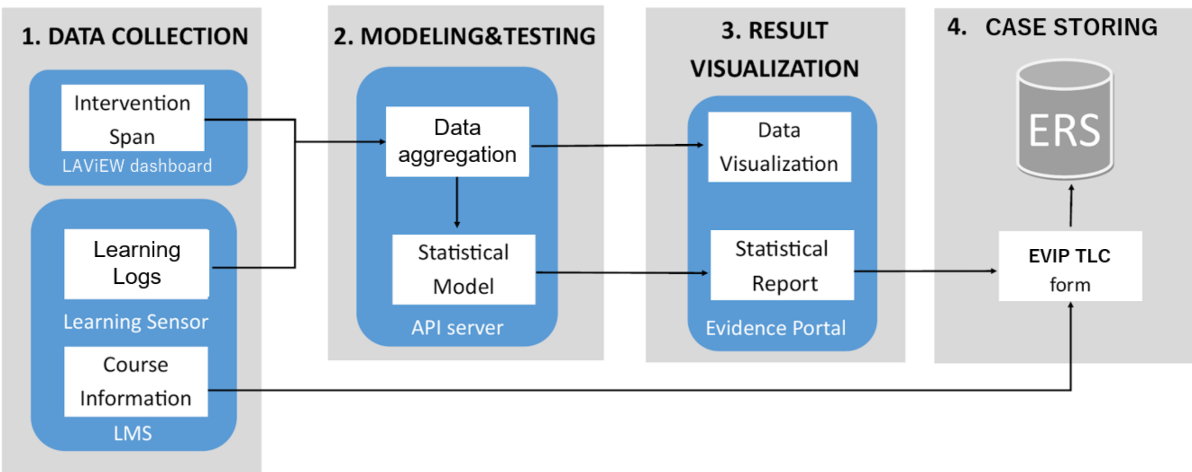

Figure 5. Case extraction process on LEAF

Table 5. Aggregated data structure in data collection step

User_id (i)	Date (j)	Intervention (X)	Indicator (y)
221	2019-02-01	0	5.5
221	2019-02-02	0	6.4
...	...	...	...
221	2019-03-01	1	8.4
221	2019-03-02	1	6.9
225	2019-02-01	0	4.1

The first step is the Data *Collection* step where learning records from LRS related to LMS activities or other learning behavior sensor logs are pre-processed in JSON format. The system aggregates the students' log data by each student and day. For example, the following data frame is generated by the system (Table 5). Here, "user_id" represents user identification, "date" denotes for the day of the measurement, "intervention" represents whether it is in the intervention period (1) or control period (0), and "y" denotes the value of outcome of the interest.

The second step is *Modeling and Testing* step, where the system computes statistical inferences about the effectiveness of the intervention based on the collected data. We implemented these computation functions as an external API for evidence extraction. This separation of the application server with evidence extraction server allows expandability of statistical computing. In our current implementation, we use a Python statistical modeling package "statsmodels" (Seabold & Josef, 2010). The specific model parameters used for evidence extraction is described as follows. Currently we have applied a Linear Mixed Effects Model. It fits the hierarchical structure data as represented like above. The modeling approach is described as follows:

$$y_{i,j} = \beta_0 + \beta_1 X_{i,j} + \gamma_{0,i} + \gamma_{1,i} X_{i,j} + \varepsilon_{i,j}$$

where $y_{i,j}$ stands for value of the indicator (the outcome of interest) of student i at date j. $X_{i,j}$ is the dummy variable for the intervention. If the intervention was conducted at time j, $X_{i,j}$ is 1 otherwise 0. β_0 and β_1 represent fixed coefficients ("global parameter") of the model, while $\gamma_{0,i}$ and $\gamma_{1,i}$ do random effects ("local parameter") of the model. At last, $\varepsilon_{i,j}$ is the error term which follows normal distribution. In this case, the fixed effects represent the overall tendency while the random effects show the characteristics of each student in the sample. Since we want to know the overall reaction of students against the intervention, we particularly pay attention to β_1. If β_1 is positive, it means the intervention has a positive effect on the indicator. If it is negative, the intervention has a negative effect on the indicator. At the statistical *Modeling and Testing* step, these coefficients and their p-value were estimated by the EM algorithms (Lindstrom & Bates, 1988).

4. Research study: Use case analysis

Based on our proposed solution in the previous section, we studied the usefulness of the system when actually implemented in a live context. We choose two case studies with two different types of data sources on LEAF. In one case (Case A) it has the learning log from an e-book reader, and in the other case (Case B) it has the log data from LMS. The two cases also focus on classrooms in two different countries and for different educational contexts. Thus, the two cases provide context variation to demonstrate to our current reader the LEAF system's application. As we consider the log data, we highlight the use case to extract effectiveness in terms of increased engagement as measured by the interaction logs.

In each case, the teacher conducted an intervention, and log-data was collected during the experiment through LEAF system. After the intervention, we used the system, which applied the linear mixed effects model and saved our interpretation as the results. Of the two cases presented, Case A checks the effectiveness of an e-mail intervention in junior high school where students were reminded about preview and review activities in an e-book reader. The data source here was the students' reading logs retrieved from the e-book reader. Case B evaluates the effectiveness of an active learning strategy, Peer Instruction, in an undergraduate physics class. The data source was Moodle logs. Our system could be used to capture the effectiveness of Peer Instruction in comparison to traditional teaching in that context.

4.1. Case A: Extracting effectiveness of E-mail reminders to boost engagement of Junior-High School Students

Our first case concerns a math class at a Junior-high school in Kyoto, Japan. We targeted 60 first-grade students and conducted the intervention to enhance students' learning engagement. The teacher sent e-mail messages to the students during the intervention period. The intervention was conducted on 4, 5, 6, and 9 December 2019. The control period (no e-mail messages were sent) was 8, 9, 10, 14 January 2020. The course topics in the experimental period were "data utilization" in statistics and "polynomial" in algebra. The reason we selected these topics was that these topics were taught by traditional lecture format which does not include much of group works, or interactive sessions. The teacher used the e-book reader in the class to upload the learning materials and create quizzes within that material. Students can then annotate with a maker and write memos while reading and attempt the set quizzes within the same environment. In this case, the teacher uploaded the math textbook and other learning materials such as teacher's notes, practice problems and its solutions. The objective of our analysis was to explore how effective e-mail messages were to enhance students' learning activities.

4.1.1. Analysis settings

First, we filled the class activity schedules. During the intervention period, the teacher sent six messages to students. The teacher noted words of encouragement, as well as the essential information, was part of the email message. Five out of six messages contained the notification of the next class's requirement, with an assumption that it would encourage students to preview and review the learning materials outside of the class. Each email was sent to all students. As the purpose of the intervention was to encourage students' previewing and reviewing motivation, the indicator we used was the average reading time. The logs which exceed 20 minutes were automatically removed by the system. It removed the data where students were just keeping an eBook open. Analysis settings for this case are shown in Figure 6A. To validate our results from the system, we conducted a brief survey to the students in order to investigate students' perception of the intervention. We asked students that (1) did you feel easy to receive email messages from the teacher? (response options: very easy, easy, not so easy, not easy) and (2) Did you think it was useful for your learning to receive the email message from the teacher? (response options: very useful, useful, not so useful, not useful).

4.1.2. Case extraction results

Figure 6B and Figure 6C represents the analysis results on the case extraction page. The overview panel (Figure 6B) showed the effectiveness of the intervention. We have attached a table of the analysis results as well in order to check the accuracy of it (Table 6). This table is not shown on the screen but logged in our analysis API server whenever the user conducts an analysis. The panel shows that email intervention increased the reading time on e-book readers by 4.74 minutes compared to the control period with moderate strength of case. Checking against the table, it does match to the results of actual mixed effects model results. The data panel (Figure 6C) shows the descriptive mean reading time both in intervention and control period. It indicates that the reading time on the e-book reader was longer (average: 27.42 min) in the intervention period than that in the control period (average: 22.77 min). For the survey response, 56 out of 60 students (93%) responded to the survey. Among the responders, 52 students (93%) responded that receiving email messages were easy (very easy or easy) and 27 students (48%) thought that the messages were useful (very useful or useful) for their learning at home.

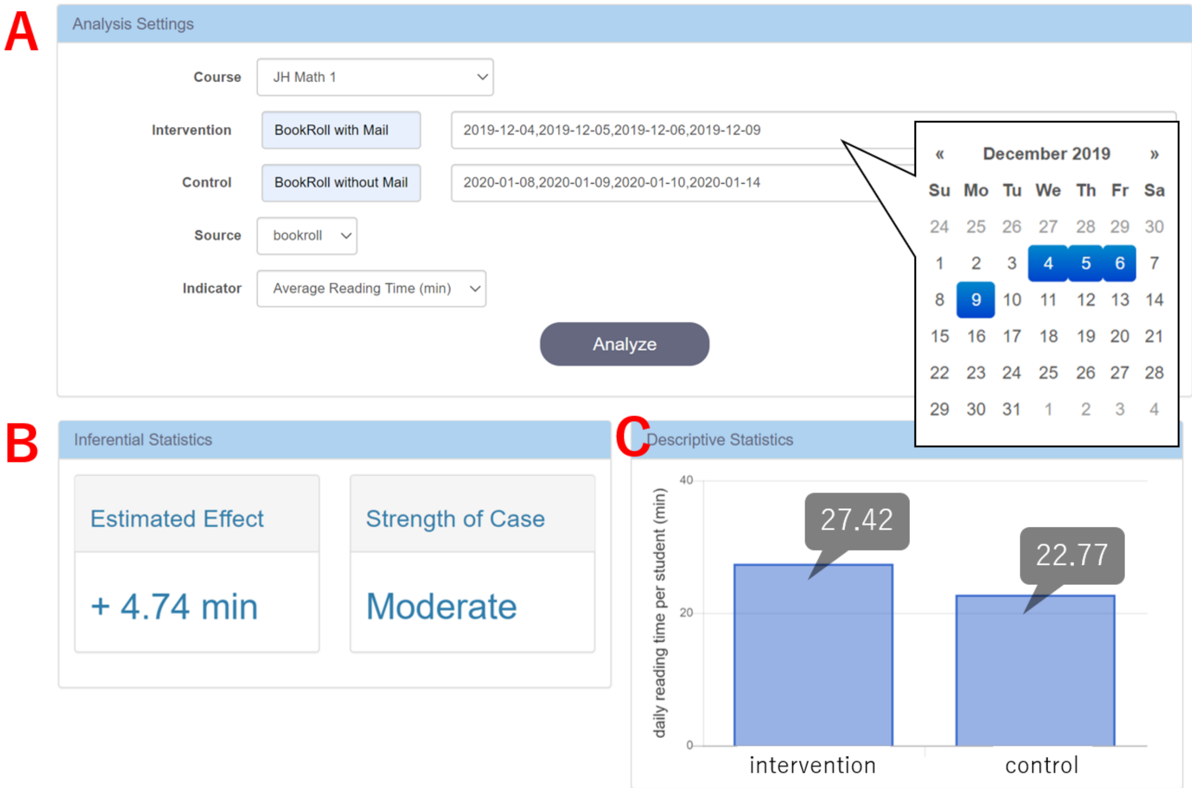

Figure 6. Analysis results on study case: Effectiveness of e-mail messages

Table 6. Mixed linear model regression results

	Coefficient	Standard Error	*p*-value
Intercept	22.55	1.73	.00
Intervention	4.74	2.17	.029
Group Variance	10.11	0.76	-

4.1.3. Findings from Case A

The results suggested that our system can be adopted to a traditional teaching-learning class in secondary education. As our platform offers a learning management system and e-book reader with learning analytics dashboard, teachers can quickly adopt technology-enhanced learning to their class. It produces a potential to extract the effectiveness of an intervention from teaching-learning logs in a face to face classroom situation.

It is beneficial for teachers because they can explore the effectiveness of the intervention through our system. In this case, although the effectiveness of email intervention to students' learning engagement were observed in many studies (Arnold & Pistilli, 2012; Heiman, 2008), few studies deal with the situation where instructors send a message via the learning analytics system to students at the junior high school level in Japan. It is often the case that a method that worked well elsewhere would be completely ineffective elsewhere.

Apart from the quantified change measured by the system, we collected the perceived improvement from the point of view of the teacher and also that of the students. Both the teacher and portion of the students perceived the message boasted their learning engagement. While such perception confirms the face validity of the system's result as verified by the stakeholders, we plan for formal validity and reliability evaluation in our future studies.

4.2. Case B: Extracting effectiveness of peer instruction strategy in a blended learning context

The second case was an introductory Physics course for first-year undergraduate students. The course instructor used LEAF platform to offer the semester-long course from 18 January to 8 April 2019. There were 64 students enrolled in the class, out of which 59 registered on the Moodle. The instructor adopted a specific in-class active learning strategy, Peer Instruction for a particular topic and Traditional lecturing for the next topic. The course was taught in a blended-learning environment using the LEAF platform. The teacher adopted LCM (Learner-Centric MOOCs) model to their class for online activities. LCM model is a prescriptive (pedagogical) model consisting of a set of guidelines, activity formats and actions for MOOC creators (Murthy, Warriem, Sahasrabudhe, & Iyer, 2018). LCM model consists of four learning components (see Figure 7). Here are typical examples of each component.

- Learning Dialogs (LeD): short videos with reflection spots
- Learning by Doing (LbD): multiple choice questions with customized and constructive feedback
- Learning Experience Interaction (LxI): peer learning through focused discussion
- Learning Extension Trajectories (LxT): various materials to address the diversity of learners

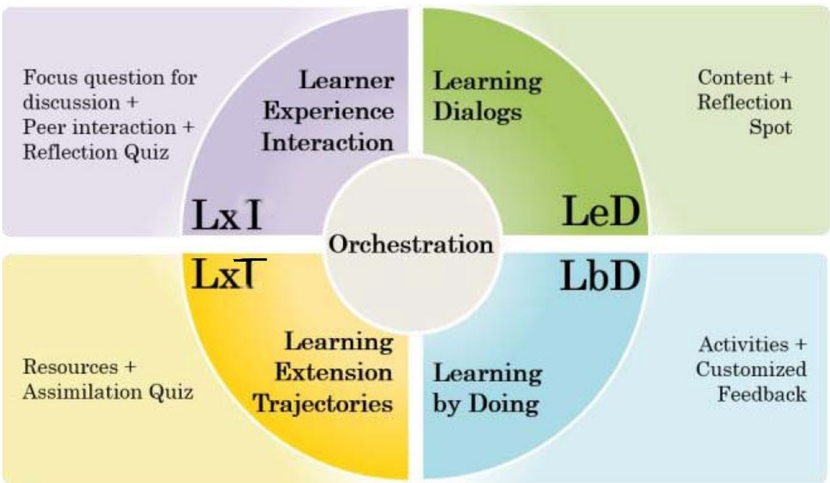

Figure 7. Four components of LCM model (Murthy et al., 2018)

In this paper, we will measure the effectiveness of Peer Instruction based on these components. We will use "the number of access to moodle contents" for LeD, "the number of access to moodle quizzes" for LbD, "the number of moodle forum" for LxI, and "the number of external resources" for LxT.

Previously the instructors reported a reflective practitioner study to elaborate the design decisions taken for the blended course activities (Kannan & Gouripeddi, 2019). We reported our initial approach of analysis and how it can help to evaluate a pedagogical model using learning logs (Kuromiya, Majumdar, Warriem & Ogata, 2019). In this paper, we further investigate the effectiveness of Peer Instruction in LCM context based on four learning components with the help of our case extraction system.

4.2.1. Analysis settings

Analysis Settings were filled like in Figure 8A. The teacher conducted the Peer Instruction method on 24, 25, 26, 27 February, and 6,7,8, 11, 12 March 2019. For comparison, the teacher held a traditional lecturing on 13, 14, 18, 19, 20, 21, 22 March 2019. As the purpose of the intervention was to enhance students' engagement, we selected four indicators which correspond to four components in the LCM model as we described in the previous section. We selected "number of access to moodle contents," "number of access to moodle quiz," "number of access to moodle forum," and "number of access to moodle resources." Each indicator corresponds to LeD, LbD, LxI, LxT components for each.

4.2.2. Case extraction results

In this case, we have conducted four analyses as mentioned. Along with it, we showed the results from "number of access to moodle contents (LeD)" in Figure 8. From the overview panel (Figure 8B), we can see that the Peer Instruction increased the number of access to moodle contents by 0.43 times. However, we can also see that the strength of the evidence was very weak, which means the results were not reliable. The result from the data panel (Figure 8C) suggested that the number of access to moodle contents in the intervention class was slightly higher than that in the control period. The average number of access to moodle contents in the intervention period was 3.57, while that in the control class was 3.17. The results from other three indicators were summarized in Table 7.

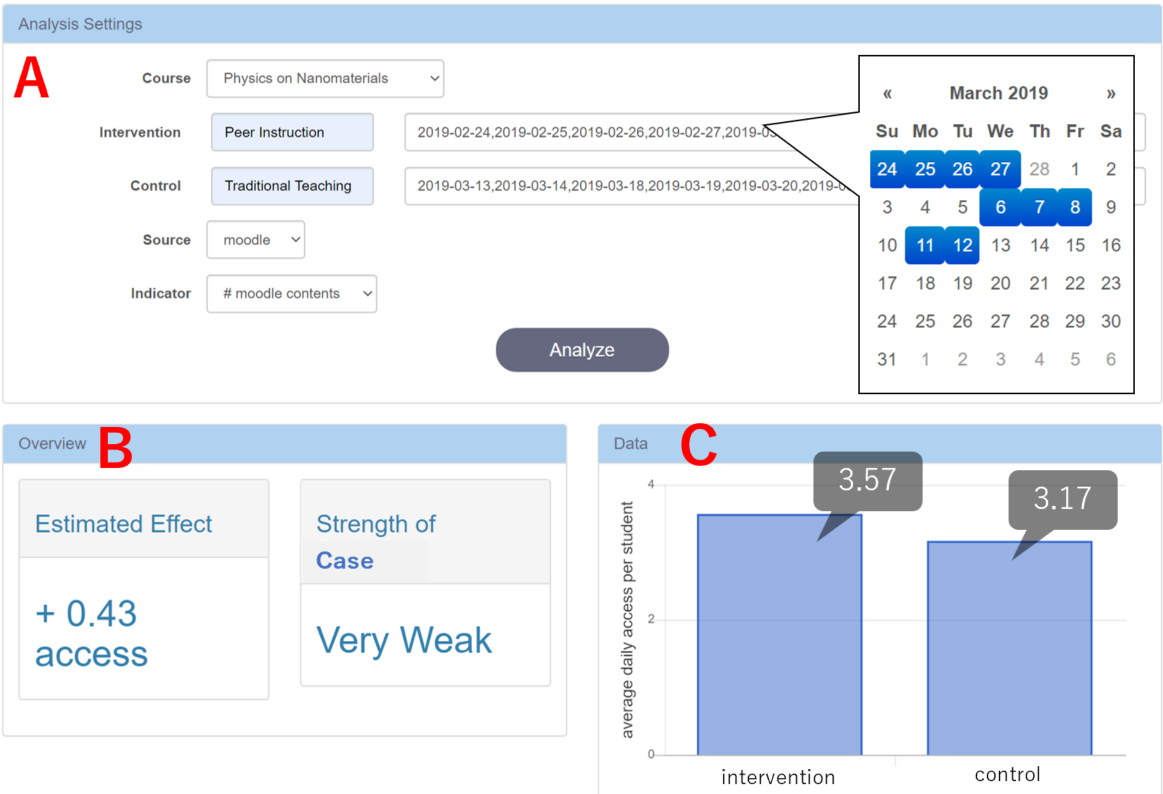

Figure 8. Analysis results on Case B: Effectiveness of the peer instruction in blended learning course

Table 7. Analysis results for each indicator

Indicator	Estimated Effect		Strength of Case
Moodle Contents (LeD)	+	0.43	Null
Moodle Quiz (LbD)	+	5.73	Low
Moodle Forum (LxI)	+	0.35	Null
External Resources (LxT)	-	1.45	Null

The results showed that Peer Instruction had no evidence of the effects to increase the number of activities on moodle compared to Traditional Teaching method. According to Table 7, only to the moodle quiz (LbD) the intervention had relatively positive effects. However, the strength of the evidence was low, which means the results were not so reliable. It suggests that we cannot say anything whether Peer Instruction activity had increased student engagement in blended online learning settings in this case.

4.2.3. Findings from Case B

This case can be considered as a demonstration of how the case extraction system was used to analyze class activity during a blended learning course. In blended learning courses, most students' learning logs are collected from the learning management system. Previous literature (Lu, Huang, Huang, Lin, Ogata, & Yang, 2018) pointed out that students' final academic performance could be predicted from only one-third of the log data in blended learning context. We can capture students' learning activities and infer the effectiveness of the intervention only from accumulated learning log by adopting our system. Also, teachers can investigate the effectiveness of the intervention from multiple aspects by selecting different indicators or sources. It is one of the advantages of using our case extraction system.

Previous literature reports that Peer Instruction encourages students' engagement in face-to-face classroom (Busebaia & John, 2020) and the online learning environment (Liu, 2019). However, no studies consider a blended learning environment. Particularly, the teacher used a specific type of blended learning strategy, Learner-Centric MOOCs (LCM) model to design the learning activities for their class. Such a combination of Peer Instruction and activities based on LCM model was not investigated yet.

5. Discussion and conclusion

5.1. Contributions to current learning analytics systems

Current learning analytics systems do not have a module for intervention evaluation. Although there are some studies where researchers tried to collect evidence about learning analytics interventions (Papamitsiou & Economides, 2014), there are no systematic solutions to collect evidence from log data. In our system, users to do reflections anytime they want by extracting the corresponding data and executing the statistical computation. It enables teachers to analyze their interventions almost in real-time. It is useful in practice because sometimes teachers have to decide whether they should continue an intervention or not. Moreover, they can retry analysis changing parameters - indicators or sources. It gives them opportunities to find unexpected things and to investigate their interventions deeper by looking at it from various aspects.

Moreover, our two case studies showed that our extracted case was essential for future teaching practice in the context where previous literature is insufficient to compare effectiveness. Traditionally, evidence-based education is for researchers and policymakers, based on well-designed RCTs. Herodotou, Heiser, and Rienties (2017) discussed the possibilities of adopting RCTs in learning analytics fields for making better evidence. Often it would be challenging to involve the teachers directly in the evidence-based practice. Our approach suggested a new type of evidence-based education based on data generated during the teaching-learning intervention which the teachers themselves orchestrate. Applying learning analytics techniques, we demonstrated how to automate data collection, analysis, and case storing for supporting evidence-based practice. It democratizes evidence-based education to practitioners and aims to make data-informed teaching practice more accessible.

5.2. Validity of our data analysis process

In both case studies, we used the indicators which can be easily retrieved from students' learning log. As we showed in case studies, these parameters can serve as a good measure for students' learning engagement.

However, we also understand that many teachers are interested in their performance as well as their learning activities. Teachers often conduct interventions in order to improve students' performance rather than their learning engagement. In that context, existing indicators are not enough. We plan to add indicators regarding students' performance such as quiz scores or exam scores.

For the *Modeling* step, we used a mixed effect model to estimate the effectiveness of the intervention in the case. By using a mixed effect model, we could consider students' differences and offer the interface which is easy to understand for users. However, we need to add models in order to deal with new types of data such as performance scores. Our project is running in the open-source spirit; we would like to allow users to define their own models for case extraction. For instance, Open Strategic Data Projects hosted by the Center for Education Policy Research, Harvard University, makes analytic tools freely available for educational data analysts (Passmore & Chae, 2019). Many analysis models published there serve as good candidates of case extraction models in our system for different analysis.

For the *Testing* step, we should mention the multiple comparison problem. Since our system allows users to do multiple adoptions of a statistical model to different indicators, users may fall into false-positives based on this repetition. That is why we decided not to use p-value as statistical testing. In both cases, we refer p-value as the strength of the case. It does not determine if the intervention worked or not but offers gradual confidence in the results. According to the statement from American Statistics Association, they noted that "Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold" (Wasserstein & Lazar, 2016, p. 131). Our approach to p-values meets this principle as we do not care about if the *p*-value exceeds a significance level or not.

5.3. Limitations

Compared to the previous studies - LACE Evidence Hub (Ferguson & Clow, 2016) and Analytics4Action (Rienties et al., 2016) (see Table 1), our system is unique to apply statistical computing automatically with learning logs. However, there are still some limitations regarding reliability and flexibility in our current approach. As our platform is primarily focused on easing the data synthesis and the statistical computation within a context using existing Learning Analytics tools, the actual selection of indicators, the context and the interpretation still remains as inputs by users. Currently the system does not evaluate the validity of that user interpretation and recorded result. However, for instance in LACE Evidence Hub, it registers only reliable meta-analysis of the literature. Similarly, regarding system flexibility, currently our system is able to extract students' engagement only based on students' behavior online. Rienties et al. (2017) pointed out that there are three levels in possible impacts of learning analytics systems - attitudes, behavior, and cognition. While Analytics4Action can collect all three types of data, our system can potentially collect all of them in the framework, however now only behavioral level data is automatically synthesized. Despite these limitations, we believe that there are benefits in democratizing evidence-based education with the practitioner by harnessing learning analytics tools and techniques to extract evidence of learning from log data.

5.4. Future research direction

In this paper, we proposed a semi-automated analysis system to the log data stored on a learning analytics platform. However, it does not guarantee the quality of the case. We only provided a place to do calculations for case registration. We should investigate our system further focusing on what context under what conditions the intervention works. Hence, we plan to develop an additional module to aggregate various teaching-learning cases into evidence. Figure 9 shows the evidence generation process on our case extraction system. By aggregating (meta-analyzing) similar teaching-learning cases, our system will be able to produce more robust evidence of the intervention.

Furthermore, we are planning to implement a case recommendation function in Evidence Portal. Once the case records keep logging in the Evidence Record Store, it might be difficult for users to find the best solution that matches their problems and situation. Case recommendation function offers the most relevant solution according to the users' context and problem. There we shall compute the context similarity between cases from the context field in the case records. It will promote users to refer to past case records rather than only to rely on their intuitive practice.

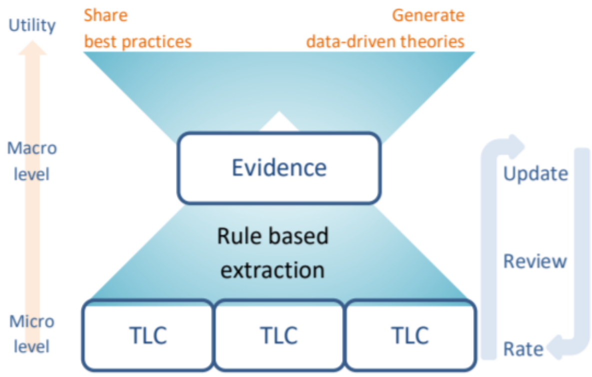

Figure 9. Evidence generation from teaching-learning cases

5.4. Conclusion

In this paper, we proposed an integrated case-extraction system based on a learning analytics platform. We showed that (1) the data processing flow of case extraction from students' learning log, (2) Mixed effects model can be used for automated case extraction, and (3) Our extraction module implemented on LEAF makes evidence-based practice accessible for practitioners by showing actual case studies.

While many implemented LA dashboards only visualize the descriptive statistics, our system design pushes that boundary to show the result of statistical modelling to users on the learning analytics platform itself. We demonstrated with two case studies of our system during a live teaching-learning scenario, where a statistical model was used to extract the effectiveness of pedagogical interventions conducted in those scenarios. In the future, we plan to build an additional module to extract evidence by meta-analyzing the cases extracted by the proposed system. We believe that integration of learning analytics techniques to automate log-data analysis, with the stakeholder's workflow has a potential to democratize evidence-based practice for different stakeholders beyond only researchers.

Acknowledgement

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304, NEDO Special Innovation Program on AI and Big Data 18102059-0, JSPS KAKENHI Grant-in-Aid for Research Activity Start-up Grant Number 19K20942, JSPS KAKENHI Grant-in-Aid for Early-Career Scientists 20K20131 and SPIRITS 2020 of Kyoto University. We sincerely thank all the students from the school in Japan and university in India. Especially, we thank Mr. Go Miyabe, Ms. Saki Maruyama, Ms. Mizuki Yamazaki, and Dr. Vijaya Nandhini Kannan, the instructors of the courses who kindly participated in our study.

References

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). doi:10.1145/2330601.2330666

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A Tutorial. *International Journal of Epidemiology*, *46*(1), 348–355.

Booth, A. (2006). Clear and present questions: formulating questions for evidence based practice. *Library Hi Tech*, *24*(3), 355–368.

Busebaia, T. J. A., & John, B. (2020). Can flipped classroom enhance class engagement and academic performance among undergraduate pediatric nursing students? A Mixed-methods study. *Research and Practice in Technology Enhanced Learning*, *15*(1), 1-16. doi:10.1186/s41039-020-0124-1

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, *33*(2), 339–360. doi:10.1007/s11245-013-9220-9

Dawson, S., Jovanovic, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 474–478). doi:10.1145/3027385.3027405

European Commission/EACEA/Eurydice. (2017). *Support Mechanisms for Evidence-based Policy-Making in Education. Eurydice Report*. Luxembourg: Publications Office of the European Union.

Ferguson, R., & Clow, D. (2016). Learning analytics community exchange: Evidence hub. In *Proceedings of the 6th International Learning Analytics and Knowledge (LAK) Conference* (pp. 520–521). doi:10.1145/2883851.2883878

Hansen, S., Schwab, F., Schneider, S., Sohr, D., Gastmeier, P., & Geffers, C. (2014). Time-series analysis to observe the impact of a centrally organized educational intervention on the prevention of central-line-associated bloodstream infections in 32 German intensive care units. *The Journal of Hospital Infection*, *87*(4), 220–226.

Heiman, T. (2008). The Effects of e-mail messages in a distance learning university on perceived academic and social support, academic satisfaction, and coping. *Quarterly Review of Distance Education*, *9*(3), 237-248.

Herodotou, C., Heiser, S., & Rienties, B. (2017). Implementing randomised control trials in open and distance learning: A Feasibility study. *Open Learning: The Journal of Open, Distance and E-Learning*, *32*(2), 147–162.

Hwang, G.-J., Spikol, D., & Li, K.-C. (2018). Guest editorial: trends and research issues of learning analytics and educational big data. *Educational Technology & Society*, *21*(2), 134–136.

Jacobs, S. K. (2008). Research guides: Health (Nursing, Medicine, Allied Health): Literature reviews. *Oncology*, *41*, 1302.

Kannan, V., & Gouripeddi, S. P. (2019). Contextualising the learner-centric MOOCs model for effective blending of flipped-classroom method in engineering physics course. In *Proceedings of the 2019 IEEE tenth international conference on technology for education (T4E)* (pp. 46–53). doi:10.1109/T4E.2019.00-51

Kuromiya, H., Majumdar, R., Warriem, J., & Ogata, H. (2019). Data-driven validation of pedagogical model - A Case of blended LCM model. In *Proceedings of the 2019 IEEE Tenth International Conference on Technology for Education (T4E)* (pp. 38–45). doi:10.1109/T4E.2019.00016

Lindstrom, M. J., & Bates, D. M. (1994). Correction to Lindstrom and Bates (1988): Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. *Journal of the American Statistical Association, 89*(428), 1572-1572. doi:10.2307/2291042

Liu, Y. (2019). Using reflections and questioning to engage and challenge online graduate learners in education. *Research and Practice in Technology Enhanced Learning*, *14*(1), 3. doi:10.1186/s41039-019-0098-z

Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220–232.

Majumdar, R., Akçapınar, A., Akçapınar, G., Flanagan, B., & Ogata, H. (2019, March). *Learning analytics dashboard widgets to author teaching-learning cases for evidence-based education*. Paper presented at the 9th International Learning Analytics and Knowledge (LAK) Conference, Tempe, AZ. Retrieved from http://hdl.handle.net/2433/244167

Morgan, R. L., Whaley, P., Thayer, K. A., & Schünemann, H. J. (2018). Identifying the PECO: A Framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environment International*, *121*(Pt 1), 1027–1031.

Mouri, K., Uosaki, N., & Ogata, H. (2018). Learning analytics for supporting seamless language learning using e-book with ubiquitous learning system. *Educational Technology & Society, 21*(2), 150–163.

Murthy, S., Warriem, J., Sahasrabudhe, S., & Iyer, S. (2018). LCM: A Model for planning, designing and conducting learner-centric MOOCs. In *Proceedings of the 2018 IEEE Tenth International Conference on Technology for Education (T4E)* (pp. 73-76). doi:10.1109/T4E.2018.00022

Organisation for Economic Co-operation and Development (OECD). (2007). *Evidence in education: Linking research and policy*. Paris, France: OECD Publishing.

Ogata, H., Majumdar, R., Akçapinar, G., Hasnine, M. N., & Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning. In *Proceedings of the International Conference on Computer in Education (ICCE 2018)* (pp. 493–496). Manila, Philippines: Asia-Pacific Society for Computers in Education (APSCE).

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *Proceedings of the International Conference on Computer in Education (ICCE 2015)* (pp. 401–406). China: Asia-Pacific Society for Computers in Education.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A Systematic literature review of empirical evidence. *Educational Technology & Society*, *17*(4), 49–64.

Passmore, D. L., & Chae, C. (2019). Potential for Meta-Scientific inquiry to improve the usefulness of HRD research outcomes for practice. *Advances in Developing Human Resources*, *21*(4), 409–420. doi:10.1177/1523422319870173

Raiola, G., & Di Tore, P. A. (2012). Statistical study on bodily communication skills in volleyball to improve teaching methods. *Journal of Human Sport and Exercise*, *7*(2), 468–488. doi:10.4100/jhse.2012.72.12

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action evaluation framework: A Review of Evidence-Based learning analytics interventions at the open university UK. *Journal of Interactive Media in Education*, 2016(1). doi:10.5334/jime.394

Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a learning analytics intervention and evaluation framework: What works? In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education: Current theory and practice* (pp. 147–166). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-06520-5_10

Seabold, S., & Josef P. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference* (Vol. 57, pp. 92-96). Retrieved from https://pdfs.semanticscholar.org/3a27/6417e5350e29cb6bf04ea5a4785601d5a215.pdf

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133.